Nami *et al. Brain Informatics* (2026) 13:28
https://doi.org/10.1186/s40708-026-00310-4

Brain Informatics

REVIEW Open Access

RDoC-informed explainable AI as a paradigm for multilevel Alzheimer's disease diagnosis and progression prediction: a systematic review

Mohammad Nami[1*], David Peebles[2], Fadi Thabtah[3] and Firuz Kamalov[4]

Abstract
Explainable Artificial Intelligence (XAI) is gaining popularity in early diagnosis and monitoring of dementia. Herein, we recommend the incorporation of the National Institute of Mental Health's Research Domain Criteria (NIMH-RDoC) framework with XAI-informed diagnostic protocols to help establish diagnosis at early stages of Alzheimer's disease (AD). RDoC has a dimensional structure that extends across units of analysis from genes and molecules to circuits, physiology, behavior, and introspection. By restructuring diverse features as inputs including apolipoprotein E (APOE) genotype, amyloid and tau biomarkers, computational neuroimaging-informed cortical atrophy, Positron Emission Tomography (PET) hypometabolism, quantitative electroencephalography (qEEG) rhythms, cognitive tests, and digital behavioral markers), onto RDoC units provides more insightful and inclusive models. In this context, data-driven approaches such as XAI can achieve not only increased interpretability but also enhance their mechanistic validity. Such an innovative approach places data-driven model outputs within neurobiologically based domains such as Cognitive Systems, Negative Valence, and Arousal/Regulatory Systems. Our synthesis suggests that a 'converging RDoC and XAI' approach may help bolster the coherence of AD biomarkers, promote model exploration in clinical decision-making. This approach is also expected to provide a strategic roadmap for translational neuroscience and personalized medicine. Another major aim of this study is to critically analyze current XAI approaches used in dementia research, particularly the diagnostic and prognostic aspects. By explicitly grounding explanations in RDoC cognitive domains and paradigms, the framework also aims to make model outputs meaningful in terms of specific mental functions (e.g., episodic memory, cognitive control), thereby supporting neuropsychologically-informed diagnosis, categorization, and communication with patients and caregivers.

Keywords Alzheimer's disease, Dementia, MCI, Biomarkers, RDoC, Explainable AI, Interpretable AI, Transparent models, Decision rules, Saliency maps

*Correspondence:
Mohammad Nami
Mohammad.nami@cud.ac.ae
[1]Cognitive Neuroscience and Neuropsychology Unit, School of Health Sciences and Psychology, Canadian University Dubai, Dubai, UAE
[2]School of Human and Health Sciences, University of Huddersfield, Huddersfield, West Yorkshire, UK
[3]Abu Dhabi School of Management, Abu Dhabi, UAE
[4]School of Engineering, Applied Science and Technology, Canadian University Dubai, Dubai, UAE

1 Introduction

Pathologically, Alzheimer's disease (AD) is characterized by the abnormal accumulation of amyloid and tau in the A/T(N) model. Here, "A" represents amyloid β (i.e., CSF Aβ42 or PET amyloid) and "T" represents aggregated tau (i.e., cerebrospinal fluid-CSF p-tau or tau PET) and "(N)" represents neurodegeneration or neuronal injury (i.e., Fluorodeoxyglucose-FDG- or PiB-PET, Magnetic Resonance Imaging (MRI) atrophy or CSF total tau) [1, 2]. While this source of biomarker classification has revolutionized diagnostic specificity, it also faces the shortcoming of limiting AD to its molecular identity.

Parallel to these advances, The Research Domain Criteria (RDoC) framework by the National Institute of Mental Health (NIMH) is a dimensional and integrative approach to mental and neurocognitive disorders that focuses on a broad range of levels of analysis [3]. Instead of traditional categorical frameworks, RDoC organizes research into fundamental functional domains. Each domain is analyzed by units of analysis, which include several variables ranging from genes and molecules to circuits, physiology, behavior, and self-report. In neurodegenerative disorders such as Alzheimer's disease, RDoC offers a clear way to associate the biological mechanisms behind such manifestations with observable cognitive and behavioral patterns, enabling us to develop more accurate, mechanistically based, and transdiagnostic models for the characterization of diseases. RDoC promotes research through six domains including Negative Valence, Positive Valence, Cognitive Systems, Social Processes, Arousal/Regulatory Systems and Sensorimotor Systems. The levels of analysis that cut across units range from genes to molecules to circuits to physiology to behavior to self-report to paradigms [4, 5].

In this context, Explainable Artificial Intelligence (XAI) is an emerging area of methodological and framework efforts to make the decision-making of artificial intelligence systems transparent, interpretable, and clinically meaningful. Following model training, post-hoc explainability methods [such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME)], are used to clarify the role of features in the prediction process of the model, without changing the underlying architecture. Meanwhile, intrinsically interpretable models, such as linear models, decision trees, and rule-based systems, are transparent by design, through which their decision-making logic can be inspected directly. In clinical settings like Alzheimer's disease diagnosis, XAI holds significance as it connects predictive accuracy with interpretability and thus increases clinician trust, assists with regulatory compliance, and promotes the translation of diverse multimodal data into actionable, cognitively meaningful insights [6, 7].

In essence, Artificial Intelligence (AI) is a broad discipline of computational methods used to model or simulate intelligent behavior, with Machine Learning (ML) specifically as a sub discipline that is trained to learn patterns using data. Deep Learning (DL) is a subset of ML that employs multi-layered artificial neural networks to model complicated, high-dimensional relationships. Although DL models have showcased impressive capabilities around neuroimaging and disease prediction, they tend to be characterized by limited interpretability. This hierarchy is key for explaining the role of XAI in improving the transparency and interpretability of these modeling methods in a clinically relevant manner [6–8].

Given recent advances in AI-driven medical research, it is timely to incorporate XAI into this paradigm. Black-box AI models based on artificial neural networks like deep learning discover patterns in high-dimensional biomarker data, but frequently lack clinically meaningful explanations [6, 7].

Beyond its underlying molecular and imaging signals, AD is principally expressed as a disruption of cognitive systems. Clinically, patients present with characteristic patterns of impairment across episodic and semantic memory, language, executive control, visuospatial skills, and social cognition. These profiles evolve over time, from preclinical stages through prodromal mild cognitive impairment (MCI) to dementia, and are influenced by factors such as cognitive reserve, comorbid pathology, and the compensatory strategies adopted by individuals. RDoC's domains, particularly Cognitive Systems, Arousal/Regulatory Systems, and Social Processes, therefore, provide a set of biological categories and concepts, together with a principled way of characterizing the cognitive features that emerge from the underlying circuit and molecular disruption.

Integrating XAI with RDoC for AD therefore provides two significant benefits: first, linking multilevel biomarkers to RDoC units of analysis, and second, enabling model outputs to be framed in terms of cognitive constructs that are directly observable in neuropsychological assessment and everyday functioning (e.g., "episodic memory retrieval", "cognitive control", "affective regulation"). This multilevel mapping is critical as it provides a shared language across neurology, psychiatry, neuropsychology, and data science, and ultimately places explanations within the clinical cognitive space, allowing model outputs to be understood as statements about changes in specific mental functions rather than less interpretable combinations of features.

The embedding of RDoC in existing data-driven models integrated with XAI methods provide two key advantages. It places discovered features within established neurobiological and behavioral constructs, and links molecular markers to the observed cognitive-behavioral

changes. Hence, AD can be conceptualized as a multilevel syndrome in which amyloid pathology (Molecules) relates to hippocampal circuit deficits (Circuits), translated into memory disturbances (Cognitive Systems/Behavior), captured by neuropsychological tests (Reports/Paradigms) [8]. The already-established biomarker-driven classification renders an accurate view of the disease in addition to the clinical symptoms and facilitates early diagnosis and classification throughout what is now recognized as the Alzheimer's continuum. Meanwhile, Alzheimer's clinical syndrome includes those individuals who demonstrate cognitive impairment either in the amnestic type, single or multiple domains. The presence of abnormal Aβ but not tau biomarker marks "Alzheimer's pathologic change," as a minimally advanced point along the continuum. BBoth 'Alzheimer's pathologic change' and 'AD' fall underthe AD spectrum including preclinical, prodromal, and demented phases of the disease [9]. This terminology comprises both the early and more severe stages of disease and is consistent with what was previously referred to as "possible" or "probable AD." Yet, clinical syndrome is not adequate to diagnose AD and neuropathological or biomarker confirmation of (amyloid and tau) pathology is ideally needed to label "AD". This emphasizes the applicability of biomarker-guided diagnostics that are based on a biological hierarchy of staging and that support more targeted approaches [10].

In the context of the A/T(N) model, the three groups of biomarkers are binarized as positive (+) or negative (-), resulting in eight possible biomarker profiles. Then, these profiles are classified in one of three main biomarker categories: (1) normal AD biomarker [A-T-(N)-], (2) Alzheimer's continuum [any A+ profile], and (3) non-Alzheimer's pathologic change, known as Suspected Non-Alzheimer's Pathophysiology or SNAP [A-T+(N)-, A-T-(N)+, or A-T+(N)+] [11, 12].

Despite the progress brought by the A/T(N) framework and the biomarker-based classification, there are still major challenges left for the implementation of these approaches into everyday clinical care. Heterogeneous sources of data vary greatly in the level of granularity and detail, and range from neuroimaging, CSF assays, cognitive scores to behavioral markers [13]. The complexity and richness of information across the modalities make the integration of findings a formidable task for clinicians. In addition, early and precise diagnosis, especially in preclinical or prodromal phases, is widely impeded by overlap between the symptomatology, restricted accessibility of advanced tests for biomarkers and lack of scalable strategies for personalized prognosis and intervention. In this regard, AI and Machine Learning (ML) methods provide promising solutions. ML can improve the accuracy of diagnosis, risk stratification, and longitudinal monitoring disease progression by integration of multimodal-heterogeneous high-dimensional data [14, 15].

While there is increasing evidence supporting explainable AI in Alzheimer's disease, its incorporation into the RDoC framework remains largely unexplored. This systematic review presents a novel synthesis that evaluates XAI applications for AD diagnosis and monitoring through the lens of RDoC domains and constructs. By aligning transparent ML techniques with mechanistic cognitive-affective processes, the review aims to promote the development of clinically relevant, interpretable models that translate multimodal biomarkers into actionable clinical insights.

The remainder of this paper is organized as follows. The clinical diagnostic process of Alzheimer's disease, including staging and biomarker-informed frameworks, is described in Sect. 2. In Sect. 3, we describe the search methodology, along with XAI approaches leveraged for dementia research. Current evidence from XAI-based studies of key neuroimaging and multimodal domains is synthesized in Sect. 4. In Sect. 5 we present the proposed RDoC-XAI integration framework, where multilevel biomarkers are mapped onto cognitive and neurobiological domains. The clinical implications, interpretability considerations, and implementation challenges of XAI in Alzheimer's disease, including limitations and reliability concerns, are discussed in Sect. 6. Finally, Sect. 7 presents conclusions and future directions for advancing explainable, clinically meaningful AI applications in neurodegenerative disease.

2 AD clinical diagnosis process

The clinical diagnosis of AD is based on a staged, structured workflow, and in accordance with recent medical recommendations. Stage I represents minimal clinical manifestation and maximal subclinical burden. It starts with a full clinical evaluation comprising cognitive testing, neurological exam, and medical history, in many cases employing standardized investigations such as the Mini-Mental State Examination (MMSE) or Montreal Cognitive Assessment (MoCA) [16]. Structural brain changes and alternative causes of cognitive impairment are excluded with the assistance of neuroimaging such as MRI or computed tomography (CT) scans. Biomarkers have increasingly been integrated into diagnosis, where levels of amyloid-beta, tau proteins can be identified by PET scans, by CSF testing and through plasma assays as well as via the biological definition of AD [17, 18]. A set of criteria for the 2024 National Institute on Aging and the Alzheimer's Association (NIA-AA) guidelines proposes that a diagnosis of AD can be established when there is clinical evidence of cognitive decline and when the biomarkers indicate AD pathology [19].

The progression of AD is a continuum and passes through a series of defined stages, as described in the NIA-AA 2024 biological and clinical framework. In the preclinical stage, there are silent pathologic changes, amyloid-beta build-up, and tau tangles, that come before obvious symptoms, and that can only be detected by biomarkers (Leon, Ghahremani et al. 2025). During the development of neurodegeneration, subjects progressively can enter the prodromal stage also known as mild cognitive impairment (MCI) in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) or Mild Neurocognitive Disorder (Mild NCD) in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR) attributable to AD, in which the onset of memory deficits can be observed, while daily life is still preserved. Lastly, in the dementia phase, the level of dependence posed by cognitive impairment is severely high and classified as mild, moderate, and severe stages based on functional deterioration. These stage definitions are based on clinical and objective biomarker measures, and represent a move toward a biologically based spectrum of AD progression [13, 20].

Of importance, cognitive status and neurobehavioral symptoms also contribute to staging. Cognitively unimpaired individuals can possess the biomarker signature of AD, even in preclinical stages. Others may present with metabolic syndrome or transitional cognitive impairment which includes subjective cognitive complaints, early subtle cognitive decline on serial testing, or neurobehavioral symptoms of anxiety, apathy or depression. Identification and classification of patients according to these dimensions are important not just for diagnosis, but for monitoring the course and evaluating treatment response [21]. Within this dynamic context, machine learning (ML) methods have emerged as an effective tool to analyze multimodal biomarker and cognitive data, and thus to help determine diagnosis and prognosis throughout the Alzheimer's continuum [15, 22] (Fig. 1).

3 Search methodology and XAI background in dementia diagnosis

To ensure a structured literature review methodology and and to assess the reporting quality of included studies, our literature search method complied with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) and STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines. Digital databases, including PubMed, Scopus, Web of Science, IEEE Xplore were analyzed, covering publications from January 2013 to October 2025. The search strategy integrated Medical Subject Headings (MeSH) and free-text terminology of AD as well as XAI. The search was conducted using the following keywords and Boolean combinations:

I. Main search terms: "Alzheimer's disease", "dementia", "MCI", "mild cognitive impairment".
II. Artificial Intelligence/Machine Learning keywords: "machine learning", "deep learning", "artificial intelligence".
III. Interpretability and Explainability: "explainable AI", "interpretable AI", "SHAP", "LIME", "transparent models", "white-box models", "decision rules", "saliency maps".

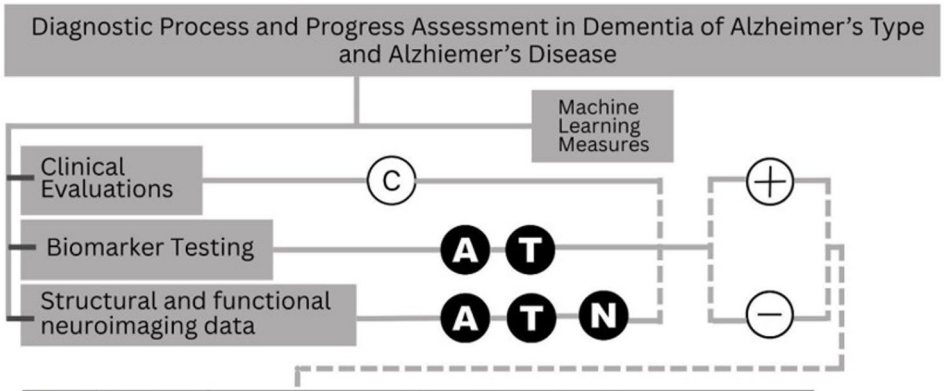
In total, 528 articles were obtained from the primary keyword search. After excluding 112 duplicates, 416 records were screened by the titles and abstracts. Among them, 297 publications were excluded for being irrelevant, in review nature, or not focusing on explainability. The other 119 full text articles were screened by the inclusion criteria including: 1-original research studies, 2-application of XAI or interpretable ML to AD dataset (Alzheimer's Disease Neuroimaging Initiative-ADNI), 3-addressing classifications, diagnoses, staging, progression, and predictions, and 4-adequacy of methodological transparency and outcome measures.

After full-text analysis, 63 articles were excluded for presenting non-interpretable methodology, poor description, or non-applicability to AD. A total of 56 studies were ultimately included in this systematic review (Fig. 2).

ML can be an effective tool to augment early diagnosis of AD as it has been shown to effectively explore hidden dependencies within high-dimensional data, which cannot necessarily be captured by classical statistical approaches. Recent investigations by Dansereau and colleagues [23, 24] have applied data-driven algorithms such as support vector machines, and others on the ADNI2 datasets to develop the HPS (Human Performance Score), which achieved high predictive accuracy performance in MCI conversion to AD diagnosis.

In the same vein, more recent works constructed an ensemble-based model that combined genomic and imaging traits to generate a multimodal dementia risk score where the model yielded robust prediction (86%) of conversion risk of cognitively normal individuals [25]. Another research study by Yamada et al. (2022) proposed a data-driven approach to investigate digital drawing patterns in tablet-based evaluation [26]. The approach achieved an AUC (area under the curve) of 0.909, and had significant correlations with clinical MMSE scores.

The clinical application of ML models is limited in part by the fact that many state-of-the-art models such as deep neural networks are "black box" models, i.e. decision-making processes are not interpretable [10]. XAI includes post-hoc explanation methods including



A/T(N) profiles	Biomarker category	
A−T−(N)−	Normal AD biomarkers	
A+T−(N)−	Alzheimer's pathologic change	Alzheimer's continuum
A+T+(N)−	Alzheimer's disease	
A+T+(N)+	Alzheimer's disease	
A+T−(N)+	Alzheimer's and concomitant suspected non-Alzheimer's pathologic change	
A−T+(N)−	Non-AD pathologic change	
A−T−(N)+	Non-AD pathologic change	
A−T+(N)+	Non-AD pathologic change	

A = Amyloid biomarker (CSF Aβ42 or PET amyloid)
T = Tau pathology biomarker (CSF p-tau or tau PET)
(N) = Neurodegeneration/neuronal injury marker (FDG-PET, MRI atrophy, or CSF total tau)

Fig. 1 The A/T(N) classification of AD was proposed by NIA-AA for the early diagnosis and tracking of AD and dementia of the Alzheimer's type. This model integrates clinical assessments (C), biomarker measures (A): amyloid, (T): tau, (N): neurodegeneration, and structural/functional imaging in the classification of individuals across the AD trajectory. The A/T(N) system, thus, allows distinguishing among normal biomarkers, Alzheimer's-related changes, and non-AD pathological changes. The combination of quantitative, functional neuroimaging (e.g. FDG-, PiB-PET, and volumetric and Voxel-based morphometric MRI) increases the early detectability of neuronal cell damage. Machine learning (ML) metrics applied across stages provide personalized prediction, classification, and staging by aggregating multi-dimensional data (cognitive, imaging, molecular). This increases accuracy of diagnosis, preclinical detection is supported, and the disease course can be monitored over time

SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), as well as prototype-based learning. The integration of XAI methods in neuroimaging analysis has recently enhanced the interpretability of complex ML models in AD research. Two popular solutions among them are for model decisions, interpretations, and learning biologically relevant features from high-dimensional imaging. SHAP and LIME are model-agnostic explainability methods that provide local interpretations of individual model predictions by estimating the contribution of input features. While LIME focuses on local surrogate approximations, SHAP offers both local and aggregated global explanations grounded in cooperative game theory [27].

SHAP is based on cooperative game theory and calculates the contribution of each input feature (e.g., brain region voxel intensity, cortical thickness) by approximating Shapley values, providing a consistent, additive feature attribution technique that can be used across model types. Applied to neuroimaging data in AD, SHAP can be used by researchers and clinicians to interrogate which brain regions have the most impact on a model's prediction of disease status or progression (e.g., hippocampal atrophy, or temporal lobe hypometabolism). By giving both global and local explanations, SHAP enables not only the verification of biologically plausible model discoveries but also the personalization of diagnostic inferences for patients. This approach has proven useful in validating the classification of ML approaches for staging AD based on MRI and PET features [28, 29].

On the other hand, LIME provides a local complementary interpretability technique that approximates the complex model by an interpretable surrogate model (usually linear or decision tree-based) in the vicinity of a prediction. It introduces data perturbation (anomalies) to the input and examines which changes in neuroimaging features most strongly drive individual predictions. LIME is especially suitable in a clinical environment when

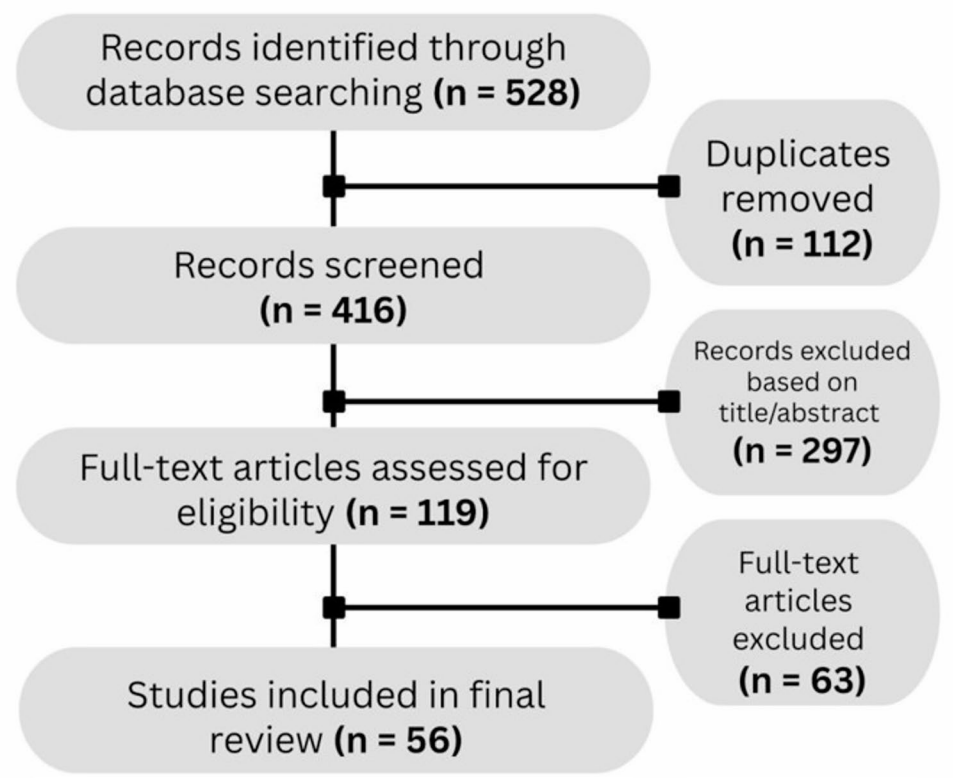


Fig. 2 PRISMA Flow Diagram : XAI in AD Diagnosis

providing explanations for individual decisions, such as why a patient's scan was labelled as MCI or early AD. Such transparency is important for establishing clinician confidence in automated systems, as well as for providing direction towards any follow-up inquiries or interventions. The model-agnostic methodology of LIME makes it applicable to a wide range of ML architectures used in brain imaging analysis [30, 31].

These approaches help the clinician to better understand which features are more responsible for diagnostic/therapeutic decisions and hence increase the personalization of machine learning predictions. The above approach potentially encourages compliance with such laws and to facilitate collaboration between AI systems and healthcare professionals [32]. Applied to the context of AD, XAI allows the transparent tracking of diagnostic decisions down to the level of cognitive, neuroimaging, and genetic risk factors. For instance, Geddes et al. (2018) and Yamada et al. (2023) have shown that interpretable models that offer accurate prediction on diagnoses can be used to describe the networks established between neurofunctional and behavioral information [26, 33].

Later in this review, we further explore the potential role of XAI techniques for early detection, disease progression, and patient stratification in AD, with emphasis on empirical work using real-world, quantitative clinical data. Particular attention is given to studies utilizing explainable or interpretable AI to interpret neuroimaging results, digital phenotyping measurements, and genetic data, frequently with the ADNI data [34]. The present report highlights the role of XAI in bridging the gap between diagnostic accuracy and clinical use in the real-world, rather than solely adding a review of AI in dementia to existing extensive supplementary literature. Through targeted analysis of experimental investigations, the review highlights the importance of interpretability in improving early screening strategies and personalized healthcare. This paper also presents existing medical diagnostic models for AD and dementia, as well as an overview of the state-of-the-art XAI approaches in the context of dementia research with the application to multimodal clinical aids. Key findings and open challenges in AI-augmented diagnosis are summarized, and recommendations are provided.

4 Research themes and discussion

4.1 MRI-based XAI studies

Several studies have applied XAI methods to structural and functional MRI, particularly using ADNI brain imaging data for the early detection and progression of dementia. For example, Coluzzi et al. (2025) performed regional volume-based analysis on T1-weighted MRI and then applied SHAP method to identify hippocampus, entorhinal cortex, and ventricles as critical predictive

markers [35]. Similarly, Nigri et al. (2020) applied Convolutional Neural Networks (CNNs) with saliency maps to visualize which brain areas impacted classification performance highlighting the significance of medial temporal lobe regions [36].

A recent study by Lozupone et al. (2024) provided interpretability to MRI data-driven models through their XAI attention framework by projecting attention weights over 3D scans to derive area specific predictions that are associated with clinical knowledge. The derived data-driven model reached an 88% accuracy for CN-MCI cohort classification [37]. In the same research arena, Anjomshoae and Pudas (2024) applied XAI methods within a graph CNN to extract predictive substructures of AD from connectome features echoing the importance of modeling inter-regional brain relationships [38].

Galazzo et al. (2022) and Amoroso et al. (2023) investigated beyond classical anatomical features and used microstructural signatures such as Accelerated Microstructure using Robust Acquisition (AMURA) and aging brainprints. These techniques when embedded into deep learning algorithms like CNNs with integrated Gradient-weighted Class Activation Mapping (Grad-CAM) can reveal more early-stage biomarkers such as frontal and parietal diffusivity patterns. The inclusion of such features represents a major advance in MRI-based AD diagnostics [39, 40].

The use of Grad-CAM is potentially useful in explaining complex class distributions and is mainly used for visualizing which segments or regions of input data significantly influence the predictions of deep learning architectures, especially CNNs. Grad-CAM exploits gradient information that enters the final convolutional layers, providing class-centered heatmaps that surface relevant features leading to that prediction. For study towards neuroimaging in the context of Alzheimer's disease, Grad-CAM allows the detection of disease-relevant brain regions, like the hippocampus or posterior cingulate cortex, which in turn allows for the interpretation of model outputs to align with well-established neurobiological signatures [40].

In longitudinal modeling studies like Bapat et al. (2024), the authors used MRI-based CNN-based algorithms trained on a 4-year follow-up dataset showing 91.3% accuracy in predicting progression from CN to AD [41]. Recently, Dolci et al. (2025) integrated genomics with MRI in a generative XAI methods. The authors maintained strong interpretability by ensuring that SHAP method scores remained clinically relevant even after fusion [42]. This model aligns well with the work of Amoroso et al. (2023) who studied brain connectivity metrics as supplementary input. The attention-based XAI proposed model identified disrupted long-range connections as early predictors of conversion to MCI thereby linking functional MRI connectivity to structural degeneration patterns noted in voxel-based studies [39].

The stability of MRI-based XAI was assessed in Lombardi et al. (2022) in which a framework was introduced to assess reliability as a metric for evaluating the robustness of saliency maps. The study particularly showed that while hippocampal features were consistently detected; however, specific regions like the precuneus fluctuating as they rely on the data split [43]. Their findings largely corroborated the work by El-Sappagh et al. who compared a few XAI methods on the same MRI features and revealed that Integrated Gradients provided more stable attribution maps compared to LIME or plain saliency [44].

Studies like Svenberg (2025) and Lozupone et al. (2024) supported visualization by creating pipelines that clinicians could interpret directly. For example, Svenberg used 3D rendering tools with XAI overlays thus allowing experts to validate model behavior, which is essential in bridging AI and neurology [37, 45]. Their findings directly align with Odusami et al. who combined MRI and PET features and showed how XAI visualizations enhance clinical confidence [46].

Overall, the MRI studies show a convergence on the hippocampus, entorhinal cortex, posterior cingulate, and lateral temporal regions as markers of early disease. These findings are consistent across 2D CNNs [47], 3D volumetric models [48], and attention-based networks [37], with explainability tools providing consensus on their importance.

4.2 PET-related XAI studies

Studies applying XAI to PET imaging in dementia focus on FDG-PET and amyloid PET with a shared aim of identifying early regional changes in brain metabolism or plaque deposition. A study by De Santi et al. applied a CNN framework on 18 F-FDG PET images and achieved an AUC of 0.92 highlighting consistent hypometabolism in the posterior cingulate, precuneus, and inferior parietal cortex. Their maps matched known early metabolic decline areas thus strengthening the model's clinical alignment [49].

Jiang et al. used deep SHAP on radiomic PET features to predict progression from early AD stages. The authors found that regional PET uptake values especially in the temporoparietal junction and medial parietal regions were important. SHAP values allowed for individualized interpretation of each patient's progression probability [8]. Further, Rasi et al. applied SHAP to FDG-PET subregions and validated their results across ADNI subsets. The reported results showed 87.4% accuracy for AD classification and confirmed that the posterior cingulate consistently emerged as a predictive subregion [50].

Odusami et al. (2023) proposed a multimodal approach fusing PET and MRI within a deep fusion network augmented by Grad-CAM. Their explainability layer revealed that PET data primarily drove classification in early stages (CN-MCI) whereas MRI contributed more prominently to later stages (MCI-AD) [46]. This temporal shift in modality utility echoes findings by Weatheritt et al. (2021) who revealed that saliency maps of PET alone captured more features in early progression than MRI-based ones [51]. Beer et al. further improved PET explainability by incorporating genetic risk via Aquaporin 4 (AQP4) polymorphisms. The proposed XAI method linked genotype variation to increased amyloid PET burden thereby mapping spatial PET changes in at-risk individuals [52].

An earlier investigation also incorporated latent multimodal features including PET, and employed layer-wise relevance scores to understand contribution of each modality. The authors revealed that PET had the highest salience scores in differentiating CN from MCI and used a modified SHAP approach to display region-wise influence [53].

Overall, PET-based XAI studies focus on posterior cingulate, precuneus, and lateral parietal cortex as core regions of early AD pathology. XAI methods such as SHAP, Grad-CAM, and integrated saliency models dominate the interpretability toolkit. PET data offers strong predictive value in preclinical and prodromal stages and, when integrated with MRI or genomics, can significantly enhance diagnostic models without compromising interpretability [54].

To ensure consistency and comparability across studies, Table 1 focuses on investigations utilizing the ADNI dataset, given its standardized data acquisition protocols, multimodal richness, and widespread use as a benchmark in Alzheimer's disease research.

The studies reviewed demonstrate a range of interpretable machine learning/deep learning methodologies for the task of AD diagnosis and progression prediction, enabled by the rich multi-modal data in the ADNI repository (Table 1). Dansereau et al. also used a SVM and logistic regression for subtype clustering reaching high precision (90%) and specificity (97%) for MCI-to-AD progression, but low sensitivity (47%) suggesting that they might struggle to identify true converters [23]. This compromise emphasizes a frequent drawback of subtype-based models: they do well on "easy cases" and not as well on less obvious clinical variation. In contrast, Jo et al. integrated 16 deep learning publications performed on the CNN, SAE, and RNN networks to obtain 96% diagnostic accuracy in AD and 84.2% in predicting MCI progression, illustrating the power of multimodal neuroimaging and nonlinear feature construction at the cost of lower interpretability in comparison to classical models [55].

Meanwhile, Pelkmans et al. applied grey matter network measures and logistic regression to categorize subjects as to likely clinical progression, highlighting structural brain network degeneration as a discriminatory measure. While performance measures are moderate (accuracy 65–72%), within the organ geometry map (OGM) framework, the approach has more interpretability and clinical relevance compared to batch learning black-box techniques [56].

Among the reviewed reports, some proposed a genomics-coupled neuroimaging model via an ensemble method and feature selection to predict dementia conversion and determined that the genetic features of the model played a more significant role in early risk stratification (86% accuracy) than the imaging features. Where biologically informative and stratified, the complexity of

Table 1 Explainable/Interpretable AI evidence Using ADNI

Title	Methods used	Data used	Evaluation metrics	Source reference
A brain signature predictive of future progression to Alzheimer's dementia	SVM (Support Vector Machine)+Logistic Regression; subtype clustering	ADNI2 dataset (sMRI+fMRI)	Precision (90%), Specificity (97%), Sensitivity (47%)	[23]
Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction	CNN, SAE, RNN (DL) RNN: Recurrent Neural Network SAE: Stacked Autoencoder SVM: Support Vector Machine	ADNI (sMRI, PET, multimodal)	Accuracy (96% AD), MCI conversion prediction (84.2%)	[55]
Grey matter network markers identify individuals with prodromal Alzheimer's disease who will show rapid clinical decline	Grey matter network analysis+ Logistic Regression	ADNI+Amsterdam Dementia Cohort	Accuracy (65–72%), AUC (Area Under the Curve), ROC (Receiver Operating Characteristic), AIC (Akaike Information Criterion)	[56]
Machine learning-based multimodal neuroimaging genomics dementia score for predicting future conversion to Alzheimer's disease	ML ensemble, feature selection, stratified risk modeling	ADNI1 (MRI + 521,017 SNPs+APOE) SNP: Single Nucleotide Polymorphism	Accuracy (86% genetics), stratified evaluation	[57]

the SNP-based models may have practical implications for routine clinical application [56, 57].

Collectively, such a body of evidence paints a clear picture of the trade-off between performance (i.e., how well models work in practice or at what level of performance) and explainability (i.e., clarity of the underlying decision logic) for models where simpler models offer greater explainability but less performance, and deep learning provides improved accuracy, suggesting that hybrid techniques that take the best of both (but not all) may be the best road forward for translational AD research (Table 2).

The NIA-AA has established guidelines for Alzheimer's Disease diagnosis which divide the disease progression into three stages: preclinical AD, mild cognitive impairment due to AD and Alzheimer's dementia [67]. The diagnosis process in clinical practice relies on neuropsychological assessments together with structural and functional neuroimaging (MRI, PET) and fluid biomarkers extracted from CSF or plasma [63]. The AT(N) framework which includes amyloid (A), tau (T) and neurodegeneration (N) markers has become widely used for distinguishing between disease stages and forecasting disease progression risk.

The standard diagnostic process includes biomarkers and clinical methods which follow current guidelines as shown in Table 3.

5 RDoC and XAI integration approach

To provide conceptual clarity, this section introduces the RDoC-XAI integration as a theoretical framework for mapping multilevel biomarkers onto interpretable cognitive and neurobiological domains. This would make it possible to design a systematic architecture where data-driven model outputs are not presented in a vacuum but closely link up with RDoC units of analysis from genes and circuits to behavior and self-report. By embedding XAI outputs within this dimensional framework, the approach facilitates a mechanistically informed interpretation of AD, linking underlying biological processes to observable cognitive and clinical phenotypes across levels of organization [4].

5.1 Genes and molecules

Genomic indices (including APOE ε4, SNP panels) and proteomic signatures of AD directly align with RDoC's molecular level of analysis. XAI methods based on SHAP or LIME can potentially identify APOE ε4 as a first-ranked feature, further substantiating its mechanistic significance in the disease conversion process. Meanwhile, recent proteomic studies demonstrate convergence with systemic metabolic pathways, even more closely linking AD risk to RDoC's Arousal/Regulatory Systems domain [8].

5.2 Cells and circuits

Based on the structural/functional MRI and PET findings, atrophy of the hippocampus and entorhinal cortex (medial temporal lobe-MTL), as well as the posterior cingulated cortex (PCC) hypometabolism and its impaired functional connectivity are amongst the key findings in AD. These model-specific neurocircuit traits correspond to RDoC's "Cognitive Systems" (declarative memory, executive function). Maps depicting saliency from XAI techniques can robustly illuminate MTL and PCC structures and bridge the black-box model outputs and biologically interpretable circuits [8].

5.3 Physiology

Electrophysiological and autonomic indices (qEEG slowing, HRV dysregulation, sleep disturbance) are evidence of abnormalities in RDoC's Arousal/Regulatory Systems. XAI when applied to EEG topographies can single out theta-beta imbalance as predictive of early AD. By establishing connections between physiology and higher-order behavior, these models move beyond static imaging and enable dynamic disease states to be included [68].

5.4 Behavior and self-report

From a cognitive science perspective, behavioral and ecological measures are not simply neutral readouts but operationalizations of specific cognitive processes. Digital drawing, speech, or gait measures are valuable because they can be collected ecologically, but they are also 'process-impure' as performance depends on a mixture of visuospatial skills, motor planning, attention, working memory, language, and motivational factors. Treating such tasks as pure indicators of a single RDoC construct (e.g., 'Cognitive Systems: memory') risks misinterpretation. A key advantage of the RDoC framework is that it encourages explicit decomposition of tasks into these underlying constructs (e.g., separating visuo-construction from planning and error monitoring within a drawing task). When XAI identifies a particular feature as predictive, mapping it to a specific RDoC cognitive construct helps to avoid task impurity and provides a more precise, mechanistic account of why behavior changes in early AD. This also creates a principled basis for selecting or designing new digital tasks whose interpretation is cognitively transparent, rather than simply maximizing predictive performance [8].

Cognitive assessments [Mini-Mental State Examination (MMSE), MoCA (Montreal Cognitive Assessment), and Repeatable battery for the Assessment of Neuropsychological Status (RBANS)] and ecological digital biomarkers (speech fluency, drawing kinematics, gait quality) represent observable behavior. Yamada et al., employed XAI for tablet-drawing data and found that hesitation time and stroke fluctuation are associated with memory

Table 2 Summary of key studies on explainable AI and biomarker-based approaches in Alzheimer's disease diagnosis and prognosis

Title-source reference	Methods used	Data used	Evaluation metrics	Results summary
Alzheimer's disease and type 2 diabetes mellitus: A systematic review of proteomic studies [58]	Systematic review of plasma/serum proteomic studies using LC-MS, MALDI-TOF, SOMAscan, iTRAQ, etc. MALDI: Matrix Assisted Laser Desorption Ionization	22 AD and 12 T2DM proteomic studies; 1,185 AD cases, 1678 controls; 834 T2DM cases, 2500 controls	QUADOMICS quality score; network biology with STRING; pathway enrichment ($p < 0.05$) QUADOMICS: Quality Assessment of Diagnostic Accuracy Studies for OMICS STRING: Search Tool for the Retrieval of Interacting Genes Proteins	Identified 17 common plasma proteomic biomarkers for AD and T2DM; high overlap in pathology
Is Alzheimer's disease a type 3 diabetes? A review [59]	Literature review based on PubMed database and meta-analyses	Epidemiological studies, meta-analyses, and biological pathways literature	Relative risk, odds ratio from studies (pooled RR 1.57); confidence intervals	T2DM increases dementia risk (RR 1.57); supports 'Type 3 Diabetes' hypothesis
Is Alzheimer's disease a Type 3 Diabetes? A critical appraisal [60]	Literature review and mechanistic analysis integrating molecular pathways, insulin resistance, and oxidative stress	Preclinical and clinical studies, molecular pathway analysis, insulin signaling, Aβ and tau pathways	Mechanistic comparison; no explicit evaluation metric - theoretical model building	AD and T2DM share insulin resistance, oxidative stress, mitochondrial dysfunction; insulin regulates Aβ and tau metabolism
Type 3 Diabetes and Its Role Implications in Alzheimer's Disease [61]	Extensive literature review of T3D and AD pathological mechanisms and insulin resistance pathways	Literature from clinical, molecular, and animal studies connecting T2DM, T3D, and AD pathology	Descriptive synthesis of molecular pathways and mechanistic links	Demonstrated bidirectional relationship of insulin resistance and Aβ pathology; Proposed T3D as a unifying concept for AD progression
A brain signature predictive of future progression to Alzheimer's dementia [23]	Machine Learning (SVM+Logistic Regression); Subtype clustering from ADNI dataset (sMRI+fMRI multimodal analysis)	ADNI2 dataset - sMRI, fMRI data of CN, MCI, and AD patients; 79 MCI cases	Precision (90%), Specificity (97%), Sensitivity (47%) for MCI to AD progression prediction; Cross-validated performance	Developed Highly Predictive Signature (HPS); 90% precision in identifying MCI likely to progress to AD within 3 years
Automated analysis of drawing process for detecting prodromal and clinical dementia [62]	Machine Learning (Logistic Regression, Random Forest, SVM) on digital drawing features from tablet tasks	145 Japanese older adults (46 CN, 67 MCI, 32 Dementia); Drawing tasks digitized via Wacom tablet	Nested Cross Validation, AUC, R2, accuracy, sensitivity, specificity, F1; permutation tests	75.1% accuracy multi-class (CN, MCI, dementia); AUC 0.909; MMSE $R2 = 0.491$; MTL atrophy $R2 = 0.293$ MTL: Medial Temporal Lobe
Biomarkers in Alzheimer's disease, role in early and differential diagnosis and recognition of atypical variants [63]	Comprehensive review of biomarkers (CSF, PET, plasma) for AD subtypes and atypical presentations	Literature review+clinical cohort data on typical and atypical AD subtypes	Narrative synthesis; highlights biomarker specificity and challenges in phenotype-based prediction	AD biomarkers essential for early diagnosis and detecting atypical variants (PCA, CBS); need for combinatorial biomarker use. CBS: Corticobasal Syndrome, PCA: Posterior Cortical Atrophy
Conceptual framework for the definition of preclinical and prodromal frontotemporal dementia [64]	Conceptual framework review for defining preclinical/prodromal Frontotemporal Dementia (FTD)	Review of genetic frontotemporal dementia (FTD) cohorts, longitudinal studies (GENFI, ARTFL/LEFFTDS), biomarkers (NfL, MRI, FDG-PET) ARTFL: Advancing Research and Treatment for Frontotemporal Lobar Degeneration GENFI: Genetic Frontotemporal Dementia Initiative LEFFTDS: Longitudinal Evaluation of Familial Frontotemporal Dementia Subjects	Conceptual model; no explicit metric; emphasizes need for biomarker standardization	Defined biological preclinical stage and prodromal FTD; emphasized lack of reliable early FTD biomarkers

Table 2 (continued)

Title-source reference	Methods used	Data used	Evaluation metrics	Results summary
Deep Learning in Alzheimer's Disease: Diagnostic classification and prognostic prediction [55]	Systematic review of 16 DL studies on AD classification/prediction using sMRI, PET, multimodal neuroimaging	ADNI dataset used in 16 DL studies; sMRI, PET, multimodal imaging datasets	Accuracy (96% AD), MCI conversion prediction (84.2%), evaluation across 16 reviewed studies	DL (CNN, SAE, RNN) achieved 96% AD classification, 84.2% MCI-to-AD prediction; multimodal imaging+ biomarkers improved accuracy
Duration of Preclinical, Prodromal, and Dementia Stages of Alzheimer's Disease in Relation to Age, Sex, and APOE Genotype [65]	Multistate modeling of disease stage durations (preclinical, prodromal, dementia) based on age, sex, APOE, CSF tau	Combined 6 cohorts ($n=3,268$) including ADNI, AIBL; APOE ε4, CSF tau biomarkers	Hazard ratios from multistate models, confidence intervals, mortality adjusted	AD total duration: 24 years at 60y, 15 years at 80y; APOE ε4, tau shorten stage duration
Grey matter network markers identify individuals with prodromal Alzheimer's disease who will show rapid clinical decline [56]	Grey matter network analysis + ROC+Logistic Regression using ADNI and Amsterdam Dementia Cohort (ADC) cohorts	Amsterdam Dementia Cohort ($n=244$)+ADNI ($n=247$); MRI, CSF p-tau, hippocampal volume	AUC, ROC, accuracy, specificity; model comparison with Akaike Information Criterion (AIC)	GM network metrics predict fast decliners (65% accuracy); combined model with p-tau, hippocampus (72% accuracy)
Machine Learning-Based Multimodal Neuroimaging Genomics Dementia Score for Predicting Future Conversion to Alzheimer's Disease [57]	Machine Learning ensemble with feature selection on multimodal (MRI+genomics) data; stratified risk modeling	ADNI1 dataset ($n=543$); 521,017 SNPs + MRI brain features; APOE genotype	Accuracy (86% genetics), stratified evaluation, ensemble learning performance	Genetics better predicted Dementia of Alzheimer's Type (DAT) in normals (86% accuracy), MRI better in sMCI; multimodal improved overall predictions
Prevalence and Risk of Progression of Preclinical Alzheimer's Disease Stages: A systematic review and meta-analysis [66]	Systematic review and meta-analysis of 36 studies on preclinical AD prevalence and progression risk	36 studies ($n=6602$); PET, CSF biomarkers, cognitive stages, SCD, MCI progression risk SCD: Subjective Cognitive Decline	Meta-analysis statistics, pooled RR, prevalence rates, I2 heterogeneity	Prevalence of preclinical AD 22%; Stage 3 preclinical AD had 73% progression risk, equal to MCI risk
Grey matter network markers identify individuals with prodromal Alzheimer's disease who will show rapid clinical decline [56]	Grey matter network analysis, Logistic Regression, ROC using ADNI and Amsterdam	ADNI+Amsterdam Dementia Cohort; MRI, CSF p-tau, hippocampus volume	AUC, ROC, accuracy, AIC model comparison	72% accuracy combining GM network + p-tau; strong predictors of rapid progression

Table 3 Diagnostic framework across preclinical, prodromal and clinical dementia stages

Stage	Diagnostic methodologies	Common biomarkers	Clinical notes
Preclinical	PET, CSF, genetic screening	Aβ, APOE ε4	Asymptomatic, risk profiling
Prodromal (MCI)	MRI, MMSE, cognitive tasks	Tau, Aβ42, hippocampal atrophy	Mild symptoms, unclear progression
Clinical dementia	Neuropsych tests, FDG-PET, plasma tests	Aβ, tau, NfL	Irreversible impairment, confirmed diagnosis

and executive functions [62]. Within RDoC, such behaviors provide a bridge from circuit-level pathology to daily functional impairment.

5.5 Reports and paradigms (Cognitive Tasks)

Experimental tasks (n-back working memory, emotional face recognition, Stroop test) present structured paradigms which are quantifiable in terms of specific RDoC domains. Combined with XAI frameworks, performance metrics can be related to circuits and molecular biomarkers which deliver interpretable cross-level predictions.

Crucially, many of these paradigms have a long history in experimental cognitive psychology and neuropsychology, and their sensitivity to specific cognitive processes such as interference control, updating, or sustained attention is well documented. Utilizing this prior knowledge within the RDoC-XAI framework allows performance in these tasks to be interpreted not only as "better" or "worse", but as reflecting selective impairment in particular computations (e.g., proactive interference resolution vs. maintenance, set-shifting vs. response inhibition). As a result, when XAI identifies n-back performance as a key feature for example, clinicians can understand whether the model is effectively tracking working memory load, speed-accuracy trade-offs, or executive control under distraction. Embedding such cognitively informed task decompositions into the mapping between features to RDoC domains strengthens construct validity and reduces the risk that highly predictive tasks are misinterpreted by clinicians.

6 Explainable AI: interpretability and clinical integration

Based on the previous section's conceptual basis, this section proposes a clinical and operational link through XAI and within the RDoC framework. An RDoC-XAI framework for Alzheimer diagnostics is summarized in Table 4. Interpretable AI outputs are demonstrated to contribute to better clinician decision-making, promote clinician trust, and improve communication with patients and caregivers. Particular emphasis has been assigned to operational issues, such as workflow integration and ethical and regulatory considerations, and on the reliability and generalizability of XAI methods for real-world clinical settings (Table 4).

The essential approach of XAI provides interpretability and feature interaction insights that traditional black-box models, such as deep learning models, cannot readily offer. The application of XAI in clinical research provides explanations about model output features which leads to improved clinician trust and enables ethical decisions based on informed choices. Common XAI methods such as SHAP and LIME analyze models to identify important features to maintain thorough insights. The approaches show particular value in AD diagnostic applications, as demonstrated by Yamada et al. who established relationships between tablet-drawing digital biomarkers including stroke pressure and hesitation and cognitive performance [62].

Further, the research by Dansereau et al. used subtype clustering and SVM to develop neurofunctional biomarkers which provided biological explanations for model prediction results. XAI methods meet the ethical standards of data privacy regulations including the Europian Union's General Data Protection Regulation (EU's GDPR) because they ensure explainability in algorithmic decision-making. More essentially, XAI serves as an essential requirement for both regulatory compliance and responsible AI deployment in medical settings beyond its role in model transparency enhancement [23].

A further level of interpretability arises when model explanations are explicitly cast in terms of cognitive constructs rather than only in terms of low-level features or neuroanatomical regions. Clinicians and neuropsychologists reason about patients using concepts such as episodic memory, semantic fluency, divided attention, or executive dysfunction. When XAI outputs are combined within RDoC domains and linked to specific paradigms, they can be translated into this cognitive language (for example, "reduced Cognitive Systems, memory domain score driven by delayed recall errors and medial temporal atrophy").

Such cognitively framed explanations align better with clinicians' existing mental models of AD and support greater trust as they allow clinicians to recognize when the model is making a plausible claim about a known pattern (e.g., memory-related AD) compared to when it suggests a more unusual cognitive pattern that may trigger the need for caution and further investigation. In addition, explanations at the cognitive level can be communicated to patients and caregivers more readily than raw biomarker values, thereby supporting shared and better-informed decision-making.

Figure 3 demonstrates a flowchart whereby the interface of RDoC and XAI can be further explored.

To operationalize the framework in Fig. 3; Table 5 shows how feature-level attributions from our models are normalized and aggregated into RDoC domains. For each unit we list example AD measures, typical model families, XAI artifacts, and an explicit aggregation rule that maps features to domains via a binary matrix.

An overarching challenge in dementia-related research is to bridge biological heterogeneity to clinical presentation. Conventional approaches, such as A/T(N), provide important biomarker stages but are insufficient for bridging molecular change and behavior. This void can be filled by RDoC, which provides a dimensional framework that extends from genes to paradigms and XAI serves as the computational means to link these levels [69].

Within this combined framework, cognitive science plays a central mediating role. RDoC domains such as Cognitive Systems and Social Processes already encode theoretical commitments about how memory, attention, language, and social cognition are organized, and how these systems interact with arousal and motivation.

Table 4 The RDoC-XAI framework for Alzheimer diagnostics. The schema describes the mapping of RDoC units of analysis (left), to clinical domains (middle), via XAI outputs (right). Arrows indicate the bridging function of XAI to connect biological variety with clinical phenotype and accordingly build a mechanistic, interpretable dementia diagnostics framework

RDoC unit of analysis	Clinical domain	XAI mapping /example output
Genes/Molecules	Genetic/Risk domain	APOE ε4, proteomics: SHAP feature ranking
Cells/Circuits	Cognitive systems	MRI hippocampal atrophy, PET hypometabolism: Saliency maps
Physiology	Arousal/Regulatory systems	qEEG slowing, HRV reduction: Attribution analysis
Behavior	Cognitive+Social processes	Tablet-drawing, speech, cognitive scores: SHAP/LIME outputs
Self-report	Negative/Positive valence	Patient/caregivers report (anxiety, sleep disturbance): Explainable classifiers
Paradigms	Cognitive control/Executive function	Task performance (n-back, Stroop): XAI paradigm analysis

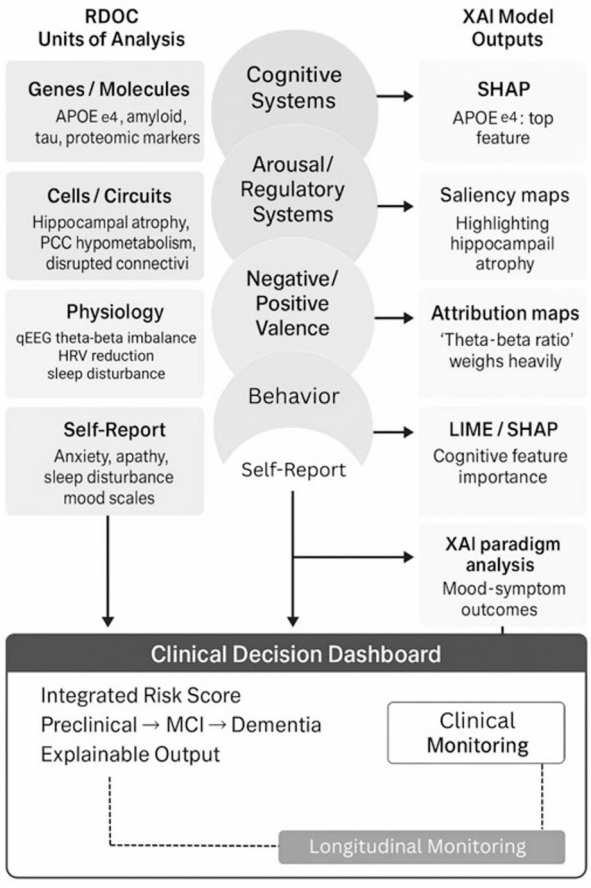


Fig. 3 A unified RDoC-XAI architecture for interpretable precision diagnostics on Alzheimer's disease. The figure shows how multi-level data (from molecular to circuit level biomarkers, physiological and behavioral assessments), are mapped onto computational transparent models (i.e. SHAP, LIME, saliency maps, and attribution maps). These outputs in turn inform a clinical decision dashboard that produces explainable risk scores and facilitates longitudinal tracking across preclinical-dementia stages. This integrative pipeline reflects the future trend of Alzheimer's study, toward explainable and multimodal personalized diagnostic systems

By aligning XAI-derived feature attributions with these domains and with specific paradigms, the proposed architecture moves beyond descriptive biomarker profiling toward mechanistic accounts of how particular circuit-level and molecular abnormalities give rise to observable cognitive attributes in AD. This is particularly important as AD is a disorder with a range of different presentations (e.g., dominated by memory, language or visuospatial deficits) and their combinations, and where categorization by cognitive profile may therefore be as clinically valuable as that by biomarker status.

Future empirical work should therefore not only test the predictive performance of RDoC-informed XAI methods, but also their ability to recover meaningful cognitive categories and measurement trajectories. Longitudinal studies combining neuroimaging, fluid biomarkers, and repeated cognitive assessments could examine whether changes in RDoC domain scores derived from XAI correspond to theoretically expected patterns of decline, and whether these trajectories improve the accuracy of prognoses compared to biomarker-only models. In parallel, evaluations including clinicians, neuropsychologists, and patients will be needed to assess whether cognitive-level explanations enhance diagnostic decisions, levels of confidence, and acceptance of AI-assisted tools in real clinical settings. These cognitive-science-informed validation strategies are crucial if RDoC–XAI frameworks are to move from conceptual description to robust, valuable clinical implementation.

For instance, if an interpretable model emphasizes posterior cingulate hypometabolism, this will load onto RDoC's Cognitive Systems domain (memory retrieval), elucidating how it relates to performance deficits on episodic recall tests. Similarly, in the SHAP analysis if SHAP values suggest less HRV to be predictive this aligns with the Arousal/Regulatory Systems domain and another pathway from autonomic physiology to patient-reported sleep [26, 62, 69, 70].

Through this tactical diagnostic scaffold, the RDoC-XAI matrix, we can expect multi-level outputs in relation to a language that clinicians already use to think about psychopathology and neurocognitive systems. This builds trust, decreases epistemic opacity and guarantees that model outputs are mechanistically based.

While the A/T(N) framework has provided a biologically anchored taxonomy of Alzheimer's disease, clinicians often struggle with the heterogeneity of clinical presentations. XAI frameworks serve as a bridge here, enabling not just prediction but the translation of biomarker constellations into clinically digestible patterns. This shift from black-box analytics to interpretable outputs is pivotal in ensuring trust and wider adoption in clinical workflows [6].

A further refinement in the current landscape is the use of hybrid models that combine symbolic reasoning with deep learning. Symbolic AI provides rule-based transparency, while neural networks excel in pattern recognition across high-dimensional data. When fused, these architectures not only enhance predictive performance but also allow clinicians to interrogate decision paths. For example, multimodal studies fusing PET, MRI, and genomics under interpretable layers have highlighted hippocampal atrophy and APOE ε4 status as synergistic predictors of disease conversion [71].

Another critical frontier concerns digital biomarkers and ecological assessments, such as speech patterns, gait dynamics, or digital drawing tests. Unlike traditional imaging and CSF measures, these are scalable, inexpensive, and more readily deployable in primary care settings. Recent evidence shows that integrating tablet-based drawing features with XAI methods achieved high

Table 5 RDoC-XAI crosswalk for AD: units, exemplar measures, model families, XAI artifacts, and the matrix M used to aggregate feature-level attributions into RDoC domain scores

RDoC unit of analysis	Exemplar AD measures	Typical model families	XAI artifacts (granularity)	Aggregation to RDoC domains	Validity checks and clinician readout
Genes and molecules	APOE genotype; polygenic risk; CSF/plasma Aβ, t/p-tau, NfL	Logistic/LASSO/ENet; GBT	SHAP or Permutation Importance (feature-level); coefficient paths	Map molecular features primarily to Arousal/Regulatory Systems (homeostasis) and Cognitive Systems-Memory via matrix M; normalize attributions before summation	Faithfulness (infidelity/feature ablation); stability (bootstrap top-k); readout: "Biomarker-weighted risk contribution"
Cells and circuits	Structural connectomes; rs-fMRI functional connectivity; EEG/MEG source power	GNN; CNN/ViT on connectivity images; SVM on graph metrics	Grad-CAM/occlusion (region/edge-level); SHAP on graph statistics	Assign ROIs/edges to Cognitive Systems subdomains (Memory, Cognitive Control); average or weighted sum by atlas parcels	Sanity checks (weight randomization); ROI plausibility audit with experts; readout: "Networks most driving the decision"
Physiology	Sleep PSG (spindles, arousals); HRV; gait variability; speech prosody	GBT; temporal CNN/Transformer; HMM	SHAP/PI (feature), temporal occlusion (segment)	Route autonomic/sleep indices to Arousal/Regulatory; map speech/gait motor indices to Sensorimotor Systems (motor action/habit)	Counterfactual perturbations (time-warp, jitter); test–retest stability; readout: "Physiological subsystems contributing"
Behavior	Digital pen/clock-drawing kinematics; reaction times; error types	Random forest/GBT; temporal CNN	TreeSHAP (feature); sequence saliency	Map error/latency features to Cognitive Systems-Working Memory/Cognitive Control and to Sensorimotor Systems when motor execution dominates	Faithfulness via feature removal; subsample Jaccard stability; readout: "Top behavioral factors and direction"
Self-report	ADL/IADL, mood/cognition items	Ordinal logistic; GBT	SHAP (item-level)	Group items to Negative/Positive Valence, Cognitive Systems, or Social Processes as appropriate; aggregate via M	Differential item functioning checks; readout: "Questionnaire items influencing risk"
Paradigms (tasks)	n-back, Stroop, Trail-Making (latencies, errors)	Mixed-effects GLM; GBT; temporal CNN	SHAP/PI (condition-level; occlusion of time bins	Map conditions to Cognitive Systems-Working Memory/Cognitive Control; aggregate over trial types	Manipulation checks (load/congruency); readout: "Task conditions most informative"

diagnostic accuracy, while retaining interpretability for clinical validation [62]. This indicates that digital phenotyping could complement conventional biomarkers, particularly in under-resourced environments.

Equally important is the longitudinal perspective. Many AI models still operate cross-sectionally, predicting conversion based on a snapshot of data. However, dementia is inherently progressive, with neurodegeneration unfolding over years or decades. Longitudinal XAI approaches that track changes in imaging, plasma, and cognitive measures are emerging as powerful tools to capture trajectories of decline. Such methods align better with clinical monitoring needs, allowing for personalized forecasts and proactive interventions [71].

From a methodological standpoint, issues of reproducibility and generalizability remain pressing. Model performance is often inflated when trained and validated on the same dataset, as seen in ADNI-centric studies. Moreover, explainability tools like SHAP and LIME must themselves be evaluated for stability, as attribution maps may vary with dataset shifts. Establishing consensus metrics for XAI reliability is therefore an urgent need in the field [72].

Ethical and regulatory considerations also come to the fore. Healthcare AI systems are increasingly scrutinized for bias, fairness, and compliance with frameworks such as GDPR and FDA regulations. Interpretable models can alleviate some of these concerns by making decision pathways auditable. However, ethical dilemmas persist, for instance, in communicating probabilistic risks of conversion to patients who may face stigma or anxiety. Embedding XAI outputs within shared decision-making models could enhance both transparency and compassion in clinical encounters [73].

Looking ahead, the integration of multi-omics data, including proteomics, metabolomics, and transcriptomics, into XAI methods may yield deeper insights into disease mechanisms. Early evidence suggests that plasma proteomic markers overlapping with type 2 diabetes pathways share significant predictive power for dementia risk. Such integrative approaches may redefine dementia not merely as a neurodegenerative entity but as a systemic, multi-organ disorder, requiring interdisciplinary management [74].

Finally, there is a need to consider the implementation science dimension: how these models can be translated from research labs to clinics. Human-in-the-loop designs, clinician training on XAI tools, and iterative feedback loops will be essential. Pilot trials embedding XAI-driven decision support in memory clinics could provide evidence for real-world feasibility and impact. Without careful translation strategies, even the most sophisticated models risk remaining academic curiosities rather than transforming dementia care [75].

6.1 Challenges and limitations: reliability of XAI methods in clinical contexts

With their increasing application, XAI approaches have limitations important to be noted when being used in clinical practice. An important challenge is the instability of attribution methods where a feature importance rank may vary across model runs, data splits, or preprocessing pipelines, limiting reproducibility. Moreover, common visualization techniques, such as saliency maps, are prone to false positives and may also identify regions that are not causally related to the model's prediction [72].

The emergence of suppressor variables and complex feature interactions can further muddy interpretation. The influence of individual predictors may rely on the wider feature space in non-intuitive manners. Most XAI techniques will give correlational explanations, rather than causal explanations, and so these can limit their mechanistic inference in clinical neuroscience. In addition, many of such models are trained and validated on specific datasets. ADNI, for instance, is subject to potential concerns about the generalizability and dataset dependency if applied to heterogeneous real-world populations. Meeting these challenges necessitates common evaluation standards for XAI reliability; integrated robustness and stability measures and more focus on evaluating externally from different cohorts that support appropriate clinically meaningful and trustworthy utilization of the technology [73, 75].

7 Conclusion and future directions

Taken together, the future development of Alzheimer's diagnostics will emerge from the combination of multiomics data integration with AI and human interpretability. The growing use of diverse modalities in ML algorithms including neuroimaging and genomics and digital behavioral assessments makes explainable systems increasingly necessary. The requirement for explainability stands as an absolute necessity for deploying systems in clinical settings. Models need to perform predictions while simultaneously providing explanations and justifications and integrating with clinical operational processes. The combination of deep learning pattern recognition with symbolic reasoning or rule-based logic through hybrid AI architectures shows great potential for developing robust clinical solutions. The advancement of XAI in neurodegenerative disease research will enable both new biomarker discoveries and better clinician decision support and enhanced patient communication. Human-in-the-loop systems that enable experts to validate and modify predictions through interactive processes will be essential for achieving real-world utility. Machine learning tools for Alzheimer's diagnostics require Explainable AI integration to reach their maximum clinical value.

In future, incorporation of RDoC into XAI-informed AD diagnostic processes offers many research and clinical fronts:

1. Cross-level dashboards: For instance, patient-specific outputs could project predictions through RDoC units (e.g., APOE ε4 into hippocampal atrophy, into episodic memory impairment and then slowed Stroop).
2. Longitudinal modeling: By following people over time across RDoC domains, longitudinal disease trajectories could be linked to interpretable XAI features to enhance personalized prediction.
3. Implementation science practitioners can be taught to understand XAI outputs in terms of RDoC-related constructs, thereby facilitating uptake and communication with patients.
4. Ethical implications: In terms of translation, the contextualization onto RDoC lessens the risk of dehumanizing predictive outputs by situating them in observable human functions (e.g., memory, emotion arousal).

In sum, by nesting RDoC within XAI, Alzheimer's diagnostics moves from a piecewise biomarker activity into a holistic, dimensional and person-centered platform. This is a strategic decision to maximize explainability, clinical validity and translational readiness.

Abbreviations List

AA	Alzheimer's association
AD	Alzheimer's disease
ADC	Amsterdam dementia cohort
ADNI/ADNI1/ADNI2	Alzheimer's disease neuroimaging initiative (Phase 1/ Phase 2)
AI	Artificial intelligence
AIC	Akaike information criterion
AMURA	Accelerated microstructure using robust acquisition
APOE	Apolipoprotein E
AQP4	Aquaporin 4
ARTFL	Advancing research and treatment for frontotemporal lobar degeneration
AT	Amyloid tau biomarker classification
AUC	Area under the curve
Aβ	Amyloid beta
CAM	Class activation map
CBS	Corticobasal syndrome
CI	Cognitive impairment
CN	Cognitively normal
CNN	Convolutional neural network
CSF	Cerebrospinal fluid
CT	Computed tomography
DAT	Dementia of the alzheimer type
DL	Deep learning
EEG	Electroencephalography
EU	European Union
FDA	U.S. food and drug administration
FDG	Fluorodeoxyglucose
fMRI	Functional magnetic resonance imaging
FTD	Frontotemporal dementia
GDPR	General data protection regulation
GENFI	Genetic frontotemporal dementia initiative

GM	Grey matter
HPS	Highly predictive signature
HRV	Heart rate variability
LC	Liquid chromatography
LEFFTDS	Longitudinal evaluation of familial frontotemporal dementia subjects
LIME	Local interpretable model agnostic explanations
MALDI	Matrix assisted laser desorption ionization
MCI	Mild cognitive impairment
ML	Machine learning
MMSE	Mini mental state examination
MoCA	Montreal cognitive assessment
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MTL	Medial temporal lobe
NIA	National institute on aging
NIMH	National institute of mental health
OGM	Organ geometry map
p tau	Phosphorylated tau
PCA	Posterior cortical atrophy
PCC	Posterior cingulate cortex
PET	Positron emission tomography
PiB	Pittsburgh compound B
PMLR	Proceedings of machine learning research
PRISMA	Preferred reporting items for systematic reviews and meta analyses
PubMed	Public publisher MEDLINE database
qEEG	Quantitative electroencephalography
QUADOMICS	Quality assessment of diagnostic accuracy studies for OMICS
RBANS	Repeatable battery for the assessment of neuropsychological status
RDoC	Research domain criteria
RNN	Recurrent neural network
ROC	Receiver operating characteristic
RR	Relative risk
SAE	Stacked autoencoder
SCD	Subjective cognitive decline
SHAP	SHapley additive explanations
sMRI	Structural magnetic resonance imaging
SNP	Single nucleotide polymorphism
STRING	Search Tool for the retrieval of interacting genes proteins
STROBE	Strengthening the reporting of observational studies in epidemiology
SVM	Support vector machine
TG	Trigeminal ganglion
XAI	Explainable artificial intelligence

Acknowledgements
The authors would like to thank the Dubai Future Foundation for its financial support, which made this research possible. The authors also extend their gratitude to the reviewers and editors for their valuable time and thoughtful feedback on this manuscript.

Author contributions
MN (conceptualization, methodology, investigation, writing the original draft), DP (investigation, writing the original draft), FT (writing, review and editing, project administration), FK (writing, review and editing, project administration).

Funding
This research was funded by the Dubai Future Foundation (2024CANAD-KAM-060).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Competing interests
The authors declare no competing interests.

Received: 28 January 2026 / Accepted: 24 May 2026
Published online: 19 June 2026

References

1. Zheng Q, Wang X (2025) Alzheimer's disease: insights into pathology, molecular mechanisms, and therapy. Protein Cell 16:83–120
2. Peretti DE, Boccalini C, Ribaldi F, Scheffler M, Marizzoni M, Ashton NJ, Zetterberg H, Blennow K, Frisoni GB, Garibotto V (2024) Association of glial fibrillary acid protein, Alzheimer's disease pathology and cognitive decline. Brain 147:4094–4104
3. Cuthbert BN (2020) The role of RDoC in future classification of mental disorders. Dialog Clin Neurosci 22:81–85
4. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, Wang P (2010) Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 167:748–751
5. Kwon KJ, Kim HY, Han S-H, Shin CY (2024) Future therapeutic strategies for Alzheimer's disease: focus on behavioral and psychological symptoms. Int J Mol Sci 25:11338
6. London AJ (2019) Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep 49:15–21
7. Quinn TP, Jacobs S, Senadeera M, Le V, Coghlan S (2022) The three ghosts of medical AI: Can the black-box present deliver? Artif Intell Med 124:102158
8. Jiang J, Li C, Lu J, Sun J, Sun X, Yang J, Wang L, Zuo C, Shi K, Initiative A (2025) s.D.N. Using interpretable deep learning radiomics model to diagnose and predict progression of early AD disease spectrum: a preliminary [18F] FDG PET study. Eur Radiol 35:2620–2633
9. Rafii MS, Aisen PS (2023) Detection and treatment of Alzheimer's disease in its preclinical stage. Nat aging 3:520–531
10. Shamrat FJM, Akter S, Azam S, Karim A, Ghosh P, Tasnim Z, Hasib KM, De Boer F, Ahmed K (2023) AlzheimerNet: An effective deep learning based proposition for alzheimer's disease stages classification from functional brain changes in magnetic resonance images. IEEE Access 11:16376–16395
11. Li J-Q, Song J-H, Suckling J, Wang Y-J, Zuo C-T, Zhang C, Song Y-Q, Xie A-M, Tan L, Yu J-T (2023) Disease trajectories in elders with suspected non-Alzheimer's pathophysiology and its comparison with Alzheimer's disease pathophysiology: a longitudinal study. Res Square. rs. 3. rs–2744271
12. Oberstein TJ, Schmidt MA, Florvaag A, Haas A-L, Siegmann E-M, Olm P, Utz J, Spitzer P, Doerfler A, Lewczuk P (2022) Amyloid-β levels and cognitive trajectories in non-demented pTau181-positive subjects without amyloidopathy. Brain 145:4032–4041
13. Zhou RZ, Wimo A, Winblad B (2024) On the 2024 Alzheimer's association criteria: still not ready for clinical use. J Prev Alzheimer's Disease 11:895–896
14. Soladoye AA, Aderinto N, Osho D, Olawade DB (2025) Explainable machine learning models for early Alzheimer's disease detection using multimodal clinical data. Int J Med Informatics 106093
15. Wang X, Ye T, Jiang D, Zhou W, Zhang J, Initiative A (2024) s.D.N. Characterizing the clinical heterogeneity of early symptomatic Alzheimer's disease: A data-driven machine learning approach. Front Aging Neurosci 16:1410544
16. Savaş S (2022) Detecting the stages of Alzheimer's disease with pre-trained deep learning architectures. Arab J Sci Eng 47:2201–2218
17. Rabinovici GD, Knopman DS, Arbizu J, Benzinger TL, Donohoe KJ, Hansson O, Herscovitch P, Kuo PH, Lingler JH, Minoshima S (2025) Updated appropriate use criteria for amyloid and tau PET: a report from the Alzheimer's association and society for nuclear medicine and molecular imaging workgroup. J Nucl Med 66:S5–S31
18. Zhao Q, Du X, Chen W, Zhang T, Xu Z (2023) Advances in diagnosing mild cognitive impairment and Alzheimer's disease using 11 C-PIB-PET/CT and common neuropsychological tests. Front NeuroSci 17:1216215
19. Leon R, Ghahremani M, Guan DX, Smith EE, Zetterberg H, Ismail Z (2025) Enhancing alzheimer disease detection using neuropsychiatric symptoms: The role of mild behavioural impairment in the revised NIA-AA research framework. J Geriatr Psychiatr Neurol 08919887251366634
20. Bieger A, Brum WS, Borelli WV, Therriault J, De Bastiani MA, Moreira AG, Benedet AL, Ferrari-Souza JP, Da Costa JC, Souza DO (2024) Influence of

different diagnostic criteria on Alzheimer disease clinical research. Neurology 103:e209753
21. Eikelboom WS, van den Berg E, Singleton EH, Baart SJ, Coesmans M, Leeuwis AE, Teunissen CE, van Berckel BN, Pijnenburg YA, Scheltens P (2021) Neuropsychiatric and cognitive symptoms across the Alzheimer disease clinical spectrum: cross-sectional and longitudinal associations. Neurology 97:e1276–e1287
22. Shah J, Rahman Siddiquee MM, Krell-Roesch J, Syrjanen JA, Kremers WK, Vassilaki M, Forzani E, Wu T, Geda YE (2023) Neuropsychiatric symptoms and commonly used biomarkers of Alzheimer's disease: A literature review from a machine learning perspective. J Alzheimer's Disease 92:1131–1146
23. Dansereau C, Tam A, Badhwar A, Urchs S, Orban P, Rosa-Neto P, Bellec P (2017) A brain signature highly predictive of future progression to Alzheimer's dementia. arXiv preprint arXiv:1712.08058
24. Tam A, Dansereau C, Iturria-Medina Y, Urchs S, Orban P, Sharmarke H, Breitner J, Bellec P, Initiative A (2019) s.D.N. A highly predictive signature of cognition and brain atrophy for progression to Alzheimer's dementia. Gigascience 8, giz055
25. Saeed A, Waris A, Fuwad A, Iqbal J, Khan J, AlQahtani D, Gilani O, Shah UH, Initiative A (2024) s.D.N. Random survival forest model for early prediction of Alzheimer's disease conversion in early and late Mild cognitive impairment stages. PLoS ONE 19:e0314725
26. Yamada Y, Shinakwa K, Kobayashi M, Nemoto M, Ota M, Nemoto K, Arai T (2023) Smartwatch-derived Acoustic Markers for Deficits in Cognitively Relevant Everyday Functioning. In Proceedings of the IEEE International Conference on Digital Health (ICDH), 2023; pp. 39–49
27. Bibi N, Courtney J, McGuinness K (2025) Enhancing brain disease diagnosis with XAI: a review of recent studies. ACM Trans Comput Healthc 6:1–35
28. Kim S, Lee P, Oh KT, Byun MS, Yi D, Lee JH, Kim YK, Ye BS, Yun MJ, Lee DY (2021) Deep learning-based amyloid PET positivity classification model in the Alzheimer's disease continuum by using 2-[18F] FDG PET. EJNMMI Res 11:56
29. Suh C, Shim W, Kim S, Roh J, Lee J-H, Kim M-J, Park S, Jung W, Sung J, Jahng G-H (2020) Development and validation of a deep learning–based automatic brain segmentation and classification algorithm for Alzheimer disease using 3D T1-weighted volumetric images. Am J Neuroradiol 41:2227–2234
30. Vimbi V, Shaffi N, Mahmud M (2024) Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. Brain Inf 11:10
31. Taiyeb Khosroshahi M, Morsali S, Gharakhanlou S, Motamedi A, Hassanbaghlou S, Vahedi H, Pedrammehr S, Kabir HMD, Jafarizadeh A (2025) Explainable artificial intelligence in neuroimaging of Alzheimer's disease. Diagnostics 15:612
32. Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1:206–215
33. Geddes MR, Mattfeld AT, de Los Angeles C, Keshavan A, Gabrieli JD (2018) Human aging reduces the neurobehavioral influence of motivation on episodic memory. NeuroImage 171:296–310
34. Better M (2023) Alzheimer's disease facts and figures. Alzheimers Dement 19(4):1598–1695
35. Coluzzi D, Bordin V, Rivolta MW, Fortel I, Zhan L, Leow A, Baselli G (2025) Biomarker investigation using multiple brain measures from MRI through explainable artificial intelligence in Alzheimer's disease classification. Bioengineering 12:82
36. Nigri E, Ziviani N, Cappabianco F, Antunes A, Veloso A (2020) Explainable deep CNNs for MRI-based diagnosis of Alzheimer's disease. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), ; pp. 1–8
37. Lozupone G, Bria A, Fontanella F, Meijer FJ, De Stefano CAXAIL (2024) Attention-based eXplainability for interpretable Alzheimer's localized diagnosis using 2D CNNs on 3D MRI brain scans. arXiv preprint arXiv:2407.02418
38. Anjomshoae S, Pudas S (2024) Explaining graph convolutional network predictions for clinicians: an explainable AI approach to Alzheimer's disease classification. *Frontiers in Artificial Intelligence 6*
39. Amoroso N, Quarto S, La Rocca M, Tangaro S, Monaco A, Bellotti R (2023) An eXplainability Artificial Intelligence approach to brain connectivity in Alzheimer's disease. Front Aging Neurosci 15:1238065
40. Galazzo IB, Cruciani F, Brusini L, Salih A, Radeva P, Storti SF, Menegaz G (2022) Explainable artificial intelligence for magnetic resonance imaging aging brainprints: Grounds and challenges. IEEE Signal Process Mag 39:99–116
41. Bapat R, Ma D, Duong TQ (2024) Predicting four-year's Alzheimer's disease onset using longitudinal neurocognitive tests and MRI data using explainable deep convolutional neural networks. J Alzheimer's Disease 97:459–469
42. Dolci G, Cruciani F, Rahaman MA, Abrol A, Chen J, Fu Z, Galazzo IB, Menegaz G, Calhoun VD, Initiative A (2025) s.D.N. An interpretable generative multimodal neuroimaging-genomics framework for decoding Alzheimer's disease. ArXiv arXiv: 2406.13292 v13293.
43. Lombardi A, Diacono D, Amoroso N, Biecek P, Monaco A, Bellantuono L, Pantaleo E, Logroscino G, De Blasi R, Tangaro S (2022) A robust framework to investigate the reliability and stability of explainable artificial intelligence markers of mild cognitive impairment and Alzheimer's disease. Brain Inf 9:17
44. El-Sappagh S, Alonso JM, Islam SR, Sultan AM, Kwak KS (2021) A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. Sci Rep 11:2660
45. Svenberg H, Explainable (2025) AI for Alzheimer's Prediction Diagnosis: Developing an Interpretable Deep Learning Pipeline with 3D Visualization
46. Odusami M, Maskeliūnas R, Damaševičius R, Misra S (2023) Explainable deep-learning-based diagnosis of Alzheimer's disease using multimodal input fusion of PET and MRI Images. J Med Biol Eng 43:291–302
47. Mahmud T, Barua K, Habiba SU, Sharmen N, Hossain MS, Andersson K (2024) An explainable ai paradigm for alzheimer's diagnosis using deep transfer learning. Diagnostics 14:345
48. Junior KJ, Carole KS, Theodore Armand TP, Kim H-C, Initiative A (2024) s.D.N. Alzheimer's multiclassification using explainable AI techniques. Appl Sci 14:8287
49. De Santi LA, Pasini E, Santarelli MF, Genovesi D, Positano V (2023) An explainable convolutional neural network for the early diagnosis of Alzheimer's disease from 18F-FDG PET. J Digit Imaging 36:189–203
50. Rasi R, Guvenis A, Initiative A.s.D.N. Subregional Biomarkers in FDG PET for Alzheimer's Diagnosis and Staging: An Interpretable and Explainable model. *bioRxiv* 2024, 2024.2012. 2027.630495
51. Weatheritt J, Palombit A, Manber R, Wolz R (2021) Alzheimer's disease detection using explainable AI on PET images. Alzheimer's Dement 17:e053831
52. Beer S, Elmenhorst D, Bischof GN, Ramirez A, Bauer A, Drzezga A, Initiative A (2024) s.D.N. Explainable artificial intelligence identifies an AQP4 polymorphism-based risk score associated with brain amyloid burden. Neurobiol Aging 143:19–29
53. Böhle M, Eitel F, Weygandt M, Ritter K (2019) Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification. Front Aging Neurosci 11:194
54. Li C, Jiao F, Wu S, Wang C, Wei M, Zhang S, Wang L, Huang Y, Yin Y, Tian R (2025) Enhancing interpretability of AI with radiomics-based deep neural network: proof of concept in the classification of Parkinsonian syndromes with 18F-FDG PET imaging. Eur J Nucl Med Mol Imaging 1–18
55. Jo T, Nho K, Saykin AJ (2019) Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Front Aging Neurosci 11:220
56. Pelkmans W, Vromen EM, Dicks E, Scheltens P, Teunissen CE, Barkhof F, van der Flier WM, Tijms BM, Initiative A (2022) s.D.N. Grey matter network markers identify individuals with prodromal Alzheimer's disease who will show rapid clinical decline. Brain Commun 4:fcac026
57. Mirabnahrazam G, Ma D, Lee S, Popuri K, Lee H, Cao J, Wang L, Galvin JE, Beg MF, Initiative A (2022) s.D.N. Machine learning based multimodal neuroimaging genomics dementia score for predicting future conversion to alzheimer's disease. J Alzheimer's Disease 87:1345–1365
58. Diniz Pereira J, Gomes Fraga V, Morais Santos AL, Carvalho MdG, Caramelli P, Braga Gomes K (2021) Alzheimer's disease and type 2 diabetes mellitus: A systematic review of proteomic studies. J Neurochem 156:753–776
59. Janoutová J, Machaczka O, Zatloukalová A, Janout V (2022) Is Alzheimer's disease a type 3 diabetes? A review. Cent Eur J Public Health 30:139–143
60. Kandimalla R, Thirumala V, Reddy PH (2017) Is Alzheimer's disease a type 3 diabetes? A critical appraisal. Biochim et Biophys Acta (BBA)-Molecular Basis Disease 1863:1078–1089
61. Nguyen TT, Ta QTH, Nguyen TKO, Nguyen TTD, Van Giau V (2020) Type 3 diabetes and its role implications in Alzheimer's disease. Int J Mol Sci 21:3165
62. Yamada Y, Kobayashi M, Shinkawa K, Nemoto M, Ota M, Nemoto K, Arai T (2022) Automated analysis of drawing process for detecting prodromal and clinical dementia. In Proceedings of the IEEE International Conference on Digital Health (ICDH), 2022; pp. 1–6
63. Dubois B, von Arnim CA, Burnie N, Bozeat S, Cummings J (2023) Biomarkers in Alzheimer's disease: role in early and differential diagnosis and recognition of atypical variants. Alzheimers Res Ther 15:175

64. Benussi A, Alberici A, Samra K, Russell LL, Greaves CV, Bocchetta M, Ducharme S, Finger E, Fumagalli G, Galimberti D (2022) Conceptual framework for the definition of preclinical and prodromal frontotemporal dementia. Alzheimer's Dement 18:1408–1423
65. Vermunt L, Sikkes SA, Van Den Hout A, Handels R, Bos I, Van Der Flier WM, Kern S, Ousset P-J, Maruff P, Skoog I (2019) Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype. Alzheimer's Dement 15:888–898
66. Parnetti L, Chipi E, Salvadori N, D'Andrea K, Eusebi P (2019) Prevalence and risk of progression of preclinical Alzheimer's disease stages: a systematic review and meta-analysis. Alzheimers Res Ther 11:7
67. Jack Jr CR, Andrews JS, Beach TG, Buracchio T, Dunn B, Graf A, Hansson O, Ho C, Jagust W, McDade E (2024) Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. Alzheimer's Dement 20:5143–5169
68. Geraldo A, Azeredo A, Pasion R, Dores AR, Barbosa F (2019) Fostering advances to neuropsychological assessment based on the Research Domain Criteria: The bridge between cognitive functioning and physiology. Clin Neuropsychol 33:327–356
69. Lorenzon G (2024) Brain heterogeneity within aging and cognitive impairment: implications for precision medicine and prevention. Karolinska Institutet
70. Nami M, Mehrabi S, Kamali A-MK, Tahamtan M, Derman S, Kosagiaharaf R (2021) Spatial distribution of K-complexes in sleep quantitative-EEG among subjects with cogniform disorder. Scandinavia J Sleep Med 1
71. Arya AD, Verma SS, Chakarabarti P, Chakrabarti T, Elngar AA, Kamali A-M, Nami M (2023) A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. Brain Inf 10:17
72. Viswan V, Shaffi N, Mahmud M, Subramanian K, Hajamohideen F (2024) Explainable artificial intelligence in Alzheimer's disease classification: A systematic review. Cogn Comput 16:1–44
73. Ursin F, Timmermann C, Steger F (2021) Ethical implications of Alzheimer's disease prediction in asymptomatic individuals through artificial intelligence. Diagnostics 11:440
74. Clark C, Rabl M, Dayon L, Popp J (2022) The promise of multi-omics approaches to discover biological alterations with clinical relevance in Alzheimer's disease. Front Aging Neurosci 14:1065904
75. Natarajan S, Mathur S, Sidheekh S, Stammer W, Kersting K (2025) Human-in-the-loop or AI-in-the-loop? Automate or Collaborate? In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, ; pp. 28594–28600

Publisher's note