



Editorial

Advancing multimodal neuroimaging: Explainable and responsible AI for early dementia detection[☆]

J.E. Arco^{1,2}, R. Alizadehsani³⁸, M. Atzmueller^{4,5}, E. Barakova⁶, L. Bloch^{44,45,46}, G. Bologna⁷, P. Bonomini^{8,9}, S.-B. Cho¹¹, J.M. Cortes^{41,42,43}, A. Elazab³⁷, M.A. Khan³⁶, E. Domínguez¹², R.J. Duro¹³, A. Fernández-Caballero^{14,15}, E. Fernández-Jover⁴⁹, J.M. Ferrández-Vicente^{10,50}, Christoph M. Friedrich^{52,45}, J. García-Rodríguez¹⁶, J. Garrigós¹⁰, A. Gómez-Rodellar^{17,18}, P. Gómez-Vilda^{17,19}, M. Graña²⁰, A. Hernández^{2,3}, Yujing Huang⁴⁰, C. Jiménez-Mesa^{2,3}, J. Levin^{33,34,35}, N. Liu⁴⁷, Q. Ma⁵³, F.J. Martínez-Murcia^{39,2}, R. Martínez-Tomás²¹, J.M. Molina²², G.J. Nalepa^{23,51}, A. Ortiz^{24,2}, D. Palacios-Alonso^{17,19}, M. Ponticorvo⁴⁸, A. Pereira²⁵, J. Ramírez^{39,2}, M. Rincón²¹, I. Rodríguez-Rodríguez^{24,2}, L. Su³⁰, J. Suckling^{30,31,32}, M. Val-Calvo²¹, P. Varona²⁶, C. Vázquez-García^{39,2}, S. Wang²⁷, Y.-D. Zhang^{28,29}, J.M. Gorriz^{2,3,30} ,*

¹ Mind, Brain and Behavior Research Center (CIMCYC), UGR, Spain

² Data Science and Computational Intelligence Institute (DaSCI), UGR, Spain

³ Signal Processing and Biomedical Applications, UGR, Spain

⁴ Semantic Information Systems Group, Osnabrück University, Osnabrück, Germany

⁵ German Research Center for Artificial Intelligence (DFKI), Osnabrück, Germany

⁶ Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands

⁷ University of Applied Sciences and Arts of Western Switzerland, Geneva, Switzerland

⁸ Instituto Argentino de Matemática Alberto Calderón y Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

⁹ Instituto Tecnológico de Buenos Aires (ITBA), Buenos Aires, Argentina

¹⁰ Department of Electronics, Computer Technology and Projects, Universidad Politécnica de Cartagena, Cartagena, Spain

¹¹ Department of Computer Science, Yonsei University, Seoul, Republic of Korea

¹² Department of Computer Science, University of Malaga, Malaga, Spain

¹³ Integrated Group for Engineering Research, CITIC, Universidade da Coruna, Spain

¹⁴ Neurocognition and Emotion Research Unit, Albacete Research Institute of Informatics, University of Castilla-La Mancha, Albacete, Spain

¹⁵ Biomedical Research Networking Center in Mental Health, Instituto de Salud Carlos III, 28016 Madrid, Spain

¹⁶ Department of Computers Technology, University of Alicante, Alicante, Spain

¹⁷ Neuromorphic Speech Processing Lab, Center for Biomedical Technology, Universidad Politécnica de Madrid, Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain

¹⁸ Usher Institute, Faculty of Medicine, University of Edinburgh, Edinburgh, UK

¹⁹ Escuela Técnica Superior de Ingeniería Informática, Universidad Rey Juan Carlos, Campus de Móstoles, 28933, Móstoles Madrid, Spain

²⁰ Department of CCIA, University of the Basque Country (UPV/EHU), Spain

²¹ Department of AI, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain

²² Computer Science and Engineering Department, Universidad Carlos III de Madrid, Colmenarejo, 28270 Madrid, Spain

²³ Department of Human-Centered Artificial Intelligence, Jagiellonian University, Kraków, Poland

²⁴ Department of Communications Engineering, University of Malaga, Malaga, Spain

²⁵ Computer Science and Communications Research Centre, School of Technology and Management, Polytechnic Institute of Leiria, Leiria, Portugal

²⁶ Grupo de Neurocomputación Biológica, Departamento de Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain

²⁷ Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215123, China

²⁸ School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, 454003, China

²⁹ Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

³⁰ Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

³¹ Centre for Human Inspired Artificial Intelligence, University of Cambridge, Cambridge, United Kingdom

³² Cambridge and Peterborough NHS Foundational Trust, Cambridge, United Kingdom

³³ Department of Neurology, Maximilians Universität München, Munich, Germany

³⁴ German Center for Neurodegenerative Diseases (DZNE), Munich, Germany

[☆] This article is part of a Special issue entitled: 'AI for Early Dementia Detection' published in NeuroImage.

* Corresponding author.

E-mail address: gorriz@ugr.es (J.M. Gorriz).

- ³⁵ Munich Cluster of Systems Neurology (SyNergy), Munich, Germany
- ³⁶ Center of AI, Prince Mohammad bin Fahd University, Alkhobar, Saudi Arabia
- ³⁷ Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China
- ³⁸ Institute for Intelligent Systems Research and Innovation (IISRI), Geelong, 3217, VIC, Australia
- ³⁹ Department of Signal Theory, Telematics and Communications, University of Granada, Spain
- ⁴⁰ Hangzhou First People's Hospital Affiliated of Westlake University School of Medicine, China
- ⁴¹ Computational Neuroimaging Lab, Biobizkaia Health Research Institute, Barakaldo, Spain
- ⁴² IKERBASQUE, the Basque Foundation for Science, Bilbao, Spain
- ⁴³ Department of Cell Biology and Histology, University of the Basque Country, Leioa, Spain
- ⁴⁴ Department of Computer Science, University of Applied Sciences and Arts Dortmund (FH Dortmund), Emil-Figge-Straße 42, 44227 Dortmund, North Rhine-Westphalia, Germany
- ⁴⁵ Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, Hufelandstraße 55, 45147 Essen, North Rhine-Westphalia, Germany
- ⁴⁶ Institute for Artificial Intelligence in Medicine (IKIM), University Hospital Essen, Hufelandstraße 55, 45147 Essen, North Rhine-Westphalia, Germany
- ⁴⁷ First Affiliated Hospital of Jinzhou Medical University, China
- ⁴⁸ University of Naples «Federico II», Naples, Italy
- ⁴⁹ Instituto de Bioingeniería, Universidad Miguel Hernández, Elche, Alicante, Spain
- ⁵⁰ ECTLab+, European University of Technology, European Union, Spain
- ⁵¹ School of Information Technology, Halmstad University, Halmstad, and ELLIIT Excellence Center, Sweden
- ⁵² Department of Computer Science, University of Applied Sciences and Arts Dortmund (FH Dortmund), Emil-Figge-Straße 42, 44227 Dortmund, North Rhine-Westphalia, Germany
- ⁵³ Laboratory of Neuromanagement, Zhejiang University, Hangzhou 310024, Zhejiang Province, China

ARTICLE INFO

Keywords:

Dementia
Neuroimaging
Artificial intelligence
Explainable AI
Multimodal fusion
Responsible AI
Medical ethics
Clinical translation
Digital biomarkers

ABSTRACT

Artificial intelligence (AI) has not seen the clinical uptake that might be expected from a technology that has received so much attention and investment. Coupled to neuroimaging, it is conceivable that AI algorithms can provide better performance in diagnosis and prognosis as well as optimize treatments in a precision medicine regimen, all of which is focused on improving both the patient experience and clinical outcomes. But in practice, little impact has been seen. Why might this be the case? This overview focuses on the possible reasons. First, technical concerns: the way in which AI algorithms are developed and validated in the research setting does not adequately prepare them for deployment in clinics and hospitals. Second, the importance of asking clinical questions with AI algorithms that have meaning and value is often underplayed or not considered. The outputs of AI algorithms mostly, but not always, also need to explain how decisions have been made. Thirdly, operational and ethical considerations loom over the integration of AI algorithms into electronic health record systems, clinical pathways, and legal frameworks. Above all these considerations is the motivation for deployment and particularly whether it is primarily for patient benefit or service economics.

1. Introduction

Artificial intelligence has drawn enormous attention across medicine, and work on dementia has moved with that current. Hundreds of systems now classify disease stages, predict progression, or extract biomarkers from neuroimaging at accuracies that would have seemed far-fetched only a decade ago. Outside the controlled conditions of a research study, however, their clinical footprint is slight. Very few have been deployed, and fewer have altered how a patient is actually assessed or managed. It is this distance between demonstrated capability and clinical reality that concerns us here, and we prefer to state it plainly at the outset rather than let it accumulate quietly beneath the technical detail.

Why the gap exists is neither simple nor a matter of engineering alone. Dementia already affects tens of millions of people, and its prevalence is climbing, from roughly 55 million in 2020 to a projected 78 million by 2030 (Alzheimer's Disease International, 2024). No cure exists. The disease-modifying drugs that have recently reached approval, among them lecanemab and donanemab for early-stage Alzheimer's disease, apply only to a narrow band of patients. The weight of the condition falls not only on those who have it but on families, caregivers, and health systems that are usually stretched thin to begin with, and in that setting anything capable of pushing detection earlier, and making it more precise, would count for a great deal. Here the appeal of AI is easy to understand: it is good at finding structure in complex, high-dimensional data that resists more conventional analysis, and deep learning in particular has shown it can do this across biomedical problems, dementia included (Cogswell and Fan, 2023; Bonomini et al., 2024; Chen et al., 2023). The difficulty is not that the technology has

nothing to offer. It is that being relevant in a research paper and being useful in a clinic are different achievements, and the field has at times treated them as one.

Neuroimaging sits near the center of this tension. Structural and functional Magnetic Resonance Imaging (MRI), together with Positron Emission Tomography (PET), provide direct information about brain structure, function, and pathology, and the high-dimensional datasets they generate fit naturally with deep learning methods. Convolutional neural networks, autoencoders, and transformer-based architectures have all been applied to biomarker extraction, dementia subtype classification, and progression prediction, often with strong results (Górriz et al., 2023; Borchert et al., 2023). But strong results on benchmark datasets are not the same as clinical trustworthiness, and this distinction matters more in medicine than almost anywhere else. A model whose reasoning is opaque is difficult to integrate into clinical decisions, regardless of its accuracy on held-out test sets (Kelly et al., 2019; Afroogh et al., 2026). Interpretability is therefore not a secondary concern that can be addressed after model development, it is one of the conditions that determines whether a model has any practical value at all. Part of what this article argues is that the field is beginning to take this more seriously, and that more genuinely explainable approaches are starting to emerge, though the evidence for their real-world impact remains limited and the deployment gap is far from closed.

There is a further structural limitation worth naming. Neuroimaging is almost always acquired once there is already clinical suspicion. A patient does not undergo MRI or PET without symptoms, family concern, or a referral. This means that imaging, however powerful analytically, tends to function as a tool for characterization and confirmation rather than early detection in any meaningful sense. It is one of the reasons the field has turned toward digital technologies. Wearable devices, mobile applications, and forms of continuous passive sensing make it possible

to collect behavioral and physiological data outside the clinic, in the course of daily life. What those devices record, sleep quality, mobility, speech, gait, can shift months or even years before anyone would think to seek a formal diagnosis (Čepukaitytė et al., 2024). None of this replaces imaging, but it lengthens the window over which a person can be observed, which imaging on its own does not do. Whether folding these signals together with neuroimaging and clinical history actually sharpens early detection is still an open question, and the evidence has only begun to accumulate (Lee et al., 2024b). The shift also brings problems that are not really methodological. As these systems edge away from the clinic and toward continuous, passive collection of data, keeping consent, privacy, and accountability on a sound footing becomes correspondingly harder. These are not soft concerns that can be appended to a technical paper. They determine whether a system is ethically defensible in the first place, and in some cases whether it should be deployed at all (Lekadir et al., 2025; Dino et al., 2025). The clinical adoption of AI requires not only that a model performs well, but that the rules governing its use are clear, that patients understand what is being collected and why, that data are protected throughout, and that responsibility for errors is unambiguous. These are questions the field is still working through, and they will not be resolved by better architectures alone (Andreoletti et al., 2024).

With that framing in mind, the sections that follow do not simply survey what AI can do in dementia research. This is also what distinguishes the present overview from the reviews that already exist. Those reviews have tended to treat the field in silos: some survey neuroimaging-based detection (Borchert et al., 2023), others the performance of explainability methods (Górriz et al., 2023), others digital and wearable biomarkers (Čepukaitytė et al., 2024), and even the most recent single-domain surveys, whether of brain-imaging foundation models (Ghamizi et al., 2026) or of explainable AI across biomedical tasks (Hettikankanamage et al., 2025), remain within one strand. The work that comes closest to an integrative view tends to treat multimodal fusion as a data-harmonization problem rather than a clinical one (Jasodanand et al., 2025). To our knowledge, none reads across neuroimaging, explainability and digital biomarkers together against a single criterion: whether the resulting systems are usable in the clinic, and whom their deployment is designed to serve. That criterion, rather than any one method, is what organizes what follows. The following sections examine what the technical literature reveals about the conditions under which these methods succeed or fall short, and what that implies for the path toward genuine clinical utility. Multimodal integration and deep learning architectures are considered first, followed by explainability and digital biomarkers, before the article turns to the translational and clinical applications and the ethical dimensions that ultimately determine whether any of this work is fit for clinical use. The argument, stated plainly, is that methodological progress is necessary but not sufficient. What the field needs now, at least as much as better models, is a clearer account of what it would mean for those models to actually work in practice.

A note on scope is in order. This article is an overview linked to the special issue it introduces, on early AI-based detection of dementia through multimodal neuroimaging and related biomarkers, rather than a systematic review, and it does not follow a PRISMA-style protocol. Readers looking for an exhaustive, reproducible enumeration of the field should treat it accordingly. Our starting point was the set of contributions gathered in the special issue, which we then situated within the broader and fast-moving literature on AI for dementia. To do so we drew on targeted, non-exhaustive searches of PubMed/MEDLINE, Scopus, IEEE Xplore and Google Scholar, supplemented by arXiv, bioRxiv and medRxiv for very recent preprints, using combinations of terms around dementia and Alzheimer's disease, neuroimaging, multimodal fusion, explainability, digital biomarkers and clinical translation. We placed particular emphasis on developments of the last few years, while retaining older studies where they remain methodologically foundational. Works were selected for their relevance to the organizing

question of the review, the gap between technical performance and clinical usability, for their representativeness of a methodological direction, and for their recency, rather than for completeness. This is, in other words, a curated and deliberately selective account: what we chose to include reflects the argument the article is making, and we have tried to be explicit about that argument throughout.

2. Multimodal integration

Combining information from multiple sources is one of the most natural responses to the complexity of dementia. No single modality captures the full picture: imaging reveals structural and metabolic changes, but says nothing about daily functioning. Cognitive tests are sensitive to early decline, but lack spatial resolution, whereas fluid biomarkers reflect pathological load, but require invasive procedures. Sociodemographic factors, including age, sex, and educational attainment, are also routinely incorporated as they modulate both disease risk and the interpretation of cognitive assessments. Multimodal models are an attempt to use all of this together. The question worth asking at the outset, however, is whether doing so actually improves things in ways that matter clinically, or whether it primarily improves benchmark scores on datasets that were never designed to reflect the messiness of real clinical populations. The simplest approach, feature concatenation, is still the most common starting point, and it does produce consistent improvements in AD classification accuracy over unimodal baselines (Feng et al., 2025; Arco et al., 2021). But concatenation also increases computational cost substantially, which is not a trivial concern when neuroimaging data alone can involve hundreds of thousands of features per subject. One concrete response to this has been to work with cortical rather than volumetric features (that is, surface-based measures such as cortical thickness and gyrification patterns), rather than raw voxel intensities or regional volumes derived from the full 3D image. Duong et al. (2025) showed that a middle-fusion approach with cross-attention over cortical representations achieves better results than the volumetric alternative for early AD identification, while being considerably more tractable computationally. That kind of trade-off (some loss of raw information, better efficiency, similar or improved clinical performance) is worth taking seriously as a design principle rather than just a computational convenience.

More recent work has moved toward fusion strategies that explicitly model the dependencies between modalities, on the reasonable assumption that imaging, biomarkers, and cognitive measures are not independent of each other, so a model does better when it represents how they interact rather than treating them as separate channels stacked together. Transformers lend themselves to this, since their attention mechanism picks up cross-modal relationships that plain concatenation leaves on the table (Zhang et al., 2023; Jasodanand et al., 2025). Wang et al. (2025) give a clear illustration: they asked how much structural MRI and polygenic risk scores add to a plasma-biomarker baseline when predicting amyloid-beta burden, and their model reached an R^2 of 0.64 against amyloid PET, holding up when moved across cohorts. What such models offer is not only better prediction. They can also expose dependencies that are invisible when each modality is read on its own, for instance the way hippocampal atrophy tracks with CSF tau abnormalities, which makes them useful as instruments for analysis and not merely as classifiers (Abdelaziz et al., 2025; Zhang et al., 2023). A parallel line of work drops below the level of the whole structure to the subfield, on the observation that hippocampal subregions are not equally vulnerable in early AD. Li et al. (2026) pair quantitative susceptibility mapping (which is sensitive to iron deposition) with structural T1 radiomics over 24 carefully segmented hippocampal subfields, and build a composite Rad-score that separates prodromal from dementia-stage AD, with an area under the ROC curve (AUC) of 0.85 on an independent external cohort. Most of that signal came from QSM texture in CA1 and the dentate gyrus granule cell layer, which points to iron-related microstructural

disorganization in exactly the subregions that fail first, information that volume measurements on their own do not capture.

Longitudinal analysis is where multimodal methods face perhaps their hardest test. This sounds straightforward, but in practice most longitudinal cohorts are heavily unbalanced toward early stages, follow-up intervals are irregular, and dropout is not random. Many studies track patients over years, collecting neuroimaging, cognitive assessments, and clinical profiles at each visit, with the aim of modeling trajectories, that is, the probability of developing further symptoms or converting to a more advanced stage. Recurrent and transformer-based architectures have both been used for this purpose. [Ding et al. \(2023\)](#) developed a multimodal Long Short-Term Memory (LSTM) model incorporating demographic, neuropsychological, and neuroimaging data across five time points in 347 Mild Cognitive Impairment (MCI) subjects from Alzheimer's Disease Neuroimaging Initiative (ADNI), and showed that the longitudinal formulation outperforms cross-sectional baselines for conversion prediction (AUC = 0.93 vs. 0.90). In a different direction, [Moghaddami et al. \(2025\)](#) proposed a transformer that processes variable-length visit histories, with particular strength in identifying subjects who will convert. Their appeal is not just a matter of accuracy. Reasoning about a patient as a trajectory, a pattern that unfolds over time rather than a single snapshot, is closer to how clinicians actually reason. There is a wrinkle, though: decline is not a one-way street. A fair number of MCI patients hold steady or drift back to normal cognition, and any model that treats conversion as the sole endpoint will get that subgroup wrong, which matters clinically. [Huang et al. \(2026\)](#) take this on directly, comparing five machine-learning algorithms across stable, convertible, and reversible trajectory groups in more than 1100 ADNI participants. Random Forest handled the heterogeneity of the continuum best, and the sharpest multimodal predictors turned out to be amyloid PET in the parahippocampus and ventricles, plasma APOE4, and neurofilament light chain.

A subtler hazard in multimodal modeling is the assumption that more data must mean better inference. Each modality brings its own biases, and stacking them together without accounting for those biases can muddy the picture as easily as sharpen it. [Condado et al. \(2025\)](#) examined this directly in BrainAge models, which are trained on healthy controls with chronological age as the target, and whose output, a predicted biological age, is subtracted from chronological age to yield a delta. Their feature-selection analysis found a division of labor: neuroimaging features such as gray matter volume and hippocampal atrophy track aging most strongly, whereas neuropsychological tests carry most of the power to discriminate clinical groups. The difficulty is that those imaging features are shaped by normal aging and by AD progression alike, so the delta conflates the two. The trade-off is explicit in their results: models optimized for age prediction yield deltas that classify poorly, and models whose deltas are optimized for classification predict age less accurately. A BrainAge model does not cleanly separate aging from disease, and its delta cannot be read as though it did. The lesson is that a multimodal model needs a hypothesis about what each input actually contributes, revealing that concatenation alone is not enough. Foundation models, pre-trained on large, heterogeneous data to learn representations that carry across modalities, are one way toward more principled fusion, and early results point to gains in temporal consistency and robustness ([Ghamizi et al., 2026](#)). Whether any of that survives contact with real clinical use, at scale, is still open.

[Table 1](#) summarizes representative approaches to multimodal fusion in the context of dementia detection, and [Fig. 1](#) provides a conceptual overview of how these systems integrate heterogeneous data.

3. Innovations in deep learning architectures

The number of deep learning architectures proposed for dementia analysis has grown steadily, but growth in the number of methods is not the same as progress toward clinical use. Many of the models described in the literature achieve strong performance on standard

benchmarks, yet few have been evaluated in conditions that resemble clinical deployment: heterogeneous populations, missing modalities, irregular follow-up, or hardware constraints in real hospital settings. This section surveys some of the most relevant recent directions, while keeping that gap in mind. Most of these architectures treat the brain as a regular array of voxels and leave it to the network to discover whatever structure lies within. Graph neural networks (GNNs) start on a different intuition: the brain is a network of interacting regions, not a set of isolated voxels. Casting it that way makes the topological relationships between regions explicit, something standard connectivity measures tend to flatten ([Zubasti et al., 2024](#)). Several groups have since folded this kind of brain organization into disease-progression models, with encouraging results for predicting how individual patients will move along the continuum. That said, GNNs remain sensitive to how the graph is constructed: a decision that is often made on computational grounds rather than neurobiological ones, and that can substantially affect what the model learns. Transformer-based architectures offer a different kind of flexibility. [Yao et al. \(2025\)](#) proposed a method that captures both local and global features in brain images within a single framework, with competitive accuracy and reduced computational cost. Speech-based approaches are also worth noting here. For example, [Pereira et al. \(2024\)](#) presented a method using large language models to process emotional content from speech, and there is growing evidence that speech patterns carry diagnostically relevant information for AD and related dementias at early stages ([Rezai et al., 2025](#); [Lv et al., 2026](#)). Hybrid architectures that combine convolutional and transformer components can consolidate some of these gains, improving robustness in classification tasks by using each component for what it does best ([Muhammad et al., 2025](#)).

A more persistent obstacle is data. Large, well-labeled, class-balanced neuroimaging datasets for dementia are difficult to assemble, and the disease stages most relevant for early intervention are typically the least represented. Class imbalance of this kind is not unique to neuroimaging, and the broader biomedical literature on feature selection in large, high-dimensional and imbalanced datasets (including work on microarray classification) offers methods the field can draw on ([Karimi et al. \(2025\)](#), [Aliyu et al. \(2025\)](#)). Self-supervised Learning (SSL) addresses part of this problem by extracting structure from unlabeled data. [Li et al. \(2025\)](#) proposed an SSL architecture for modeling the temporal dynamics of AD from neuroimaging that maintains strong performance despite limited labels and high inter-patient variability, two conditions that are the norm rather than the exception in clinical data. Generative approaches such as Generative Adversarial Networks (GAN) and variational autoencoders offer complementary tools. They can synthesize realistic neuroimages to augment training sets, fuse the features of existing modalities for classification (as in the coupled-GAN of [Choudhury et al. \(2024\)](#), which combines MRI and PET rather than generating one from the other), or perform genuine cross-modal translation, synthesizing PET from MRI when a tracer scan is unavailable ([Wang et al., 2026a](#)). The latter is of particular interest for lower-resource settings, where multi-tracer amyloid and tau PET are least accessible, though, as discussed in the section on clinical applications, the mapping from structure to molecular pathology is biologically ill-posed and the synthesized images must be read with corresponding caution ([Baron et al., 2026](#)).

Uncertainty quantification is a related concern that is underrepresented in the literature relative to its practical importance. A model that provides a prediction without any estimate of its own reliability is difficult to use responsibly in clinical settings. Bayesian and evidential approaches offer a way to attach confidence intervals to predictions, which matters more as these systems get closer to informing real decisions ([Tanveer et al., 2024](#); [Gao et al., 2024](#)). A complementary challenge is the validity of the evaluation itself. Standard cross-validation is known to produce optimistic estimates in small-sample regimes, which are the norm rather than the exception in neuroimaging. [Hernández et al. \(2026\)](#) address this directly through a

Table 1

Comparison of recent studies for dementia detection using multimodal fusion technologies.

Ref.	Method	Cohort Size	Modalities	Fusion strategy	Task and main results
Jasodanand et al. (2025)	Transformer-based cross-modal attention	12,185 (7 cohorts: NACC, A4, OASIS-3, AIBL, FHS, ADNI, HABS)	Multi-sequence sMRI, demographics, cognitive tests, genetics, clinical history	Intermediate fusion using modality-specific embeddings and a pretrained transformer	$A\beta/\tau$ status classification: AUC 0.79 ($A\beta$), AUC 0.84 (τ); findings consistent with postmortem neuropathology.
Han et al. (2025)	AD-Diff: 3D diffusion with Mamba classifier	416 (OASIS) + ADNI (sample size not specified)	sMRI, synthesized FDG-PET, tabular cognitive assessments	Cross-modal MRI→PET generation followed by BiCross intermediate pixel-level fusion	AD progression prediction: Accuracy 94.40% (1-year), 88.77% (3-year).
Zhang et al. (2023)	Multimodal cross-attention network (MCAD)	364 (ADNI)	sMRI, FDG-PET, CSF	Intermediate cross-attention fusion	HC/MCI/AD classification: Accuracy 71.26% (HC vs. MCI), 91.07% (AD vs. HC), 64.03% (three-class classification).
Rallabandi and Seetharaman (2023)	Inception-ResNet wrapper with DWT fusion	1098 (OASIS)	T1-weighted sMRI, FDG-PET	Image-level fusion using discrete wavelet transform	HC/MCI/AD classification: Accuracy 95.5% (HC vs. MCI), 94.6% (MCI vs. AD), 95.7% (AD vs. HC).
Odusami et al. (2023)	Modified ResNet18	412 (ADNI)	sMRI, FDG-PET	Early feature fusion by concatenation	MCI stage classification: Accuracy 73.9% (eMCI vs. IMCI); demonstrated good control compared with oversized models.
Yu et al. (2024a)	AD-Transformer: unified attention-based framework	1651 (ADNI)	sMRI, clinical assessments, APOE genotype	Intermediate fusion via Patch-CNN and unified transformer	AD/MCI classification: AUC 0.993 (AD vs. HC), AUC 0.845 (MCIs vs. MCIC).
Ahmed et al. (2025)	Dual-stream 3D CNN with attention	ADNI, AIBL, OASIS-3 (sample size not reported)	sMRI, FDG-PET	Attention-guided late fusion with gating mechanism	AD classification: Accuracy up to 98.7% (AD vs. HC); robust cross-cohort performance; inference time ≤ 160 ms/subject.
Sheng et al. (2024)	Hybrid KELM with Harris Hawks Optimization (ILHHO-KELM)	202 (ADNI)	sMRI, Amyloid-PET (FBB), CSF	Intermediate fusion with simultaneous metaheuristic optimization	HC/MCI/AD classification: Accuracy 99.2% (AD vs. HC), 92.0% (AD vs. MCI), 94.6% (MCI vs. HC).
Lee et al. (2024a)	Graph convolutional network (GCN-ED-com)	556 (DAUH + ADNI)	Amyloid-PET (FBB), clinical indicators	Intermediate graph-based fusion using feature similarity population graphs	HC/MCI/AD classification: Accuracy 99.68% (AD vs. HC), 93.82% (MCI vs. HC), 90.43% (three-class classification).
Arco et al. (2023)	Siamese neural network with self-attention	251 (ADNI)	sMRI, FDG-PET	Intermediate fusion via self-attention in a coupled latent space	HC/MCI/AD classification: Accuracy 96.0% (AD vs. HC), 94.0% (MCIC vs. HC), 86.0% (MCIs vs. HC).

(continued on next page)

Table 1 (continued).

Abuhantash et al. (2025)	Fast random forest for survival analysis	814 (ADNI + AIBL)	Clinical assessments, cognitive tests, comorbidities	Feature-level fusion	Time-to-conversion prediction (HC to MCI): C-index 0.84; age and comorbidities identified as key predictors.
Saeed et al. (2024)	Random Survival Forest (RSF)	837 (291 eMCI + 546 IMCI; ADNI)	Cognitive tests, CSF, sMRI, FDG-PET	Feature-level fusion	Prediction of conversion to AD: C-index 0.90 (eMCI), C-index 0.82 (IMCI); cognitive assessments provided the highest predictive contribution.
Liu et al. (2025)	Multimodal Joint Attention (MJA) network	89 (ADNI + public dataset)	sMRI, resting-state EEG	Intermediate fusion through joint attention	AD/HC classification: Accuracy 94.75%; note that ADNI does not routinely acquire EEG data, indicating heterogeneous cohort fusion.
Liu et al. (2024)	InterFusion deep learning framework	307 (BioFIND)	sMRI, resting-state MEG	Intermediate fusion using the InterFusion module with spatial-channel cross-attention	MCI/HC classification: Accuracy 82.70%; significantly outperformed conventional fusion architectures.
Arco et al. (2025)	Deep Siamese autoencoder	251 (ADNI)	sMRI, FDG-PET	Intermediate fusion in the encoded latent space	HC/AD classification: AUC 0.88; biological interpretability confirmed through occlusion analysis.
Trinh et al. (2024)	Supervised encoder with dimensionality reduction	1419 (NACC)	sMRI, CSF, clinical assessments	Intermediate fusion using a supervised encoder for dimensionality reduction	HC/MCI/AD classification: Accuracy 89.50% (HC vs. MCI+AD), 84.1% (three-class classification).
Liu et al. (2023)	TriFormer: triple-branch transformer framework	1025 (ADNI1, ADNI2)	sMRI, clinical assessments	Intermediate fusion via three coupled parallel transformers	MCI/AD classification: Accuracy 91.47% (MCI vs. AD).

statistically agnostic regression (SAR) framework that bounds expected risk without parametric assumptions, demonstrating that conventional validation pipelines can overstate model reliability on ADNI data.

Naming the problem is not the same as fixing it, and it is fair to ask what a more defensible practice would actually look like. Three requirements do most of the work. The first is that data be split at the level of the subject, never the scan: when 2D slices or 3D patches from the same brain fall on both sides of the train–test boundary, the model can memorize individual anatomy and scanner signatures rather than disease, and the reported accuracy becomes an artefact. The systematic evaluation by Wen et al. (2020) found that more than half of the CNN-for-AD studies it examined were exposed to some form of this leakage, a sobering figure for a mature subfield. The second is nested cross-validation, with hyperparameter selection confined to an inner loop and the outer loop touched only once, for estimation. Without that separation, the test set quietly becomes part of the training procedure. The third is that generalization be assessed across sites rather than within one. Site-stratified or leave-one-site-out evaluation (ideally combined with feature harmonization such as ComBat Fortin et al., 2018) is the clearest way to show that a model has learned biology rather than the idiosyncrasies of a single scanner (Kelly et al., 2019). The SAR framework is useful here precisely because it reports

what a point estimate hides: a bound on expected risk rather than one optimistic number (Hernández et al., 2026).

Because these requirements are easy to endorse and easy to omit, the field would benefit from stating them as an explicit minimum. Table 2 sets out a short reporting standard for AI studies in dementia neuroimaging, distilled from the established guidelines (TRIPOD+AI, CLAIM, DECIDE-AI, CONSORT-AI and FUTURE-AI) but pared down to the items that most often separate a result that will replicate from one that will not (Collins et al., 2024; Tejani et al., 2024; Vasey et al., 2022; Liu et al., 2020; Lekadir et al., 2025). None of this proposes a new statistic; the point is only that these be treated as conditions for taking a result seriously, not as optional good practice.

These safeguards matter because the failure modes they guard against are not hypothetical, and the recent enthusiasm for foundation models and self-supervised pretraining has tended to outrun the evidence that they behave well under clinical conditions. Several documented cases are worth stating plainly. The most uncomfortable concerns shortcut learning, where a model arrives at the right answer from the wrong cue: what Lapuschkin et al. (2019) call a Clever Hans predictor. Tinauer et al. (2025) trained CNNs to classify AD on 1980 ADNI scans from 360 subjects, then stripped the images of their gray–white matter contrast by binarising them. Accuracy barely moved:

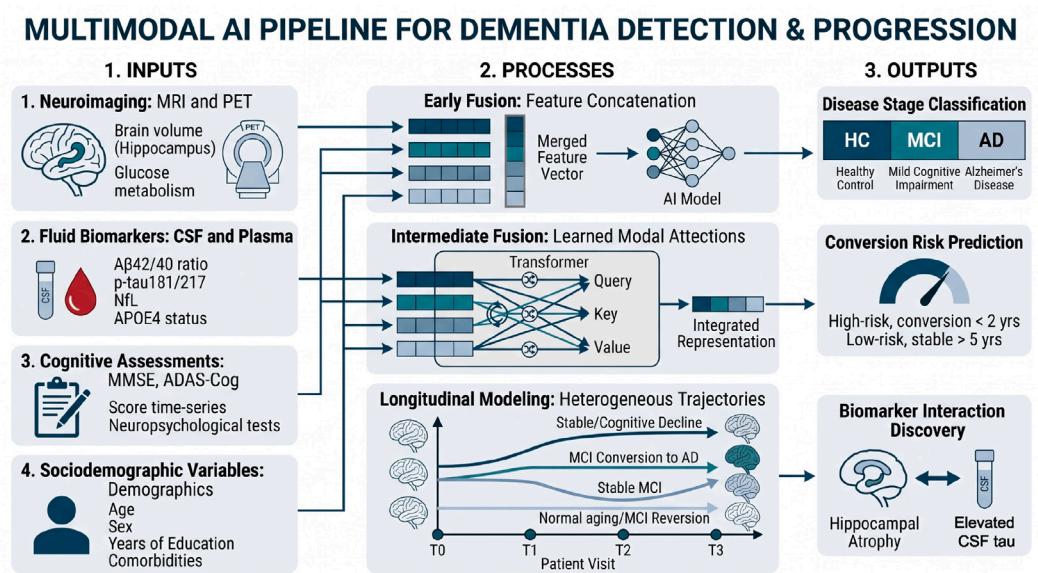


Fig. 1. Multimodal AI pipeline for dementia detection. Heterogeneous data sources (neuroimaging, fluid biomarkers, cognitive assessments, and sociodemographic variables) are integrated through early concatenation, cross-modal attention, and longitudinal trajectory modeling. Disease trajectories are shown as heterogeneous: from a diagnosis of MCI, conversion, stabilization, and reversion to normal cognition are all possible outcomes. The output represents risk stratification in support of clinical decision-making, not autonomous diagnosis.

Table 2
Minimum reporting standard proposed for AI studies in dementia neuroimaging, distilled from established reporting guidelines. Anchors: (Wen et al., 2020) (leakage), Fortin et al. (2018) (harmonization), Hernández et al. (2026) (SAR), Kelly et al. (2019) (multi-site generalization), and the guidelines TRIPOD+AI (Collins et al., 2024), CLAIM (Tejani et al., 2024), DECIDE-AI (Vasey et al., 2022), CONSORT-AI (Liu et al., 2020) and FUTURE-AI (Lekadir et al., 2025).

#	Reporting item	What to report	Guideline anchor
1	Subject-level data splitting	No scans, slices or longitudinal visits from the same subject across train/validation/test sets	CLAIM 19–20; Wen et al. (2020)
2	Nested cross-validation	Inner loop for tuning; outer loop used once, only for performance estimation	TRIPOD+AI 12c
3	External validation/testing	Independent cohort or leave-one-site-out, not a held-out split of the same sample	CLAIM 33; Kelly et al. (2019)
4	Site/scanner reporting and harmonization	Manufacturer, field strength and <i>n</i> per site; harmonization method (e.g. ComBat) and protected covariates	CLAIM 13; Fortin et al. (2018)
5	Uncertainty over point estimates	Confidence intervals or risk bounds (e.g. SAR), not a single accuracy or AUC	TRIPOD+AI 23a; Hernández et al. (2026)
6	Subgroup performance	Sensitivity and specificity by sex, age, and where possible ethnicity	TRIPOD+AI 23a, 14; FUTURE-AI (Fairness)
7	Missing-modality handling	How incomplete multimodal cases were treated, and the effect on the reported result	TRIPOD+AI 11
8	Quantitative explanation evaluation	If saliency or XAI is shown, report faithfulness and stability metrics, not only example maps	Metsch and Hauschild (2025), Gomez and Mouchère (2023); FUTURE-AI (Explainability)
9	Code and model availability	Repository with preprocessing, training pipeline and final model weights	CLAIM 43; FUTURE-AI (Traceability)

the binarised skull-stripped variants scored between 78% and 82%, against 81.63% for the reference model, whereas images that had not been skull-stripped fell to between 62% and 78%. Layer-wise relevance propagation showed the networks were keying on the brain contour

introduced by the skull-stripping itself rather than on any neurobiology. A model can score well and be learning nothing of clinical interest. Missing modalities produce a different and equally concrete failure. The multimodal OmniBrain framework reached 92.2% accuracy with

all modalities present but fell to 70.4% when only MRI was available at test time (Sharshar et al., 2025), a drop of more than twenty points that reflects how fusion layers co-adapt to the exact combination of inputs seen during training. Self-supervised methods have their own characteristic pathology: because brain scans are far more homogeneous than natural images, contrastive objectives can drive representations into a narrow, low-rank subspace, a dimensional collapse in which apparently trained embeddings have quietly lost their discriminative power (Hassanpour et al., 2024).

Scale does not automatically resolve any of this. In the FOMO25 clinical challenge, nineteen foundation models pretrained on 60,000 clinical brain volumes were evaluated on data drawn straight from hospital workflows; the strongest entries were among the smaller models, and neither added capacity nor longer pretraining bought a reliable gain (Munk et al., 2026). In a separate benchmark, a tabular model over handcrafted radiomics features (AUROC 0.92) matched or beat every visual foundation model tested within cohort, BiomedCLIP (0.85) and BrainDINO (0.82) among them, at predicting molecular status from MRI (Hollet et al., 2026). Nor is the danger only lower accuracy: under scanner shift, segmentation models for white-matter hyperintensities become confidently wrong, producing high-probability false-positive lesions unless explicitly regularized toward appropriate uncertainty (Matzkin et al., 2025). None of them argues against foundation models or self-supervised learning as such. They argue against deploying either on the assumption that benchmark performance transfers, and they are the concrete reason the reporting standard above insists on external, cross-scanner evaluation rather than a single held-out split (see Table 3).

A fourth direction, less about architecture and more about what models are built on, involves incorporating prior neurobiological knowledge into the learning process. The underlying intuition is simple: models that ignore what we already know about the brain are leaving signal on the table. The challenge is that this prior knowledge is itself incomplete and sometimes contradictory (Tuan et al., 2024), so the question is not whether to use it but how. Yin et al. (2023) developed a framework for brain age estimation that maps individual brain region contributions explicitly, recovering sex- and disease-specific patterns of neurological aging with stronger association to early AD signs than chronological age alone. Zhang et al. (2025b) took a different route, pre-training on over 9000 unlabeled MRIs to learn structural brain priors before transferring that anatomical knowledge to downstream cognitive impairment prediction. Yang et al. (2025) incorporated connectivity constraints derived from known AD pathophysiology into a graph convolutional network, achieving over 94% accuracy in AD classification while producing results that are consistent with established neuropathological patterns. Wang et al. (2024) proposed an approach that extracts domain knowledge from the AD literature automatically, using it to guide graph learning without manual curation. These studies differ in their design choices, but they share the view that domain knowledge and learned representations are more useful together than separately. Whether that combination will hold up across diverse clinical populations (which look rather different from the ADNI cohorts that dominate evaluation in this literature) is a question that has not been answered adequately yet.

Several architectural directions sit at the edge of this literature and are worth flagging, not because they are established but because they address limitations the methods above do not. Vision–language models are the most visible: rather than fusing imaging with tabular data through concatenation, they align image-derived features with clinical text and neuropsychological findings in a shared representation, which can help resolve the common situation where imaging and cognition point in different directions (Nie et al., 2026). Neural ordinary differential equations take a different problem seriously, namely that disease does not progress in the discrete steps that recurrent and transformer models assume. By modeling morphological change as a continuous trajectory, they handle irregular follow-up more naturally and can

Table 3
Documented failure modes of foundation models and self-supervised learning in neuroimaging. Evidence is drawn from empirical studies in brain MRI where available.

Ref.	Failure mode	Source of the problem
Tinauer et al. (2025)	Shortcut learning (Clever Hans)	Model keys on the skull-stripping contour rather than gray–white matter texture; binarised skull-stripped images classify AD as well as full-contrast scans (ADNI; 1980 scans, 360 subjects). In OmniBrain, fusion layers co-adapt to the training modality set: 92.2% with all modalities (ANMerge) against 70.4% on MRI alone (ADNI), so modality loss and cohort shift compound.
Sharshar et al. (2025)	Missing-modality collapse	High anatomical similarity drives embeddings into a low-rank subspace, losing discriminative power. Small pretrained models match or beat larger ones, and scaling capacity or pretraining brings no reliable gain; within cohort, a tabular model over radiomics beats every visual foundation model (AUROC 0.92 vs. 0.85 BiomedCLIP/0.82 BrainDINO).
Hassanpour et al. (2024)	Dimensional collapse (SSL)	Confident false-positive WMH lesions on unseen scanners unless entropy-regularized.
Munk et al. (2026), Hollet et al. (2026)	Scale ≠ generalization	
Matzkin et al. (2025)	Overconfidence under shift	

forecast progression from a single scan (Park et al., 2025). Object-detection and weakly-supervised localization methods offer a route to explanation that does not depend on post-hoc saliency, learning to localize the atrophy that drives a prediction as part of the model itself (Lian et al., 2022). And test-time adaptation addresses the scanner-shift problem from the opposite end to harmonization, adjusting a trained model to each incoming scan without retraining (Valanarasu et al., 2024). None of these is mature: most have been demonstrated on a single cohort, and their behavior under the conditions described earlier in this section, heterogeneous populations, missing modalities, real clinical noise, is largely untested. But they are the directions most likely to matter next, precisely because each targets a limitation that better benchmark accuracy alone will not fix.

Table 4 summarizes representative works using recent DL architectures for AD detection.

4. Explainability and interpretability

Interpretability is sometimes presented as a technical feature that can be added to a model once it performs well enough. That framing understates the problem. In clinical settings, a model whose reasoning is opaque is not just less convenient: it is genuinely difficult to integrate into a decision process where the clinician is accountable for the outcome (Afroogh et al., 2026). Whether an AI output should be trusted, overridden, or flagged for review requires some understanding of how it was reached. This is especially true in dementia, where the boundary

Table 4
Recent works using innovative DL architectures for the study of Alzheimer's detection.

Ref.	Model architecture	Modalities	Innovative feature and results
Sheikh et al. (2025)	EfficientNetV2B0	sMRI (T1-weighted, 2D slices)	Lightweight architecture (6.08M parameters) for real-time inference; Acc 88.0% tri-class (CN/EMCI/LMCI) on 102 ADNI subjects (5-fold CV) ; Grad-CAM++ and Guided Grad-CAM++.
Alsaleh (2025)	NeuroFusion-ADNet: dual-path cross-attention network	sMRI, FDG-PET	Bi-directional cross-attention with modality-specific encoders (PWResNet, CADDLFCNet) for joint ROI segmentation and classification; Acc 99.48%, Dice 0.985 (NC vs. AD) on 381 ADNI subjects.
AbdelAziz et al. (2024)	SECNN-RF	sMRI	Squeeze-and-Excitation channel recalibration with a Random Forest classifier; Acc 99.89% (test) on balanced 4-class dementia stages, Kaggle dataset. ^a
Olaimat et al. (2023)	PPAD (RNN-based)	Cognitive scores, MRI-derived biomarkers, demographics (20 features; ADNI)	Handles irregular visit intervals by using patient age as an indicator of elapsed time; MCI-to-AD conversion validated on ADNI and NACC.
Mahim et al. (2024)	ViT-GRU hybrid model	sMRI (T1-weighted)	ViT for spatial dependencies combined with GRU for temporal correlations; Acc 99.53% (4-class, Kaggle) and 99.26% (3-class, ADNI); LIME and SHAP. ^a
Zahedipour et al. (2026)	GASHAP (3D-CNN + GA)	sMRI (3D T1-weighted)	Genetic algorithm ($\alpha=0.025$, $\beta=0.975$; 200 individuals, 500 generations) refines Deep SHAP heatmaps into anatomically interpretable brain masks; 85% test accuracy (96% training), CN vs. AD on ADNI.
Zubair et al. (2025)	EfficientNet + U-Net + SVM	sMRI + segmentation masks	Multi-layer U-Net gray-matter segmentation (Dice 91.52%) with multi-scale EfficientNet features and an SVM classifier; Acc $97.78\% \pm 0.54\%$ for 3-class staging on ADNI.
Alex et al. (2025)	3D-CNN + Fuzzy Logic	sMRI (hippocampal volume) + CSF biomarkers	Two-stage framework: 3D-CNN hippocampal segmentation plus fuzzy clustering of CSF biomarkers to handle clinical uncertainty; Acc 93.6% (NC vs. AD), 93.7% (MCI vs. AD) on ADNI.
Elmannai et al. (2025)	SSA-Net (self-supervised)	sMRI	SimCLR-style contrastive pretraining with a ResNet-50 backbone using attentive activation; Acc 99.7% vs. 98.1% for standard ResNet-50 on KACD. ^a
Priyadharshini et al. (2026)	ETT-SSL (TabTransformer)	Tabular clinical data (cognitive assessments, demographics)	Self-supervised TabTransformer with tokenization, SMOTE class balancing and SHAP-based feature selection to mitigate class imbalance; Acc 95.8%.

^a Results on Kaggle-hosted datasets should be interpreted with caution: data provenance and written informed consent are often not traceable for medical data shared via such platforms.

between normal aging and early pathology is not sharp, and where the same imaging finding can carry very different implications depending on context. Explainable AI (XAI) grew up under exactly this pressure, as a set of techniques for prising open the reasoning of machine- and deep-learning models so that a human can inspect it (Górriz et al., 2023, 2020). The open question is whether what these techniques return is

genuinely informative, or only the look of transparency without much behind it.

The two most common post-hoc methods, Grad-CAM (Selvaraju et al., 2020) and SHAP (Sylvester et al., 2024), both point at the parts of the input responsible for a prediction, one as a spatial map, the other as feature attributions. On brain scans that usually takes

the form of a color-coded overlay drawn over regions of interest. When they are compared head to head, Grad-CAM tends to be the more consistent and stable, locating hippocampal atrophy and medial temporal structures reliably. SHAP resolves attributions down to the voxel but pays for it in computation, whereas LIME (Local Interpretable Model-agnostic Explanations), whose explanations hinge on how the image is carved into supervoxels, is the most variable of the three (Qin, 2025). When these explanations point to regions that align with established neuropathological knowledge (e.g., as Song and Yoshida, 2024 demonstrated for the precuneus and hippocampus using 3D Grad-CAM on resting-state fMRI) they provide some reassurance that the model is learning something real rather than exploiting dataset artefacts. But biological plausibility in one cohort does not guarantee that the same regions will be highlighted consistently across different sites, scanners, or patient populations, which is a limitation that comparative evaluations do not always examine carefully enough.

More recent work has moved toward embedding interpretability into the model architecture rather than adding it afterwards. Ye et al. (2024) proposed MAD-Former, a transformer with a traceable interpretability module that uses attention-based selection and receptive field tracing to identify the 3D regions driving its decisions, including the fusiform gyrus as a contributor to AD recognition. Counterfactual reasoning represents a different approach. Oh et al. (2025a) synthesized structural MRIs depicting what a patient's brain would look like under a different diagnostic label, and used these to construct an AD-relatedness index for each brain region, a more intuitive form of explanation than a saliency map for many clinical users. Leonardsen et al. (2024) showed that layer-wise relevance propagation can produce individual-level explanation maps that carry prognostic value in longitudinal MCI cohorts, going beyond classification output to characterize how the disease manifests differently across patients. These are genuine advances. It is also important to be clear that most of them have been evaluated in controlled research settings, and their behavior in clinical workflows (with real time pressure, mixed-quality data, and users who did not design the system) remains largely untested.

Explainability is not confined to imaging. Zhang et al. (2025a) proposed a multi-task speech-based framework fusing acoustic, linguistic, and expert-knowledge features, leading to a classification accuracy of 89.58%. SHAP analysis revealed that slower speech rate, reduced syntactic complexity, and increased pause fillers are the most discriminative features for AD. Pan et al. (2025) introduced a two-step cross-attention system for speech-based detection that achieves competitive results across DementiaBank, ADReSS, and IVA, showing that self-supervised acoustic representations can outperform simpler fusion strategies when combined with linguistic features. On the longitudinal side, Kang et al. (2024) used a deep self-supervised framework to disentangle AD-specific neuroanatomical patterns from normal aging effects, identifying four subtypes with distinct biomarker profiles and two separate progression pathways. Work like this suggests that interpretability tools can do more than justify individual predictions: they can generate hypotheses about disease heterogeneity that purely predictive models would not surface. How robust those hypotheses are across different cohorts and acquisition settings is, again, a question the field has not answered systematically.

Recognizing that explanation methods need to be evaluated as rigorously as the models they are applied to, several groups have developed standardized benchmarking frameworks. Metsch and Hauschild (2025) built BenchXAI, a package evaluating fifteen post-hoc XAI methods across clinical, imaging, and biomolecular data, finding that Integrated Gradients, DeepLift, and GradientShap perform consistently well, while Guided Backpropagation and LRP- α_1 - β_0 struggle in certain tasks. Gomez and Mouchère (2023) provided a tutorial-style evaluation of saliency methods covering both faithfulness metrics and user protocols, with an open-source implementation. Foundational work on layer-wise relevance propagation (Montavon et al., 2018) continues to underpin much of this evaluation literature. The value of this kind of systematic

comparison is not only methodological: it is also a prerequisite for any serious conversation about which explanation methods are suitable for clinical use and under what conditions. Without it, the choice of XAI method remains largely arbitrary.

To ground these observations rather than leave them at the level of impression, Table 5 collects the quantitative evidence that is actually available for the three methods the debate usually centers on. The most directly relevant benchmark is that of Qin (2025), who ran Grad-CAM, Deep-SHAP and LIME through a single 3D ResNet-50 pipeline on structural brain MRI under identical preprocessing. Grad-CAM came out most reproducible, with a consistency score of 0.764 and a stability score of 0.684 on the dementia task (both above 0.90 on the easier tumour task), while LIME showed the most pronounced variability, a direct consequence of its reliance on super-voxel segmentation that does not respect cortical boundaries. Deep-SHAP produced the most anatomically detailed attributions but at a computational cost that makes it impractical on full 3D volumes. A complementary analysis by Martin et al. (2026) on the NACC cohort adds a twist worth noting: across network architectures LIME and plain backpropagation produced the most similar maps (Dice 0.67 and 0.64 between the coronal CNN and ViT models), while for models trained on sagittal slices no method retained appreciable overlap at all. The two findings are not contradictory, but together they caution against treating any single stability figure as definitive: a method can be volatile from run to run and yet comparatively robust to a change of backbone, and which property matters depends on how the tool is deployed. A third comparison reaches a similar conclusion from another angle: (Persson et al., 2026) map the attributions of DeepLIFT, Layer Gradient $\times$ Activation, occlusion sensitivity and XGrad-CAM onto anatomical regions in ADNI, and find that the methods converge on hippocampus and amygdala for cognitively normal subjects but diverge once the target class is Alzheimer's disease, while regional relevance correlates only weakly with amyloid and tau PET. On that evidence explanations are both method- and class-dependent, and are better treated as complementary than interchangeable.

One caveat should be made clear. Even the most systematic of these benchmarks rely in part on augmented or repository-hosted MRI collections rather than curated clinical cohorts, so the absolute scores in Table 5 are better read as relative rankings than as values that would transfer unchanged to a hospital population. That limitation is itself part of the point: the evidence base needed to choose an explanation method on clinical grounds is thinner than the sheer volume of XAI papers would suggest.

Table 6 summarizes representative studies applying interpretability approaches to AD detection and staging. Fig. 2 illustrates the principal XAI strategies used in dementia research, from gradient-based visualization to hybrid explainable frameworks.

5. Digital biomarkers and wearables

What is compelling about digital biomarkers is less the extra data than where the observation happens. Neuroimaging catches a patient at a fixed moment, under tightly controlled conditions. On the other hand, a wearable or a phone records behavior and physiology as they play out in ordinary life. Heart rate, respiration, gait, sleep, activity, speech, the way someone uses a touchscreen or finds their way around, these run continuously, sit close to real behavior, and can in principle register changes that a single clinic visit would miss. In dementia that matters, because early decline tends to be subtle and to come and go, and because what a patient manages in a testing room and what they manage at home are not always the same thing. The evidence that such signals start to move years ahead of diagnosis is by now fairly solid (Kouritis et al., 2019). Within RADAR-AD, remote monitoring has picked up changes in physical activity, sleep, and gait that may mark prodromal Alzheimer's disease (Lentzen et al., 2025). Speech gives another route in: slowing, prosodic shifts, and altered rhythm seem

Table 5
Quantitative comparison of the three most widely used post-hoc XAI methods on brain MRI, drawn from head-to-head neuroimaging benchmarks. Consistency and stability scores for structural MRI are from [Qin \(2025\)](#) (3D ResNet-50 trained on 2D slice-based datasets under identical volumetric preprocessing); cross-architecture agreement is from [Martin et al. \(2026\)](#) (NACC, CNN vs. ViT); anatomical failure modes draw additionally on [\(Chakroborty et al., 2025\)](#). Absolute scores partly reflect augmented or repository-hosted data and should be read as relative rankings.

Method	Consistency /stability on sMRI (Qin, 2025)	Cross-architecture agreement (Martin et al., 2026)	Run-time	Dominant failure mode in 3D neuroimaging
Grad-CAM	Highest of the three: consistency 0.764, stability 0.684 (dementia); >0.90 (tumour) (Qin, 2025)	Not among the most concordant; only LIME and backpropagation reached appreciable CNN/ViT overlap (Martin et al., 2026)	Negligible (single backward pass)	Coarse maps; activations spread into non-brain regions; Grad-CAM++ unstable under Gaussian-noise perturbation in DenseNet (Chakroborty et al., 2025)
Deep-SHAP	Anatomically detailed attributions, but not the most reproducible (Qin, 2025)	Below LIME and backpropagation, the only methods with appreciable CNN/ViT overlap (Martin et al., 2026)	High; exact Shapley values scale exponentially, forcing kernel approximations on 3D volumes	Computationally impractical at full 3D resolution; the kernel variant requires very large perturbation counts and failed to yield stable explanations on 3D MRI (Chakroborty et al., 2025)
LIME	Most pronounced within-run variability, attributed to super-voxel segmentation (Qin, 2025)	Highest, with backpropagation: Dice 0.67 vs. 0.64 (coronal CNN/ViT) (Martin et al., 2026)	Moderate (repeated perturbation sampling)	Thin super-voxel segmentation with linear local approximation yields fragmented, unstable regions across slices (Chakroborty et al., 2025)

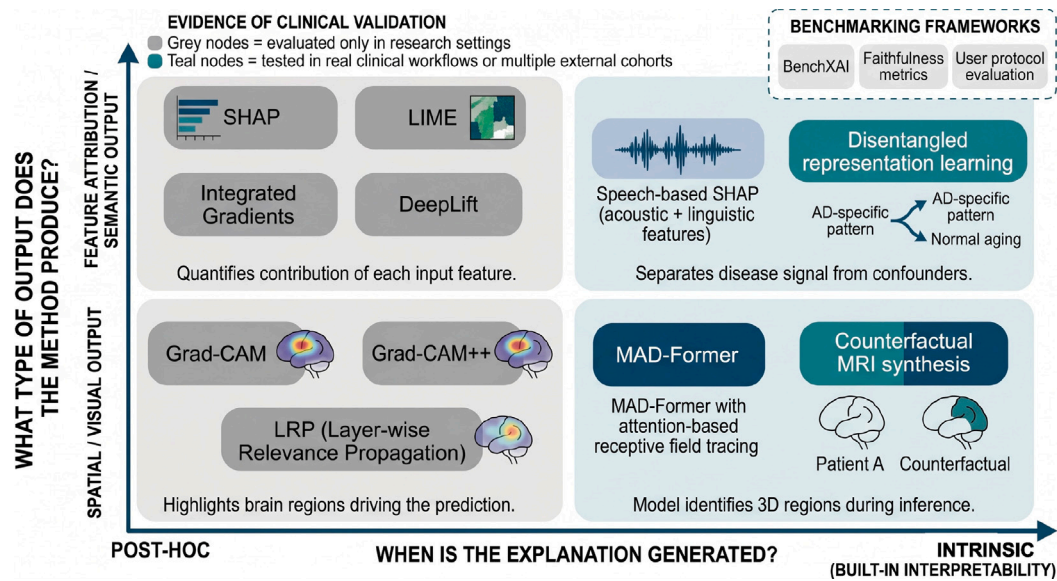


Fig. 2. Landscape of XAI methods for dementia detection, organized by timing (post-hoc vs. intrinsic) and output type (spatial visualization vs. feature attribution). Color intensity reflects clinical validation status: most methods have been evaluated in research settings only (gray); limited evidence exists from real-world clinical workflows (teal). Benchmarking frameworks are shown as a prerequisite for principled method selection.

to capture facets of decline that conventional tests measure poorly ([Oh and Park, 2025](#)), and can occasionally tell dementia subtypes apart ([Oh et al., 2023](#)). Longitudinal cohorts tell the same story from a different angle: activity, sleep, and circadian measures track with incident dementia risk ([Raichlen et al., 2023](#)), accelerometer patterns carry predictive weight for later diagnosis ([Zhao et al., 2025](#)), and UK Biobank work suggests wearable-derived features can add to, and

sometimes beat, the usual risk factors ([Chan et al., 2025](#)). That the picture holds together as well as it does, across such varied devices and populations, is itself worth something.

Smartphones push this further still, since they need no dedicated hardware. If you can read cognitive change off the way someone moves through a virtual space, handles a timed task, or taps a screen, then monitoring at scale over long stretches becomes far more practical. Sea

Hero Quest is still the clearest case. Built as a navigation game and released widely, it showed that spatial navigation can flag preclinical change and separate people by genetic risk for Alzheimer's disease even when standard neuropsychological testing turns up nothing (Coughlan et al., 2019). What earns that result attention is not the headline accuracy but what it says about sampling: cognition measured again and again, in natural conditions, over time, may hold information that a formal assessment misses for the very reason that it is formal. Nicosia et al. (2023) reported reliable and valid cognitive measures from smartphone-based assessment in at-risk older adults, which supports the practical feasibility of this kind of monitoring outside clinical settings, though feasibility and clinical utility are not the same thing and the distinction matters. Beyond assessment, a broad ecosystem of mobile-health applications has been developed to support people with dementia and their carers in daily life, and surveys of this space convey both its breadth and its unevenness (Yousaf et al., 2019). The machine-learning frameworks underlying continuous health-care monitoring are themselves an active area of development (Kurdi, 2026).

Still, this is one of those areas where the enthusiasm can outrun the evidence, and we think it is worth being precise about the actual state of the field. Standardization is limited, validation frameworks are still developing, and results across studies are difficult to compare because acquisition conditions, populations, and outcome definitions vary considerably (Iulita et al., 2025; Lott et al., 2024). Signals acquired in daily life are noisy and shaped by many factors that have nothing to do with neurodegeneration, which means that distinguishing the disease-relevant component from contextual variation is harder than the reported performance figures sometimes suggest. Most current digital biomarkers are better understood as promising candidates than as clinically established markers, and translational claims should be proportionate to that (Lentzen et al. (2025), Kourtis et al. (2019)). Progress will require not only better preprocessing and harmonization, but also analytical approaches that connect digital features to interpretable cognitive or behavioral mechanisms rather than treating the signal as a black box at a different level of abstraction. There is also a representational problem that deserves more attention than it typically gets. Many digital biomarker studies draw on relatively homogeneous populations, and a model trained under those conditions may not transfer to groups with different social, cultural, or technological backgrounds (Azizan et al., 2025). In smartphone- and wearable-based research specifically, access, adherence, and patterns of device use are themselves unevenly distributed, which means the populations most likely to be underserved by the healthcare system may also be the least represented in the training data. For us, the central question is not whether these technologies can contribute to dementia monitoring (they clearly can). The harder question is whether they will do so in a way that is robust, interpretable, and equitable enough to deserve clinical trust. That is not yet answered.

Assessing added value fairly means comparing these signals with established neuroimaging on equal terms, and that turns out to be harder than it looks. Most studies report an AUC from a single cohort, a single task and a single validation scheme, and placing those numbers side by side invites a false equivalence that is easy to fall into and difficult to justify. Table 7 keeps what can be compared apart from what cannot. The clearest evidence comes from studies that acquire several modalities in the same participants. In one such study, Caro et al. (2026) evaluated 33 patients with AD and 33 controls and found that a speech-based classifier reached an AUC of 0.84, statistically indistinguishable from structural MRI (0.90) and from episodic memory testing (0.90), and clearly above functional connectivity (0.65). Within a single sample, then, speech carried roughly as much diagnostic information as a structural scan. Across studies the picture is more measured: pooling 21 cohorts, Jafari et al. (2025) report a speech AUC of 0.78 (95% CI 0.70–0.86) for separating MCI from unimpaired controls, moderate rather than striking, but obtained without imaging, contrast agents or a hospital visit.

Gait and wearable signals belong in a different column, and it is worth resisting the temptation to merge them. Their evidence is prognostic rather than diagnostic: they indicate who is likely to decline, not who already carries a given pathology. Slow gait speed is associated with incident all-cause dementia (hazard ratio 1.53, 95% CI 1.42–1.65) but only weakly with Alzheimer's disease in particular (HR 1.03), tracking vascular and non-AD dementias more closely (Beauchet et al., 2016). Wearable-derived activity behaves similarly: in more than 90,000 UK Biobank participants, the amount and timing of daily movement predicted incident dementia over almost eight years of follow-up, with a hazard ratio of 0.44 for AD in the most active morning group (Ning et al., 2025). These are real associations, but they are hazard ratios accrued over years rather than areas under a diagnostic curve, and reading them as interchangeable with an imaging AUC would misstate both. The relevant non-imaging benchmark for diagnosis is molecular: plasma p-tau217 now outranks other plasma assays across the disease continuum (Chen et al., 2026), and sets the standard any digital biomarker with diagnostic ambitions will eventually have to meet.

Taken together, these numbers suggest that digital signals do not displace neuroimaging, and for the most part are not intended to. Their contribution is complementary and temporal: speech approaches imaging-level diagnostic performance at a small fraction of the cost, while gait and wearables extend the window of observation backwards, marking risk before a scan would ordinarily be requested. What remains missing, and what the gaps in Table 7 make visible, is a body of work that measures digital and imaging markers in the same individuals under a common validation protocol. Until that exists, comparisons across modalities are better treated as indicative than as settled.

Table 8 summarizes representative recent studies in this field, and Fig. 3 provides an overview of the main wearable- and mobile-derived sources of digital biomarkers for dementia monitoring and prediction.

6. Clinical applications and translational impact

It is already known how to build dementia AI models that look convincing on paper. What remains unsolved is how to make them useful in practice. That distinction is not a minor one. A tool is clinically useful not when it achieves a high AUC on a held-out test set, but when it answers a question a clinician actually has, in a form they can act on, within the constraints of how care is actually delivered. By that standard, most dementia AI systems are still at an early stage. In most countries, these tools remain confined to pilot studies or research settings, and in many they have not yet entered serious implementation discussions (Rahamtalla et al., 2025). That gap between technical achievement and clinical adoption has been documented repeatedly, with regulatory, organizational, and workflow-related barriers consistently identified as the reasons why uptake remains slow (Ahmed et al., 2023; Nair et al., 2024).

The gap persists partly because research and clinical deployment reward different things. A tool that lightens a radiologist's load and raises diagnostic rates is plainly worth having, but only when it is built to back the clinician's judgment rather than stand in for it. The distinction is anything but academic. Flag a subset of scans for a second human look and the clinician stays inside the decision loop, with outcomes improving at the margin. Have the model issue the primary diagnosis and leave the clinician to catch its mistakes afterward, and the relationship flips: the human now works for the system, carries its liability, and the system keeps the efficiency for itself (Suleman et al., 2025). That second arrangement can be attractive on the balance sheet and still be wrong clinically and ethically, and it goes some way toward explaining why clinicians remain wary of AI tools even when the numbers look good.

The pull of earlier detection in Alzheimer's disease and related dementias is genuine. Catching risk sooner can improve the timing of treatment, allow care to be planned ahead rather than in reaction, and

Table 6
Comparative analysis of high-impact XAI research in the study of AD.

Ref.	XAI technique	Modalities	Main results and notes
Lozupone et al. (2024)	Soft Attention	sMRI	AXIAL: Attention-based voxel-level maps from 2D CNNs on 3D MRI; Acc 85.60% and MCC 0.712
Alsaleh (2025)	LIME/Attention	MRI + FDG-PET	(HC vs. AD) on ADNI; highlights hippocampus, amygdala, parahippocampal gyrus, and inferior lateral ventricles; demonstrates high spatial consistency across cross-validation runs. NeuroFusion-ADNet: Dual-path fusion network with joint ROI segmentation and classification; Acc 99.48% and Dice 0.985 (HC vs. AD); attention heatmaps consistently focus on hippocampus, entorhinal cortex, and basal ganglia.
Song and Yoshida (2024)	Grad-CAM (3D)	fMRI	Grad-CAM applied to 3D-VGG16-GAP on resting-state fMRI activity maps (ALFF, ReHo, fALFF, VMHC); maximum Acc 96.4% (average combined Acc 87.90%); highlighted precuneus and hippocampus as key ROIs for AD subjects; GAP layer preserves spatial coordinate integrity.
Al-Bakri et al. (2025)	SHAP + Grad-CAM	sMRI + 14 clinical features	HXAI framework combining clinical variables (MMSE, nWBV, SES, age) and sMRI ResNet50 features via logistic regression meta-model ensemble; 83.61% explainability accuracy for SHAP feature-based removal tests and 58.16% for Grad-CAM visual explanations.
AbdelAziz et al. (2024)	Adaptive attention	sMRI	SECNN-RF: Squeeze-and-excitation blocks for channel-wise feature prioritization combined with Random Forest classifier; Acc 99.89% on RGB normalized images; integrates saliency maps for local visual explanations.
Leandrou et al. (2023)	SHAP	MRI radiomics	XGBoost with SHAP on hippocampal and entorhinal texture features; F1-score 0.949 (NC vs. AD), 0.818 (NC vs. MCI), 0.810 (MCI vs. AD); capturing texture patterns beyond standard volumetric measurements.
Jahan et al. (2023)	SHAP/LIME	Clinical, Psychological, MRI	Five-class prediction (Cognitively Normal, AD, Non-AD Dementia, Uncertain Dementia, Others) using data-level fusion on OASIS-3; RF classifier achieved 10-fold CV Acc 98.81%; local explanations compared via LIME and SHAP to ensure attribution stability.

(continued on next page)

Table 6 (continued).

Yi et al. (2023)	SHAP	Clinical, neuropsychological, sMRI, genetics	XGBoost-SHAP framework for NC vs. MCI vs. AD classification incorporating APOE- ϵ 4 status, CDRSB, and ADAS scores; achieved Acc 87.57% (AUC 0.91) on ADNI and Acc 80.52% (AUC 0.88) on external validation (NACC) under extreme class imbalance.
AlMansoori et al. (2024)	SHAP	Genetic + clinical + blood biomarkers	Support Vector Machine with Mutual Information feature selection trained on blood gene expression, SNPs, and clinical data; achieved Acc 95.00% (0.95) and AUC 0.94 (CN vs. AD/MCI spectrum) on 623 ADNI participants; SHAP identified genetic biomarkers linked to axon myelination.
Chattopadhyay et al. (2024)	Grad-CAM/OSA	T1-weighted MRI	DenseNet architecture with Grad-CAM and Occlusion Sensitivity Analysis (OSA) on pre-processed T1w MRI scaled to $91 \times 109 \times 91$ isometric voxel grid; feature localization consistent across different cohorts; resolves features at different scales.
Bhattarai et al. (2024)	Deep-SHAP	MRI + PET	Deep-SHAP mapping multivariate relationships; sMRI identified insula, lateral occipital, temporal pole, and occipital fusiform gyrus; PiB-PET identified inferior temporal, parahippocampal, and lateral frontal gray matter as key predictors of cognitive decline.
Zahedipour et al. (2026)	Genetic Algorithm SHAP (GASHAP)	sMRI	Novel GASHAP framework integrating Genetic Algorithms with Deep SHAP to generate localized brain masks for 3D-CNN; genetic algorithm initialized with SHAP scores; 3D-CNN classification of AD vs. cognitively normal controls reached 85% test accuracy (96% on training data), narrowing the explanation to 11 brain regions.
Chakroborty et al. (2025)	Multiple gradient-based and model-agnostic interpretability methods	sMRI	Structured evaluation of 8 XAI methods (gradient-based and model-agnostic) on 3D DenseNet, ResNet-18 and a Swin hybrid; assessed via faithfulness (model-reinitialization sanity check), robustness (Gaussian noise and blur), and alignment with known AD regions; Guided Backpropagation confined saliency to the brain, whereas Grad-CAM++ proved unstable.
Bloch et al. (2026)	SHAP, High-level radiomics explanations	MRI radiomics + clinical	Time-to-event survival models with Radiomics features; SHAP revealed complex microstructural associations including tissue texture in left entorhinal cortex, irregular shape of right amygdala, and fine texture in left middle temporal gyrus; validated on ADNI, AIBL, and OASIS-3.

Table 7
Quantitative comparison of digital biomarkers against established neuroimaging and molecular references. AUCs from different cohorts are not directly comparable; hazard ratios (HR) are prognostic (future risk) rather than diagnostic. A P-score is a ranking probability from a network meta-analysis and is not a measure of diagnostic accuracy. The gaps reflect the scarcity of same-sample, same-protocol head-to-head studies.

Modality	Representative quantitative result	Task/cohort	Metric and comparability
Speech (same- sample benchmark)	AUC 0.84 vs. structural MRI 0.90 (n.s., $p > 0.14$); fMRI 0.65	AD vs. HC; 33/33 (Caro et al., 2026)	AUC — directly comparable (same sample)
Speech (meta-analytic)	AUC 0.78 (95% CI 0.70–0.86)	MCI vs. CU; 804/1134, 21 studies (Jafari et al., 2025)	AUC — pooled; cohorts differ from imaging
Gait speed	HR 1.53 (95% CI 1.42–1.65) any dementia; 1.03 (95% CI 1.01–1.05) AD; 1.79 (95% CI 1.51–2.12) vascular	Incident dementia (Beauchet et al., 2016)	HR — prognostic, not diagnostic; weak for AD
Wearable activity	HR 0.44 (95% CI 0.22–0.86) AD; 0.60 (95% CI 0.40–0.90) all-cause (morning activity)	Incident dementia; UK Biobank $n = 90,329$ (Ning et al., 2025)	HR — prognostic over ~7.9-year median follow-up
Structural MRI (reference)	AUC 0.90 (95% CI 0.81–0.95)	AD vs. HC; same sample as speech (Caro et al., 2026)	AUC — imaging reference
Plasma p-tau217 (non-imaging reference)	Top-ranked assay (mass spectrometry): P-score 0.86 ($A\beta$ detection), 0.91 (Tau-PET staging), 0.82 (progression)	AD continuum; network meta-analysis (Chen et al., 2026)	Ranking (P-score) — not comparable with AUC or HR; molecular benchmark

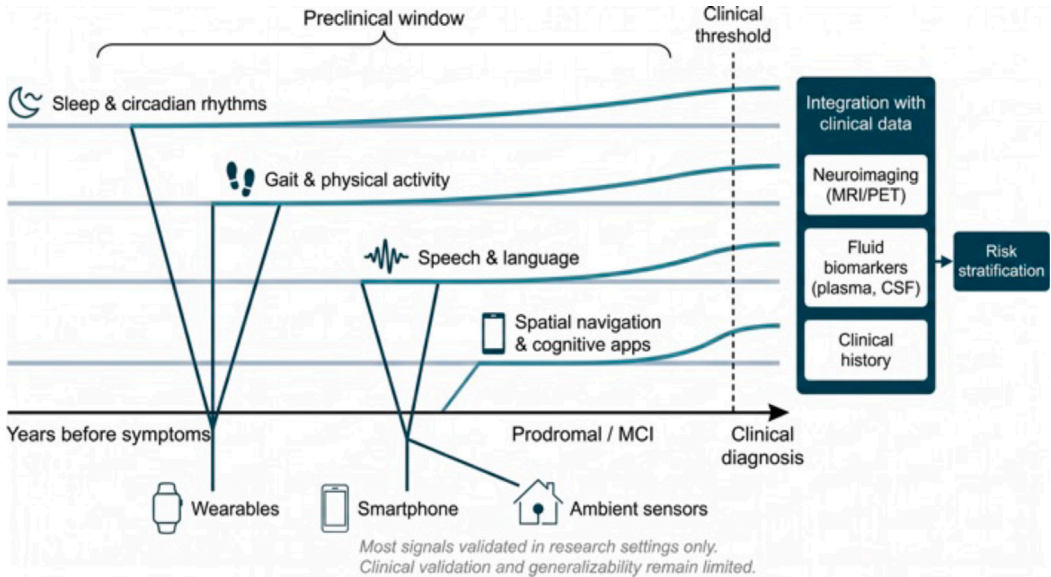


Fig. 3. Temporal landscape of digital biomarkers for early dementia detection. Signal tracks illustrate when behavioral and physiological measures begin to deviate from baseline relative to clinical diagnosis. The output is framed as risk stratification in support of clinical decision-making. Most signals have been validated in research settings only; clinical generalizability and equity of access remain unresolved.

make better use of scarce specialist time, all the more so as disease-modifying therapies mature (Khude and Shende, 2025; Davis et al., 2023). Tie AI into electronic health records and other patient-level data and it can help shape more individualized care pathways and better-informed decisions at the level of the whole system (Saxena and Khan, 2025b; Khachaturian et al., 2023). Socially assistive robotics and affective computing show a different kind of worth again, using emotion-aware models to lift engagement and quality of life where the relational side of care counts for as much as diagnostic accuracy (Hsieh et al., 2023). Every one of these gains, though, rests on a single

condition: the system has to bend to the clinical setting it serves, not the reverse (Wang et al., 2023). Tools built without real input from the clinicians meant to use them, or delivered in a form that does not match how decisions actually get made, seldom get past the demonstration stage whatever their technical merit. That is why clinicians cannot be brought in only at the end. Leave physicians out of the design and the result can be accurate and still fail, because it answers the wrong question, presents its output unusably, or does not fit the tempo of routine work (Kapadia et al., 2024). Trust depends on clinicians grasping what the system is doing and why it reached

Table 8
Comparison of recent studies for dementia prediction using digital biomarkers and wearables.

Ref.	Biomarker domain	Technology	Main results and notes
Butler et al. (2025)	Cognitive/activity	Apple Watch/iPhone	Remote observational study of 23,004 adults over 24 months; proof-of-concept MCI classification from routine device use, with ~52% adherence to cognitive assessments.
Rigny et al. (2024)	Sleep architecture	Under-mattress sensor	Unsupervised clustering of NREM, REM, light-sleep and WASO features identified three sleep phenotypes, one predominantly of people living with dementia.
Lima et al. (2025)	Voice/speech	Smartphone audio	Lexical markers of analytical thinking and disfluency used for automated risk stratification.
Badawi et al. (2025)	Physiological/motion	EmbracePlus wristband	Real-time prediction of agitation episodes at least 6 min in advance via multimodal biometric fusion.
Lentzen et al. (2025)	Multimodal functional	Fitbit/Axivity AX3	Tracking of real-world functional decline; AUC 73.0% for prodromal AD identification.
Wei et al. (2025)	Cardiovascular/HRR	ECG (UK Biobank)	Heart rate response/recovery index from a submaximal exercise test; higher HRR associated with reduced all-cause dementia risk (HR 0.72, 95% CI 0.54–0.94) in 46,348 participants.
Tian et al. (2025)	Spatial navigation	Serious game (iPad)	Mini-SPACE serious game for unsupervised screening; prediction of MoCA scores improved with repeated measures.
Bafaloukou et al. (2025)	Multi-sensor passive	Sleep/light/temperature sensors	Interpretable ML (LightGBM); key predictors of agitation were low nocturnal respiratory rate, heightened alertness during sleep and increased indoor illuminance.
Chernov (2025)	Handwriting/motor	Digital tablet	Longitudinal analysis of 41 handwriting features identified markers for early detection of CI/AD.
Bayat et al. (2021)	Behavioral/driving	GPS sensors	GPS-based driving patterns serve as a scalable and non-invasive digital biomarker for preclinical AD.

a given conclusion (Dymm and Khalid, 2025): a model that cannot be questioned is hard to fold into a decision process that stays, rightly, a clinical responsibility. And deployment is not the finish line. A system that performs well at validation can drift as documentation habits, acquisition protocols, or institutional conditions shift, so it has to be watched for that drift before it does harm (Wells et al., 2025).

Framing AI as a source of diagnostic readouts also understates the range of roles it can play. Some of the most plausible near-term contributions are supportive rather than diagnostic, and they fit the “system for the clinician” model more naturally than autonomous classification does. Triage is the clearest example: convolutional models that flag likely-abnormal head MRIs immediately after acquisition can reorder the reporting queue, and in one large deployment halved the reporting time for abnormal scans (from 28 to 14 days at one network, 9 to 5 at another) across more than 70,000 examinations (Wood et al., 2022). In primary care, low-cost speech screening can play the analogous role of deciding who warrants confirmatory imaging rather than making a diagnosis at all. Conversational assistants are a second role: large language models can summarize records for clinicians and translate

instructions for caregivers, though evaluations in dementia care find accuracy and safety still under-characterized and hallucination a real hazard for vulnerable users (Oh et al., 2025b). Automated report generation is a third: on a multicentre benchmark of 4293 brain-MRI reports, a large language model placed the correct diagnosis among its top three suggestions for 97.6% of patients when working from structured findings, though that figure drops to 87.1% once it must commit to a single diagnosis (Wang et al., 2026b). The same class of models is diagnostically unreliable when asked to read the scans themselves: a systematic audit of seven vision-language models found fabricated or misidentified content in 74.4% of readings, and MRI was the modality they handled worst, at 13.9% diagnostic accuracy (von der Stück et al., 2026). The pattern is consistent: these systems are useful when they support a human who remains accountable, and unsafe when trusted to act alone.

Cross-modal synthesis deserves separate mention because it is where the supportive framing is most tempting and most fragile. Generating PET from structural MRI would, if it worked, bring molecular readouts to settings without a cyclotron, and recent diffusion-based models

produce visually convincing multi-tracer images (structural similarity above 0.87) with measurable gains when combined with MRI for classification (Wang et al., 2026a). But amyloid and tau accumulate years before the structural changes MRI captures, which makes the MRI-to-PET mapping mathematically ill-posed: a model trained across disease stages tends to regress toward an average that can invent or erase focal pathology, and standard image-quality metrics do not detect this (Baron et al., 2026). The risk is concrete enough that nuclear-medicine bodies have begun to formalize it, defining AI-fabricated abnormalities that are visually plausible but factually false as a distinct failure class requiring governance (Xia et al., 2025). Synthetic modalities may still have a place, particularly when conditioned on independent molecular signals such as plasma biomarkers, but they illustrate the general point of this section: the value of these tools depends less on what they can technically produce than on whether their outputs are trustworthy in the hands of the people who must act on them.

It is worth being concrete about what has actually reached the clinic, because the regulatory record gives a sharper picture of AI's current role than performance figures do. The AI tools cleared for use in dementia are, almost without exception, quantitative and assistive rather than diagnostic. A number of automated brain-volumetry packages (NeuroQuant, icobrain, Quantib ND and others) hold FDA clearance or CE marking. They segment structures such as the hippocampus and report volumes against age- and sex-matched normative databases, and a systematic review of these commercial tools confirms that their value lies in standardizing measurement rather than rendering a diagnosis (Pemberton et al., 2021). Lesion-detection software has followed the same pattern: icobrain aria, cleared by the FDA in 2024, is the first tool approved to detect and grade amyloid-related imaging abnormalities in patients on anti-amyloid antibodies, and in its validation study it raised radiologists' sensitivity for ARIA-E from 71% to 87%, without removing the radiologist from the decision (Sima et al., 2024). In every case the approved system measures or flags, but the clinician still decides.

Direct evidence of the disease itself still comes from methods that are not AI at all. Amyloid PET tracers (florbetapir, flutemetamol and florbetaben) and the tau tracer flortaucipir have been approved for a decade and bind the pathology directly, and their diagnostic impact is well documented (Jovalekic et al., 2024). The most consequential recent change is not an imaging algorithm but a fluid assay: in May 2025 the FDA cleared the first blood test for Alzheimer's diagnosis, a plasma pTau217/A β 42 ratio with reported concordance around 92% with amyloid PET (Hu et al., 2025). Where AI has won approval, its job is to quantify structure and help with interpretation. Where what is needed is a direct read of the pathology, the approved tools are molecular. That is less a failing of AI than an honest map of where each approach has so far earned clinical trust, and it echoes the through-line of this section: the systems that make it into practice are the ones that support the clinician, not the ones that try to supplant the evidence clinicians work from.

What holds back routine use is rarely one thing; it is several, and they feed on each other. Generalizability draws the most comment: a model tuned to one set of technical and organizational conditions can behave quite differently once it is moved to a center with other imaging pipelines, other diagnostic protocols, or a different patient mix (Kelly et al., 2019). In dementia that kind of variability is the rule, not the exception. Everyone agrees multicenter validation is needed, and yet governance rules, legal constraints, and poor interoperability between systems keep getting in its way (Rieke et al., 2020). Regulation piles on another layer: showing that a system is safe, robust, and reproducible across settings is increasingly the expectation, but it is hard to pull off anywhere except a tightly controlled environment (Liu et al., 2020). The honest reading is that translation is not a technical problem waiting on a technical fix. It turns on institutional coordination, on regulatory clarity, and on a willingness to build systems around how clinical practice actually works rather than around the logic of a research

pipeline (Topol, 2019). In the end, how health systems choose to deploy these tools, and whose interests that deployment mainly serves, will count for as much as anything the tools can technically do.

Table 9 summarizes representative studies that address the gap between AI research and real-world clinical application in Alzheimer's disease and related dementias.

7. Ethical and translational challenges

Primary among all the ethical issues around the deployment of AI tools for AD diagnosis, or indeed any other predictive technology based on machine learning, is the manner in which it is implemented. First is the context. MRI scans are a relatively expensive examination involving a high-technology resource of limited capacity. For this reason alone, MRI would not be used as a population screening technique to identify pre-symptomatic individuals, something that the typical case-control comparison approximates and commonly appears in research articles of prototype AI systems. Instead, in the neurodegenerative context, MRI is usually reserved for patients who already have cognitive difficulties to the extent that have prompted them to seek medical advice. Here MRI might be used to exclude serious immediately life-threatening illness, such as brain tumours or stroke, for differential diagnosis, or prognosis of cognitive outcomes. When developing an AI system, researchers should articulate the application in mind and what value this might have to patients within existing health services and design their experiments accordingly.

Success at increasing diagnostic rates comes with its own considerations. We can all recognize the appeal of estimating future dementia risk, especially years before it becomes distressing. The emotional impact and influence on everyday life that such a prediction might have on a patient and their family is greater than articles published in technical journals acknowledge. A probabilistic prediction is only an estimate, yet for patients and families it may alter how the future is imagined, planned, and feared, and researchers should also provide measures of uncertainty in the prediction to allow accurate and proportionate communication of an elevated risk of further cognitive decline in some future period. These problems are compounded when there is no prospect of a disease-modifying treatment, as is the current case with dementia, or when waiting-lists for treatments are already at capacity, with further referrals, not all of whom will develop cognitive deficits, only serving to increase pressure on health services. Excessive emphasis on early detection for prevention can also threaten the provision of reactive capacity to treat those with disease (Nuffield Council on Bioethics and Predicting, 2024).

What truly sets AI systems apart from statistical models of empirical data already used to make diagnoses and prognoses is their absolute dependency on the datasets on which they are trained. The likelihood of developing dementia is best assessed from multi-domain data combining peripheral blood biomarkers, genetics, cognitive testing, and imaging where available. In a hospital or clinic, these disparate sources are accessible and cross-referenced by electronic health record systems (EHRS) designed and procured to actively promote data security and patient confidentiality. AI systems need free access to large datasets for training. To make clinical data available requires its transfer from the intentionally restrictive EHRS to more accessible storage. And this is not a single event. Procedures evolve and acquisition devices are upgraded or replaced, leading to data drift where models become unreliable over time due to a divergence between the statistical properties of the training dataset and the day-to-day data on which clinical decisions are made. Detecting data drift by a worsening of patient outcomes is obviously undesirable and continuous monitoring and model retraining on the most recent data is the safest, albeit most resource-intensive, approach. The risk to patient privacy is therefore elevated by the frequent transfer of patient data away from the designed-in security of the EHRS.

Table 9

Comparison of recent studies for clinical application and translational impact in dementia care.

Ref.	Application focus	Setting	Key translational outcome
Saxena and Khan (2025b)	Clinical decision support	Hospital workflow	Multi-omic, clinical, and environmental data integration framework supporting personalized care delivery and administrative resource mobilization.
Alex et al. (2025)	Multi-biomarker detection	Clinical memory care	Hybrid model combining 3D-CNN segmented hippocampal sMRI and fuzzified CSF biomarkers achieved 93.7% accuracy for MCI vs AD classification.
Bafaloukou et al. (2025)	Agitation monitoring	Home-based care	LightGBM model tracking passive in-home sensors combined with a traffic-light alert system achieved 90.3% specificity, supporting non-pharmacological interventions.
Lukkien et al. (2024)	Responsible co-design	Collaborative R&D	Identified six conditions for responsible design and deployment: protecting privacy, preventing cognitive overload, providing transparency, empowering caregivers to be in control, safeguarding accuracy, and training users.
Lekadir et al. (2025)	FUTURE-AI guidelines	Multi-institutional collaboration	International consensus establishing six core principles and 30 best practices for safe, trustworthy, and deployable AI in general healthcare.
Khude and Shende (2025)	Personalized drug therapy	Healthcare IT systems	Narrative review of AI-driven clinical decision support for medication selection and personalized drug therapy, integrating multi-omic data to predict adverse drug reactions and reduce trial-and-error prescribing.
Lima et al. (2025)	Spoken language screening	Telemedicine/home monitoring	Random Forest on linguistic features enabling scalable screening and risk-stratified triage; 69.4% sensitivity and 83.3% specificity on the main cohort ($N = 291$), falling to 70.0%/52.5% on a small in-residence pilot ($N = 22$).
Zajac et al. (2025)	Clinician-in-the-loop AI	Global radiology practice	Identification of five visions of AI support grounded in radiology practice across the Global South and North, prioritizing professional expertise and contextual awareness.
Borty et al. (2025)	IoMT-based diagnosis	Remote monitoring	Optimized ensemble of ViT-B16, ResNet152V2 and EfficientNetV2B3 achieving 97.31% accuracy on the OASIS MRI database (obtained via Kaggle) for IoMT-based services. ^a
Saxena and Khan (2025a)	Intelligent chatbot	Patient-centered care	Practicality assessment of a proposed NLP conversational agent to support patients and informal caregivers in daily care.

^a Results on Kaggle-hosted datasets should be interpreted with caution: data provenance and written informed consent are often not traceable for medical data shared via such platforms.

Anonymization of patient data does not mitigate against triangulation attacks when the datasets contain such diverse and extensive sources of information which can be combined in ways that are unique to individuals. In short, AI systems cannot operate safely long-term if they do not have continuous access to recently acquired clinical data, requiring a fundamental rethinking of the principles of EHRS design and associated data access procedures.

Arguably, most critical to the successful and sustainable deployment of AI systems is how they are embedded within the existing care pathways and the multidisciplinary team that manages the patient's care. It is important to recognize here the potential tension between marketing and sales imperatives of AI system vendors and the best interest of patients and practitioners. Vendors have a legitimate need to sell systems to customers (mostly healthcare providers) and so must find the most effective means of motivating the purchase of an AI system over other possible options. Consider the following fictional vignette: a hospital is offered an AI system for the automated reading of MRI to determine dementia. The system might be implemented in a way that it asks the human radiologist to take a second look at certain scans and in doing so improves diagnostic rates by 5% but reduces the number of scans seen per day by 1%–2%. Most people would agree that this is desirable and is an improvement over current practice. However, the system could also be implemented in a way in which the AI makes the diagnoses, and it is the human radiologist who must a posteriori re-read scans when there are requests for a second opinion or other information contradicts the outcome. Now the human is working for the AI and taking the responsibility for its mistakes, rather than the vendor.

In times of financial insecurity, there is clear reasoning for the second approach in that a smaller number of highly paid human specialists are needed, which more than offsets the cost of the AI system, in contrast to the augmentation of human specialists which improves patient outcomes but increases overall spending. Of course, there are situations where an AI system might offer new capabilities beyond those currently achievable with human specialists, but the manner of implementation remains crucial for sustainable changes to procedures that lead to better outcomes for patients and no deterioration in the working conditions of practitioners. Researchers have some culpability in fostering the conditions where these kinds of decisions are motivated by the replacement rather than the augmentation of humans (Anderson et al., 2024). Research outputs on prototype AI systems are frequently focused on achieving a higher detection rate or equivalent than an “expert” human operator, usually in an opportunistic dataset without the demographic complexity of the target clinical population. This adds to a general sense that AI systems easily surpass the performance of human operators and the subsequent fallacy of an appeal to authority is difficult to avoid during procurement. Relatively little work has been done on combining AI system outputs with human operators who are also susceptible to appeal to authority fallacy when interacting with AI systems (Bernstein et al., 2023; Yu et al., 2024b).

Accountability for the decisions of AI systems remains unresolved, and the rhetoric is often cleaner than the reality. A frequent narrative is that AI will “support” clinicians rather than replace them, although such statements do not solve the practical question of who takes responsibility when an error occurs (Gerke et al., 2020; Sujun et al., 2023; Price et al., 2024; Borchert et al., 2023). How responsibility for decisions is apportioned is unclear but is shared by the attending clinician, vendor, and the institution in which it is operating. Clarity is needed here to ease the translation pathway. The Medical Device Regulation (MDR, 2017/745) and the Artificial Intelligence Act (Regulation (EU) 2024/1689) have established stricter obligations for high-risk systems, including clinical decision-support tools, particularly in relation to data governance, human oversight, and post-market surveillance (Busch et al., 2024; Aboy et al., 2024). Nevertheless, an AI system compliant with regulatory requirements may still have failure modes, and it is

essential that researchers effectively communicate these limitations to build and then sustain confidence in the application of the AI system.

Outright failures, though important to clearly articulate, are not necessarily the most insidious errors generated by AI systems. These systems are trained on data, and from the data they may discover patterns that the human eye is incapable of detecting. The danger is that these patterns may not be related to the disease, but to artefacts or other physical characteristics that covary with its incidence. More subtly, training datasets, even when obtained from the clinics in which the AI system might subsequently be deployed, will not accurately represent the demographic profile of those who suffer from the disease (Seyyed-Kalantari et al., 2021). Socioeconomic status, ethnicity, age and gender are all factors in how late an individual may seek help when concerned about symptoms (Scott and Hoskin, 2024; Ho et al., 2025; Wilm et al., 2024; Hogg et al., 2023), and this is reflected in the datasets. Furthermore, the demographic composition of communities geographically varies. AI systems trained on datasets where there is, for example, an underrepresentation of early disease associated with particular demographics can result in false-negatives in these individuals, exacerbating existing health inequalities. Solutions require identifying uneven health outcomes across communities and then sampling larger datasets to produce training datasets suitable locally, although the long-term solution is the promotion of health literacy for everyone, but particularly in hard-to-reach communities.

Fig. 4 draws these threads together across the deployment lifecycle, marking where risk enters the pipeline, the structural tensions that follow from it, and the conditions that responsible deployment requires. AI systems have enormous potential in narrow, specialist roles to improve the health outcomes of patients. Ageing populations make dementia and other neurodegenerative disease a priority. Careful consideration of the purpose of a putative AI system, the target populations, and how it is embedded into existing care pathways is necessary at the very beginning of algorithm development and should be part of the academic reporting.

8. Conclusion

The literature reviewed here suggests that dementia research no longer lacks AI methods. What it still lacks, in many cases, is a clear path from methodological success to dependable clinical use. That is perhaps the most important conclusion to draw, and it is worth stating plainly rather than burying it in caveats. AI is now applied well beyond isolated classification tasks: multimodal models, longitudinal architectures, digital biomarker research, and decision-support systems are all active areas, and the scope of what is being attempted has broadened considerably. In a disease that unfolds over time, affects multiple domains, and is only partially captured by any single source of data, that broadening is not a research indulgence, but a reasonable response to the nature of the problem.

The field has also reached a point where better numbers on benchmark datasets are no longer sufficient to count as progress in any meaningful sense. The literature repeatedly presents models with strong performance under controlled conditions, while questions of external validation, interpretability, and routine usability remain partly unresolved. A model that performs well in a curated cohort but cannot be trusted across sites, or cannot be understood by the clinicians expected to use it, has limited value regardless of its accuracy. Robustness, clinical usability, and interpretability should now be treated as central criteria for evaluating work in this field, not as secondary refinements to be addressed once the model is built.

Multimodality deserves a closer look, not for solving these problems but for being truer to what dementia is. Neuroimaging is not going anywhere, yet speech, gait, sleep, wearable signals, and everyday behavior are stretching both the timeline and the settings over which the disease can be observed. The aim is not to crown one dominant biomarker but to work out how these different kinds of evidence fit together in

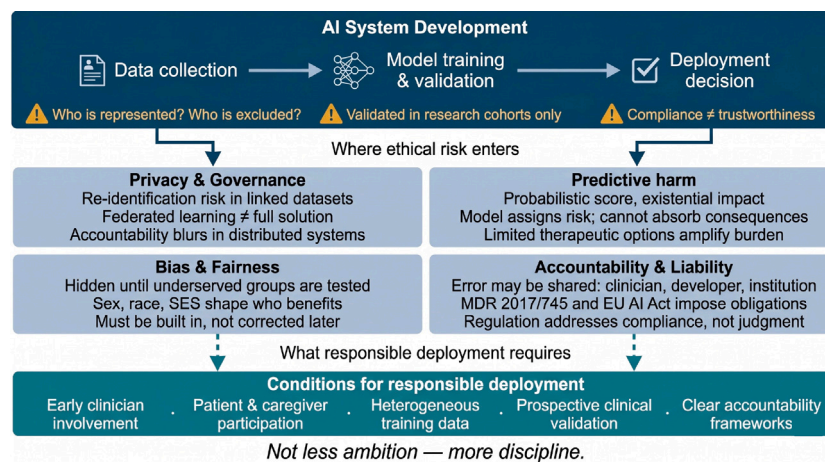


Fig. 4. Ethical and translational challenges across the AI deployment lifecycle in dementia research. Warning indicators mark where risk enters the pipeline: representational gaps in data, overconfidence in research only validation, and conflation of regulatory compliance with trustworthiness. Four structural tensions are shown: privacy and governance, predictive harm, bias and fairness, and accountability and liability. Required conditions for responsible deployment are listed at the base. Closing statement reflects the normative position of this article.

ways a clinician can use. Plenty stands in the way: standardization is thin, sharing data across centers is still awkward, acquisition protocols differ, and in most places the frameworks for validation are not mature enough to underwrite confident translation. The ethical and regulatory questions cannot be peeled away from the methodological ones. Privacy, fairness, accountability, and the responsible handling of data collected around the clock are not bolt-on concerns; they are part of what decides whether any of this is credible enough to put into practice.

Beneath all of it sits a question the field tends to skirt: who is the deployment of these tools mainly meant to benefit? A system that helps clinicians decide better is not the same as one that trims institutional costs by shifting clinical responsibility onto the people least able to push back. Medicine could go either way with AI, and the algorithm does not settle which; that is settled by how deployment is organized, governed, and judged. If success is measured by patient benefit rather than throughput, the evidence bar ought to be set accordingly. Our optimism about what AI can bring to dementia research and care is real, but it comes with conditions: models that survive outside controlled settings, explanations clinicians can genuinely use, translational strategies that take the messiness of real health systems seriously, and deployment choices that put patient welfare, not the economics of the service, first. Whether those conditions are met is not, in the end, a technical question, but one of priorities.

CRediT authorship contribution statement

J.E. Arco: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Conceptualization. **R. Alizadehsani:** Writing – review & editing, Writing – original draft. **M. Atzmueller:** Writing – review & editing, Writing – original draft. **E. Barakova:** Writing – review & editing, Writing – original draft. **L. Bloch:** Writing – review & editing, Writing – original draft, Funding acquisition. **G. Bologna:** Writing – review & editing, Writing – original draft. **P. Bonomini:** Writing – review & editing, Writing – original draft, Funding acquisition. **S.-B. Cho:** Writing – review & editing, Writing – original draft. **J.M. Cortes:** Writing – review & editing, Writing – original draft. **A. Elazab:** Writing – review & editing, Writing – original draft. **M.A. Khan:** Writing – review & editing, Writing – original draft. **E. Domínguez:** Writing – review & editing, Writing – original draft. **R.J. Duro:** Writing – review & editing, Writing – original draft. **A. Fernández-Caballero:** Writing – review & editing, Writing – original draft, Funding acquisition. **E. Fernández-Jover:** Writing – review & editing, Writing – original draft, Funding

acquisition. **J.M. Ferrández-Vicente:** Writing – review & editing, Writing – original draft. **Christoph M. Friedrich:** Writing – review & editing, Writing – original draft. **J. García-Rodríguez:** Writing – review & editing, Writing – original draft. **J. Garrigós:** Writing – review & editing, Writing – original draft. **A. Gómez-Rodellar:** Writing – review & editing, Writing – original draft. **P. Gómez-Vilda:** Writing – review & editing, Writing – original draft. **M. Graña:** Writing – review & editing, Writing – original draft. **A. Hernández:** Writing – review & editing, Writing – original draft. **Yujing Huang:** Writing – review & editing, Writing – original draft. **C. Jiménez-Mesa:** Writing – review & editing, Writing – original draft, Funding acquisition. **J. Levin:** Writing – review & editing, Writing – original draft. **N. Liu:** Writing – review & editing, Writing – original draft. **Q. Ma:** Writing – review & editing, Writing – original draft. **F.J. Martínez-Murcia:** Writing – review & editing, Writing – original draft. **R. Martínez-Tomás:** Writing – review & editing, Writing – original draft. **J.M. Molina:** Writing – review & editing, Writing – original draft, Funding acquisition. **G.J. Nalepa:** Writing – review & editing, Writing – original draft. **A. Ortiz:** Writing – review & editing, Writing – original draft. **D. Palacios-Alonso:** Writing – review & editing, Writing – original draft. **M. Ponticorvo:** Writing – review & editing, Writing – original draft. **A. Pereira:** Writing – review & editing, Writing – original draft. **J. Ramírez:** Writing – review & editing, Writing – original draft. **M. Rincón:** Writing – review & editing, Writing – original draft. **I. Rodríguez-Rodríguez:** Writing – review & editing, Writing – original draft, Funding acquisition. **L. Su:** Writing – review & editing, Writing – original draft. **J. Suckling:** Writing – review & editing, Writing – original draft, Funding acquisition. **M. Val-Calvo:** Writing – review & editing, Writing – original draft. **P. Varona:** Writing – review & editing, Writing – original draft, Funding acquisition. **C. Vázquez-García:** Writing – review & editing, Writing – original draft. **S. Wang:** Writing – review & editing, Writing – original draft. **Y.-D. Zhang:** Writing – review & editing, Writing – original draft. **J.M. Gorriz:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work of L. Bloch was partially funded by a PhD grant from the University of Applied Sciences and Arts Dortmund, Dortmund, Germany.

P. Bonomini was supported (partially) by the Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED), through Red 225RT0169.

A. Fernández-Caballero was supported by grant PID2023-149753OB-C21 funded by Spanish MCIU/AEI/10.13039/501100011033/ERDF, EU.

E. Fernández-Jover was supported in part by grants PDC2022-133 952-100, PID2022-141606OB-I00 and AIA2025-164108-C31 from the Spanish ‘Ministerio de Ciencia, Innovación y Universidades’, and by grant CIPROM/2023/25 from the Generalitat Valenciana. The Principal Investigator for all these grants is E. Fernández.

The work by C. Jiménez-Mesa is part of the grant JDC2023-051807-I funded by MICIU/AEI/10.13039/501100011033 and by ESF+.

J. M. Molina was funded by public research projects of the Spanish Ministry of Science and Innovation (CACTUS), reference.

Data availability

No data was used for the research described in the article.

References

- AbdelAziz, N.M., Said, W., AbdelHafeez, M.M., Ali, A.H., 2024. Advanced interpretable diagnosis of alzheimer's disease using secnn-rf framework with explainable ai. *Frontiers Artificial Intelligence* 7, 2024. <http://dx.doi.org/10.3389/frai.2024.1456069>.
- Abdelaziz, M., Wang, T., Anwaar, W., Elazab, A., 2025. Multi-scale multimodal deep learning framework for Alzheimer's disease diagnosis. *Comput. Biol. Med.* 184, 109438. <http://dx.doi.org/10.1016/j.cmpbiomed.2024.109438>.
- Aboy, M., Minssen, T., Vayena, E., 2024. Navigating the EU AI Act: implications for regulated digital medical products. *npj Digital Medicine* 7 (1), 237. <http://dx.doi.org/10.1038/s41746-024-01232-3>.
- Abuhantash, F., Welsch, R.E., Finkelstein, S., AlShehhi, A., 2025. Alzheimer's disease risk prediction using machine learning for survival analysis with a comorbidity-based approach. *Sci. Rep.* 15, 28723. <http://dx.doi.org/10.1038/s41598-025-14406-0>.
- Afroogh, S., Ahmed, S.I., Ahrweiler, P., Alvarez-Melis, D., Arief, M.M., Barakova, E., Bargagli-Stoffi, F.J., Biyik, E., Chen, H., Chen, X.A., Clements, R.A., Crockett, K., Dhurandhar, A., Dogan, F.I., Dollinger, M., Eslami, M., Faisal, A.A., Farahi, A., Pradier, M.F., Gabriel, S., Garcia-Olano, D., Ghassemi, M., Ghosh, S., Gunes, H., Hajiramezani, E., Haufe, S., Huang, B., Hwang, A., Islam, M.T., Jiao, J., Karimi, A.-H., Kazeminasab, S., Kuzminykh, A., Cava, W.L., Lim, B.Y., Liu, X., Mofrad, M.R.K., Parrish, A., Perez-Ortiz, M., Raj, S., Swayamdipta, S., Talebi, S., Varshney, K.R., Vorvoreanu, M., Weng, L., Xiang, A., Xu, Y., Zhao, D., Zhao, J., 2026. Beyond explainable AI (xAI): An overdue paradigm shift and post-XAI research directions. <http://dx.doi.org/10.48550/arXiv.2602.24176>, arXiv: 2602.24176. <https://arxiv.org/abs/2602.24176>.
- Ahmed, S., Hadi, A., Baddour, L., Ali, H. Mohammed, Asaad, A., Fashakh, A., Khames, Q., Nassrullah, A., Miah, A.S.M., 2025. Alzheimer's disease prediction using MRI-PET fusion through multi-modal deep learning with late fusion techniques. In: 2025 3rd International Conference on Cyber Resilience (ICCR). pp. 1–8. <http://dx.doi.org/10.1109/ICCR67387.2025.11292326>.
- Ahmed, M.I., Spooner, B., Isherwood, J., Lane, M., Orrock, E., Dennison, A., 2023. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. *Cureus* 15 (10), e46454. <http://dx.doi.org/10.7759/cureus.46454>, eCollection 2023 oct.
- Al-Bakri, F.H., Bejuri, W.M.Y.W., Al-Andoli, M.N., Ikram, R.R.R., Khor, H.M., Mispan, M.S., Yunos, N.M., Yusof, N.F.A., Fauadi, M.H.F.M., Jaya, A.S.M., Mokhtar, N.A., Yusop, N., Burhanudin, K., Marindah, T.P., Bustamin, A., Zainuddin, Z., Wahyuni, D., Ariffin, U.K., 2025. A hybrid explainable AI framework (HXAI) for accurate and interpretable diagnosis of Alzheimer's disease. *Diagnostics* 15 (24), 3118. <http://dx.doi.org/10.3390/diagnostics15243118>.
- Alex, J.S.R., Roshini, R., Maneesha, G., Aparajeeta, J., Priyadarshini, B., Lin, C.Y., Lung, C.W., 2025. Enhanced detection of mild cognitive impairment in alzheimer's disease: a hybrid model integrating dual biomarkers and advanced machine learning. *BMC Geriatr.* 25 (1), 54. <http://dx.doi.org/10.1186/s12877-025-05683-5>, published 2025-01-23.
- Aliyu, B.S., Isuwa, J., Abdulrahim, A., Abdullahi, M., Hassan, I.H., Momi, T.S., 2025. Enhanced feature selection for imbalanced microarray cancer gene classification using chaotic salp swarm algorithm. *Int. J. Theor. Appl. Comput. Intell.* 259–283. <http://dx.doi.org/10.65278/IJTACI.2025.16>.
- AlMansoori, M.E., Jemimah, S., Abuhantash, F., AlShehhi, A., 2024. Predicting early Alzheimer's with blood biomarkers and clinical features. *Sci. Rep.* 14, <http://dx.doi.org/10.1038/s41598-024-56489-1>.
- Alsaleh, A., 2025. An interpretable multimodal deep learning framework for Alzheimer's disease diagnosis. *Digital Health* 11, <http://dx.doi.org/10.1177/20552076251390281>.
- Alzheimer's Disease International, 2024. Dementia statistics. <https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/>. (Accessed 01 April 2026).
- Anderson, P.G., Tarder-Stoll, H., Alpaslan, M., Keathley, N., Levin, D.L., Venkatesh, S., Bartel, E., Siclar, S., Howell, S., Lindsey, R.V., Jones, R.M., 2024. Deep learning improves physician accuracy in the comprehensive detection of abnormalities on chest X-rays. *Sci. Rep.* 14 (1), 25151. <http://dx.doi.org/10.1038/s41598-024-76608-2>.
- Andreoletti, M., Haller, L., Vayena, E., Blasimme, A., 2024. Mapping the ethical landscape of digital biomarkers: A scoping review. *PLOS Digit. Health* 3 (5), e0000519. <http://dx.doi.org/10.1371/journal.pdig.0000519>.
- Arco, J.E., Ortiz, A., Gallego-Molina, N., Górriz, J., Ramírez, J., 2023. Enhancing multimodal patterns in neuroimaging by siamese neural networks with self-attention mechanism. *Int. J. Neural Syst.* 33 (4), 2350019. <http://dx.doi.org/10.1142/S0129065723500193>.
- Arco, J.E., Ramírez, J., Górriz, J.M., Ruz, M., 2021. Data fusion based on searchlight analysis for the prediction of alzheimer's disease. *Expert Syst. Appl.* 185, 115549. <http://dx.doi.org/10.1016/j.eswa.2021.115549>.
- Arco, J.E., rez, J. Ramí, Salas, D., nez Murcia, F.J. Martí, Górriz, J.M., 2025. Interpretable multimodal brain image fusion for Alzheimer's diagnosis via a siamese autoencoder. In: 2025 IEEE Nuclear Science Symposium (NSS), Medical Imaging Conference (MIC) and Room Temperature Semiconductor Detector Conference (RTSD). pp. 1–2. <http://dx.doi.org/10.1109/NSS/MIC/RTSD57106.2025.11287936>.
- Azizan, A., Cao, S., Dahlan, A., Endrini, S., 2025. Mapping knowledge landscapes and emerging trends in digital biomarkers for dementia in older adults: A scoping and bibliometric analysis. *Arch. Gerontol. Geriatr. Plus* 2 (2), 100148. <http://dx.doi.org/10.1016/j.aggp.2025.100148>, <https://www.sciencedirect.com/science/article/pii/S295030782500030X>.
- Badawi, A., Elmoghazy, S., Choudhury, S., Elgazzar, S., Elgazzar, K., Burhan, A.M., 2025. Multimodal detection of agitation in people with dementia in clinical settings: Observational pilot study. *JMIR Aging* 8, e68156. <http://dx.doi.org/10.2196/68156>.
- Bafaloukou, M., Schalkamp, A.-K., Fletcher-Lloyd, N., Capstick, A., Walsh, C., Sandor, C., Kouchaki, S., Nilforooshan, R., Barnaghi, P., 2025. An interpretable machine learning tool for in-home monitoring of agitation episodes in people living with dementia: a proof-of-concept study. *EclinicalMedicine* 80, 103032. <http://dx.doi.org/10.1016/j.eclinm.2024.103032>.
- Baron, L.E.G., Callaghan, R., Cash, D.M., Weston, P.S.J., Azadbakht, H., Zhang, H., 2026. When brains disagree: Biological ambiguity underlies the challenge of amyloid PET synthesis from structural MRI. [arXiv:2605.11867](https://arxiv.org/abs/2605.11867). <https://arxiv.org/abs/2605.11867>.
- Bayat, S., Babulal, G.M., Schindler, S.E., Fagan, A.M., Morris, J.C., Mihailidis, A., Roe, C.M., 2021. GPS driving: a digital biomarker for preclinical alzheimer disease. *Alzheimer's Res. Ther.* 13 (1), 115. <http://dx.doi.org/10.1186/s13195-021-00852-1>.
- Beauchet, O., Annweiler, C., Callisaya, M.L., Cock, A.-M.D., Verghese, J., et al., 2016. Poor gait performance and prediction of dementia: Results from a meta-analysis. *J. Am. Med. Dir. Assoc.* 17 (6), 482–490. <http://dx.doi.org/10.1016/j.jamda.2015.12.092>.
- Bernstein, M.H., Atalay, M.K., Dibble, E.H., Baird, G.L., Lourenco, A.P., 2023. Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography. *Eur. Radiol.* 33, 8263–8269. <http://dx.doi.org/10.1007/s00330-023-09747-1>.
- Bhattarai, P., Thakuri, D.S., Nie, Y., Chand, G.B., 2024. Explainable AI-based DeepSHAP for mapping the multivariate relationships between regional neuroimaging biomarkers and cognition. *Eur. J. Radiol.* 174, 111403. <http://dx.doi.org/10.1016/j.ejrad.2024.111403>.
- Bloch, L., Borys, K., Nensa, F., Friedrich, C.M., 2026. MRI-based radiomics and volumetrics for predicting the onset of Alzheimer's disease with explainable machine learning. *NeuroImage* 329, <http://dx.doi.org/10.1016/j.neuroimage.2026.121813>.
- Bonomini, M.P., Ghiglioni, E., Rios, N.B., 2024. Connectivity patterns in alzheimer disease and frontotemporal dementia patients using graph theory. In: *Artificial Intelligence for Neuroscience and Emotional Systems*. Springer Nature Switzerland, pp. 390–397.
- Borchert, R.J., Azevedo, T., Badhwar, A., Bernal, J., Betts, M., Bruffaerts, R., et al., 2023. Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: A systematic review. *Alzheimer's Dement.* 19 (12), 5885–5904. <http://dx.doi.org/10.1002/alz.13412>.

- Borty, J. Chakra, Chakraborty, G.S., Noman, I.R., Batra, S., Das, J., Bishnu, K.K., Tarafder, M.T.R., Islam, A., 2025. A novel diagnostic framework with an optimized ensemble of vision transformers and convolutional neural networks for enhanced alzheimer's disease detection in medical imaging. *Diagnostics* 15 (6), <http://dx.doi.org/10.3390/diagnostics15060789>.
- Busch, F., Kather, J.N., Johner, C., Moser, M., Truhn, D., Adams, L., Bressen, K., 2024. Navigating the European Union Artificial Intelligence Act for healthcare. *npj Digital Medicine* 7, 261. <http://dx.doi.org/10.1038/s41746-024-01213-6>.
- Butler, P.M., Yang, J., Brown, R., Hobbs, M., Becker, A., Penalver-Andres, J., Syz, P., Muller, S., Cosne, G., Juraver, A., Song, H.H., Saha-Chaudhuri, P., Roggen, D., Scotland, A., Silveira, N., Demircioglu, G., Gabelle, A., Hughes, R., Erkinen, M.G., Langbaum, J.B., Lingler, J.H., Price, P., Quiroz, Y.T., Sha, S.J., Sliwinski, M., Porsteinsson, A.P., Au, R., Bianchi, M.T., Lenyoun, H., Pham, H., Patel, M., Belachew, S., et al., 2025. Smartwatch- and smartphone-based remote assessment of brain health and detection of mild cognitive impairment. *Nature Med.* 31 (3), 829–839. <http://dx.doi.org/10.1038/s41591-024-03475-9>.
- Caro, I., Pérez, G., Bize, J.V., Ponferrada, J., Ferrante, F.J., Welford, A.S., Gauder, L., Olavarria, L., Henríquez, F., Ramos, T., Besnier, C., Ferrer, L., Gorno-Tempini, M.L., Slachevsky, A., Ibañez, A., García, A.M., 2026. Benchmarking speech biomarkers of Alzheimer's against cognitive and neural measures. *Alzheimer's Dement.* 22, e71365. <http://dx.doi.org/10.1002/alz.71365>.
- Čepukaitytė, G., Newton, C., Chan, D., 2024. Early detection of diseases causing dementia using digital navigation and gait measures: A systematic review of evidence. *Alzheimer's Dement.* 20 (5), 3054–3073. <http://dx.doi.org/10.1002/alz.13716>.
- Chakraborty, T., Colafranceschi, A., Liu, Y., 2025. Beyond accuracy: Explainable deep learning for Alzheimer's disease detection using structural MRI data. *Information* 16 (12), 1058. <http://dx.doi.org/10.3390/info16121058>.
- Chan, L.L., Espinoza Cerda, M.T., Brodie, M.A., Lord, S.R., Taylor, M.E., 2025. Daily-life walking speed, running duration and bedtime from wrist-worn sensors predict incident dementia: A watch walk – UK biobank study. *Int. Psychogeriatr.* 37 (3), 100031. <http://dx.doi.org/10.1016/j.inpsyc.2024.100031>.
- Chattopadhyay, T., Joshy, N.A., Jagad, C., Gleave, E.J., Thomopoulos, S.I., Feng, Y., Villalón-Reina, J.E., Laltoo, E., Joshi, H., Venkatasubramanian, G., John, J.P., Steeg, G.V., Ambite, J.L., Thompson, P.M., 2024. Comparison of explainable AI models for MRI-based Alzheimer's disease classification. *BioRxiv preprint*. <http://dx.doi.org/10.1101/2024.09.17.613560>.
- Chen, X., Huang, T., Shi, C., Xu, S., Fan, S., 2026. Diagnostic performance of plasma p-Tau217, p-Tau181, and p-Tau231 across the Alzheimer's disease continuum: a network meta-analysis. *Frontiers in Aging Neuroscience* 18, 2026. <http://dx.doi.org/10.3389/fnagi.2026.1834591>.
- Chen, X., Li, Y., Li, R., Yuan, X., Liu, M., Zhang, W., Li, Y., 2023. Multiple cross-frequency coupling analysis of resting-state EEG in patients with mild cognitive impairment and alzheimer's disease. *Frontiers Aging Neuroscience* 15, 2023. <http://dx.doi.org/10.3389/fnagi.2023.1142085>.
- Chernov, Y., 2025. Formalized assessment of handwriting deterioration caused by dementia. *Medinformatics* 3 (1), 34–45. <http://dx.doi.org/10.47852/bonviewMEDINS2025601>.
- Choudhury, C., Goel, T., Tanveer, M., 2024. A coupled-GAN architecture to fuse MRI and PET image features for multi-stage classification of alzheimer's disease. *Inf. Fusion* 109, 102415. <http://dx.doi.org/10.1016/j.inffus.2024.102415>.
- Cogswell, P.M., Fan, A.P., 2023. Multimodal comparisons of QSM and PET in neurodegeneration and aging. *NeuroImage* 273, 120068. <http://dx.doi.org/10.1016/j.neuroimage.2023.120068>.
- Collins, G.S., Moons, K.G.M., Dhiman, P., Riley, R.D., Beam, A.L., Calster, B. Van, Ghassemi, M., Liu, X., Reitsma, J.B., van Smeden, M., Boulesteix, A.-L., Camaradov, J.C., Celi, L.A., Denaxas, S., Dennison, A.K., Glocker, B., Golub, R.M., Harvey, H., Heinze, G., Hoffman, M.M., Kengne, A.P., Lam, E., Lee, N., Loder, E.W., Maier-Hein, L., Mateen, B.A., McCradden, M.D., Oakden-Rayner, L., Ordish, J., Parnell, R., Rose, S., Singh, K., Wynants, L., Logullo, P., 2024. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 385, e078378. <http://dx.doi.org/10.1136/bmj-2023-078378>.
- Condado, J. García, Recuero, I. Verdugo, n. Tellaetxe Elorriaga, I., Birkenbihl, C., Carrigan, M., Diez, I., Buckley, R.F., Erramuzpe, A., Cortes, J.M., 2025. Aging as an active player in alzheimer's disease classification: Insights from feature selection in BrainAge models. *NeuroImage* 322, 121548. <http://dx.doi.org/10.1016/j.neuroimage.2025.121548>.
- Coughlan, G., Coutrot, A., Khondoker, M., Minihi, A.-M., Spiers, H., Hornberger, M., 2019. Toward personalized cognitive diagnostics of at-genetic-risk alzheimer's disease. *Proc. Natl. Acad. Sci.* 116 (19), 9285–9292. <http://dx.doi.org/10.1073/pnas.1901600116>.
- Davis, M.A., Lim, N., Jordan, J., Yee, J., Gichoya, J.W., Lee, R., 2023. Imaging artificial intelligence: A framework for radiologists to address health equity, from the AJR special series on DEI. *AJR Am. J. Roentgenol.* 221 (3), 302–308. <http://dx.doi.org/10.2214/AJR.22.28802>, epub 2023-02-22.
- Ding, H., Wang, B., Hamel, A.P., Melkonian, M., Ang, T.F.A., Au, R., Lin, H., 2023. Prediction of progression from mild cognitive impairment to Alzheimer's disease with longitudinal and multimodal data. *Front. Dement.* 2, 1271680. <http://dx.doi.org/10.3389/frdem.2023.1271680>.
- Dino, F.R., Pressman, P.S., Cohen, K.B., Dubljevic, V., Jarrold, W., Foltz, P.W., DeCamp, M., Mahoor, M.H., Hunter, L.E., 2025. Ethics in digital phenotyping: considerations regarding alzheimer's disease, speech and artificial intelligence. *J. Med. Ethics* <http://dx.doi.org/10.1136/jme-2024-110252>.
- Duong, Q.A., Tran, S.D., Gahm, J.K., 2025. Multimodal surface-based transformer model for early diagnosis of alzheimer's disease. *Sci. Rep.* 15 (1), 5787. <http://dx.doi.org/10.1038/s41598-025-90115-y>.
- Dymm, B., Khalid, A., 2025. Artificial intelligence in neurology and stroke education: Current applications and future directions. *Seminars Neurology* 46, <http://dx.doi.org/10.1055/a-2713-6622>.
- Elmannai, H., Saleem, N., Bourouis, S., Alkanhel, R.I., 2025. Self-supervised attentive feature learning for Alzheimer's disease detection. *Int. J. Interact. Multimed. and Artif. Intell.* 9 (5), 119–127. <http://dx.doi.org/10.9781/ijimai.2025.09.002>.
- Feng, J., Zhao, X., Liu, Z., Ding, Y., Wang, F., 2025. A multi-view multimodal deep learning framework for alzheimer's disease diagnosis. *Front. Neurosci.* 19, 1658776. <http://dx.doi.org/10.3389/fnins.2025.1658776>.
- Fortin, J.-P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooper, C., Fava, M., McGrath, P.J., McInnis, M., Phillips, M.L., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., 2018. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage* 167, 104–120. <http://dx.doi.org/10.1016/j.neuroimage.2017.11.024>.
- Gao, J., Chen, M., Xiang, L., Xu, C., 2024. A comprehensive survey on evidential deep learning and its applications. <http://dx.doi.org/10.48550/arXiv.2409.04720>, <https://arxiv.org/abs/2409.04720>.
- Gerke, S., Minssen, T., Cohen, G., 2020. Ethical and legal challenges of artificial intelligence-driven healthcare. In: *Artificial Intelligence in Healthcare*. Elsevier, pp. 295–336. <http://dx.doi.org/10.1016/B978-0-12-818438-7.00012-5>.
- Ghamizi, S., Kanli, G., Deng, Y., Palissot, V., Perquin, M., Keunen, O., 2026. Foundation models for brain imaging: A systematic review. *NeuroImage* 333, 121877. <http://dx.doi.org/10.1016/j.neuroimage.2026.121877>.
- Gomez, T., Mouchère, H., 2023. Computing and evaluating saliency maps for image classification: a tutorial. *J. Electron. Imaging* 32 (2), 020801. <http://dx.doi.org/10.1117/1.JEI.32.2.020801>.
- Górriz, J., Illán, I., Álvarez, Marquina, A., Álvarez, Arco, J., Atzmueller, M., Bal-larini, F., Barakova, E., Bologna, G., Bonomini, P., Castellanos-Dominguez, G., Castillo-Barnes, D., Cho, S., Contreras, R., Cuadra, J., E. Domínguez, F. Domínguez-Mateos, Duro, R., Elizondo, D., Fernández-Caballero, A., Fernández-Jover, E., Formoso, M., Gallego-Molina, N., Gamazo, J., J.G. González, J. García-Rodríguez, Garre, C., J. Garrigós, A. Gómez-Rodellar, Gómez-Vilda, P., M. Graña, B. Guerrero-Rodríguez, Hendrikse, S., Jimenez-Mesa, C., Jodra-Chuan, M., Julian, V., Kotz, G., Kutt, K., Leming, M., J. de Lope, B. Macas, Marrero-Aguai, V., Martinez, J., Martínez-Murcia, F., R. Martínez-Tomás, J. Mekyska, Nalepa, G., Novais, P., Orellana, D., Ortiz, A., Palacios-Alonso, D., Palma, J., Pereira, A., Pinacho-Davidson, P., Pinninghoff, M., Ponticorvo, M., Psarrou, A., J. Ramírez, V. Rodellar-Biarge, I. Rodríguez-Rodríguez, P. Roelofsma, Santos, J., Salas-Gonzalez, D., Salcedo-Lagos, P., Segovia, F., Shoeibi, A., Silva, M., Simic, D., Suckling, J., Treur, J., Tsanas, A., Varela, R., Wang, S., Wang, W., Zhang, Y., Zhu, H., Zhu, Z., Ferrández-Vicente, J., 2023. Computational approaches to explainable artificial intelligence: Advances in theory, applications and trends. *Inf. Fusion* 100, 101945. <http://dx.doi.org/10.1016/j.inffus.2023.101945>.
- Górriz, J.M., Ramírez, J., Ortiz, A., Martínez-Murcia, F.J., Segovia, F., Suckling, J., Leming, M., Zhang, Y.-D., J.R. Álvarez Sánchez, G. Bologna, Bonomini, P., Casado, F.E., Charte, D., Charte, F., Contreras, R., Cuesta-Infante, A., Duro, R.J., Fernández-Caballero, A., Fernández-Jover, E., Gómez-Vilda, P., M. Graña, F. Herrera, Iglesias, R., Lekova, A., J. de Lope, E. López-Rubio, R. Martínez-Tomás, M.A. Molina-Cabellero, Montemayor, A.S., Novais, P., Palacios-Alonso, D., Pantrigo, J.J., Payne, B.R., F. de la Paz López, M.A. Pinninghoff, M. Rincón, J. Santos, Thurnhofer-Hemsi, K., Tsanas, A., Varela, R., Ferrández, J.M., 2020. Artificial intelligence within the interplay between natural and artificial computation: Advances in data science, trends and applications. *Neurocomputing* 410, 237–270. <http://dx.doi.org/10.1016/j.neucom.2020.05.078>.
- Han, L., Wang, M., Li, H., Chen, Y., Zhang, H., Liu, X., Sun, Q., 2025. AD-Diff: enhancing alzheimer's disease prediction accuracy through multimodal fusion. *Front. Comput. Neurosci.* 19, <http://dx.doi.org/10.3389/fncom.2025.1484540>.
- Hassanpour, J., Srivastav, V., Mutter, D., Padoy, N., 2024. Overcoming dimensional collapse in self-supervised contrastive learning for medical image segmentation. In: *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. pp. 1–5. <http://dx.doi.org/10.1109/isbi56570.2024.10635740>.
- Hernández, A., Illán, I., Ramírez, J., Segovia, F., Martínez-Murcia, F., Levin, J., Górriz, J., 2026. Robust validation of neuroimaging and clinical models via the sar method: A case study based on the adni dataset. *NeuroImage* 333, 121917. <http://dx.doi.org/10.1016/j.neuroimage.2026.121917>.
- Hettikankanamge, N., Shafiabady, N., Chateaur, F., Wu, R.M.X., Din, F. Ud, Zhou, J., 2025. Explainable artificial intelligence (XAI): A systematic review for unveiling the black box models and their relevance to biomedical imaging and sensing. *Sensors* 25 (21), <http://dx.doi.org/10.3390/s25216649>.
- Ho, N.H., Charisis, S., Brandigampala, S.R., Wang, D., Heckbert, S.R., Martinez, D.M., Hughes, T.M., Honnorat, N., Archer, D.B., Hohman, T.J., Seshadri, S., Davatzikos, C., Habes, M., 2025. Addressing ai bias in MRI-based dementia classification among hispanic individuals. *Alzheimer's Dement.* 21 (Suppl. 2), e105856. <http://dx.doi.org/10.1002/alz70856.105856>.

- Hogg, H.D.J., Al-Zubaidy, M., Group, T.E.M.S.S.R., Talks, J., Denniston, A.K., Kelly, C.J., Malawana, J., Papoutsis, C., Teare, M.D., Keane, P.A., Beyer, F.R., Maniopoulos, G., 2023. Stakeholder perspectives of clinical artificial intelligence implementation: Systematic review of qualitative evidence. *J. Med. Internet Res.* 25, e39742. <http://dx.doi.org/10.2196/39742>.
- Hollet, N., Robinson, E., Georgiou, E., Ermis, E., Nahum, U., Brünink, S., 2026. A benchmark of (MRI -) foundation models to predict IDH mutational status in glioma. <https://arxiv.org/abs/2606.23172>.
- Hsieh, C.-J., Li, P.-S., Wang, C.-H., Lin, S.-L., Hsu, T.-C., Tsai, C.-T., 2023. Socially assistive robots for people living with dementia in long-term facilities: A systematic review and meta-analysis of randomized controlled trials. *Gerontology* 69 (8), 1027–1042. <http://dx.doi.org/10.1159/000528799>.
- Hu, S., Yu, H., Gao, J., 2025. The pTau217/a β 1-42 plasma ratio: The first FDA-cleared blood biomarker test for diagnosis of Alzheimer's disease. *Drug Discoveries & Therapeutics* 19 (3), 208–209. <http://dx.doi.org/10.5582/ddt.2025.01055>.
- Huang, Y., Zhang, H., Ma, B., Yu, Z., Dai, S., Cheng, L., Su, L., Yang, G., Ma, Q., 2026. Converse or reverse? machine-learning modeling for disease progression: A study based on alzheimer's disease continuum cohort. *NeuroImage* 327, 121754. <http://dx.doi.org/10.1016/j.neuroimage.2026.121754>.
- Iulita, M.F., Streel, E., Harrison, J., 2025. Digital biomarkers: Redefining clinical outcomes and the concept of meaningful change. *Alzheimer's Dement.: Transl. Res. Clin. Interv.* 11 (2), e70114. <http://dx.doi.org/10.1002/trc2.70114>.
- Jafari, Z., Andrew, M.K., Rockwood, K.J., 2025. Diagnostic utility of speech-based biomarkers in mild cognitive impairment: a systematic review and meta-analysis. *Age Ageing* 54 (10), afaf316. <http://dx.doi.org/10.1093/ageing/afaf316>.
- Jahan, S., Taher, K.A., Kaiser, M.S., Mahmud, M., Rahman, M.S., Hosen, A., Ra, I.-H., 2023. Explainable AI-based Alzheimer's prediction and management using multimodal data. *PLoS ONE* 18 (11), e0294253. <http://dx.doi.org/10.1371/journal.pone.0294253>.
- Jasodanand, V.H., Kowshik, S.S., Puducheri, S., Romano, M.F., Xu, L., Au, R., Kolachalama, V.B., 2025. AI-driven fusion of multimodal data for alzheimer's disease biomarker assessment. *Nat. Commun.* 16, 7407. <http://dx.doi.org/10.1038/s41467-025-62590-4>.
- Jovalekic, Aleksandar, Bullich, Santiago, Roé-Vellvé, Núria, Kolinger, Guilherme Domingues, Howard, Lorelei R., Elsholz, Floriana, Lagos-Quintana, Mariana, Blanco-Rodríguez, Beatriz, Pérez-Martínez, Esther, Gismondi, Rossella, Perrotin, Audrey, Chapeau, Marianne, Keegan, Richard, Mueller, Andre, Stephens, Andrew W., Koglin, Norman, 2024. Experiences from clinical research and routine use of florbetaben amyloid PET—a decade of post-authorization insights. *Pharmaceuticals* 17 (12), 1648. <http://dx.doi.org/10.3390/ph17121648>.
- Kang, S., Kim, S.-W., Seong, J.-K., 2024. Disentangling brain atrophy heterogeneity in Alzheimer's disease: A deep self-supervised approach with interpretable latent space. *NeuroImage* 297, 120737. <http://dx.doi.org/10.1016/j.neuroimage.2024.120737>.
- Kapadia, K., Ruwali, S., Malav, T., Seshadri, S., Seidmann, A., Press, D.Z., Kolachalama, V.B., 2024. Enhancing efficiency with an AI-augmented clinician in neurology. *Aging Dis.* 16 (5), 2498–2503. <http://dx.doi.org/10.14336/AD.2024.1249>.
- Karimi, M., Karimi, Z., Khosravi, M., Delaram, Z., Dehsheikh, M.H., Najafabadi, S.A., Aliabadi, M.A., Tavakoli, N., 2025. Feature selection methods in big medical databases: A comprehensive survey. *Int. Journal Theoretical Applied Computational Intelligence* 181–209. <http://dx.doi.org/10.65278/IJTACI.2025.21>.
- Kelly, C.J., Karthikesalingam, A., Suleyman, M., Corrado, G., King, D., 2019. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 17 (1), 195. <http://dx.doi.org/10.1186/s12916-019-1426-2>.
- Khachaturian, A.S., Cassin, B., Finney, G.R., 2023. Using clinical decision intelligence applications to improve pathways for earlier detection of underrecognized cognitive disorders. *The Journal of Aging Research & Lifestyle* 12, 14–17. <http://dx.doi.org/10.14283/jarlife.2023.3>, published 2023-05-11.
- Khude, H., Shende, P., 2025. AI-driven clinical decision support systems: Revolutionizing medication selection and personalized drug therapy. *Adv. Integr. Med.* 12, 100529. <http://dx.doi.org/10.1016/j.aimed.2025.100529>.
- Kourtis, L.C., Regele, O.B., Wright, J.M., Jones, G.B., 2019. Digital biomarkers for alzheimer's disease: the mobile/wearable devices opportunity. *npj Digit. Med.* 2 (1), 9. <http://dx.doi.org/10.1038/s41746-019-0084-2>.
- Kurdi, S.Z., 2026. Machine learning-based classification framework for human health care monitoring. *Int. J. Theor. Appl. Comput. Intell.* 2026, 1–15. <http://dx.doi.org/10.65278/IJTACI.2026.1>.
- Lapushkin, S., S. Wäldchen, A. Binder, Montavon, G., Samek, W., Müller, K.-R., 2019. Unmasking Clever Hans predictors and assessing what machines really learn. *Nat. Commun.* 10, 1096. <http://dx.doi.org/10.1038/s41467-019-08987-4>.
- Leandrou, S., Lammis, D., Bougias, H., Stogiannos, N., Georgiadou, E., Achilleos, K.G., Pattichis, C.S., 2023. A cross-sectional study of explainable machine learning in alzheimer's disease: diagnostic classification using MR radiomic features. *Front. Aging Neurosci.* 15, <http://dx.doi.org/10.3389/fnagi.2023.1149871>.
- Lee, G.-B., Jeong, Y.-J., Kang, D.-Y., Yun, H.-J., Yoon, M., 2024a. Multimodal feature fusion-based graph convolutional networks for Alzheimer's disease stage classification using F-18 florbetaben brain PET images and clinical indicators. *PLoS ONE* 19 (12), e0315809. <http://dx.doi.org/10.1371/journal.pone.0315809>.
- Lee, M.-W., Kim, H.W., Choe, Y.S., Yang, H.S., Lee, J., Lee, H., Yong, J.H., Kim, D., Lee, M., Kang, D.W., et al., 2024b. A multimodal machine learning model for predicting dementia conversion in alzheimer's disease. *Sci. Rep.* 14 (1), 12276. <http://dx.doi.org/10.1038/s41598-024-60134-2>.
- Lekadir, K., Feragen, A., Fofanah, A.J., Frangi, A.F., Buys, A., Duminy, A., Mihalef, V., Navab, N., Ogier, A., Schmainda, K., et al., 2025. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* 388, e081554. <http://dx.doi.org/10.1136/bmj-2024-081554>.
- Lentzen, M., Vairavan, S., Muurling, M., Alepoulos, V., Atreya, A., Boada, M., de Boer, C., Conde, P., Curcic, J., Frisoni, G., Galluzzi, S., Gjesten, M.T., Gkioka, M., Grammatikopoulou, M., Hausner, L., Hinds, C., Lazarou, I., de Mendonça, A., Nikolopoulos, S., Religa, D., Scebb, G., Visser, P.J., Wittenberg, G., Narayan, V.A., Coello, N., Brem, A.-K., Aarsland, D., Fröhlich, H., 2025. RADAR-AD: assessment of multiple remote monitoring technologies for early detection of Alzheimer's disease. *Alzheimer's Research Therapy* 17 (1), 29. <http://dx.doi.org/10.1186/s13195-025-01675-0>.
- Leonardsen, E.H., Persson, K., Grødem, E., Dinsdale, N., Schellhorn, T., Roe, J.M., Vidal-Piñeiro, D., Sørensen, Ø., Kaufmann, T., Westman, E., Marquand, A., Selbæk, G., Andreassen, O.A., Wolfers, T., Westlye, L.T., Wang, Y., 2024. Constructing personalized characterizations of structural brain aberrations in patients with dementia using explainable artificial intelligence. *npj Digit. Med.* 7, 110. <http://dx.doi.org/10.1038/s41746-024-01123-7>.
- Li, D., He, J., Liu, B., Zhu, L., Yang, Y., Peng, Y., Nie, L., Wang, R., 2026. Multimodal radiomics of precisely segmented hippocampal subfields: Iron deposition and structural biomarkers for early diagnosis of alzheimer's disease. *NeuroImage* 332, 121900. <http://dx.doi.org/10.1016/j.neuroimage.2026.121900>.
- Li, Y., Yang, Y., Chen, Y., Ye, C., Ma, T., 2025. Self-supervised learning to unveil brain dysfunctional signatures in brain disorders: Methods and applications. *Health Data Sci.* 5, 0282. <http://dx.doi.org/10.34133/hds.0282>.
- Lian, C., Liu, M., Wang, L., Shen, D., 2022. Multi-task weakly-supervised attention network for dementia status estimation with structural MRI. *IEEE Trans. Neural Netw. Learn. Syst.* 33 (8), 4056–4068. <http://dx.doi.org/10.1109/TNNLS.2021.3055772>.
- Lima, M.R., Capstick, A., Geranmayeh, F., Nilforooshan, R., Matarić, M., Vaidyanathan, R., Barnaghi, P., 2025. Evaluating spoken language as a biomarker for automated screening of cognitive impairment. *Commun. Med.* 6 (6), 14. <http://dx.doi.org/10.1038/s43856-025-01263-1>.
- Liu, L., Lyu, J., Liu, S., Tang, X., Chandra, S.S., Nasrallah, F.A., 2023. Tri-former: a multi-modal transformer framework for mild cognitive impairment conversion prediction. <http://dx.doi.org/10.1109/isbi53787.2023.10230709>, <https://arxiv.org/abs/2307.07177>.
- Liu, X., Rivera, S. Cruz, Moher, D., Calvert, M.J., Denniston, A.K., 2020. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nature Med.* 26 (9), 1364–1374. <http://dx.doi.org/10.1038/s41591-020-1034-x>.
- Liu, Y., Wang, L., Ning, X., Gao, Y., Wang, D., 2024. Enhancing early Alzheimer's disease classification accuracy through the fusion of sMRI and rsMEG data: a deep learning approach. *Front. Neurosci.* 18, 1480871. <http://dx.doi.org/10.3389/fnins.2024.1480871>.
- Liu, J., Wu, S., Fu, Q., Luo, X., Luo, Y., Qin, S., Huang, Y., Chen, Z., 2025. Multimodal diagnosis of Alzheimer's disease based on resting-state electroencephalography and structural magnetic resonance imaging. *Front. Physiol.* 16, 1515881. <http://dx.doi.org/10.3389/fphys.2025.1515881>.
- Lott, S.A., Streel, E., Bachman, S.L., Bode, K., Dyer, J., Fitzer-Attas, C., Goldsack, J.C., Hake, A., Jannati, A., Fuertes, R.S., Fromy, P., 2024. Digital health technologies for Alzheimer's disease and related dementias: Initial results from a landscape analysis and community collaborative effort. *J. Prev. Alzheimer's Dis.* 11 (5), 1480–1489. <http://dx.doi.org/10.14283/jpad.2024.103>.
- Lozupone, G., Bria, A., Fontanella, F., Meijer, F.J.A., De Stefano, C., 2024. AXIAL: attention-based eXplainability for interpretable Alzheimer's localized diagnosis using 2D CNNs on 3D MRI brain scans. <http://dx.doi.org/10.48550/arXiv.2407.02418>, arXiv:2407.02418.
- Lukken, D.R.M., Askari, S.I., Stolwijk, N.E., Hofstede, B.M., Nap, H.H., Boon, W.P.C., Peine, A., Moors, E.H.M., Minkman, M.M.N., 2024. Making co-design more responsible: Case study on the development of an AI-based decision support system in dementia care. *JMIR Hum. Factors* 11, e55961. <http://dx.doi.org/10.2196/55961>.
- Lv, J., Rani, S., Li, K., Liu, N., 2026. Neural-linguistic analysis for alzheimer's detection: A deep learning approach informed by cognitive neuroscience. *NeuroImage* 327, 121739. <http://dx.doi.org/10.1016/j.neuroimage.2026.121739>.
- Mahim, S.M., Ali, M.S., Hasan, M.O., Nafi, A.A.N., Sadat, A., Hasan, S.A., Shareef, B., Ahsan, M.M., Islam, M.K., Miah, M.S., Niu, M.-B., 2024. Unlocking the potential of XAI for improved Alzheimer's disease detection and classification using a ViT-GRU model. *IEEE Access* 12, 8390–8412. <http://dx.doi.org/10.1109/ACCESS.2024.3351809>.
- Martin, S.A., Zhao, A., Qu, J., Imms, P., Irimia, A., Barkhof, F., Cole, J.H., 2026. Investigating the utility of explainable artificial intelligence for neuroimaging-based dementia diagnosis and prognosis. *Hum. Brain Mapp.* 47 (2), e70456. <http://dx.doi.org/10.1002/hbm.70456>.

- Matzkin, F., Larrazabal, A., Milone, D.H., Dolz, J., Ferrante, E., 2025. Towards reliable WMH segmentation under domain shift: An application study using maximum entropy regularization to improve uncertainty estimation. 196, <http://dx.doi.org/10.1016/j.combiomed.2025.110639>.
- Metsch, J.M., Hauschild, A.-C., 2025. BenchXAI: Comprehensive benchmarking of post-hoc explainable AI methods on multi-modal biomedical data. *Comput. Biol. Med.* 191, 110124. <http://dx.doi.org/10.1016/j.combiomed.2025.110124>.
- Moghaddami, M., Schubring, C., Siadat, M.-R., 2025. Transformer model for Alzheimer's disease progression prediction using longitudinal visit sequences. In: *Proceedings of the ACM Conference on Health Inference, and Learning (CHIL)*. <http://dx.doi.org/10.48550/arXiv.2507.03899>.
- Montavon, G., Samek, W., Müller, K.-R., 2018. Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.* 73, 1–15. <http://dx.doi.org/10.1016/j.dsp.2017.10.011>.
- Muhammad, A., Jin, Q., Elwasila, O., Gulzar, Y., 2025. Hybrid deep learning architecture with adaptive feature fusion for multi-stage alzheimer's disease classification. *Brain Sci.* 15 (6), <http://dx.doi.org/10.3390/brainsci15060612>.
- Munk, A., Cerri, S., Nersesjan, V., Krag, C.H., Ambsdorf, J., García, P.R., Machnio, J., Liu, P., Ahn, S., Akbari, N., Khalil, Y.A., Amador, K., Amirrajab, S., Arbel, T., Cuadra, M.B., Baid, U., Baheti, B., Banus, J., Barbierik, K., Brune, C., Bu, Y., Callard, B., Chen, Y., Crijnen, C., Dancette, C., Drotar, P., Dutande, P., Forkert, N.D., Garg, S., Gazda, J., Gazda, M., Gérin, B., Ghosh, P., Gong, W., Goraliza, P.M., Hashemi, S., Heumann, T., Jia, F., Jiang, J., Kaczmarek, E., Kang, C., Kang, S.K., Khazaei, M., Khlaut, J., Koutsouvelis, P., Lee, J.S., Li, Y., Lyu, M., Ma, M., Madabhushi, A., Maier-Hein, K.H., Manceron, P., Mora, A.M., Mazher, M., Meister, F., Molchanova, N., Niederer, S.A., Nürnberg, L., Park, J., Qayyum, A., Richiardi, J., Saporta, A., Setlak, B., Shen, N., Szeto, J., Ulrich, C., Vaish, P., Vigneshwaran, V., Volmer, L., Wang, Z., Wei, S., Winder, A., Wolterink, J.M., Wynen, M., Yang, C., Yie, S.Y., Ghazi, M.M., Pai, A., Solem, E.J., Llambias, S.N., Boesen, M., Benros, M.E., Iglesias, J.E., Nielsen, M., 2026. Towards brain MRI foundation models for the clinic: Findings from the FOMO25 challenge. *arXiv preprint. arXiv:2604.11679* [arXiv:2604.11679](https://arxiv.org/abs/2604.11679).
- Nair, M., Svedberg, P., Larsson, I., Nygren, J.M., 2024. A comprehensive overview of barriers and strategies for AI implementation in healthcare: Mixed-method design. *PLOS ONE* 19 (8), e0305949. <http://dx.doi.org/10.1371/journal.pone.0305949>.
- Nicosia, J., Aschenbrenner, A.J., Balota, D.A., Sliwinski, M.J., Tahan, M., Adams, S., Stout, S.S., Wilks, H., Gordon, B.A., Benzinger, T.L.S., Fagan, A.M., Xiong, C., Bateman, R.J., Morris, J.C., Hasselstab, J., 2023. Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for alzheimer's disease. *J. Int. Neuropsychol. Soc.* 29 (5), 459–471. <http://dx.doi.org/10.1017/S135561772200042X>.
- Nie, Y., Cui, Q., Li, W., Lü, Y., Yu, W., 2026. A vision-language model for enhanced MCI identification in alzheimer's disease through neuropsychological and neuroimaging data integration. *Neural Netw.* 197, 108415. <http://dx.doi.org/10.1016/j.neunet.2025.108415>.
- Ning, Y., Chen, M., Yang, H., Jia, J., 2025. Accelerometer-measured physical activity timing with incident dementia. *Alzheimer's Dement.* 21 (2), e14452. <http://dx.doi.org/10.1002/alz.14452>.
- Nuffield Council on Bioethics and Predicting, 2024. The future of health? nuffield council on bioethics. <https://www.nuffieldbioethics.org/publication/predicting-the-future-of-health/>. (Accessed 2025).
- Oduami, 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. Biological Eng.* 43, 291–302. <http://dx.doi.org/10.1007/s40846-023-00801-3>.
- Oh, K., Heo, D.-W., Mulyadi, A.W., Jung, W., Kang, E., Lee, K.H., Suk, H.-I., 2025a. A quantitatively interpretable model for alzheimer's disease prediction using deep counterfactuals. *NeuroImage* 309, 121077. <http://dx.doi.org/10.1016/j.neuroimage.2025.121077>.
- Oh, K.M., Hong, S., Zhu, Z., Zhou, H., Lee, J.A., 2025b. Evaluating large language models to support dementia caregivers: Identifying opportunities for improvement. *Innovation Aging* 9 (Supplement_2), <http://dx.doi.org/10.1093/geroni/igaf122.4109>, igaf122.4109.
- Oh, C., Morris, R., Wang, X., Raskin, M.S., 2023. Analysis of emotional prosody as a tool for differential diagnosis of cognitive impairments: a pilot research. *Front. Psychol.* 14, <http://dx.doi.org/10.3389/fpsyg.2023.1129406>.
- Oh, C., Park, M.S., 2025. Unveiling 30 years of research on speech biomarker of dementia using text mining. *Digital Health* 11, 20552076251360901. <http://dx.doi.org/10.1177/20552076251360901>.
- Olaimat, M. Al, Martinez, J., Saeed, F., Bozdag, S., 2023. PPAD: a deep learning architecture to predict progression of alzheimer's disease. *Bioinformatics* 39 (Supplement_1), i149–i157. <http://dx.doi.org/10.1093/bioinformatics/btad249>.
- Pan, Y., Mirheidari, B., Blackburn, D., Christensen, H., 2025. A two-step attention-based feature combination cross-attention system for speech-based dementia detection. *IEEE Trans. Audio Speech Lang. Process.* 33, 806–907. <http://dx.doi.org/10.1109/TASLP.2025.3533363>.
- Park, W., Ahn, S., Hernandez, M.C.V., Maniega, S.M., Park, J., 2025. BrainODE: Neural shape dynamics for age- and disease-aware brain trajectories. In: *Advances in Neural Information Processing Systems (NeurIPS)*.
- Pemberton, H.G., Zaki, L.A.M., Goodkin, O., Das, R.K., Steketee, R.M.E., Barkhof, F., Vernooij, M.W., 2021. Technical and clinical validation of commercial automated volumetric MRI tools for dementia diagnosis—a systematic review. *Neuroradiology* 63 (11), 1773–1789. <http://dx.doi.org/10.1007/s00234-021-02746-3>.
- Pereira, R., Mendes, C., Costa, N., L. Frazão, A. Fernández-Caballero, Pereira, A., 2024. Human-computer interaction approach with empathic conversational agent and computer vision. In: *Artificial Intelligence for Neuroscience and Emotional Systems. Springer Nature Switzerland, Cham*, pp. 431–440.
- Persson, S., Svenberg, H., Rial, A.M., Ferreira, D., Moreno, R., 2026. Explainable AI for MRI alzheimer's disease classification: A comparative analysis. *NeuroImage* 338, 122043. <http://dx.doi.org/10.1016/j.neuroimage.2026.122043>.
- Price, II, W.N., Gerke, S., Cohen, I.G., 2024. Liability for use of artificial intelligence in medicine. In: Solaiman, B., Cohen, I.G. (Eds.), *Research Handbook on Health, AI and the Law*. Edward Elgar Publishing, Cheltenham, UK, pp. 150–166. <https://www.ncbi.nlm.nih.gov/books/NBK613216/>. published 2024-07-16.
- Priyadarshini, M., Muruges, V., Rybin, O., 2026. Enhancing alzheimer's disease classification with a transformer-based model using self-supervised learning. *Scientific Reports* 16 (1), <http://dx.doi.org/10.1038/s41598-025-33957-w>.
- Qin, X., 2025. Consistency and stability benchmarking of Grad-CAM, SHAP, and LIME for diffuse and focal brain MRI classification. In: *Fifth International Conference on Biomedicine and Bioinformatics Engineering (ICBBE 2025)*. <http://dx.doi.org/10.1117/12.3095478>.
- Rahamtalla, B., Medani, I., Abdelhag, M., Eltigani, S., Rajan, S.K., Falgy, E., Hassan, N.M., Fadailu, M., Khudhayr, H., Abdalla, A., 2025. The AI-powered healthcare ecosystem: Bridging the chasm between technical validation and systemic integration — a systematic review. *Futur. Internet* 17 (12), 550. <http://dx.doi.org/10.3390/fi17120550>.
- Raichlen, D.A., Aslan, D.H., Sayre, M.K., Bharadwaj, P.K., Ally, M., Maltagliati, S., Lai, M.H.C., Wilcox, R.R., Klimentidis, Y.C., Alexander, G.E., 2023. Sedentary behavior and incident dementia among older adults. *JAMA* 330 (10), 934–940. <http://dx.doi.org/10.1001/jama.2023.15231>.
- Rallabandi, V.P., Subramanyam, Seetharaman, K., 2023. Deep learning-based classification of healthy aging controls, mild cognitive impairment and Alzheimer's disease using fusion of MRI-PET imaging. *Biomed. Signal Process. Control.* 80, 104312. <http://dx.doi.org/10.1016/j.bspc.2022.104312>.
- Rezaii, N., Wong, B., Aisen, P., Beckett, L., Dage, J.L., Eloyan, A., Foroud, T., Womack, K., Carrillo, M.C., Kramer, J.H., Jack, C.R., Koeppe, R., Kukull, W.A., Nudelman, K., Polsinelli, A.J., Raman, R., Rumbaugh, M., Taurone, A., Thangarajah, M., Toga, A., Vemuri, P., Atri, A., Clark, D., Day, G.S., Duara, R., Graff-Radford, N.R., Grant, I., Honig, L.S., Johnson, E.C.B., Jones, D.T., Masdeu, J.C., Mendez, M.F., Musiek, E., Onyike, C.U., Riddle, M., Rogalski, E., Salloway, S., Sha, S., Turner, R. Scott, Wingo, T.S., Wolk, B.A., Hammers, D., Kirby, K., Rabinovici, G.D., Apostolova, L.G., Dickerson, B.C., Joie, R. La, Blazhenets, G., Touroutoglou, A., Ghetti, B., Grinberg, L.T., Murray, M.E., Newell, K., the LEADS Consortium, 2025. Voiceprints of cognitive impairment: analyzing digital voice for early detection of Alzheimer's and related dementias. *npj Dementia* 1 (1), 35. <http://dx.doi.org/10.1038/s44400-025-00040-0>.
- Rieke, N., Hancox, J., Li, W., F. Milletari, H.R. Roth, Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R.M., Trask, A., Xu, D., Baust, M., Cardoso, M.J., 2020. The future of digital health with federated learning. *npj Digital Medicine* 3 (1), 119. <http://dx.doi.org/10.1038/s41746-020-00323-1>.
- Rigny, L., Fletcher-Lloyd, N., Capstick, A., Nilforooshan, R., Barnaghi, P., 2024. Assessment of sleep patterns in dementia and general population cohorts using passive in-home monitoring technologies. *Commun. Med.* 4 (1), 222. <http://dx.doi.org/10.1038/s43856-024-00646-0>.
- Saeed, A., Waris, A., Fuwad, A., Iqbal, J., Khan, J., AlQahtani, D., Gilani, O., Shah, U.H., 2024. Random survival forest model for early prediction of Alzheimer's disease conversion in early and late mild cognitive impairment stages. *PLoS ONE* 19 (12), e0314725. <http://dx.doi.org/10.1371/journal.pone.0314725>.
- Saxena, R.R., Khan, A., 2025a. Assessing the practicality of designing a comprehensive intelligent conversation agent to assist in dementia care. In: *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF.* pp. 655–663. <http://dx.doi.org/10.5220/0013191600003911>.
- Saxena, R.R., Khan, A., 2025b. Machine learning-based clinical decision support systems in dementia care. In: *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025)-Volume 2: HEALTHINF.* pp. 664–671. <http://dx.doi.org/10.5220/0013194100003911>.
- Scott, E.C.S., Hoskin, P.J., 2024. Health inequalities in cancer care: a literature review of pathways to diagnosis in the United Kingdom. *eClinicalMedicine* 76, 102864. <http://dx.doi.org/10.1016/j.eclinm.2024.102864>.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2020. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* 128, 336–359. <http://dx.doi.org/10.1007/s11263-019-01228-7>.
- Seyyed-Kalantari, L., Zhang, H., McDermott, M.B.A., Chen, I.Y., Ghassemi, M., 2021. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Med.* 27 (12), 2176–2182. <http://dx.doi.org/10.1038/s41591-021-01595-0>.

- Sharshar, A., Ashraf, Y., Bakr, T., Hassan, S., Elgendy, H., Yaqub, M., Guizani, M., 2025. Not only grey matter: OmniBrain for robust multimodal classification of Alzheimer's disease. In: *Proceedings of the ICCV Workshop on Computer Vision for Automated Medical Diagnosis*. CVAMD, pp. 1311–1320. <http://dx.doi.org/10.1109/iccvw69036.2025.00140>.
- Sheikh, F., Marouf, A. Al, Rokne, J.G., Alhaji, R., 2025. Lightweight deep learning models with explainable AI for early alzheimer's detection from standard MRI scans. *Diagnostics* 15 (21), 2709. <http://dx.doi.org/10.3390/diagnostics15212709>.
- Sheng, J., Zhang, Q., Zhang, Q., Wang, L., Yang, Z., Xin, Y., Wang, B., 2024. A hybrid multimodal machine learning model for detecting Alzheimer's disease. *Comput. Biol. Med.* 170, 108035. <http://dx.doi.org/10.1016/j.compbiomed.2024.108035>.
- Simá, D.M., Phan, T.V., Eyndhoven, S. Van, Vercruyssen, S., Magalhães, R., Liseune, A., Brys, A., Frenyo, P., Terzopoulos, V., Maes, C., Guo, J., Hughes, R., Gabr, R.E., Huijbers, W., Saha-Chaudhuri, P., Curiale, G.G., Becker, A., Belachew, S., Heck, W. Van, Ribbens, A., Smeets, D., 2024. Artificial intelligence assistive software tool for automated detection and quantification of amyloid-related imaging abnormalities. *JAMA Netw. Open* 7 (2), e2355800. <http://dx.doi.org/10.1001/jamanetworkopen.2023.55800>.
- Song, B., Yoshida, S., 2024. 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* 19 (5), <http://dx.doi.org/10.1371/journal.pone.0303278>.
- von der Stüick, M.S., Vuskov, R., Westfechtel, S., Siepmann, R., Kuhl, C., Truhn, D., Nebelung, S., 2026. Visual large language models in radiology: A systematic multimodel evaluation of diagnostic accuracy and hallucinations. *Life* 16 (1), <http://dx.doi.org/10.3390/life16010066>.
- Sujan, M., Smith-Frazier, C., Malamateniou, C., Connor, J., Gardner, A., Unsworth, H., Husain, H., 2023. Validation framework for the use of AI in healthcare: overview of the new British standard BS30440. *BMJ Health Care Inform.* 30 (1), e100749. <http://dx.doi.org/10.1136/bmjhci-2023-100749>.
- Suleman, M.U., Mursaleen, M., Khalil, U., Khan, S.A., Saboor, A., Hussnain, M.A., Zahir, M., Khan, U.A., Rehman, D.U., Tabassum, S.N., 2025. Integration of AI diagnostic tools into clinical practice for alzheimer's disease: barriers and solutions. *Ann. Med. Surg.* 88 (2), 1488–1494. <http://dx.doi.org/10.1097/MS9.0000000000004505>.
- Sylvester, S., Sagehorn, M., Gruber, T., Atzmueller, M., Schöne, B., 2024. SHAP value-based ERP analysis (SHERPA): Increasing the sensitivity of EEG signals with explainable AI methods. *Behav. Res. Methods* 56 (6), 6067–6081. <http://dx.doi.org/10.3758/s13428-023-02335-7>.
- Tanveer, M., Sajid, M., Akhtar, M., Quadir, A., Goel, T., Aimen, A., Mitra, S., Zhang, Y.-D., Lin, C.T., Ser, J.D., 2024. Fuzzy deep learning for the diagnosis of Alzheimer's disease: approaches and challenges. *IEEE Trans. Fuzzy Syst.* 32 (10), 5477–5492. <http://dx.doi.org/10.1109/TFUZZ.2024.3409412>.
- Tejani, A.S., Klontzas, M.E., Gatti, A.A., Mongan, J.T., Moy, L., Park, S.H., Kahn, Jr., C.E., 2024. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. *Radiol. Artif. Intell.* 6 (4), e240300. <http://dx.doi.org/10.1148/ryai.240300>.
- Tian, N., Colombo, G., Schinazi, V., 2025. From play to detection: Mini-SPACE as a serious game for unsupervised cognitive impairment screening. <https://arxiv.org/abs/2511.12068>.
- Tinauer, C., Sackl, M., Stollberger, R., Schmidt, R., Ropele, S., Langkammer, C., 2025. Skull-stripping induces shortcut learning in MRI-based Alzheimer's disease classification. *Insights Into Imaging* 16, 283. <http://dx.doi.org/10.1186/s13244-025-02158-4>.
- Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature Med.* 25 (1), 44–56. <http://dx.doi.org/10.1038/s41591-018-0300-7>.
- Trinh, M.-N., Shahbaba, R., Stark, C.E.L., Ren, Y., 2024. Alzheimer's disease detection using data fusion with a deep supervised encoder. *Front. Dement.* 3, 1332928. <http://dx.doi.org/10.3389/frdem.2024.1332928>.
- Tuan, P.M., Adel, M., Trung, N.L., Guedj, E., 2024. Does brain atlas choice matter? An empirical study in Alzheimer's diagnosis using FDG-PET images. In: *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. <http://dx.doi.org/10.1109/APSIPAASC63619.2025.10848822>.
- Valanarasu, J.M.J., Guo, P., VS, V., Patel, V.M., 2024. On-the-fly test-time adaptation for medical image segmentation. In: *Proceedings of Machine Learning Research (Medical Imaging with Deep Learning, MIDL)*, Vol. 227, pp. 586–598.
- Vasey, B., Nagendran, M., Campbell, B., Clifton, D.A., Collins, G.S., Denaxas, S., Denniston, A.K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B.A., Mathur, P., McCradden, M.D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D.S.W., Watkinson, P., Weber, W., Wheatstone, P., McCulloch, P., 2022. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* 377, e070904. <http://dx.doi.org/10.1136/bmj-2022-070904>.
- Wang, Z., Bao, R., Wu, Y., Liu, G., Yang, L., Zhan, L., Zheng, F., Jiang, W., Zhang, Y., 2024. Self-guided knowledge-injected graph neural network for alzheimer's diseases. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024: 27th International Conference, Marrakesh, Morocco, October 6–10, 2024, Proceedings, Part II*. Springer-Verlag, Berlin, Heidelberg, pp. 378–388. http://dx.doi.org/10.1007/978-3-031-72069-7_36.
- Wang, Y., Chen, H.-J., Cheng, Y., Xie, Y., Cheng, Y., Zhao, S., Jiang, Y., Bai, T., Huo, Y., Wang, K., Zhang, M., Huang, W., Feng, G., Han, Y., Shu, N., 2025. Multimodal integration of plasma biomarkers, MRI, and genetic risk to predict cerebral amyloid burden in Alzheimer's disease. *NeuroImage* 322, 121550. <http://dx.doi.org/10.1016/j.neuroimage.2025.121550>.
- Wang, C., Piao, S., Wang, J., Li, Z., Chen, T., Cui, M., Zhao, J., Guo, Q., Zhang, J., Xie, F., Li, Y., Shan, H., 2026a. Manifold diffusion transformer enables diagnostic multi-tracer PET synthesis from MRI for early detection of alzheimer's disease. *Innov. Inform.* 2 (2), 100048. <http://dx.doi.org/10.59717/j.xinn-inform.2026.100048>.
- Wang, L., Zhang, Z., Wang, D., Cao, W., Zhou, X., Zhang, P., Liu, J., Fan, X., Tian, F., 2023. Human-centered design and evaluation of AI-empowered clinical decision support systems: a systematic review. *Front. Comput. Sci.* 5, 1187299. <http://dx.doi.org/10.3389/fcomp.2023.1187299>.
- Wang, M.-L., Zhang, R.-P., Wu, W.-J., Lu, Y., Wei, X.-E., Sun, Z., Guan, B.-H., Zhang, J.-J., Wu, X., Zhang, L., Wang, T.-L., Li, Y.-H., 2026b. Evaluation of large language models for diagnostic impression generation from brain MRI report findings: A multicenter benchmark and reader study. *npj Digital Medicine* 9 (1), 187. <http://dx.doi.org/10.1038/s41746-026-02380-4>.
- Wei, Y., Shan, Y., Fan, W., Huang, X., Ding, J., Mao, M., Liu, Q., Tian, M., Li, X., Lu, J., Chang, H., Dong, Y., Du, Y., Qiu, C., Han, X., Wang, Y., 2025. Heart rate response and recovery during exercise and dementia risk: a prospective UK biobank study. *Sci. Rep.* 15 (1), 38362. <http://dx.doi.org/10.1038/s41598-025-22299-2>.
- Wells, B.J., Nguyen, H.M., McWilliams, A., Pallini, M., Bovi, A., Kuzma, A., Kramer, J., Chou, S.H., Hetherington, T., Corn, P., Taylor, Y.J., Cuisson, A., Gagen, M., Isreal, M., Consortium, F.-A., 2025. A practical framework for appropriate implementation and review of artificial intelligence (FAIR-AI) in healthcare. *npj Digital Medicine* 8 (1), 514. <http://dx.doi.org/10.1038/s41746-025-01900-y>.
- Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., J. Samper-González, A. Routier, Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O., 2020. Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. *Med. Image Anal.* 63, 101694. <http://dx.doi.org/10.1016/j.media.2020.101694>.
- Wilm, F., Reimann, M., Taubmann, O., A. Mühlberg, K. Breininger, 2024. Appearance-based debiasing of deep learning models in medical imaging. In: Maier, A., Deserno, T.M., Handels, H., Maier-Hein, K., Palm, C., Tolxdorff, T. (Eds.), *Bildverarbeitung Für Die Medizin 2024*. Springer Fachmedien Wiesbaden, Wiesbaden, pp. 19–24. http://dx.doi.org/10.1007/978-3-658-44037-4_9.
- Wood, D.A., Kafiabadi, S., Busaidi, A.A., Guilhem, E., Montvila, A., Lynch, J., Townend, M., Agarwal, S., Mazumder, A., Barker, G.J., Ourselin, S., Cole, J.H., Booth, T.C., 2022. Deep learning models for triaging hospital head MRI examinations. *Med. Image Anal.* 78, 102391. <http://dx.doi.org/10.1016/j.media.2022.102391>.
- Xia, M., Bayerlein, R., Chemli, Y., Liu, X., Ouyang, J., Lin, M., Fakhri, G. El, Badawi, R.D., Li, Q., Liu, C., 2025. On hallucinations in artificial intelligence-generated content for nuclear medicine imaging (the DREAM report). *J. Nucl. Med.* <http://dx.doi.org/10.2967/jnumed.125.270653>.
- Yang, Z., Cheng, L., Gan, H., Shi, M., Huang, Z., Cheng, B., 2025. PKFA-GCN: Prior knowledge-guided feature alignment graph convolutional network with effective connectivity for alzheimer's disease classification. In: *2025 IEEE International Conference on Bioinformatics and Biomedicine*. BIBM, pp. 1925–1928. <http://dx.doi.org/10.1109/BIBM66473.2025.11356238>.
- Yao, Z., Xie, W., Chen, J., Zhan, Y., Wu, X., Dai, Y., Pei, Y., Wang, Z., Zhang, G., 2025. IT: An interpretable transformer model for alzheimer's disease prediction based on PET/MR images. *NeuroImage* 311, 121210. <http://dx.doi.org/10.1016/j.neuroimage.2025.121210>.
- Ye, J., Zeng, A., Pan, D., Zhang, Y., Zhao, J., Chen, Q., Liu, Y., 2024. MAD-Former: A traceable interpretability model for Alzheimer's disease recognition based on multi-patch attention. *IEEE J. Biomed. Health Inform.* 28 (6), 3637–3648. <http://dx.doi.org/10.1109/JBHI.2024.3368500>.
- Yi, F., Yang, H., Chen, D., Qin, Y., Han, H., Cui, J., Bai, W., Ma, Y., Zhang, R., Yu, H., 2023. XGBoost-SHAP-based interpretable diagnostic framework for Alzheimer's disease. *BMC Med. Inform. Decis. Mak.* 23 (1), <http://dx.doi.org/10.1186/s12911-023-02238-9>.
- Yin, C., Imms, P., Cheng, M., Amgalan, A., Chowdhury, N.F., Massett, R.J., Chaudhari, N.N., Chen, X., Thompson, P.M., Bogdan, P., Irimia, A., 2023. Anatomically interpretable deep learning of brain age captures domain-specific cognitive impairment. *Proc. Natl. Acad. Sci.* 120 (2), e2214634120. <http://dx.doi.org/10.1073/pnas.2214634120>.
- Yousaf, K., Mehmood, Z., Saba, T., Rehman, A., Munshi, A.M., Alharbey, R., Rashid, M., 2019. Mobile-health applications for the efficient delivery of health care facility to people with dementia (PwD) and support to their carers: A survey. *BioMed Res. Int.* 2 019, 7151475. <http://dx.doi.org/10.1155/2019/7151475>.
- Yu, Q., Ma, Q., Da, L., Li, J., Wang, M., Xu, A., Li, Z., Li, W., 2024a. A transformer-based unified multimodal framework for Alzheimer's disease assessment. *Comput. Biol. Med.* 180, 108979. <http://dx.doi.org/10.1016/j.compbiomed.2024.108979>.
- Yu, F., Moehring, A., Banerjee, O., Salz, T., Agarwal, N., Rajpurkar, P., 2024b. Heterogeneity and predictors of the effects of AI assistance on radiologists. *Nature Med.* 30 (3), 837–849. <http://dx.doi.org/10.1038/s41591-024-02850-w>.

- Zahedipour, M., Abadeh, M., Saniee, Shojaei, S., 2026. Alzheimer's disease prediction via an explainable CNN using genetic algorithm and SHAP values. *PLoS ONE* 21, <http://dx.doi.org/10.1371/journal.pone.0337800>.
- Zajac, H.D., Andersen, T.O., Kwasa, E., Wanjohi, R., Onyinkwa, M.K., Mwaniki, E.K., Gitau, S.N., Yaseen, S.S., Carlsen, J.F., Fraccaro, M., Nielsen, M.B., Chen, Y., 2025. Towards clinically useful AI: From radiology practices in Global South and North to visions of AI. *ACM Transactions on Computer-Human Interaction* 32 (2), 1–38. <http://dx.doi.org/10.1145/3715115>.
- Zhang, J., He, X., Liu, Y., Cai, Q., Chen, H., Qing, L., 2023. Multi-modal cross-attention network for alzheimer's disease diagnosis with multi-modality data. *Comput. Biol. Med.* 162, 107050. <http://dx.doi.org/10.1016/j.compbiomed.2023.107050>.
- Zhang, Z., Wang, T., Hu, Z., Yang, L.-Z., Li, H., 2025a. DEMENTIA: A hybrid attention-based multimodal and multi-task learning framework with expert knowledge for alzheimer's disease assessment from speech. *IEEE J. Biomed. Health Inform.* 29 (4), 2957–2968. <http://dx.doi.org/10.1109/JBHI.2024.3509620>.
- Zhang, L., Wu, J., Wang, L., Wang, L., Steffens, D.C., Qiu, S., Potter, G.G., Liu, M., 2025b. Brain anatomy prior modeling to forecast clinical progression of cognitive impairment with structural MRI. *Pattern Recognit.* 165, 111603. <http://dx.doi.org/10.1016/j.patcog.2025.111603>.
- Zhao, A., Cui, E., Leroux, A., Zhou, X., Muschelli, J., Lindquist, M.A., Crainiceanu, C.M., 2025. Objectively measured physical activity using wrist-worn accelerometers as a predictor of incident alzheimer's disease in the UK biobank. *Journals Gerontol.: Ser. A* 80 (2), glae287. <http://dx.doi.org/10.1093/gerona/glae287>.
- Zubair, M., Jaffar, A., Hussain, S., Akram, S., 2025. Alzheimer's disease classification using a hybrid deep learning approach with multi-layer U-net segmentation and XAI driven analysis. *PLoS One* 20 (9), e0332572. <http://dx.doi.org/10.1371/journal.pone.0332572>.
- Zubasti, P., Berlanga, A., Patricio, M.A., Molina, J.M., 2024. Assessing the interplay of attributes in dementia prediction through the integration of graph embeddings and unsupervised learning. In: *Artificial Intelligence for Neuroscience and Emotional Systems*. Springer Nature Switzerland, Cham, pp. 371–380. http://dx.doi.org/10.1007/978-3-031-61140-7_35.