Journal on
Internet of Things

Tech Science Press

https://doi.org/10.32604/jiot.2026.090447

ARTICLE

An IoT-Oriented Smart Healthcare Framework for Alzheimer's Disease Detection: Integrating MRI and Electronic Health Records Using Multimodal Deep Learning

Fahim Faysal Sakib[1,*], **Md. Iftekharul Mobin**[1], **Pretom Chandra Roy**[1] **and Al-Sakib Khan Pathan**[2,*]

[1]Department of Computer Science and Engineering, American International University-Bangladesh, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh

*Corresponding Authors: Fahim Faysal Sakib. Email: fahim.sakib.10@gmail.com; Al-Sakib Khan Pathan. Email: sakib.pathan@gmail.com

Received: 05 August 2026; Accepted: 26 October 2026; Published: 13 November 2026

ABSTRACT: Diagnosing Alzheimer's Disease (AD) begins with neuroimaging, followed by corroboration from medical history and comorbidities, physiological measurement, and both genetic and brain imaging markers. Numerous artificial intelligence approaches reported to date only leverage a single modality. This study presents a clinically applicable multimodal method that uses three independent data streams: (i) magnetic resonance imaging (MRI) neuroimaging; (ii) physiological variables (e.g., heart rate, oxygen saturation, blood pressure, tremor indices, and blood glucose), represented by structured electronic health record (EHR) proxies for Internet of Things (IoT)-acquired signals, together with longitudinal clinical history (comorbidities and cognitive scores); and (iii) single-nucleotide polymorphism (SNP) genetic markers. A customized convolutional neural network (CNN) that implements a 1×1 bottleneck and progressively scales filter depth while batch-normalizing each layer achieves 98% accuracy, 98% precision, 98% recall, and F1 score of 98% in separating cognitively normal (CN), mild cognitive impairment (MCI), and AD severity in an imaging-only experiment, outperforming the other imaging architectures considered. When used to classify structured clinical variables in a binary AD screening task, the light gradient-boosting machine (LightGBM) achieved 96% accuracy, 96% precision, 95% recall, and F1 score of 96%. A systematic scenario analysis (unimodal, pairwise, and full trimodal fusion) compared the contribution of each data stream; EHR and SNP combined (without imaging) achieved 78% accuracy, 78% precision, 79% recall, and F1 score of 78%. These results suggest the potential development of point-of-care AD diagnosis within an IoT-oriented framework.

KEYWORDS: Alzheimer's disease; clinical AI; deep learning; electronic health records; IoT-oriented physiological monitoring; MRI neuroimaging; multimodal data fusion

1 Introduction

1.1 Alzheimer's Disease: Clinical Overview

Alzheimer's Disease (AD) is a progressive neurodegenerative disease that accounts for 60%–70% of dementia cases [1]. It develops from mild cognitive symptoms (e.g., word-finding difficulties and visuospatial impairment) to severe symptoms, including memory loss, loss of functional independence, and complete dependence in later stages. The disease most often afflicts those over 65 years of age. Today, more than 55 million people in the world suffer from dementia, and by 2050, the figure may grow to 153 million [2], necessitating systems which can facilitate early and scalable decentralized AD diagnostics. Pathologically,

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the disease involves the accumulation of extracellular β-amyloid plaques and intracellular neurofibrillary tangles of hyperphosphorylated tau, together with progressive neuronal loss and brain atrophy, especially in the hippocampus and entorhinal cortex.

1.2 Challenges in Early Detection of Alzheimer's Disease

Because the clinical and pathological features of AD vary widely between patients, the changes it produces are distributed across several distinct biological systems. Each of these systems is captured by a different measurement modality, so no single modality is sufficient on its own for a confident diagnosis. Accordingly, diagnosis in practice rests on a multisource assessment carried out by a specialist clinician, typically a neurologist, geriatrician, or old-age psychiatrist. Clinical evaluation should include cognitive screening scores (Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR), and Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog)), comorbidities (hypertension, diabetes, and cardiovascular disease history), physiological parameters (heart rate, blood pressure, oxygen saturation, and tremor), genetic markers (apolipoprotein E (APOE) $\varepsilon 4$ and single-nucleotide polymorphisms (SNPs)), cerebrospinal fluid (CSF) markers, and structural imaging of the brain.

Early detection is further complicated by three factors. Firstly, the pathological changes of AD begin years before cognitive symptoms become measurable, so biomarker abnormalities precede clinical diagnosis by a considerable interval [3]. Secondly, the confirmatory biomarkers are invasive or costly, since CSF analysis requires lumbar puncture and amyloid positron emission tomography (PET) requires specialized radiotracers [4]. Thirdly, the transition from cognitively normal (CN) aging to mild cognitive impairment (MCI) is gradual, which blurs the boundary between the two stages at precisely the point where intervention is most valuable [5].

1.3 Role of Smart Healthcare and IoT in Alzheimer's Diagnosis

Smart healthcare systems combine wearable sensors, wireless connectivity, and cloud or edge analytics to monitor patients continuously outside the clinic [6]. In such systems, Internet of Things (IoT) devices acquire physiological signals, transmit them through secure gateways, and integrate them with clinical records for artificial intelligence (AI)-driven decision support [7]. For AD, continuous monitoring is attractive because it captures fluctuations between clinical appointments and can indicate deterioration earlier than episodic assessment [8].

The framework proposed in this study is IoT-oriented. It is designed for a setting in which remote devices continuously monitor heart rate, peripheral oxygen saturation (SpO_2), temperature, tremor frequency, blood glucose, and blood pressure, and in which these streams are combined with the patient's electronic health record (EHR) and magnetic resonance imaging (MRI) data in the manner that a physician integrates them. As a living cohort equipped with IoT devices could not be recruited for this study, the framework is instantiated on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [9], in which structured EHR components serve as proxies for the IoT physiological streams. The IoT layer is therefore conceptual in the present work, and the mapping between each intended signal and its ADNI equivalent is presented in Section 4.2.

1.4 MRI-Based Detection of Alzheimer's Disease

Structural MRI is the most widely used imaging modality for AD assessment, since it is non-invasive, widely available, and sensitive to the hippocampal and cortical atrophy that accompanies disease progression [10]. Deep learning (DL), and convolutional neural networks (CNNs) in particular, has become the dominant approach for learning discriminative features directly from MRI, and systematic reviews report

high accuracy for the classification of AD stages [11]. However, MRI captures only the structural consequence of the disease and cannot by itself convey the clinical context in which a scan is interpreted.

1.5 Electronic Health Records for Clinical Decision-Making

EHRs contain demographic information, comorbidities, vital signs, laboratory results, medication histories, and cognitive assessments collected during routine care. Because these data already exist, they incur no additional acquisition cost and remain available in settings where MRI is not [12]. Machine learning (ML) models trained on structured clinical variables have been used to predict AD status and future cognitive decline [9,13], and EHR platforms provide the natural point of integration for IoT-acquired measurements [7].

1.6 Need for Multimodal Deep Learning

Each of the modalities above observes a different aspect of the disease: MRI reflects neuroanatomical change, EHR variables reflect clinical and physiological context, and SNP markers reflect genetic risk. Multimodal DL offers a means of combining these complementary views within a single model, and recent reviews identify multimodal fusion as the principal direction of AD detection research [14,15]. Venugopalan et al. [16] showed that integrating imaging, EHR, and genetic data can outperform single-modality models for AD staging. Motivated by these observations, this study develops an IoT-oriented multimodal framework that combines MRI, EHR-derived physiological and clinical variables, and SNP markers for AD detection and severity classification.

1.7 Core Contributions

1. **Lightweight Customized Convolutional Neural Network (CNN):** A compact convolutional architecture is presented that combines established design elements, namely a 1×1 bottleneck layer, depth-scaled filters (8–256), and per-block Batch Normalization, in a configuration tailored to ADNI MRI slices. The model achieves **98% multiclass accuracy** for CNN-based Alzheimer's Disease (AD) severity classification on the ADNI data used in this study.
2. **Multimodal Information-Theoretic Formulation:** We derive a set of formal equations in Section 3 that describe how combining three modalities changes diagnostic uncertainty. These equations quantify how much residual uncertainty trimodal fusion removes, and how much mutual information with the diagnostic label it gains, relative to any single modality.
3. **Systematic Three-Scenario Analysis:** Unimodal, pairwise multimodal, and full trimodal fusion of MRI, EHR, and SNP data are compared systematically across seven classifier families, which shows how much each modality and each combination contributes to AD classification.
4. **IoT-to-Clinical Translation:** A detailed feature mapping (Section 4.2) and a proposed layered deployment architecture (Section 5.5) outline a pathway from IoT-acquired physiological data to AI-informed diagnosis.

1.8 Structure of the Paper

The structure of this paper is as follows: Section 2 provides a review of the existing literature. Section 3 presents the multimodal theoretical formulation. Section 4 describes the dataset and IoT mapping used in this study. Section 5 outlines the proposed framework, its layered architecture, and the proposed deployment architecture. Section 6 presents the experimental results for all scenarios. Section 7 discusses the

results. Section 8 addresses explainability and clinical trust, and Section 9 discusses security, privacy, and ethical considerations. Section 10 addresses the limitations and suggests directions for future research. Section 11 concludes the paper.

2 Related Works

2.1 Alzheimer's Disease Detection Techniques

Computational approaches to AD detection fall into six broad families according to the modality they exploit: neuropsychological and clinical assessment, structural and molecular neuroimaging, fluid biomarkers obtained from cerebrospinal fluid or blood, genomic markers, behavioral and speech signals, and continuously acquired physiological measurements. Recent surveys establish that each family has been studied extensively in isolation and that multimodal fusion has become the dominant direction of the field [14,15,17]. Comprehensive reviews of machine learning (ML) applied to multimodal neuroimaging [10,11] and of explainability in AD classification [18] reach the same conclusion. This section reviews each family in turn, beginning with traditional clinical assessment and concluding with the convolutional architectures on which the present work builds.

2.2 Traditional Clinical and Neurological Assessment

Clinical diagnosis of AD rests on neuropsychological instruments administered by a specialist, supported where available by fluid and imaging biomarkers [19]. Petersen et al. [5] characterize the ADNI cohort using exactly this battery, reporting that ADAS-Cog scores moved by 1.1 points for MCI subjects and 4.3 points for AD subjects over twelve months, while normal controls showed minimal movement, and that baseline CSF Aβ-42 separated the three diagnostic groups and predicted subsequent cognitive change. Schaap et al. [3] extend this over two decades of ADNI follow-up and establish that biomarker abnormalities emerge in an ordered sequence well before cognitive decline becomes measurable.

Three limitations of this pathway motivate computational alternatives. Firstly, the cognitive instruments are subject to practice effects, educational confounding, and ceiling effects at the preclinical stage, at which the disease is most amenable to intervention. Secondly, the confirmatory biomarkers are invasive or costly, since CSF assay requires lumbar puncture and amyloid PET requires cyclotron-dependent radiotracers, a constraint which has driven sustained effort towards blood-based alternatives [4]. Thirdly, assessment is episodic rather than continuous, so signals that fluctuate between appointments are not captured at all.

2.3 MRI-Based Machine Learning and Deep Learning Methods

Currently, extensive studies on the diagnosis of Alzheimer's disease using AI have been performed, focusing primarily on neuroimaging techniques that involve structural MRI and PET images. According to Arafa et al. [20], CNN-based methods for analyzing images from the Kaggle MRI dataset achieved an accuracy of more than 97%. The research utilized the Visual Geometry Group 16-layer network (VGG16) architecture, which completed the diagnosis with an accuracy of 97.44%. However, the limitation of these purely scanning approaches is that they cannot provide sufficient clinical interpretation of the results, especially in borderline and conditional situations. Similarly, De Silva and Kunz [21] developed an MRI-based CNN classifier using the Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) dataset. The results revealed that the accuracy of the classification results reached 89% with an area under the curve (AUC) of 92%. Nevertheless, it was designed for binary classification only and did not consider MCI cases. In their study, Faisal and Kwon [22] successfully classified patients diagnosed with Alzheimer's disease, mild cognitive impairment (MCI), and cognitively normal (CN) using a modified convolutional neural network.

The accuracy achieved was 96.12%, while the F1 score was 95.23%. However, their model was limited to the use of imaging data.

Transfer learning has been applied widely to the same task. Jraba et al. [23] combine transfer-learned deep CNN backbones with data augmentation for AD classification, and Alwakid et al. [24] report that image preprocessing applied ahead of the network materially affects detection performance. Amrutesh et al. [25] compare transfer-learned models against classical learners and obtain 92.14% binary accuracy. Across these studies, the recurring limitation is the one identified above: accuracy is high on curated collections, but the models observe a single modality and therefore reproduce only one component of the clinical assessment described in Section 2.2.

2.4 IoT Wearables for Continuous Patient Monitoring

The episodic nature of conventional assessment has motivated a body of work on continuous monitoring through the Internet of Things (IoT). Sathish et al. [8] construct an intelligent decision-making system that couples wearable sensing with deep learning (DL) for early AD detection, establishing that consumer-grade devices can supply diagnostically relevant streams outside a clinical setting. Yuan et al. [26] approach the same problem from the hardware side, presenting a configurable system-on-chip that processes multi-modal physiological signals for AD detection at the edge, which removes the requirement for continuous transmission of patient data to a central server. Ullah et al. [27] demonstrate contactless monitoring using electromagnetic radar data, classifying patients without requiring any device to be worn at all. At the system level, Ghadi et al. [6] review the integration of wearable technology and AI for remote patient care and identify data privacy, interoperability with existing health systems, and regulatory ambiguity as the principal barriers to adoption.

These studies establish feasibility, but two gaps persist. None integrates the continuously acquired physiological stream with structural neuroimaging, and none maps the acquired signals onto the structured clinical variables against which a diagnostic model would have to be trained.

2.5 Wearable Sensor Data for Cognitive and Behavioral Analysis

Beyond physiological measurement, wearable and ambient devices capture behavioral signals that carry cognitive information, of which speech is the most extensively studied. The detection of AD using speech has been investigated by Ying et al. [28], employing Support Vector Machines (SVM) with a combination of acoustic (IS10) and linguistic features, yielding an accuracy of 89.1% on the National Conference on Man-Machine Speech Communication 2021 (NCMMSC2021) database and 83.7% on the Alzheimer's Dementia Recognition through Spontaneous Speech only (ADReSSo) dataset. Thus, speech was shown to be an affordable IoT monitoring signal. Yang et al. [29] employed a similar approach by incorporating an additional early diagnosis of Alzheimer's disease based on a multi-attention mechanism (EDAMM) architecture that uses attention-based fusion of acoustic and linguistic representations.

2.6 EHR-Based Alzheimer's Risk Prediction

Structured electronic health record (EHR) data offer a complementary route, since they are already collected during routine care and therefore incur no additional acquisition cost. El-Sappagh et al. [12] construct an early-fusion model over cost-effective multimodal clinical variables and report 90.51% binary accuracy using SVM, random forest (RF), and k-nearest neighbors. Baglat et al. [30] apply logistic regression, RF, and adaptive boosting (AdaBoost) to the Open Access Series of Imaging Studies (OASIS) cohort and obtain 86.8% binary accuracy. Kavitha et al. [13] reach comparable performance on structured cognitive and demographic variables for early-stage prediction. Velazquez and Lee [31] assemble a multimodal ensemble

for conversion prediction from early MCI, and Karaman et al. [9] conduct an empirical study of multi-modal prediction of future decline across the ADNI cohort.

Two observations follow. Tree-based ensembles remain strongly competitive on tabular clinical data, which is consistent with the characteristics of the modality, namely heterogeneous variable types, irregular sampling intervals, missing longitudinal values, and pronounced class imbalance. However, the reported accuracies fall consistently below those obtained on imaging, which indicates that structured data alone carry less discriminative information than neuroanatomy for severity staging, even while supplying context that imaging cannot.

2.7 Multimodal Learning in Healthcare

The difficulties encountered during the exploitation of unimodal imaging techniques led to the emergence of multimodal fusion. Wang et al. [32] proposed a hybrid DenseNet3D-support vector regression (SVR) framework to predict cognitive decline using MRI scans, biomarkers, and cognitive scores based on the North American ADNI (NA-ADNI) dataset (506 contributions), demonstrating a 22.4% reduction in allocation bias and a 37% cut in sample size compared with standard methods. Subsequently, models proposed by Abdelaziz et al. [33] relied on a nominally similar supervised CNN architecture with modality-specific pathways for MRI, PET, SNPs, and clinical data from ADNI-805, achieving positive outcomes in regression and classification tasks over single-modality baselines. Likewise, Lee et al. [34] analyzed demographic and multimodal imaging data, including T1 and T2 fluid-attenuated inversion recovery (T2-FLAIR) MRI, as well as amyloid PET, and successfully created machine learning models that demonstrated the highest efficiency possible for machine learning modeling in the Gradient Boosting Machine (GBM) with demographic data and amyloid feature profile, as T2-FLAIR had almost no impact on performance. Agostinho et al. [35] implemented an SVM-based ensemble fusion technique on structural MRI (sMRI), fluorodeoxyglucose PET (FDG-PET), florbetapir PET (AV45-PET), and diffusion tensor imaging (DTI) data obtained from 486 MCI patients and determined that AV45-PET was the single most effective application of a technique and that, despite the SVM-based approach, performance may not always be improved by fusion methods.

Wang et al. [36] integrated MRI, clinical, and genetic data from 252 MCI patients, achieving 92.92% accuracy and a 0.962 AUC with multimodal fusion. Through utilizing Recurrent Neural Networks (RNNs) on certain longitudinal multimodal data, Lee et al. [37] were able to accomplish an accuracy level of 81% along with an AUC of 0.86 in predicting the conversion from MCI to AD. Young et al. [38] showed that the Gaussian process classification method is superior to SVM and unimodal approaches in terms of MCI conversion prediction and demonstrated the necessity of advanced multimodal fusion methods. Akkaya and Kalkan [39] introduced a gene-to-image fusion approach that integrates genomic data with its 2D image representation, thereby enhancing the AUC from 0.86 to 0.91 for AD vs. cognitively normal (CN) controls and from 0.76 to 0.88 for CN vs. Mild Cognitive Impairment (MCI). Kim et al. [40] combined MRI, FDG-PET, amyloid-PET, and tau-PET in a middle-fusion model, reporting a balanced accuracy of 1.00 for AD vs. CN and 0.76 for MCI vs. CN on their evaluation set. Ravichandran et al. [41] demonstrated that the combination of retinal and plasma biomarkers yields high performance (AUC 0.97, 93.8% sensitivity), supporting the efficacy of multimodal non-brain AD detection. A recurring constraint across this literature is that modality coverage is rarely complete in practice. Gravina et al. [42] address this directly with a multi-input, multi-output 3D CNN designed to operate under incomplete multimodal data, which reflects the clinical reality that patients seldom present with every measurement available.

2.8 CNN-Based Medical Image Analysis

The architectural choices underlying the results above warrant separate consideration, since the present work proposes a customized CNN. Turkan and Tek [43] pair a convolutional attention network for MRI-based classification with an explicit interpretability analysis, establishing that attention weights can be inspected for clinical plausibility rather than accepted on accuracy alone. Yu et al. [44] introduce high-order pooling into a tensorized generative adversarial network for AD assessment. Song et al. [45] fuse MRI and PET at the image level and report 94% binary and 74.5% multiclass accuracy, which illustrates the characteristic gap between binary and multiclass performance that motivates the severity-staging focus of this study. Muksimova et al. [46] report 94% binary accuracy with a FusionNet architecture incorporating longitudinal analysis. Adarsh et al. [47] developed a Convolutional Neural Network (CNN)-based multimodal framework with kernel discriminative dictionary learning, achieving 95.5% accuracy across AD, MCI, and cognitively normal classes.

The architectures surveyed here achieve their performance largely through depth and parameter count inherited from general-purpose backbones such as VGG16 and DenseNet, rather than through design choices matched to the characteristics of AD neuroimaging. This observation motivates the customized architecture presented in Section 5.3.

2.9 Temporal Models for Sequential and Sensor Data

Physiological streams from wearable devices and longitudinal clinical records are sequential, so models that capture temporal dependencies are natural candidates for their analysis. Recurrent neural networks (RNNs) process a sequence through a hidden state that is updated at each time step, and Lee et al. [37] used RNNs on longitudinal multimodal ADNI data to predict conversion from MCI to AD with 81% accuracy. Long short-term memory (LSTM) networks extend RNNs with input, forget, and output gates that mitigate vanishing gradients over long sequences [48], and Goyal et al. [49] combined a transfer-learned AlexNet with an LSTM to reach 96.83% multiclass accuracy. Transformer models replace recurrence with self-attention, which relates all time steps to one another directly and permits parallel training over long sequences [50]. Sathish et al. [8] couple wearable sensing with DL for early AD detection. In the present study, the ADNI proxies for physiological signals are static tabular variables rather than raw time series; temporal models therefore become relevant once the IoT acquisition layer supplies continuous sensor streams.

2.10 Multimodal Fusion Strategies

Multimodal fusion methods are commonly grouped according to the stage at which the modalities are combined [51]. Early fusion concatenates raw or lightly processed features before a single model is trained, as El-Sappagh et al. [12] do for cost-effective clinical variables. Intermediate fusion first learns a representation for each modality and then combines these representations within the network; Venugopalan et al. [16] and Zhou et al. [52] learn modality-specific latent features before joint classification, and Kim et al. [40] apply middle fusion to MRI and PET. Late fusion trains a separate model for each modality and combines their outputs or embeddings at the decision stage, as in the SVM-based ensemble of Agostinho et al. [35]. Attention-based fusion learns weights that express the relative contribution of each modality; Lu et al. [53] propose a hierarchical attention-based framework for predicting AD progression, and Chen et al. [54] combine neuroimaging and medical examination data through an attention-combined fusion network.

Early fusion is simple but sensitive to differences in scale and dimensionality between modalities. Late fusion accommodates heterogeneous learners, such as a CNN for images and gradient-boosted trees for tabular data, at the cost of modeling fewer cross-modal interactions. The present framework adopts two-stage late fusion (Section 3) for this reason.

2.11 Explainable AI for Alzheimer's Detection

Clinical adoption of AD classifiers depends on whether their decisions can be inspected and justified. Viswan et al. [18] systematically review explainable artificial intelligence (XAI) methods applied to AD classification. Gradient-weighted class activation mapping (Grad-CAM) highlights the image regions that drive a CNN prediction [55], while Shapley additive explanations (SHAP) assign each input feature a contribution to an individual prediction [56]. Jahan et al. [1] apply XAI to multimodal data for AD prediction and management, Kamal et al. [57] combine image and gene-expression data with XAI to present genes associated with AD, and Adarsh et al. [47] incorporate transparent decision-making into a CNN-based multimodal classifier. Viswan et al. [15] further examine how explainability can be maintained when several modalities are fused.

2.12 Existing Smart Healthcare Frameworks

Smart healthcare frameworks organize sensing, communication, storage, and analytics into layered architectures. Ghadi et al. [6] describe how wearable devices, cloud platforms, and AI models combine to support remote patient care, and propose federated learning as a remedy for privacy constraints. Mazhar et al. [7] examine the combination of generative AI, IoT, and blockchain in healthcare, and argue that blockchain improves EHR security, privacy, and interoperability, while standardized protocols such as Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR) ease integration with legacy systems. For AD specifically, Sathish et al. [8] and Yuan et al. [26] present wearable and edge-hardware systems, respectively. These frameworks address data acquisition and transport, but they seldom specify how continuously acquired signals are combined with neuroimaging and genetic data for severity staging.

2.13 Comparative Analysis of Existing Studies

Table 1 summarizes representative studies reviewed in this section according to the modalities used, the method, the reported result, and the principal limitation relative to the objectives of this study.

Table 1: Comparative analysis of representative Alzheimer's disease (AD) detection studies. Reported results are as stated by the original authors and were obtained on different datasets and protocols. ADReSSo, Alzheimer's Dementia Recognition through Spontaneous Speech only; AUC, area under the curve; AV45, florbetapir; CN, cognitively normal; CNN, convolutional neural network; DL, deep learning; DTI, diffusion tensor imaging; FDG, fluorodeoxyglucose; KNN, k-nearest neighbors; LR, logistic regression; MCI, mild cognitive impairment; MIRIAD, Minimal Interval Resonance Imaging in Alzheimer's Disease; MRI, magnetic resonance imaging; NCMMSC2021, National Conference on Man-Machine Speech Communication 2021; OASIS, Open Access Series of Imaging Studies; PET, positron emission tomography; RF, random forest; RNN, recurrent neural network; sMRI, structural MRI; SVM, support vector machine; VGG16, Visual Geometry Group 16-layer network. –, no quantitative result reported.

Study	Modalities	Method	Reported Result	Limitation
[8]	Wearable signals	DL	–	No neuroimaging
[12]	Clinical variables	SVM, RF, KNN	90.51% binary accuracy	No imaging
[20]	MRI (Kaggle)	VGG16	97.44% accuracy	Single modality
[21]	MRI (MIRIAD)	CNN	89% accuracy, AUC 0.92	Binary only; MCI not considered

(Continued)

Table 1 (continued)

Study	Modalities	Method	Reported Result	Limitation
[22]	MRI	Modified CNN	96.12% accuracy (3-class)	Imaging only
[28]	Speech	SVM	89.1% (NCMMSC2021), 83.7% (ADReSSo)	Single behavioral signal
[30]	Clinical (OASIS)	LR, RF, AdaBoost	86.8% binary accuracy	Binary only; no imaging
[35]	sMRI, FDG-PET, AV45-PET, DTI	SVM ensemble	Fusion did not always improve performance	Imaging modalities only
[36]	MRI, clinical, genetic	Multimodal DL	92.92% accuracy, AUC 0.962	MCI cohort only
[37]	Longitudinal multimodal	RNN	81% accuracy, AUC 0.86	MCI-to-AD conversion only
[40]	MRI, FDG/amyloid/tau PET	Middle fusion	Balanced accuracy 1.00 (AD vs. CN), 0.76 (MCI vs. CN)	Requires several PET tracers
[41]	Retinal, plasma biomarkers	Multimodal model	AUC 0.97, sensitivity 93.8%	No neuroimaging
[45]	MRI, PET	CNN with image fusion	94% binary; 74.5% multiclass	Imaging only

2.14 Research Gaps

The comparative analysis in Table 1 identifies the following research gaps, which motivate this work.

1. **Modality coverage:** Most studies analyze a single modality or combine imaging modalities only, and few combine MRI, EHR data (including the physiological variables mapped from IoT signals), and SNP genetic markers within one framework.
2. **IoT-to-clinical mapping:** Wearable and IoT studies establish acquisition feasibility, but they neither map the acquired signals onto the structured clinical variables required for model training nor combine them with neuroimaging.
3. **Architectural efficiency:** High MRI accuracy is typically obtained with large general-purpose backbones, whereas lightweight architectures configured for AD MRI, such as a 1×1 bottleneck combined with progressive depth scaling, are less explored.
4. **Theoretical justification:** Fusion is usually motivated empirically, and an information-theoretic account of how additional modalities affect diagnostic uncertainty is seldom provided.
5. **Task coverage:** Few studies evaluate both binary screening and multiclass severity staging across unimodal, pairwise, and trimodal combinations of the same data streams.

3 Multimodal Integration Framework

Let $\mathbf{x}^{(1)} \in \mathbb{R}^{d_1}$ (MRI features), $\mathbf{x}^{(2)} \in \mathbb{R}^{d_2}$ (EHR/physiological features), and $\mathbf{x}^{(3)} \in \mathbb{R}^{d_3}$ (SNP features) denote the three input modalities, and let $y \in \{CN, MCI, AD\}$ denote the diagnostic label.

3.1 Mutual Information Gain from Fusion

The mutual information between the fused representation $\mathbf{z} = [\mathbf{x}^{(1)}; \mathbf{x}^{(2)}; \mathbf{x}^{(3)}]$ and the label y decomposes as:

$$I(\mathbf{z}; y) = I(\mathbf{x}^{(1)}; y) + I(\mathbf{x}^{(2)}; y \mid \mathbf{x}^{(1)}) + I(\mathbf{x}^{(3)}; y \mid \mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \tag{1}$$

Each conditional term $I(\mathbf{x}^{(k)}; y \mid \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k-1)}) \geq 0$, so $I(\mathbf{z}; y) \geq I(\mathbf{x}^{(m)}; y)$ for any single modality m. This establishes the fact that fusion can never reduce mutual information with the label and increases it whenever modalities provide complementary information.

3.2 Conditional Entropy Reduction

The residual uncertainty about the label cannot increase when a modality is added:

$$H(y \mid \mathbf{z}) = H(y) - I(\mathbf{z}; y) \leq H(y \mid \mathbf{x}^{(m)}) \tag{2}$$

where $H(y)$ denotes the prior label entropy. Through Fano's inequality, a lower residual entropy lowers the bound on the minimum achievable classification error [58]; it does not, however, guarantee that a particular trained classifier attains that bound.

3.3 Bayesian Posterior Formulation

Under a Bayesian view, the multimodal posterior is:

$$p(y \mid \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}) \propto p(\mathbf{x}^{(1)} \mid y)\, p(\mathbf{x}^{(2)} \mid y)\, p(\mathbf{x}^{(3)} \mid y)\, p(y) \tag{3}$$

assuming conditional independence of modalities given y, an approximation that is clinically reasonable because MRI, physiological signals, and genetics measure distinct biological phenomena. Under this assumption, Eq. (3) shows that each informative modality contributes an additional likelihood term that can sharpen the posterior and reduce the probability mass assigned to incorrect classes. In practice, the gain depends on how accurately each likelihood is estimated from finite data.

3.4 Late-Fusion Feature Concatenation

Our late-fusion strategy extracts modality-specific embeddings $\phi_k(\mathbf{x}^{(k)})$ and forms:

$$\mathbf{h} = \phi_1(\mathbf{x}^{(1)}) \oplus \phi_2(\mathbf{x}^{(2)}) \oplus \phi_3(\mathbf{x}^{(3)}) \tag{4}$$

where $\oplus$ denotes the concatenation. The joint classifier $f(\mathbf{h})$ is trained in two stages: each modality-specific extractor ϕ_k is first trained independently to produce a fixed embedding, after which the fusion head alone is trained using the categorical cross-entropy loss:

$$\mathcal{L} = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{ic} \log \hat{p}_{ic} \tag{5}$$

with $\hat{p}_{ic} = \text{softmax}(f(\mathbf{h}_i))_c$. The Customized CNN implements $\phi_1(\cdot)$ for MRI and is trained via backpropagation; gradient-boosted trees implement $\phi_2(\cdot)$ and $\phi_3(\cdot)$ for the tabular modalities and are trained by

stagewise splitting, and their outputs (leaf-index or probability embeddings) are frozen before fusion. During Stage 2, ϕ_1 (the CNN embedding) is likewise held frozen, so only the fusion head's weights are updated by backpropagation on $\mathcal{L}$.

4 Dataset and IoT Mapping

4.1 ADNI Dataset

We used ADNI phases 1, GO, and 2, comprising 503 MRI scans, 808 SNP samples, and 2004 EHR records from 2220 unique participants. Labels follow three severity classes: Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD). The class distributions per ADNI phase are shown in Table 2. The phase-wise class counts in the table summarize 1674 labeled entries of the clinical (EHR) stream, that is, the labeled participant-visit pool before cross-modal linkage; only a subset of these entries has a matching MRI acquisition and genotyping record. Each entry is assigned to the ADNI phase under which the visit was recorded, so the ADNI-GO column reflects the design of that phase: ADNI-GO continued the follow-up of ADNI-1 participants with CN and MCI and newly enrolled participants with early MCI, whereas ADNI-1 participants with AD were not followed in ADNI-GO and no new AD participants were recruited. The ADNI-GO column therefore contains CN and MCI entries but no AD entries.

Table 2: Class distribution across Alzheimer's Disease Neuroimaging Initiative (ADNI) phases. ADNI-3 was held out as it remains ongoing. Entries are counted under the phase in which the visit was recorded.

Class	ADNI-1	ADNI-GO	ADNI-2
Cognitively Normal (CN)	200	200	150
Mild Cognitive Impairment (MCI)	274	200	300
Alzheimer's Disease (AD)	200	–	150

Note: –, no AD entries in this phase.

The three streams were linked at the participant level through the ADNI Roster Identifier (RID), each MRI acquisition being paired with the clinical visit closest to it in time and with the participant's genotyping record. Participants for whom any of the three streams was absent were excluded at this linkage step. The aligned multimodal cohort that results from this procedure is therefore a subset of the label pool in Table 2. Because each MRI acquisition is paired with exactly one clinical visit, the cohort is anchored on the MRI stream and comprises 503 participant-visits, each carrying an MRI acquisition, an EHR record, and a genotype. All seven fusion scenarios (unimodal, bimodal, and trimodal; Section 6.4) are evaluated on this same set of 503 participant-visits under the same participant-level partition, so that the rows differ only in the modalities supplied to the fusion head. The imaging-only multiclass experiment (Section 6.3) draws on the same 503 MRI acquisitions, but operates on the axial slices extracted from them rather than on one record per participant-visit; the consequences of this difference are examined in Section 6.4.

4.2 IoT-to-ADNI Feature Mapping

Table 3 maps our intended IoT physiological signals to their ADNI structural equivalents, motivating the use of ADNI as a proxy for IoT-augmented clinical data. The ADNI values are discrete measurements recorded during clinical visits, which stand in for the aggregated summary features that an edge layer would derive from continuous wearable streams (Section 5.5); the equivalence between the two has not been tested in this study.

Table 3: Mapping between Internet of Things (IoT)-derived physiological signals and Alzheimer's Disease Neuroimaging Initiative (ADNI) structured equivalents used in this study. BPM, beats per minute; EHR, electronic health record; MRI, magnetic resonance imaging; SpO_2, peripheral oxygen saturation.

IoT Signal	ADNI Source Form/Variable	Clinical Significance
Heart rate (BPM)	Vital Signs form (seated pulse); cardiovascular history from the Medical History form	Cardiovascular comorbidity and vascular contribution to dementia
Body temp. (°C)	Vital Signs form (body temperature)	Infection-related delirium and cognitive fluctuation
Tremor frequency	Neurological and Physical Examination forms (tremor finding)	Supports differentiation of AD from Parkinson's disease dementia
SpO_2	No ADNI counterpart; design element only, not instantiated	Chronic hypoxemia is associated with accelerated cognitive decline
Blood glucose	Laboratory Data panel (serum glucose); diabetes status from the Medical History form	Type 2 diabetes is an independent AD risk factor
Blood pressure	Vital Signs form (systolic and diastolic pressure); hypertension status from the Medical History form	Midlife hypertension is an established modifiable AD risk factor
EHR narrative	Longitudinal Medical History and adverse-event records	Prior neurological events are associated with conversion
MRI scan	ADNI structural MRI	Hippocampal atrophy biomarker

Note: the ADNI entries are discrete measurements recorded at scheduled study visits, not continuous monitoring traces. Peripheral oxygen saturation is not acquired anywhere in ADNI, so that row belongs to the framework design and was not instantiated in any experiment reported here.

5 Proposed Framework and Customized CNN

5.1 Framework Overview

Our framework (Fig. 1) processes (1) MRI images using a Customized CNN, (2) EHR/physiological data using gradient-boosted classifiers, and (3) SNP data using principal component analysis (PCA)-reduced features with ensemble learners. Each modality-specific extractor is trained independently; the resulting frozen embeddings are then fused through late fusion (Eq. (4)), followed by a fully connected layer trained on the concatenated features for AD severity classification.

5.2 Layered System Architecture

The framework is organized into ten functional layers, which are summarized in Table 4. Each layer is marked as implemented in this study, represented by an ADNI proxy, or proposed for a future deployment, so that the experimentally evaluated scope is distinguished from the intended system design.

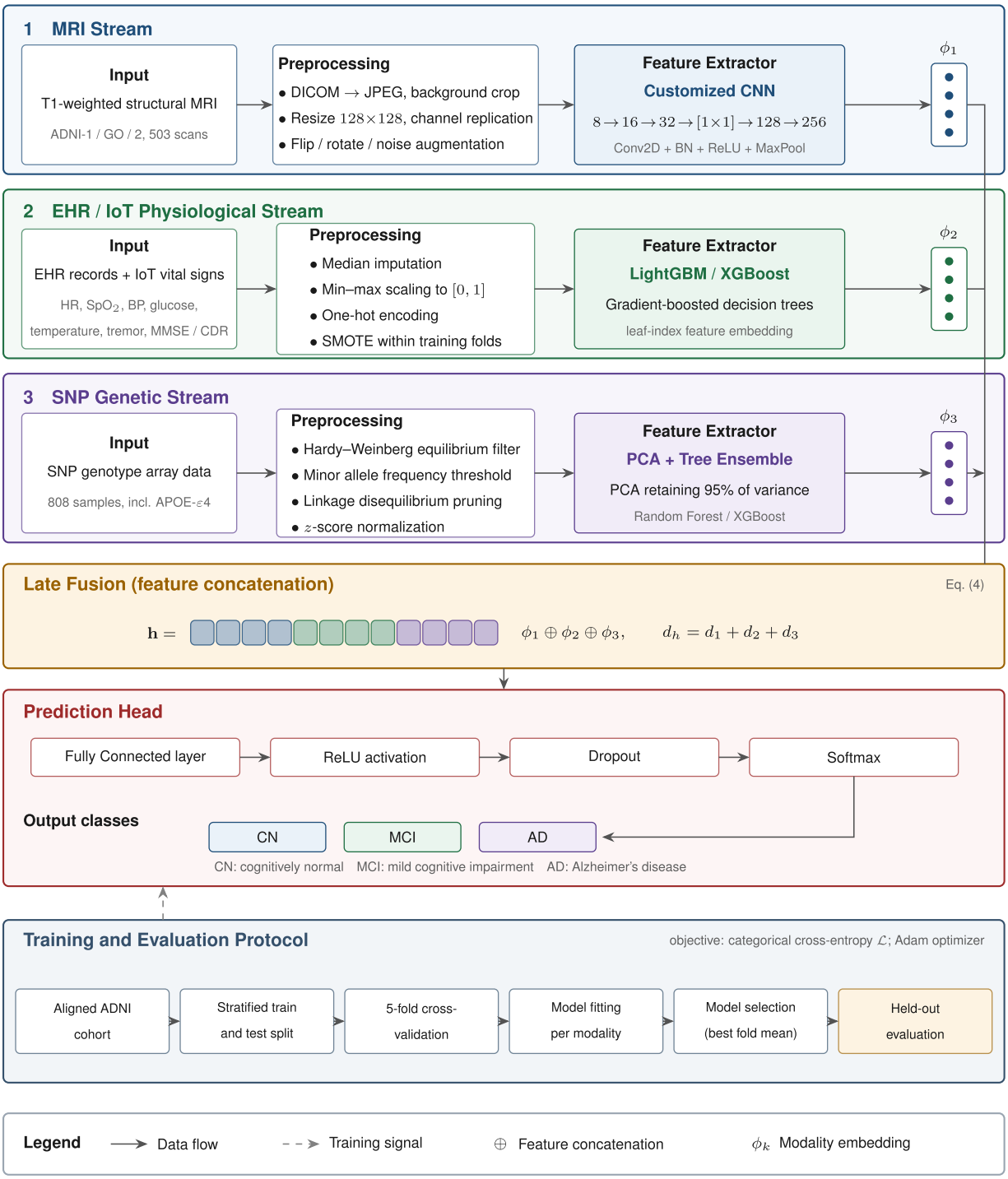


Figure 1: Proposed multimodal Alzheimer's disease (AD) framework with three feature streams: a Customized convolutional neural network (CNN) for magnetic resonance imaging (MRI), gradient-boosted ensembles for electronic health record (EHR)/physiological data, and a principal component analysis (PCA)-based ensemble for single-nucleotide polymorphism (SNP) features. The extracted representations are fused via late fusion (Eq. (4)) for cognitively normal (CN), mild cognitive impairment (MCI), and AD classification.

Table 4: Functional layers of the proposed framework and their status in the present study. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; CN, cognitively normal; CNN, convolutional neural network; EHR, electronic health record; Grad-CAM, gradient-weighted class activation mapping; IoT, Internet of Things; MCI, mild cognitive impairment; MRI, magnetic resonance imaging; SHAP, Shapley additive explanations; SNP, single-nucleotide polymorphism; SpO_2, peripheral oxygen saturation.

Layer	Function	Status in This Study
IoT wearable data acquisition	Continuous capture of heart rate, SpO_2, temperature, tremor, blood glucose, and blood pressure	ADNI proxy (Table 3)
MRI data acquisition and processing	Acquisition of T1-weighted MRI and conversion to normalized 2D slices	Implemented
EHR integration	Structured clinical history, comorbidities, and cognitive assessments	Implemented (ADNI)
Preprocessing and quality control	Imputation, normalization, encoding, augmentation, and SNP quality control (Section 5.4)	Implemented
Feature extraction and modality-specific models	Customized CNN for MRI; gradient-boosted trees for EHR and SNP features	Implemented
Multimodal feature representation and fusion	Concatenation of frozen modality-specific embeddings (Eq. (4))	Implemented
AD classification	Three-class severity staging (CN/MCI/AD) and binary AD screening	Implemented
Risk prediction and decision support	Presentation of class probabilities and risk alerts to clinicians	Proposed
Explainability and clinical interpretability	Grad-CAM for MRI and SHAP for tabular features (Section 8)	Proposed
Privacy and security	Encryption, access control, and data governance (Section 9)	Proposed

5.3 Customized CNN Architecture

The Customized CNN is a lightweight network configured for AD MRI classification. It combines three established design elements, which are arranged specifically for this task rather than inherited from a general-purpose backbone such as VGG16:

1. **1 × 1 Bottleneck Layer:** After the third convolutional block (32 filters), a 1 × 1 convolution reduces feature redundancy and enables channel-wise recombination before the deep convolutional stack, a mechanism analogous to an Inception module but computationally lighter.

2. **Progressive Depth Scaling:** Filter counts scale: 8 → 16 → 32 → 128 → 256, followed by a final reduction to 2 filters before flattening, matching progressive abstraction of hippocampal morphology (edges → textures → regional shape → global atrophy patterns).
3. **Per-Block Batch Normalization:** Each convolutional layer is followed by Batch Normalization before activation, stabilizing gradients across the convolutional stack and enabling convergence on ADNI data without ImageNet pre-training used by VGG16/EfficientNetB3.

The full layer specification is given in Table 5. For multiclass output (K = 3 classes: CN/MCI/AD), the final fully connected layer outputs K logits with a softmax activation. Fig. 2 illustrates the dual-path fusion architecture.

Table 5: Customized convolutional neural network (CNN) layer-by-layer architecture. Input: 128 × 128 × 3 (red-green-blue (RGB)-replicated magnetic resonance imaging (MRI) slice). The 1 × 1 bottleneck at row 6 reduces channel redundancy before deep processing. AD, Alzheimer's disease; BN, batch normalization; CN, cognitively normal; FC, fully connected; MCI, mild cognitive impairment; ReLU, rectified linear unit. –, not applicable.

Layer	Filters	Kernel	Output Shape
Conv2D + BN + ReLU	8	3 × 3	128 × 128 × 8
MaxPooling	–	2 × 2	64 × 64 × 8
Conv2D + BN + ReLU	16	3 × 3	64 × 64 × 16
MaxPooling	–	2 × 2	32 × 32 × 16
Conv2D + BN + ReLU	32	3 × 3	32 × 32 × 32
Conv2D + BN (1 × 1 bottleneck)	16	1 × 1	32 × 32 × 16
MaxPooling	–	2 × 2	16 × 16 × 16
Conv2D + BN + ReLU	128	3 × 3	16 × 16 × 128
Conv2D + BN + ReLU	256	3 × 3	8 × 8 × 256
MaxPooling	–	2 × 2	4 × 4 × 256
Conv2D + BN	2	3 × 3	4 × 4 × 2
Flatten	–	–	32
FC (Softmax)	–	–	K = 3 (CN/MCI/AD)

Computational complexity: From the layer specification in Table 5, the Customized CNN contains approximately 0.33 million parameters, including 1832 Batch Normalization parameters, and a forward pass for a single 128 × 128 × 3 input requires approximately 37 million multiply-accumulate operations. These figures assume same padding and a stride of two in the 256-filter convolution, which reduces the feature map from 16 × 16 to 8 × 8. For comparison, the convolutional base of VGG16 alone contains approximately 14.7 million parameters, so the Customized CNN is roughly 45 times smaller, which makes it suitable for execution on edge hardware. The evidence offered here for this configuration is therefore architecture-level: it is the comparison against VGG16 and EfficientNetB3 in Tables 6 and 7 together with the parameter and operation counts above, rather than a component-wise decomposition. A controlled ablation that removes the 1 × 1 bottleneck, the per-block normalization, and the progressive filter schedule in turn would attribute the observed accuracy among the three elements individually, and is identified as future work in Section 10.

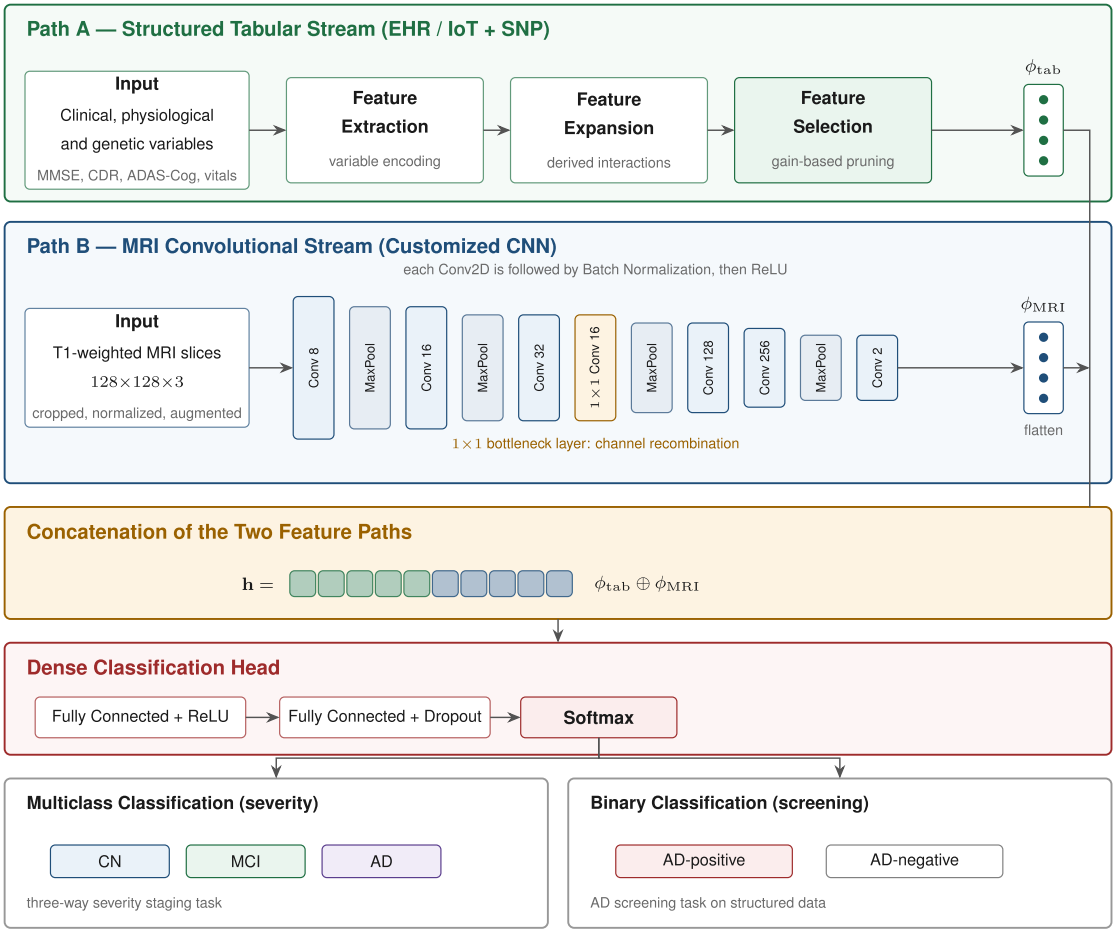


Figure 2: Dual-path classification workflow. Path A applies successive feature extraction, expansion, and selection to the structured electronic health record (EHR)/Internet of Things (IoT) and single-nucleotide polymorphism (SNP) variables; Path B applies the Customized convolutional neural network (CNN) of Table 5 to the magnetic resonance imaging (MRI) slices. The two embeddings, each produced by an independently trained and then frozen extractor, are concatenated and passed to a dense head that is trained in a second stage (Eq. (4)); the head supports both the three-class severity task (cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD)) and the binary AD screening task. ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive Subscale; CDR, Clinical Dementia Rating; MMSE, Mini-Mental State Examination.

Table 6: Binary classification performance. The classes were Alzheimer's disease (AD)-positive (AP) and AD-negative (AN). Macro-average Precision/Recall/F1 are reported for conciseness. The light gradient-boosting machine (LightGBM) achieves the best overall accuracy. XGBoost, extreme gradient boosting. The value in bold indicates the highest accuracy among the compared models.

Model	Precision	Recall	F1 Score	Accuracy
LightGBM	0.96	0.95	0.96	**96%**
XGBoost	0.95	0.94	0.95	95%
Random Forest	0.94	0.90	0.92	93%
EfficientNetB3	0.84	0.81	0.82	83%

Table 7: Multiclass AD severity analysis. Classes: CN (Cognitively Normal), MCI (Mild Cognitive Impairment), AD (Alzheimer's Disease). Performance reported as macro-average across all 3 classes. The row in bold indicates the proposed customized CNN, which achieves the highest value for every metric. CNN, convolutional neural network; VGG16, Visual Geometry Group 16-layer network.

Model	Precision	Recall	F1	Accuracy
VGG16	0.89	0.89	0.89	89%
Decision Tree	0.89	0.86	0.86	86%
Random Forest	0.62	0.64	0.46	64%
Customized CNN (Ours)	**0.98**	**0.98**	**0.98**	**98%**

5.4 Data Preprocessing

MRI: Digital Imaging and Communications in Medicine (DICOM) to Joint Photographic Experts Group (JPEG) conversion, channel replication, 128 × 128 resizing, background cropping, and comprehensive augmentation (flips, rotations at 15°/30°/45°/90°, Gaussian noise, elastic deformation, histogram equalization, zoom/shear). The synthetic minority oversampling technique (SMOTE) is not applied to the MRI images or to the Customized CNN; it is used only for the tabular EHR/physiological data described below. The conversion to an 8-bit three-channel raster was adopted so that the Customized CNN and the two ImageNet-pretrained baselines, VGG16 and EfficientNetB3, receive an identical input representation: those backbones expect three-channel 8-bit input, and supplying the same tensors to all three networks keeps the architectural comparison in Table 7 free of any confound arising from the input format. The attendant loss of native intensity depth is stated as a limitation in Section 10.

EHR/Physiological: Median imputation for missing longitudinal values, min-max normalization to $[0,1]$, one-hot encoding of categorical variables, and SMOTE oversampling within training folds.

SNP: Hardy-Weinberg equilibrium filtering, minor allele frequency thresholding, linkage disequilibrium pruning, z-score normalization, and PCA retaining 95% explained variance.

5.5 Proposed Deployment Architecture

This subsection describes how the framework is intended to operate in a clinical IoT setting. The deployment architecture has not been implemented or evaluated in this study; it is presented to specify the requirements that a future deployment must satisfy.

Hardware and software requirements: Firstly, the patient side requires wearable sensors, namely a photoplethysmography (PPG) sensor for heart rate and SpO_2, a skin temperature sensor, an inertial measurement unit for tremor, a blood pressure monitor, and a glucose meter, together with a smartphone or home gateway. The clinical side requires an edge node for signal processing and a server with a graphics processing unit (GPU) for MRI inference, running a Python-based DL and gradient-boosting software stack.

IoT wearable device configuration and sensor data transmission: Secondly, each sensor streams its readings over Bluetooth Low Energy to the gateway, which forwards them to the edge node through a lightweight publish-subscribe protocol such as Message Queuing Telemetry Transport (MQTT) secured with Transport Layer Security (TLS).

Edge and cloud healthcare infrastructure: Thirdly, the edge node filters noise, removes motion artifacts, and aggregates the raw streams into windowed summary features (for example, mean and variability over a monitoring period) that match the structured variables on which the tabular models are trained. The cloud tier hosts the Customized CNN, the fusion head, and the patient record store.

MRI processing pipeline: Fourthly, MRI scans are retrieved in DICOM format from the hospital picture archiving and communication system, de-identified, and processed with the preprocessing steps in Section 5.4 before CNN inference produces the imaging embedding.

EHR integration and interoperability: Fifthly, vital signs, diagnoses, and cognitive assessments are exchanged with hospital information systems as HL7 FHIR resources, which provide a standardized route for combining IoT measurements with existing records [7].

Real-time monitoring, prediction, and dashboard: Sixthly, the tabular prediction is updated whenever a new aggregated reading arrives, and the fused prediction is recomputed when an MRI embedding is available. A clinician-facing dashboard displays physiological trends, predicted class probabilities, and the explanation outputs described in Section 8, and raises an alert when the predicted risk changes.

Security and privacy implementation: Lastly, data are encrypted in transit and at rest, and access is restricted by role-based authentication, as detailed in Section 9.

6 Experimental Results

This study assessed seven distinct classifiers, including the proposed Customized CNN, LightGBM, extreme gradient boosting (XGBoost), Random Forest, Decision Tree, VGG16, and EfficientNetB3. The Customized CNN, VGG16, and EfficientNetB3 were evaluated on the ADNI MRI subset, whereas LightGBM, XGBoost, Random Forest, and Decision Tree were evaluated on EHR, SNP, or fused features.

Partitioning protocol: The imaging models operate on two-dimensional axial slices extracted from the 503 ADNI MRI acquisitions, several informative slices being retained per acquisition. The division into training and test sets is performed on the participant identifier and not on the extracted images: participants are first assigned by a stratified 80:20 split, and slices are then drawn only from the acquisitions belonging to each side of that split. Every slice originating from a given participant therefore lies wholly within one partition, and no participant contributes to both training and testing. The resulting held-out test set comprises 1357 slices (450 AD, 461 MCI, 446 CN), from which the confusion matrix and the per-class metrics reported in Section 7.1 are computed. The augmentation operations listed in Section 5.4 were applied after the split and to the training partition only, so neither an augmented image nor a slice from a training participant enters the test partition. The reported imaging accuracy is thus an estimate of performance on participants unseen during training, subject to the demographic composition of ADNI discussed in Section 10. The tabular and fusion streams hold one record per participant-visit, the fusion stream consisting of the 503 trimodally aligned participant-visits of Section 4, and are partitioned on the same participant identifier, and five-fold cross-validation over the training portion was used for model selection in these streams; Fig. 3 reports the single held-out imaging test partition rather than a cross-validated aggregate.

Training and optimization protocol: The Customized CNN was trained from random initialization by backpropagation on the categorical cross-entropy loss of Eq. (5) with an adaptive first-order optimizer, mini-batch updates, and early stopping governed by the validation loss; VGG16 and EfficientNetB3 were fine-tuned from ImageNet weights under the same loss and stopping rule. The gradient-boosted and tree-based models were tuned by grid search over the learning rate, the number of estimators, and the maximum tree depth, the retained configuration in each case being the one with the highest mean validation accuracy across the five folds. The fusion head of Eq. (4) was trained in the same manner on the concatenated frozen embeddings. All experiments were conducted in Python on a single general-purpose graphics processing unit, using a standard deep-learning framework for the imaging models and established gradient-boosting and machine-learning libraries for the tabular models; the implementation is available in the repository cited in Section 7.3.

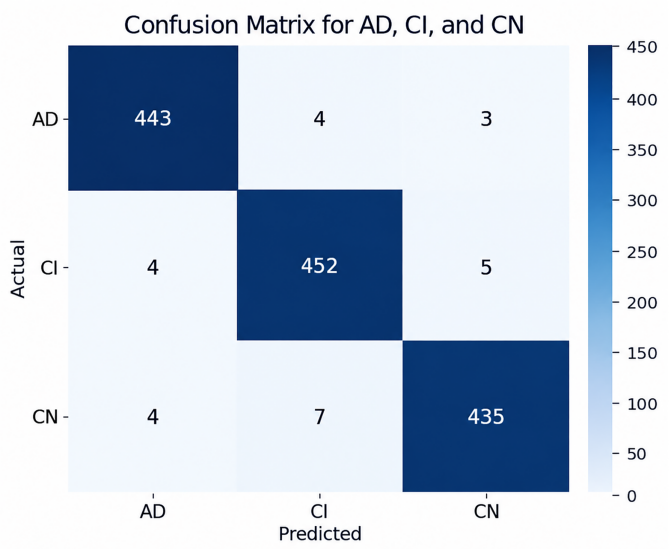


Figure 3: Confusion matrix of the customized convolutional neural network (CNN) on the cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) multiclass severity classification task, showing near-perfect diagonal discrimination and the expected concentration of residual error between the MCI and CN classes. In the figure, CI (cognitive impairment) denotes the MCI class.

6.1 Evaluation Metrics

The classifiers are evaluated with accuracy, precision, recall (sensitivity), specificity, and the F1 score, computed from the confusion matrix as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \qquad \text{Recall} = \frac{TP}{TP + FN}, \tag{7}$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \qquad \text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{8}$$

where TP, TN, FP, and FN denote the true positives, true negatives, false positives, and false negatives, respectively. For the multiclass task, each class is treated in turn as the positive class, and the per-class values are macro-averaged. The Matthews correlation coefficient (MCC) is additionally reported for the Customized CNN, since it summarizes all cells of the multiclass confusion matrix in a single value that remains informative under class imbalance:

$$\text{MCC} = \frac{c\,s - \sum_k p_k t_k}{\sqrt{\left(s^2 - \sum_k p_k^2\right)\left(s^2 - \sum_k t_k^2\right)}}, \tag{9}$$

where c is the number of correctly classified samples, s is the total number of samples, and p_k and t_k are the numbers of samples predicted as, and truly belonging to, class k.

Each macro-averaged quantity reported in Tables 6–8 is the unweighted arithmetic mean of the corresponding per-class quantity, the F1 column therefore being the mean of the per-class F1 scores rather than the harmonic mean of the macro-averaged precision and recall. Because the harmonic mean in Eq. (8) is not a linear function of its arguments, these two quantities do not coincide, and they separate appreciably whenever performance is uneven across classes: a class on which a model fails contributes a near-zero term

to the mean of the per-class F1 scores while still contributing a moderate precision and recall to their own means. This is the reason why the F1 entry for Random Forest in Table 7, and for the SNP-only row in Table 8, lies below the value that Eq. (8) would return if it were applied to the macro-averaged precision and recall listed in the same row.

Table 8: Comparison of unimodal and multimodal fusion performance across cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's disease (AD) classes, showing that EHR + SNP matches full fusion while EHR provides dominant clinical context and SNP + Imaging underperforms. EHR, electronic health record; MRI, magnetic resonance imaging; SNP, single-nucleotide polymorphism. All rows use the two-stage late-fusion pipeline; the Imaging (MRI) row uses the frozen MRI embedding and is not the standalone customized CNN of Table 7. All rows are evaluated on the same 503 trimodally aligned participant-visits.

Type	Data	Precision	Recall	F1	Accuracy
Unimodal	EHR only	0.76	0.77	0.76	76%
	Imaging (MRI embedding)	0.83	0.83	0.83	83%
	SNP only	0.66	0.57	0.53	56%
Multimodal	EHR + SNP +MRI	0.77	0.78	0.78	78%
	EHR + SNP	0.78	0.79	0.78	78%
	EHR + MRI	0.76	0.77	0.77	77%
	SNP + MRI	0.62	0.55	0.57	63%

6.2 Binary AD Classification

Binary classification (AD-positive vs. AD-negative) was evaluated using structured EHR and imaging data. The results in Table 6 show that LightGBM achieved the best binary accuracy (96%), closely followed by XGBoost (95%) and Random Forest (93%). EfficientNetB3 (83%) underperformed because MRI alone cannot distinguish early AD from CN without clinical context, precisely the gap that multimodal fusion addresses.

6.3 Multiclass AD Severity Classification

The multiclass classification task distinguishes between CN, MCI, and AD, making it a complex undertaking that is vital for determining a course of treatment. The overall performance of the model has been shown in Table 7. Our unique CNN achieved a remarkable score of 98%, ahead of all models by 9–34 percentage points. The CNN model has shown impressive levels of precision, recall, and F1 scores across CN, MCI, and AD, meaning there was no bias toward the majority class. The results for VGG16 and Decision Tree were also acceptable (89% and 86%, respectively), but compared to the CNN model they still possess certain drawbacks related to certain classes. Random forest was the weakest performer in identifying CN and MCI classes (64%), indicating its inefficacy in recognizing the other stages.

6.4 Unimodal vs. Multimodal Fusion

Table 8 presents the central experimental comparison, which is examined against the theoretical arguments of Section 3. Several key findings have emerged. The results in Table 8 come from the two-stage late-fusion pipeline of Section 3, in which each modality is reduced to a frozen embedding before a dense classification head is trained. The Imaging (MRI) row therefore reports this fusion head evaluated on the frozen MRI embedding alone, and it is a different experiment from the standalone Customized CNN evaluated in Table 7.

(1) **Imaging is the strongest single modality** (83%), establishing the superior status of MRI as an Alzheimer's disease biomarker. Nevertheless, even this best unimodal baseline still cannot match the 98% accuracy of the Customized CNN, trained independently on MRI in the complete imaging-only multiclass experiment (Table 7).

The distance between these two imaging figures is a consequence of three deliberate protocol differences and not of a degradation of the imaging model. Firstly, the evaluation units differ: both experiments draw on the same 503 MRI acquisitions and the same participant-level partition described in Section 4, but the imaging-only result is computed over 1357 held-out slices, whereas each fusion scenario is scored once per held-out participant-visit. Secondly, the unit of prediction differs: the fusion head issues one decision per aligned participant-visit from a single pooled imaging embedding, so the slice-level redundancy that the standalone CNN exploits, several slices per acquisition each contributing evidence, is absent by construction. Thirdly, the imaging extractor is frozen in Stage 2, so the fusion head adapts only its own weights and cannot refine the imaging representation for the fused objective. The 83% and the 98% are therefore measurements of different quantities under different protocols, and the appropriate comparison for the 83% figure is with the other rows of Table 8, all of which were obtained under identical conditions.

(2) **EHR alone achieved 76%**—indicating clinical relevance, as it demonstrates that the application of IoT streaming of physiological data, even in the absence of imaging, makes significant contributions to AD detection in situations with limited resources.

The composition of the tabular feature set should be borne in mind when this row, and the binary result of Table 6, are interpreted. The structured stream includes the cognitive instruments MMSE, CDR, and ADAS-Cog, which are themselves part of the adjudication procedure by which ADNI assigns its diagnostic labels. Their presence means that the tabular and fused accuracies measure how closely a model reproduces the established clinical assessment pathway from routinely recorded variables, rather than how well it predicts a diagnosis from evidence independent of that pathway. This is the intended reading for a decision-support instrument that is meant to operate alongside, and not in place of, clinical assessment; a variant trained with the cognitive instruments withheld would isolate the contribution of the physiological and comorbidity variables alone, and is identified as the sixth limitation in Section 10. The imaging-only result of Table 7 is unaffected, as the Customized CNN receives no clinical or cognitive variable of any kind.

(3) **EHR + SNP (78%) matches complete trimodal fusion (78%)** demonstrating that clinical context (EHR) is the crucial factor needed to contextualize genetic data, which SNP + Imaging (63%) alone fails to achieve.

(4) **SNP + Imaging underperforms EHR-only (63% vs. 76%)**, offering a finding that is contrary to common sense but makes clinical sense: it is impossible to place genetic risk and neuroimaging information in order correctly without the explanatory intermediate of EHR data.

This apparent tension with Eq. (1) is a finite-sample estimation effect rather than a violation of it. The non-negativity of $I(\mathbf{x}^{(\mathrm{SNP})}; y \mid \mathbf{x}^{(\mathrm{MRI})})$ is a population-level property that holds for a Bayes-optimal decoder operating on the true joint distribution; it does not guarantee that any specific trained classifier, evaluated on a small held-out cohort, will recover all of that information. With only 808 SNP samples, of which only the 503 linked to an MRI acquisition enter the fusion cohort, and a PCA-reduced genetic embedding concatenated late with CNN-derived MRI features, the conditional information the fusion classifier actually extracts can be driven toward zero, that is, become empirically negligible, by estimation noise, PCA dimensionality reduction, and a fusion architecture that lacks the clinical context (EHR) needed to disambiguate genetic risk from imaging findings. The true conditional mutual information does not become negative; rather, the achieved classifier underutilizes it. This distinction between population-level information content and

achieved classifier performance is consistent with the broader limitation noted in Section 10 regarding the modality-independence assumption of late fusion.

7 Discussion

7.1 Performance of the Customized CNN

The Customized CNN achieves a 98% accuracy rate in CN/MCI/AD classification, representing the primary empirical finding of this study. Its superiority over VGG16, which achieves 89% on the same multiclass task, is plausibly attributable to the way in which established design elements are configured for the AD MRI context: (i) the 1×1 bottleneck layer keeps the filter count low in the middle layers, thus limiting overfitting due to small ADNI dataset sizes; (ii) progressive depth scaling creates the right feature hierarchy from sulcal edge detection to global atrophy pattern identification; and (iii) per-block Batch Normalization makes it possible to train a deep network from scratch on ADNI without ImageNet pre-training, thereby reducing domain-transfer-related performance degradation when pre-trained architectures are adapted to grayscale MRI.

The confusion matrix for the Customized CNN (Fig. 3) shows that 1330 of the 1357 test samples were classified correctly, which corresponds to an accuracy of 98.0%. The per-class metrics derived from this matrix using Eqs. (6)–(8) are given in Table 9. The AD class achieved a precision of 0.982, a recall of 0.984, and an F1 score of 0.983; the MCI class achieved 0.976, 0.980, and 0.978; and the CN class achieved 0.982, 0.975, and 0.979, respectively. Specificity exceeded 0.98 for all three classes, and the MCC computed with Eq. (9) was 0.970. Of the 27 misclassified samples, 12 lie between MCI and CN, compared with 8 between AD and MCI and 7 between AD and CN. The main type of misclassification is therefore between MCI and CN, since the actual clinical reality at this initial phase of the disease makes it difficult to differentiate between these two conditions. Because SMOTE is applied only to the tabular EHR/physiological stream (Section 5.4), this result involves no synthetic oversampling.

Table 9: Per-class performance of the customized convolutional neural network (CNN) derived from the confusion matrix in Fig. 3. AD, Alzheimer's disease; CN, cognitively normal; MCI, mild cognitive impairment.

Class	Support	Precision	Recall	Specificity	F1
AD	450	0.982	0.984	0.991	0.983
MCI	461	0.976	0.980	0.988	0.978
CN	446	0.982	0.975	0.991	0.979
Macro average	1357	0.980	0.980	0.990	0.980

7.2 Multimodal Fusion as Robust Clinical Intelligence

The fusion experiments are examined here against the theoretical framework presented in Section 3. Among the non-imaging modalities, adding SNP features to EHR raised accuracy from 76% to 78%, which is consistent with the non-negative conditional information term in Eq. (1). Electronic Health Records (EHR) alone achieved an accuracy of 76%. This is well above Single Nucleotide Polymorphism (SNP) data alone, which reached 56%, and it also exceeds the SNP + MRI fusion at 63%. Taken together, these figures support the clinical intuition that patient context, namely vital signs, comorbidities, and cognitive trajectory, carries more diagnostic information than genetic data considered in isolation, although MRI remained the strongest single modality (83%, Table 8). In the context of an IoT-oriented deployment, this result is promising: structured EHR variables, for which the IoT physiological signals in Table 3 serve as proxies,

achieved an accuracy of 76% without MRI, which indicates their potential as a first-line screening aid in resource-constrained environments.

7.3 Comparison with Published Results

Table 10 compares the results of this study with those reported in recent literature. The Customized CNN attained a multiclass accuracy of 98%, which is the highest among the listed studies, and the LightGBM binary classifier reached 96%, which exceeds the structured-data binary accuracies listed. However, as these studies used different datasets, cohorts, and evaluation protocols, the comparison is indicative rather than conclusive, and a direct comparison would require evaluation under a common protocol. Finally, the implementation code supporting the experiments is available in the GitHub repository[1]. All experiments reported in this paper were conducted on ADNI data, which can be accessed at https://adni.loni.usc.edu/. OASIS-3, which appears in Table 1 as the corpus used by one of the cited studies and is a natural candidate for the external validation identified in Section 10, is available at https://sites.wustl.edu/oasisbrains/.

Table 10: Comparison with published Alzheimer's disease (AD) detection results. The listed studies differ in dataset, cohort, and evaluation protocol, so the figures indicate relative standing rather than a direct head-to-head comparison. –, not reported for that task. Rows in bold indicate the models proposed in this study. CNN, convolutional neural network; DT, decision tree; ELM, extreme learning machine; KNN, k-nearest neighbors; LightGBM, light gradient-boosting machine; LR, logistic regression; LSTM, long short-term memory; RF, random forest; RNN, recurrent neural network; SVM, support vector machine.

Study	Model	Binary	Multiclass
[12]	SVM, RF, KNN, etc.	90.51%	–
[25]	RF, SVM, DT, etc.	92.14%	–
[30]	LR, RF, AdaBoost	86.8%	–
[37]	RNN	75%	81%
[45]	CNN	94%	74.5%
[46]	FusionNet	94%	–
[49]	Alexnet and LSTM	–	96.83%
[59]	CNNs	–	93.45%
[60]	3D CNN	86%	89.21%
[61]	LR, RF	91.6%	86.7%
[62]	MobileNetV3-Small	–	96%
[62]	ELM	–	95%
Ours	**Customized CNN**	–	**98%**
Ours	**LightGBM**	**96%**	–

8 Explainable and Trustworthy Alzheimer's Detection

This section specifies how the predictions of the framework are to be explained to clinicians. Explanation outputs were not generated in the present experiments; the methods below define the explainability layer of the framework (Table 4).

Need for explainable AI in healthcare: A diagnostic model that reports only a class label gives a clinician no basis on which to accept or reject its output. Explanations allow the clinician to verify that a

[1] https://github.com/Fahimsakib1/Alzheimer-Disease-Paper

prediction rests on clinically plausible evidence, to detect failures caused by artifacts or data errors, and to communicate the basis of a decision to patients and caregivers [18].

Grad-CAM-based MRI interpretation: For the Customized CNN, Grad-CAM [55] computes, for class c and feature map A^k of a chosen convolutional layer, an importance weight and a class-discriminative localization map:

$$\alpha_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A_{ij}^k}, \qquad L^c_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right), \tag{10}$$

where y^c is the score for class c before the softmax, ReLU denotes the rectified linear unit, and Z is the number of spatial positions in A^k. The map $L^c_{\text{Grad-CAM}}$ is upsampled to the input resolution, so that a clinician can check whether the regions driving an AD or MCI prediction coincide with the hippocampal and entorhinal atrophy described in Section 1.1.

SHAP-based EHR interpretation: For the gradient-boosted models on EHR and SNP features, SHAP [56] decomposes each prediction $f(\mathbf{x})$ into additive feature contributions:

$$f(\mathbf{x}) = \phi_0 + \sum_{j=1}^{M} \phi_j, \tag{11}$$

where ϕ_0 is the expected model output over the training data and ϕ_j is the Shapley value of feature j among M features. Tree-based models admit efficient exact computation of ϕ_j, which makes SHAP suitable for patient-level explanation of the tabular stream.

Feature importance from wearable data: Once the IoT acquisition layer is operational, the SHAP values in Eq. (11) for the physiological variables in Table 3 (heart rate, SpO_2, blood pressure, temperature, tremor, and blood glucose) indicate how strongly each continuously monitored signal contributes to a given prediction, which allows changes in risk to be traced to specific physiological trends.

Multimodal decision explanation: Because the fusion head operates on concatenated embeddings (Eq. (4)), attribution can also be computed at the level of the modality blocks ϕ_1, ϕ_2, and ϕ_3, which indicates the share of a decision contributed by MRI, EHR, and SNP evidence. A combined report then presents the Grad-CAM map, the leading tabular contributions, and the modality shares together.

Model transparency and interpretability: The framework has three properties that support transparency. Firstly, the Customized CNN is compact (Section 5.3), which keeps its intermediate feature maps few enough to inspect. Secondly, the gradient-boosted models expose split-based feature importance directly. Thirdly, the information-theoretic formulation in Section 3 states explicitly how each modality is expected to affect diagnostic uncertainty.

Clinical trust and reliability: Trust also depends on how a model errs. The residual errors of the Customized CNN are concentrated between MCI and CN (Fig. 3), which is the boundary that clinicians themselves find most difficult, rather than between AD and CN. Reporting per-class metrics, specificity, and MCC (Table 9) alongside accuracy gives clinicians a fuller view of reliability than a single headline figure.

9 Security, Privacy, and Ethical Considerations

The proposed framework handles neuroimaging, clinical, genetic, and physiological data, all of which are sensitive personal health information, and de-identification alone does not protect such data adequately [63]. Firstly, genetic data need particular care, because an SNP profile cannot be revoked once it is disclosed and it also reveals information about the patient's biological relatives. Secondly, MRI data can

be used to re-identify patients. DICOM headers carry patient identifiers, and Schwarz et al. [64] reported that commercial face-recognition software matched faces reconstructed from research head MRI to the photographs of 70 of 84 participants (83%). In the proposed deployment, header identifiers are therefore removed and the scans are defaced before they leave the hospital network. Thirdly, EHR repositories hold diagnoses, medications, and cognitive assessments in one place, which makes them attractive targets for attackers. Mazhar et al. [7] show that a blockchain-based audit trail, which records every access to a record and every change made to it in a tamper-evident ledger, improves EHR security, privacy, and interoperability.

The IoT acquisition layer adds risks that a hospital-based pipeline does not face. Wearable devices and gateways are open to eavesdropping, spoofed sensor readings, device tampering, and denial-of-service attacks, and their limited computing resources often rule out strong built-in protection [6,65]. A spoofed physiological stream could change a risk prediction without any visible fault, so the proposed deployment authenticates each device and checks the integrity of each message at the point of acquisition. Data are encrypted with TLS on every link, from wearable to gateway, from gateway to edge node, and from edge node to cloud, and stored records are encrypted at rest. Role-based access control limits each user to the data that the role requires; for example, a caregiver receives alerts but cannot view genetic results. Clinical accounts additionally require multi-factor authentication.

Automated AD assessment also raises ethical and regulatory questions. A false-positive prediction can cause considerable distress to a patient and family, while a false-negative prediction can delay care. The framework is therefore designed as a decision-support tool, and the final diagnosis remains the responsibility of a clinician. Bias must also be examined. Obermeyer et al. [66] found that a widely used clinical algorithm underestimated the health needs of Black patients because it predicted healthcare costs rather than illness. The ADNI population is not demographically representative (Section 10), so the performance of the framework across population groups has to be verified before deployment. Continuous monitoring further requires informed consent that states which signals are collected, how long they are kept, and who can access them, and that allows the patient to withdraw at any time. Federated learning, in which each site trains the model locally and shares only the model parameters, reduces the need to pool raw patient data in one place [67,68]. The present study used de-identified ADNI data under the ADNI data use agreement. A deployed system would also have to comply with the data protection law of each jurisdiction, such as HIPAA in the United States and the GDPR in the European Union. As software that informs diagnosis, it would further be regulated as a medical device and would require clinical validation before approval.

10 Limitations and Future Work

Despite showing gains, the proposed framework has seven main shortcomings. First, the lack of diversity in the ADNI population means that this framework needs to be validated in other types of populations. Second, the assumption of modality independence made by the late fusion technique may not be valid; other options, such as cross-attention and tensor fusion, should be explored to find inter-modality dependencies in cases where patients were categorized as CN/MCI. Third, the Internet of Things (IoT) component is still in its conceptual phase because it does not handle the noise, drift, and missing values of information that come from actual wearable devices. Fourth, deploying the framework in clinical IoT environments will require an intermediate signal-filtering and feature-aggregation layer, as outlined in Section 5.5, to bridge raw continuous sensor streams with the structured inference pipeline. Fifth, the MRI slices were converted from DICOM to 8-bit JPEG images before training, which reduces the original intensity depth of the scans and may introduce compression artifacts; training on native-intensity images is therefore planned for future work. Sixthly, the structured feature set contains the cognitive instruments used in ADNI diagnostic adjudication, so the tabular and fused accuracies express concordance with the clinical assessment pathway rather than

prediction independent of it, and a variant trained with those instruments withheld remains to be evaluated. Seventhly, the results are reported as single summary values obtained under one evaluation protocol; per-fold dispersion, confidence intervals, paired significance testing between classifiers, and a component-wise ablation of the Customized CNN were outside the scope of the present study and constitute the natural next stage of its evaluation. Future research on this topic will aim to make the IoT component fully operational.

11 Conclusions

This study presented an IoT-oriented multimodal framework for Alzheimer's disease detection that combines MRI, EHR-derived physiological and clinical variables, and SNP markers. In the imaging-only experiment, the lightweight Customized CNN, built from a 1×1 bottleneck layer, progressively increasing filter depth, and per-block Batch Normalization, classified CN, MCI, and AD with an accuracy of 98%, a macro-averaged precision, recall, and F1 score of 0.98, and an MCC of 0.970. For binary AD screening on structured clinical and physiological data, LightGBM achieved the highest accuracy (96%). Among the non-imaging streams, EHR data alone reached 76%, well above SNP data alone (56%), which shows that routinely collected clinical variables carry substantial diagnostic information and supports the prospect of IoT-oriented screening where MRI is not available. The information-theoretic formulation in Section 3 shows that, at the population level, adding a modality cannot reduce the mutual information with the diagnostic label; whether a trained model realizes this gain depends on the fusion strategy and the available sample size. The proposed deployment architecture, explainability layer, and security measures set out the work required to move the framework from the present ADNI-based proxy evaluation toward clinical use.

Acknowledgement: The authors thank the Alzheimer's Disease Neuroimaging Initiative (ADNI) for providing access to data used in this study. The authors also acknowledge the OASIS-3 project for making its data openly available.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, study design, and framework development, Fahim Faysal Sakib; experiments, Fahim Faysal Sakib; data preprocessing and results analysis, Md. Iftekharul Mobin and Pretom Chandra Roy; supervision and clinical/methodological guidance, Al-Sakib Khan Pathan. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The dataset analyzed in this study is publicly available. The ADNI data on which all reported experiments were conducted can be obtained from https://adni.loni.usc.edu/. OASIS-3, referenced in this paper as the corpus of a cited study and as a candidate for future external validation, is available from https://sites.wustl.edu/oasisbrains/. The implementation code is available at https://github.com/Fahimsakib1/Alzheimer-Disease-Paper.

Ethics Approval: Not applicable. This study used de-identified, publicly available secondary data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), for which ethics approval and participant consent were obtained by the original data-collecting institutions.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AD	Alzheimer's Disease
MCI	Mild Cognitive Impairment
CN	Cognitively Normal
ADNI	Alzheimer's Disease Neuroimaging Initiative
EHR	Electronic Health Record

IoT	Internet of Things
SNP	Single-Nucleotide Polymorphism
CNN	Convolutional Neural Network
BN	Batch Normalization
ReLU	Rectified Linear Unit
FC	Fully Connected (layer)
PCA	Principal Component Analysis
SMOTE	Synthetic Minority Oversampling Technique
MMSE	Mini-Mental State Examination
CDR	Clinical Dementia Rating
CSF	Cerebrospinal Fluid
APOE	Apolipoprotein E
ADAS-Cog	Alzheimer's Disease Assessment Scale-Cognitive Subscale
AUC	Area Under the (ROC) Curve
SpO_2	Peripheral Oxygen Saturation
DL	Deep Learning
ML	Machine Learning
AI	Artificial Intelligence
XAI	Explainable Artificial Intelligence
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
PET	Positron Emission Tomography
MCC	Matthews Correlation Coefficient
Grad-CAM	Gradient-weighted Class Activation Mapping
SHAP	Shapley Additive Explanations
FHIR	Fast Healthcare Interoperability Resources
MQTT	Message Queuing Telemetry Transport
TLS	Transport Layer Security
PPG	Photoplethysmography
GPU	Graphics Processing Unit
GDPR	General Data Protection Regulation
MRI	Magnetic Resonance Imaging
OASIS	Open Access Series of Imaging Studies
DICOM	Digital Imaging and Communications in Medicine
JPEG	Joint Photographic Experts Group
RGB	Red-Green-Blue
BPM	Beats Per Minute
RID	ADNI Roster Identifier
SVM	Support Vector Machine
SVR	Support Vector Regression
RF	Random Forest
KNN	k-Nearest Neighbors
LR	Logistic Regression
DT	Decision Tree
GBM	Gradient Boosting Machine
LightGBM	Light Gradient-Boosting Machine
XGBoost	Extreme Gradient Boosting
AdaBoost	Adaptive Boosting
ELM	Extreme Learning Machine

VGG16	Visual Geometry Group 16-layer network
sMRI	Structural MRI
FDG-PET	Fluorodeoxyglucose PET
AV45-PET	Florbetapir PET
DTI	Diffusion Tensor Imaging
T2-FLAIR	T2 Fluid-Attenuated Inversion Recovery
MIRIAD	Minimal Interval Resonance Imaging in Alzheimer's Disease
NA-ADNI	North American ADNI
NCMMSC2021	National Conference on Man-Machine Speech Communication 2021
ADReSSo	Alzheimer's Dementia Recognition through Spontaneous Speech only
EDAMM	Early Diagnosis of AD based on a Multi-attention Mechanism
HL7	Health Level Seven
AP/AN	AD-Positive/AD-Negative
TP/TN	True Positive/True Negative
FP/FN	False Positive/False Negative
HIPAA	Health Insurance Portability and Accountability Act

References

1. Jahan S, Abu Taher K, Kaiser MS, Mahmud M, Rahman MS, Hosen AS, et al. Explainable AI-based Alzheimer's prediction and management using multimodal data. PLoS One. 2023;18(11):e0294253.
2. Nichols E, Steinmetz JD, Vollset SE, Fukutaki K, Chalek J, Abd-Allah F, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of disease study 2019. Lancet Public Health. 2022;7(2):e105–25. doi:10.1016/s2468-2667(21)00249-8.
3. Schaap T, Thropp P, Tosun D. Alzheimer's Disease Neuroimaging Initiative. Timing of Alzheimer's disease biomarker progressions: a two-decade observational study from the Alzheimer's disease neuroimaging initiative (ADNI). Alzheimer's Dement. 2024;20(12):9060–7. doi:10.1002/alz.14306.
4. Teunissen CE, Verberk IMW, Thijssen EH, Vermunt L, Hansson O, Zetterberg H, et al. Blood-based biomarkers for Alzheimer's disease: towards clinical implementation. Lancet Neurol. 2022;21(1):66–77. doi:10.4103/nrr.nrr-d-25-00569.
5. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. Neurology. 2010;74(3):201–9. doi:10.1212/wnl.0b013e3181cb3e25.
6. Ghadi YY, Shah SFA, Waheed W, Mazhar T, Ahmad W, Saeed MM, et al. Integration of wearable technology and artificial intelligence in digital health for remote patient care. J Cloud Comput. 2025;14(1):39. doi:10.1186/s13677-025-00759-4.
7. Mazhar T, Khan S, Shahzad T, Khan MA, Saeed MM, Awotunde JB, et al. Generative AI, IoT, and blockchain in healthcare: application, issues, and solutions. Discov Internet Things. 2025;5(1):5. doi:10.1007/s43926-025-00095-8.
8. Sathish R, Muthukumar R, Kumaran KM, Murugan SP. Intelligent decision-making systems for early detection of Alzheimer's disease using wearable technologies and deep learning. Sci Rep. 2026;16(1):6025. doi:10.1038/s41598-026-36895-3.
9. Karaman BK, Mormino EC, Sabuncu MR, Alzheimer's Disease Neuroimaging Initiative. Machine learning based multi-modal prediction of future decline toward Alzheimer's disease: an empirical study. PLoS One. 2022;17(11):e0277322. doi:10.1371/journal.pone.0277322.
10. Sharma S, Mandal PK. A comprehensive report on machine learning-based early detection of Alzheimer's disease using multi-modal neuroimaging data. ACM Comput Surv. 2022;55(2):1–44. doi:10.1145/3492865.
11. Odusami M, Maskeliūnas R, Damaševičius R, Misra S. Machine learning with multimodal neuroimaging data to classify stages of Alzheimer's disease: a systematic review and meta-analysis. Cogn Neurodyn. 2024;18(3):775–94. doi:10.1007/s11571-023-09993-5.

12. El-Sappagh S, Saleh H, Sahal R, Abuhmed T, Islam SMR, Ali F, et al. Alzheimer's disease progression detection model based on an early fusion of cost-effective multimodal data. Future Gener Comput Syst. 2021;115:680–99.
13. Kavitha C, Mani V, Srividhya SR, Khalaf OI, Tavera Romero CA. Early-stage Alzheimer's disease prediction using machine learning models. Front Public Health. 2022;10:853294. doi:10.3389/fpubh.2022.853294.
14. Elazab A, Wang C, Abdelaziz M, Zhang J, Gu J, Gorriz JM, et al. Alzheimer's disease diagnosis from single and multimodal data using machine and deep learning models: achievements and future directions. Expert Syst Appl. 2024;255(2):124780. doi:10.1016/j.eswa.2024.124780.
15. Viswan V, Shaffi N, Malathy E, Chemmalar Selvi G, Kavitha BR, Abdesselam A, et al. Multimodal fusion and explainability of artificial intelligence models in Alzheimer's disease detection. Brain Inform. 2026;13(1):5.
16. Venugopalan J, Tong L, Hassanzadeh HR, Wang MD. Multimodal deep learning models for early detection of Alzheimer's disease stage. Sci Rep. 2021;11(1):3254. doi:10.1038/s41598-020-74399-w.
17. Kaur A, Mittal M, Bhatti JS, Thareja S, Singh S. A systematic literature review on the significance of deep learning and machine learning in predicting Alzheimer's disease. Artif Intell Med. 2024;154(1):102928. doi:10.1016/j.artmed.2024.102928.
18. Viswan V, Shaffi N, Mahmud M, Subramanian K, Hajamohideen F. Explainable artificial intelligence in Alzheimer's disease classification: a systematic review. Cogn Comput. 2024;16(1):1–44. doi:10.1007/s12559-023-10192-x.
19. Jin J. Alzheimer disease. JAMA. 2015;313(14):1488. [cited 2025 Jan 14]. Available from: https://jamanetwork.com/journals/jama/fullarticle/2247146.
20. Arafa DA, Moustafa HED, Ali HA, Ali-Eldin AMT, Saraya SF. A deep learning framework for early diagnosis of Alzheimer's disease on MRI images. Multimed Tools Appl. 2024;83(2):3767–99. doi:10.1007/s11042-023-15738-7.
21. De Silva K, Kunz H. Prediction of Alzheimer's disease from magnetic resonance imaging using a convolutional neural network. Intell-Based Med. 2023;7:100091. doi:10.21203/rs.3.rs-1419743/v1.
22. Faisal FUR, Kwon GR. Automated detection of Alzheimer's disease and mild cognitive impairment using whole brain MRI. IEEE Access. 2022;10:65055–66.
23. Jraba S, Elleuch M, Ltifi H, Kherallah M. Alzheimer disease classification using deep CNN methods based on transfer learning and data augmentation. Int J Comput Inf Syst Ind Manag Appl. 2024;16(3):17–7. doi:10.1007/978-3-031-64813-7_37.
24. Alwakid G, Tahir S, Humayun M, Gouda W. Improving Alzheimer's detection with deep learning and image processing techniques. IEEE Access. 2024;12:153445–56. doi:10.1109/access.2024.3481238.
25. Amrutesh A, Bhat CGG, Amruthamsh A, Rani KPA, Gowrishankar S. Alzheimer's disease prediction using machine learning and transfer learning models. In: Proceedings of the 2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS); 2022 Dec 21–23; Bangalore, India. p. 1–6.
26. Yuan Y, Sheng L, Chen Z, Zhang Y, Li Q, Chen J, et al. Design of a configurable SoC for Alzheimer's disease detection based on multimodal signals. BMC Bioinform. 2026;27(1):88. doi:10.1186/s12859-026-06410-6.
27. Ullah R, Dong Y, Arslan T, Chandran S. A machine learning-based classification method for monitoring Alzheimer's disease using electromagnetic radar data. IEEE Trans Microw Theory Tech. 2023;71(9):4012–26.
28. Ying Y, Yang T, Zhou H. Multimodal fusion for Alzheimer's disease recognition. Appl Intell. 2023;53(12):16029–40.
29. Yang X, Hong K, Zhang D, Wang K. Early diagnosis of Alzheimer's disease based on multi-attention mechanism. PLoS One. 2024;19(9):e0310966. doi:10.1371/journal.pone.0310966.
30. Baglat P, Salehi AW, Gupta A, Gupta G. Multiple machine learning models for detection of Alzheimer's disease using OASIS dataset. In: Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation (TDIT 2020). Cham, Switzerland: Springer; 2020. p. 614–22.
31. Velazquez M, Lee Y. Multimodal ensemble model for Alzheimer's disease conversion prediction from early mild cognitive impairment subjects. Comput Biol Med. 2022;151(2 14):106201. doi:10.1016/j.compbiomed.2022.106201.
32. Wang C, Tachimori H, Yamaguchi H, Sekiguchi A, Li Y, Yamashita Y, et al. A multimodal deep learning approach for the prediction of cognitive decline and its effectiveness in clinical trials for Alzheimer's disease. Transl Psychiatry. 2024;14(1):105. doi:10.1038/s41398-024-02819-w.

33. Abdelaziz M, Wang T, Elazab A. Alzheimer's disease diagnosis framework from incomplete multimodal data using convolutional neural networks. J Biomed Inform. 2021;121(11):103863. doi:10.1016/j.jbi.2021.103863.
34. Lee MW, Kim HW, Choe YS, Yang HS, Lee J, Lee H, et al. A multimodal machine learning model for predicting dementia conversion in Alzheimer's disease. Sci Rep. 2024;14(1):12276. doi:10.1038/s41598-024-60134-2.
35. Agostinho D, Simões M, Castelo-Branco M. Predicting conversion from mild cognitive impairment to Alzheimer's disease: a multimodal approach. Brain Commun. 2024;6(4):fcae208.
36. Wang Y, Gao R, Wei T, Johnston L, Yuan X, Zhang Y, et al. Predicting long-term progression of Alzheimer's disease using a multimodal deep learning model incorporating interaction effects. J Transl Med. 2024;22(1):265. doi:10.1186/s12967-024-05025-w.
37. Lee G, Nho K, Kang B, Sohn KA, Kim D. Predicting Alzheimer's disease progression using multi-modal deep learning approach. Sci Rep. 2019;9(1):1952.
38. Young J, Modat M, Cardoso MJ, Mendelson A, Cash D, Ourselin S, et al. Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. NeuroImage Clin. 2013;2:735–45.
39. Akkaya UM, Kalkan H. A new approach for multimodal usage of gene expression and its image representation for the detection of Alzheimer's disease. Biomolecules. 2023;13(11):1563. doi:10.3390/biom13111563.
40. Kim SK, Duong QA, Gahm JK. Multimodal 3D deep learning for early diagnosis of Alzheimer's disease. IEEE Access. 2024;12(1):46278–89. doi:10.1109/access.2024.3381862.
41. Ravichandran S, Snyder PJ, Alber J, Murchison CF, Chaby LE, Jeromin A, et al. Association and multimodal model of retinal and blood-based biomarkers for detection of preclinical Alzheimer's disease. Alzheimer's Res Ther. 2025;17(1):19. doi:10.1186/s13195-024-01668-5.
42. Gravina M, García-Pedrero A, Gonzalo-Martín C, Sansone C, Soda P. Multi input–multi output 3D CNN for dementia severity assessment with incomplete multimodal data. Artif Intell Med. 2024;149(6):102774. doi:10.1016/j.artmed.2024.102774.
43. Turkan Y, Tek FB. Convolutional attention network for MRI-based Alzheimer's disease classification and its interpretability analysis. In: Proceedings of the 2021 6th International Conference on Computer Science and Engineering (UBMK); 2021 Sep 15–17; Ankara, Turkey. p. 1–6.
44. Yu W, Lei B, Ng MK, Cheung AC, Shen Y, Wang S. Tensorizing GAN with high-order pooling for Alzheimer's disease assessment. IEEE Trans Neural Netw Learn Syst. 2022;33(9):4945–59. doi:10.1109/tnnls.2021.3063516.
45. Song J, Zheng J, Li P, Lu X, Zhu G, Shen P. An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis. Front Digit Health. 2021;3:637386. doi:10.3389/fdgth.2021.637386.
46. Muksimova S, Umirzakova S, Baltayev J, Cho YI. Multi-modal fusion and longitudinal analysis for Alzheimer's disease classification using deep learning. Diagnostics. 2025;15(6):717.
47. Adarsh V, Gangadharan GR, Fiore U, Zanetti P. Multimodal classification of Alzheimer's disease and mild cognitive impairment using custom MKSCDDL kernel over CNN with transparent decision-making for explainable diagnosis. Sci Rep. 2024;14(1):1774. doi:10.1038/s41598-024-52185-2.
48. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi:10.1162/neco.1997.9.8.1735.
49. Goyal P, Rani R, Singh K. A multilayered framework for diagnosis and classification of Alzheimer's disease using transfer learned Alexnet and LSTM. Neural Comput Appl. 2024;36(7):3777–801. doi:10.1007/s00521-023-09301-6.
50. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:5998–6008.
51. Baltrušaitis T, Ahuja C, Morency LP. Multimodal machine learning: a survey and taxonomy. IEEE Trans Pattern Anal Mach Intell. 2019;41(2):423–43. doi:10.1109/tpami.2018.2798607.
52. Zhou T, Thung KH, Zhu X, Shen D. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Hum Brain Mapp. 2019;40(3):1001–16.
53. Lu P, Hu L, Mitelpunkt A, Bhatnagar S, Lu L, Liang H. A hierarchical attention-based multimodal fusion framework for predicting the progression of Alzheimer's disease. Biomed Signal Process Control. 2024;88:105669.

54. Chen H, Guo H, Xing L, Chen D, Yuan T, Zhang Y, et al. Multimodal predictive classification of Alzheimer's disease based on attention-combined fusion network: integrated neuroimaging modalities and medical examination data. IET Image Process. 2023;17(11):3153–64.
55. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22–29; Venice, Italy. p. 618–26.
56. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765–74.
57. Kamal MS, Northcote A, Chowdhury L, Dey N, Crespo RG, Herrera-Viedma E. Alzheimer's patient analysis using image and gene expression data and explainable-AI to present associated genes. IEEE Trans Instrum Meas. 2021;70:1–7. doi:10.1109/tim.2021.3107056.
58. Cover TM, Thomas JA. Elements of information theory. 2nd ed. Hoboken, NJ, USA: Wiley-Interscience; 2006.
59. Awarayi NS, Twum F, Hayfron-Acquah JB, Owusu-Agyemang K. A bilateral filtering-based image enhancement for Alzheimer disease classification using CNN. PLoS One. 2024;19(4):e0302358. doi:10.1371/journal.pone.0302358.
60. Tufail AB, Anwar N, Othman MTB, Ullah I, Khan RA, Ma YK, et al. Early-stage Alzheimer's disease categorization using PET neuroimaging modality and convolutional neural networks in the 2D and 3D domains. Sensors. 2022;22(12):4609. doi:10.3390/s22124609.
61. Hamdani H, Septiarini A, Pawitra TA, Sumampouw A, Tejawati A, Kurniawan TB. Effective Alzheimer's stage classification using transfer learning based on MRI images. In: Proceedings of the 2024 Seventh International Conference on Vocational Education and Electrical Engineering (ICVEE); 2024 Oct 30–31; Malang, Indonesia. p. 63–9.
62. Chauhan N, Choi BJ. Classification of Alzheimer's disease using maximal information coefficient-based functional connectivity with an extreme learning machine. Brain Sci. 2023;13(7):1046.
63. Price WN II, Cohen IG. Privacy in the age of medical big data. Nat Med. 2019;25(1):37–43. doi:10.1038/s41591-018-0272-7.
64. Schwarz CG, Kremers WK, Therneau TM, Sharp RR, Gunter JL, Vemuri P, et al. Identification of anonymous MRI research participants with face-recognition software. N Engl J Med. 2019;381(17):1684–6.
65. Hathaliya JJ, Tanwar S. An exhaustive survey on security and privacy issues in healthcare 4.0. Comput Commun. 2020;153:311–35.
66. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53. doi:10.1530/ey.17.12.7.
67. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nat Mach Intell. 2020;2(6):305–11. doi:10.1038/s42256-020-0186-1.
68. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. npj Digit Med. 2020;3(1):119. doi:10.1038/s41746-020-00323-1.