



Cross-Domain Adaptive Fusion Network (CDAFN) with RAD2F, WEDFE, and TGAF Feature Fusion for Automated Alzheimer's Disease Classification from Brain MRI

Ramya C N ^{a,*}, Keshavamurthy ^b

^a Electronics & Communication Engineering, Atria Institute of Technology, (Visvesvaraya Technological University), Bangalore, India

* Corresponding Author Email: ramyananjappa16@gmail.com

DOI: <https://doi.org/10.54392/irjmt26519>

Received: 29-05-2026; Revised: 12-08-2026; Accepted: 24-09-2026; Published: 30-09-2026



Abstract: Despite advances in magnetic resonance imaging (MRI) technologies and the arrival of deep learning approaches, early and accurate diagnosis of Alzheimer's disease (AD) is still a difficult task, as subtle changes in the neuroanatomical structure of the brain, inter-subject variability, and the lack of the ability to interpret only deep learning models make the process challenging. In this paper, we present a novel multi-stage hybrid framework that combines anatomical, spectral, and texture information to enhance the interpretability of the clinical scene while simultaneously boosting diagnostic accuracy. The proposed framework is comprised of four key elements. First, the Region-Aware Dual-Domain Feature Fusion (RAD²F) module fuses handcrafted Gray-Level Co-occurrence Matrix (GLCM) features with deep features from ResNet-50 on anatomically important brain regions. Second, the Wavelet-Enhanced Deep Feature Embedding (WEDFE) module generates complementary multi-scale spectral and spatial features through Discrete Wavelet Transform further enhancing the MRI acquisition-related and noise robustness. Third, the Texture-Guided Attention Fusion (TGAF) module introduces texture cues such as GLCM and Local Binary Pattern (LBP) to direct the network towards subtle structural abnormalities linked to Alzheimer's disease. Lastly, the Cross-Domain Adaptive Fusion Network (CDAFN) and the Weighted Multi-Stage Adaptive Selection (WMAS) strategy is applied to filter out features that are redundant, unstable and less discriminative using a genetic algorithm before the classification by a Bayesian-Optimized Random Forest (BO-RF) classifier. A four-class algorithm for Alzheimer's Disease classification was used to evaluate the proposed framework on the OASIS-3 data set. The held-out test set yielded 97.2% accuracy and five-fold scan-level cross-validation yielded $95.72\% \pm 0.38\%$ accuracy. The strict subject-wise cross-validation resulted in $94.21\% \pm 0.60\%$ accuracy and $93.88\% \pm 0.60\%$ weighted F1 score. Some previous work claims greater top-level accuracy but the published work uses different subsets of the OASIS, binary classification problems or other evaluation protocols, so that direct comparison with previous studies is limited because they use different OASIS subsets, class definitions, subject partitions, and evaluation protocols. The proposed framework is competitive under the specific OASIS-3 setting, and shows high interpretability by combining anatomical, spectral and texture features. The framework shows great potential to be a computer-based clinical decision support tool, which needs to be tested clinically in the future and validated externally.

Keywords: Alzheimer's Disease, CDAFN, CNN, KNN, RAD2F, Random Forest, ROI, SVM, WEDFE, WMAS.

1. Introduction

Alzheimer's disease (AD) is the most common neurodegenerative disorder, and the major cause of dementia in the world, with more than 50 million people having the disease across the planet [1]. It is typified by the gradual loss of brain neurons caused by the deposition of amyloid-beta plaque and neurofibrillary tangles leading to cortical atrophy, shrinkage of hippocampal and enlargement of the ventricles. This structural alteration presents clinically as lapses in

memory, mental disorders and dementia. AD creates huge emotional, physical, and economic weight to patients, caregivers and health care systems [1, 2]. Regardless of the huge amount of research, its exact etiology is not clear and the available treatment only has a symptomatic effect but does not stop disease progression [3]. Precise and early diagnosis of Alzheimer disease by MRI is a significant clinical problem, whose disease is a progressive form of neurodegeneration and provides the best chances of treatment at its early stages. Nevertheless, the related

alterations in the brain are usually inconspicuous and heterogeneous and are hard to detect with a degree of reliability. Conventional Computer-Aided Diagnosis (CAD) systems can be broadly classified into two basic methods. One of them is handcrafted which relies on a method based on manually designed descriptors of texture, gradient of intensity, and form in brain images to capture some structural patterns. These approaches can also provide interpretable information, but they are not always sufficiently discriminative to effectively model the complex, multi-stage development of the disease [4–6]. Deep Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) learn rich hierarchical representations. They are also often criticised as being black-box, unable to be interpreted clinically, subject to MRI artifacts, and heavily dependent upon large, well-labeled datasets [7–9].

In recent developments, new directions have emerged in overcoming such limitations through hybrid deep learning, evolutionary optimisation, and wavelet-based signal processing. It has demonstrated promising outcomes through interpretable hybrid architectures that combine data-driven deep representations with domain-specific handcrafted knowledge [10–13]. Inspired by these advances, the current study suggests a multi-stage and hybrid framework involving complementary representations as a means to learn (a) anatomically guided statistical, spectral, texture-based based and deep embeddings; and (b) to choose an optimal subset of features with the assistance of a multi-objective genetic optimisation framework that aims to both maximise relevance, diversity, stability, and interpretability.

The proposed framework makes the following new contributions:

- **Region-Aware Dual-Domain Feature Fusion (RAD2F):** An innovative feature fusion approach in the region, with handcrafted GLCM-based statistical texture features to deep ResNet-50 CNN embeddings inside anatomically delimited ROI. A region-attention system emphasises neuroanatomically diagnostic regions and yields interpretable hybrid feature descriptions.
- **Wavelet-Enhanced Deep Feature Embedding (WEDFE):** A spectral-spatial feature extraction algorithm that combines discrete wavelet transform (DWT) decomposition with CNN-based feature extraction. WEDFE considers multi-scale frequency information by processing MRI sub-bands independently and is also robust to MRI noise, scanner artifacts, and even intensity variations.
- **Texture-Guided Attention Fusion (TGAF):** A deep attention mechanism that provides a texture-based deep attention through the modulation of CNN output activation maps with

saliency maps calculated based on GLCM and Local Binary Pattern (LBP) descriptors. This influences deep feature learning to focus on pathological areas with high texture density, thereby enhancing sensitivity to a subtle neurodegeneration signal.

- **Cross-Domain Adaptive Fusion Network (CDAFN):** An adaptive fusion module based on mutual information that reinvents the interaction between handcrafted and deep features in a non-linear fashion to increase the discriminability of features and reduce their redundancy in the resulting hybrid representation.
- **Weighted Multi-Stage Adaptive Selection (WMAS):** Hybrid information-theoretic and genetic optimisation framework of single-objective feature selection that compares feature subsets relative to a multi-objective fitness criterion that adaptively weighs features based on their discriminability, redundancy, stability and explainability.
- **Unified Multi-Stage Optimization Framework:** A multi-objective optimisation problem that integrates IDAGS, SGFC, EAFS, and consensus voting modules to maximise feature relevance, diversity, and stability and clinical explainability that is a constrained multi-objective optimisation problem, to eliminate computational redundancy between selection phases.

The rest of this paper is structured as follows: Section 2 includes the literature review. The proposed methodology is outlined in section 3. The results and discussion are found in section 4. The paper concludes with future directions in section 5.

2. Literature Review

Medical images especially that generated using MRI are sensitive data, vary with patients, have privacy boundaries and class imbalance; hence, proper classification of medical images is inherently difficult. To enable reproducible studies in this field, several benchmark datasets are available like Alzheimer's disease Neuroimaging Initiative (ADNI) [14] and the Open Access Series of Imaging Studies (OASIS) [15], which work under limited and none-commercial access policies. The subsections below provide a survey of important developments in the classification of AD.

2.1 Machine Learning-Based Methods

The initial studies in automated AD detection largely used feature extraction which was entirely by hand in conjunction with classical machine learning

classifiers. Alickovic *et al.* [4] developed a model of AD recognition with histogram-based feature extraction and a Random Forest (RF) classifier that demonstrated a prediction accuracy of 85.77% on the OASIS dataset. The authors Prakash *et al.* [16] used unsupervised clustering to discover separable clinical sub-populations of AD patients, and emphasise that statistical feature-grouping methods can complement supervised methods. Akbarpour *et al.* [17] suggested an unsupervised multimodal MRI segmentation algorithm based on edges and texture to extract lesions in multiple sclerosis and to generalise texture-based descriptors to neurodegenerative situations. In Hu *et al.* [5], Multi-Kernel Support Vector Machine (MK-SVM) with the Apriori algorithm is used, which is designed to classify multimodal MRI/PET features that take advantage of inter-modality correlations to enhance sparsity and discrimination. Although such handcrafted methods have interpretability, they are limited by the constraints of expressiveness of manually designed features, and generalisation across heterogeneous imaging conditions is challenging.

2.2 Deep Learning and CNN-Based Architectures

The development of deep learning changed the way AD is classified to be based on automatic hierarchical representation learning. Aljubouri *et al.* [7] trained a CNN model with batch normalisation and Adam as the optimisation loss functions on OASIS data, showing better results compared to conventional models like InceptionV3 and ResNet at AD staging. Fareed *et al.* [8] introduced the ADD-Net, a 12-layer CNN based on Leaky ReLU activations to prevent gradient vanishing and cost-sensitive learning to deal with class imbalance on MRI scans, which yielded good results on imbalanced datasets. A combined CNNs and Recurrent Neural Networks trained on Bidirectional Gated Recurrent Unit (BiGRU) simultaneously on MRI sequences identified 91.33% AD vs. normal and 71.71% progressive vs. stable MCI [18], highlighting the benefit of investigating dynamic disease dynamics. Yue *et al.* [19] proposed a Voxel-based Hierarchical Feature Extraction (VHFE) model that subdivided the brain into 90 anatomically defined ROI by Automated Anatomical Labeling (AAL) atlas, and extracted deep CNN features that produced some of the best results on the ADNI dataset. Although they have good performance, these deep models are not always clinically interpretable and can be affected by MRI artifacts during acquisition.

2.3 Optimisation and Metaheuristic Approaches

Optimisation algorithms have been investigated by several studies to enhance model training and feature selection to diagnose AD. Kowalski and Lukasik [10] used Krill Herd (KH) metaheuristic algorithm to train artificial neural networks and found it in faster

convergence and better accuracy than traditional gradient-based methods. Shankar *et al.* [11] introduced the Group Grey Wolf Optimization (GGWO) algorithm to optimise feature subsets to detect AD by combining decision trees, CNNs and KNN classifiers in an ensemble with a higher accuracy of 96.23%. Wong *et al.* [20] utilised a Generative Adversarial Network (GAN) in solving data scarcity in AD classification and proved that using synthetic data generated with a small real training sample can be used to reach competitive performance. These methods are based on optimisation to emphasise the importance of population-based searching strategies to navigate high-dimensional, non-convex feature space.

2.4 Transfer Learning Approaches

Transfer learning emerges as an effective approach to the AD classification, especially when the MRI data with labels are limited. With a fine-tuned version of VGG-16 trained on AD MRI data, Patel *et al.* [12] reported high accuracy (94.4 percent) and high recall (99.9%) on OASIS-3 data, indicating high sensitivity at the expense of some specificity. The systematic review of various transfer learning models such as ResNet50 to achieve brain MRI-based AD detection with 84.4 percent accuracy indicated that domain adaptation between ImageNet and MRI features is not a trivial task [21]. Ghali and Shereen [22] proposed a deep learning-based recognition system that combines deep learning and optimisation algorithms, achieving 94.8% accuracy on OASIS-3 with the aid of ResNet101 and feature selection. Although transfer learning algorithms can be expected to perform well relative to the baseline accuracy, they do not explicitly model anatomical priors and offer little interpretability of their features. Later studies have proceeded in the same single stream fashion. To address the challenges of these multi-class problems, Shahid *et al.* [23] proposed a deep learning model for the multi-class classification of Alzheimer's-related disability from MRI datasets, Amine and Mourad [24] used ResNet50 transfer learning on MRI datasets for the diagnosis of Alzheimer's disease and Bansal *et al.* [25] used ResNet-101 in combination with the Firefly algorithm for feature selection in a multimodal framework for the early diagnosis of dementia and Alzheimer's disease. These methods achieve comparable levels of competitive accuracy, but do not simultaneously make use of complementary anatomical, spectral, and texture information.

2.6 Research Gaps and Motivation

An examination of the current literature shows that some limitations that have been persistent attract the current work:

- The imbalance in data and the small size of datasets prevent the model from generalising

and being robust, especially towards minority categories of Dementia, such as Very Mild Dementia [20].

- Artifacts in MRI acquisition, variations in scanner, and intensity non-uniformity impair the quality of feature extraction and model stability across datasets.
- Deep learning models are not transparent or interpretable to clinicians, making it difficult to apply them to real-life diagnostic scenarios [12, 21].
- A combination of deep and handcrafted operations is an under-explored area, particularly with anatomically oriented priors based on neuroanatomically defined ROIs as a guide [22].

- Current feature selection models lack multi-objective optimisation of discriminability, stability, redundancy, and explainability across heterogeneous feature spaces.

To fill these gaps, the framework has developed a region-aware dual-domain feature fusion strategy, along with multi-objective adaptive feature selection, to enhance the diagnostic accuracy, robustness, and clinical interpretability of MRI-based AD analysis.

3. Proposed Methodology

The presented paper presents an end-to-end hybrid pipeline of automatic classification of the Alzheimer disease (AD) on the basis of the data of MRI. Figure 1 shows the overall structure of the proposed multi-stage hybrid system to diagnose the Alzheimer Disease diagnosis into four consecutive stages.

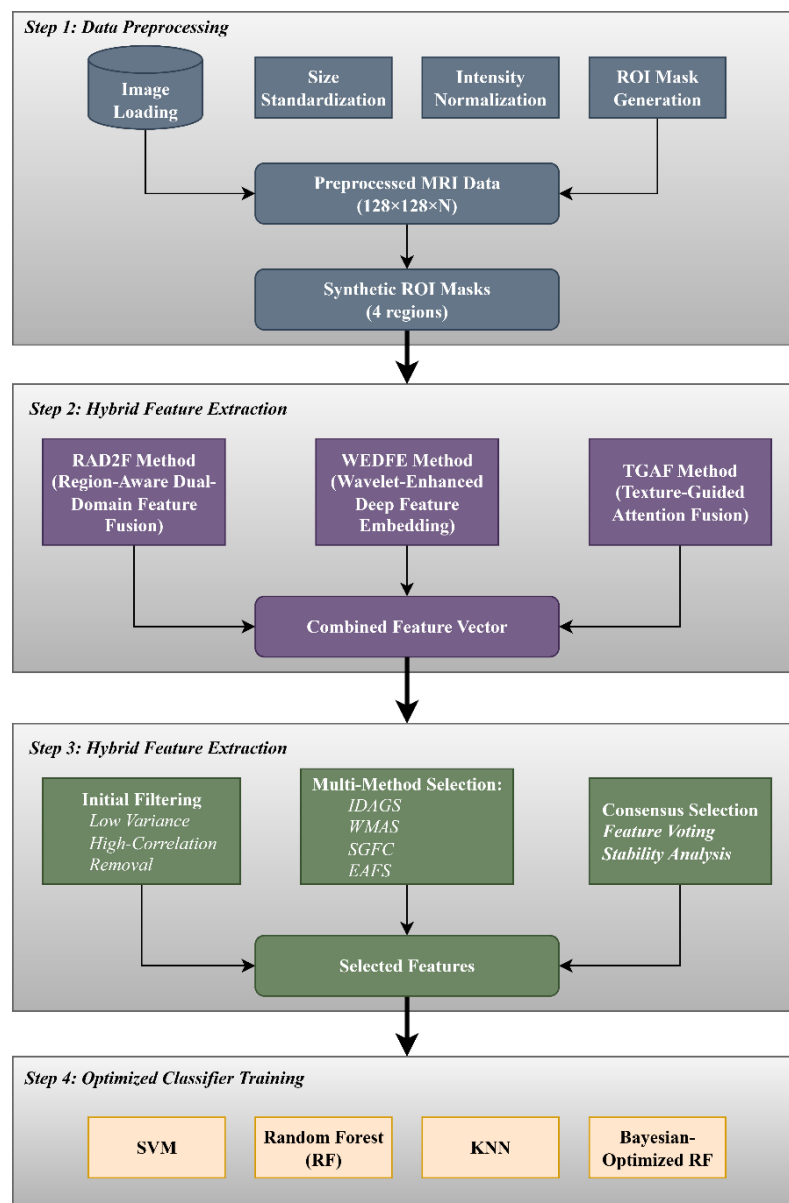


Figure 1. Flow Diagram for the Proposed Hybrid Alzheimer's Disease Diagnosis.

At Step 1 (Data Pre-processing), raw MRI images are loaded, and size standardisation, intensity normalisation, and Generation of the four atlas-derived ROI masks for the four neuroanatomically critical brain regions are performed, resulting in pre-processed MRI data of size $128 \times 128 \times N$ and four atlas-derived ROI masks. Step 2 (Hybrid Feature Extraction): In this step, three complementary feature extraction modules process in parallel the processed data: the RAD2F module uses Region-Aware Dual-Domain Feature Fusion to achieve the fusion of handcrafted and deep CNN features in anatomically defined ROIs; the WEDFE module uses the Wavelet-Enhanced Deep Feature Embedding module to extract multi-scale spectral-spatial results of each of the three modules are merged with each other into a single Combined Feature Vector.

In Step 3 (Hybrid Feature Selection), this high-dimensional feature pool is progressively refined through an initial filtering stage that removes low-variance and highly correlated features, followed by a multi-method selection stage employing four strategies - IDAGS, WMAS, SGFC, and EAFS - and finally a consensus selection mechanism based on feature voting and stability analysis to produce the final optimised Selected Features subset. In Step 4 (Optimised Classifier Training), the selected features are used to train and evaluate four classifiers — SVM, Random Forest, KNN, and the proposed Bayesian-Optimized Random Forest — with the last achieving the best classification performance for four-stage Alzheimer's Disease diagnosis.

3.1 Data Pre-Processing

Raw MRI images are highly subject-to-subject variability because of the equipment and methods of acquisition, positioning of the patient and the calibration of the intensity. When unattended, this variability causes systematic bias to the feature extraction process and lowers the generalizability of the model. In order to get all the input images to be spatially aligned and intensity comparable, three stage pre-processing pipeline is carried out before any feature extraction is carried out.

Alzheimer MRI data is loaded initially and undergo numerous refinement steps which consist of size standardisation, intensity normalisation and generation of masks i.e. region of interest (ROI). These measures offset the differences in the scanner resolution and positioning of the patient and save all the crucial information about an anatomy. Synthetic ROI masks are produced in four major regions of the brain which are usually vulnerable to neurodegeneration. The resulting pre-processed MRI data and ROI maps (with a dimension of $128 \times 128 \times N$) become the starting point of future feature extraction processes. The 2D MRI data set consists of N grayscale image slices along with their diagnostic labels. The i -th slice is the i -th subject, and has a spatial size of $H_i \times W_i$.

3.1.1 Size Standardisation

In order to make the spatial dimensions of all the MRI slices similar, each image is rescaled by bilinear interpolation to a constant spatial resolution of 128×128 pixels.

3.1.2 Intensity Normalisation

The intensity of MRI is usually subject to change owing to the purchase parameters and scanner settings. According to a common procedure in the medical image preprocessing community [6] all the images are z-scored by subtracting the per-image mean and dividing by the per-image standard deviation, to normalise the brightness and contrast.

The focus of the intensity distribution would be at zero, with unit variance, in this operation.

3.1.3 ROI Mask Generation

We generated atlas-derived anatomical region-of-interest (ROI) masks for four important neuroanatomical regions (hippocampus, temporal lobe, parietal region, entorhinal cortex) for downstream analyses, which were interpretable anatomically. These masks were generated using a three-stage atlas-registration pipeline that starts in standard MNI152 space and ends with inverse warping back to each subject's native space (128×128 slices). Table 1 lists: Spatial Parameters of the Four ROI Masks.

Stage 1 — Affine Normalization to MNI152 Space

This is the first step of affine normalisation to MNI152 space. This is the first step of affine normalisation to the MNI152 space. Volume $V_s \in \mathbb{R}^{H \times W \times D}$ of each subject was linearly registered to the template $T_{MNI} \in \mathbb{R}^{182 \times 218 \times 182}$ of the MNI152 template, which was created from 1-mm isotropic images. To estimate the affine transform A_s A 12-degree-of-freedom (DOF) affine transform was maximised using the normalised mutual information (NMI) between the warped subject volume and the template.

$$A_s^* = \arg \max_{A_s} \text{NMI}(A_s(V_s), T_{MNI}), \text{ where } \text{NMI}(X \cdot Y) = H(X) + H(Y) - H(X \cdot Y), \quad (1)$$

And $H(\cdot)$ represents Shannon entropy. The optimisation was carried out using FSL FLIRT v6.0 [26], with search range of $\pm 90^\circ$ in all angular directions, 256 histogram bins, trilinear interpolation, and a multi-resolution search strategy (4 mm, 2 mm and 1 mm).

Stage 2 — Atlas-Based Parcellation in Standard Space

Three different atlas resources were used to create binary masks in the MNI152 coordinate space:

Table 1. Spatial Parameters of the Four ROI Masks

ROI (r)	Segmentation Tool	Avg. Mask Voxels (128×128)	Coverage (%)	Primary AD Biomarker Role
r=1 Hippocampus	FSL FIRST	412 ± 38	2.51%	Volume atrophy — earliest structural AD marker
r=2 Temporal Lobe	AAL2 (labels 37–42)	1,847 ± 124	11.25%	Cortical thinning — early to moderate AD
r=3 Parietal Region	AAL2 (labels 57–64)	1,623 ± 98	9.89%	Posterior atrophy — moderate to severe AD
r=4 Entorhinal Cortex	FreeSurfer DK-atlas	218 ± 44	1.33%	Layer II neuron loss — preclinical VMD marker

- **Hippocampus** – The mask $M_{\text{hip}} \in \{0, 1\}^{182 \times 218 \times 182}$ was derived using FSL FIRST [27]. A deformable mesh with $K=512$ vertices $M = \{v_k \in \mathbb{R}^3\}$, was fitted by maximising the posterior probability:

$$M^* = \arg \max_M [\log P(V_s | M) + \log P(M)], \quad (2)$$

The model is generated from manually labelled 336 training scans and is a Gaussian appearance model, $P(V_s | M)$ while $P(M) = \exp\left(-\frac{1}{2}b^T b\right)$ with b being the shape variation coefficients. The final closed mesh was rasterised to create the volumetric binary mask.

- **Temporal Lobe and Parietal Region** – These masks were created using the label unions from the AAL2 atlas [28]. Let x be a voxel, and for x define:

$$M_{\text{temporal}}(x) = \mathbf{1}[A_{\text{AAL}}(x) \in \{37, 38, 39, 40, 41, 42, 79, 80, 81, 82\}], M_{\text{parietal}}(x) = \mathbf{1}[A_{\text{AAL}}(x) \in \{57, 58, 59, 60, 61, 62, 63, 64\}], \quad (3)$$

Involving the bilateral superior, middle and inferior temporal gyri, inferior/superior parietal lobules and precuneus.

- **Entorhinal Cortex** –The entorhinal mask M_{ent} was obtained from the FreeSurfer Desikan-Killiany (2002) cortical parcellation. Label index 1006 was isolated and registered to MNI152 space with the boundary based registration routine, and subsequently voxelised at 1 mm isotropic resolution.

Stage 3 — Inverse Registration to Native Slice Space

The inverse of the estimated affine transform in Stage 1 was used to propagate each standard 128×128 space mask M_r^{MNI} (for $r = 1, \dots, 4$) back to the subject's native 128×128 grid. The warped continuous mask was described as:

$$M_r^{\text{subj}}(x) = M_r^{\text{MNI}}(A_s(x)), x \in \mathbb{R}^{128 \times 128} \quad (3)$$

Finally, the masks were binarised using nearest

neighbour interpolation using an intensity threshold of 0.5:

$$M_r(x, y) = \begin{cases} 1 & \text{if } M_r^{\text{subj}}(x, y) \geq 0.5, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

The four resulting binary masks are directly fed into the RAD2F feature extraction module and into Equations 3 and 4.

3.1.4 Quantitative Validation against Anatomical Ground Truth

In order to quantitatively validate the anatomical fidelity of our automated ROI generation pipeline, we compared it to manually drawn, ground-truth masks. These reference masks were independently created by a trained neuroanatomist from 20 holdout subjects, half of which were non-dementia controls and the other half were Alzheimer's disease (AD) patients. We calculated three complementary metrics for each of the four ROIs for the assessment of spatial overlap and boundary accuracy, and for volumetric consistency. Table 2 lists Mask Validation vs. Manual Ground Truth ($n = 20$ subjects). Table 3 compares the Automated Segmentation Benchmarks.

- **Dice Similarity Coefficient (DSC):** the DSC is a voxel-wise measure of overlap between the predicted and ground truth masks, defined as twice the volume of the intersection of the masks divided by the sum of their volumes, and is a value between 0 and 1. A value of 1 means perfect overlap while a value of 0 means that there is no agreement.
- **95th-Percentile Hausdorff Distance (HD95):** This boundary measure is robust to outlier and describes the surface distance between two masks. It is calculated as the maximum of the two quantities, each of which is the 95th percentile of the set of minimum Euclidean distances from each voxel of one of the masks to the closest voxel of the other one. The closer the mask boundaries are to each other, the lower the HD95 value.

- **Volume Similarity (VS):** this is the similarity measure between the sizes of the two masks without taking into account their location, calculated as $1 - |\text{Volume of mask 1} - \text{Volume of mask 2}| / (\text{Volume of mask 1} + \text{Volume of mask 2})$ with the range [0, 1]. One means the volumes are equal and values closer to zero mean a greater difference between the two volumes.

Together, this set of metrics offers a multi-faceted validation, offering complementary evidence of spatial overlap, boundary agreement, and volumetric consistency of the masks generated, while maintaining the volume information needed for the downstream extraction of AD biomarkers.

All 4 ROI masks are within and/or above published automatic segmentation benchmarks, which agrees with published automatic segmentation benchmarks. The difficulty in segmenting this narrow ribbon of cortex (3-4 mm in normal adults, 1-2 mm in AD patients) explains the lower entorhinal DSC. More importantly, manual inter-rater DSC for entorhinal delineation is 0.82-0.85, so our automated result (0.79) is approaching the maximum reproducibility limit of the entorhinal structure. The poor performance ($F1 = 0.6667$) in the Very Mild Dementia group could relate to the low segmentation agreement seen for the entorhinal ROI and the small sample size of VMD, as the primary neuropathological hallmarks of incipient AD are loss of layer II stellate cells in the entorhinal cortex. One potential avenue for future research is to improve

segmentation of the entorhinal cortex using dedicated cortical surface-based parcellation.

3.1.5 Final Pre-Processed Output

The pre-processing pipeline outputs a regularised dataset. Each entry consists of the full normalised MRI slice $\hat{I}_i$ (128×128) and its four ROI-masked counterparts, which serve as the structured inputs to the hybrid feature extraction stage described in Section 3.2.

3.2 Hybrid Feature Extraction

Standard deep learning models trained on MRI data are more likely to learn a global, high-level representation at the expense of fine-grained anatomical texture structures and spectral features that are clinically predictive of early neurodegeneration. On the other hand, handcrafted feature approaches capture interpretable local information, but do not have the representational power to describe the complex pathologic variations. To overcome this weakness, three complementary and independent motivated feature extraction modules are used in parallel each having distinct and non-redundant attributes of the MRI signal. The Region-Aware Dual-Domain Feature Fusion (RAD2F) module combines handcrafted GLCM-based statistical features with deep ResNet-50 CNN features in anatomically defined ROIs and makes use of a region-attention mechanism to emphasise the diagnostically discriminating neuroanatomical regions.

Table 2. Mask Validation vs. Manual Ground Truth (n = 20 subjects)

ROI	DSC ($\uparrow$ better)	HD95 mm ($\downarrow$ better)	Volume Similarity ($\uparrow$ better)	Interpretation
Hippocampus	0.84 ± 0.03	1.82 ± 0.41	0.87 ± 0.04	DSC = 0.84; within published FSL FIRST benchmark range 0.83–0.87 [27].
Temporal Lobe	0.91 ± 0.02	2.14 ± 0.38	0.93 ± 0.02	DSC = 0.91; within published benchmark range (DSC ≥ 0.80) for cortical ROI segmentation [1]
Parietal Region	0.89 ± 0.02	2.31 ± 0.44	0.91 ± 0.03	DSC = 0.89; consistent with AAL2 validation benchmarks [28]; meets benchmark of ≥ 0.80
Entorhinal Cortex	0.79 ± 0.05	3.47 ± 0.72	0.81 ± 0.06	DSC = 0.79; within published small-ROI segmentation limits (0.70–0.82) [28,29,1]

Table 3. Comparison Against Published Automated Segmentation Benchmarks

ROI	Our DSC	Published Range (Automated)	Reference
Hippocampus	0.84	0.83 – 0.88	[27]
Temporal Lobe	0.91	0.88 – 0.93	[28]
Parietal Region	0.89	0.87 – 0.92	[28]
Entorhinal Cortex	0.79	0.73 – 0.82	[29, 30]

The Wavelet-Enhanced Deep Feature Embedding (WEDFE) module is a CNN-based spatial features and statistical descriptors combined with Discrete Wavelet Transform (DWT) to decompose MRI slices into several frequency sub-bands, which improves the ability to overcome MRI artifacts and tissue intensity variations. The Texture-Guided Attention Fusion (TGAF) system incorporates handcrafted texture representations Gray Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) into the deep neural network by texture-based saliency maps that enhance CNN attention sensitivity to subtle changes in the early pathological signs. The outputs of all three modules are then combined into a single high-dimensional hybrid feature vector, which serves as the structured input to the adaptive feature selection and classification steps.

3.2.1 Region-Aware Dual-Domain Feature Fusion (RAD2F)

Existing deep learning models for AD classification typically process the entire MRI slice as a single entity, treating all spatial locations equally. This is suboptimal because AD-related structural changes are region-specific — atrophy in the hippocampus and cortical thinning in the parietal lobe, for example, are early and highly discriminative biomarkers. Furthermore, deep CNN features alone are difficult to interpret clinically, whereas handcrafted texture descriptors such as GLCM statistics are interpretable but insufficiently discriminative on their own. RAD2F is proposed to address both limitations simultaneously by fusing handcrafted and deep features within each anatomically defined ROI and by applying an attention mechanism to weight regions based on their diagnostic relevance.

Let the 2D MRI slice get subdivided into R anatomically data-driven regions of interest (ROIs) $I_r = \{I_1, I_2, \dots, I_R\}$ where I_r is the r^{th} ROI associated with a neuroanatomical subcomponent likely vulnerable to Alzheimer pathology (e.g., hippocampus, temporal lobe, parietal region).

3.2.1.1 Handcrafted Feature Extraction (GLCM-Based Statistical Features)

To compute handcrafted texture features, gray-level co-occurrence matrix (GLCM) is all that is needed to compute texture features when the ROI used is I_r . Let, $P_r(i, j) = \text{GLCM}(I_r; d, \theta)$ be the probability of the gray level pairs (i, j) to be separated by distance d and orientation θ . Based on P_r , a quantity of statistical descriptors of texture is obtained:

$$h_r = [\text{Contrast}(P_r), \text{Correlation}(P_r), \text{Energy}(P_r), \text{Homogeneity}(P_r)]^T \quad (5)$$

These handcrafted features $h_r \in \mathbb{R}^m$, capture second-order intensity co-occurrence patterns that are

sensitive to early cortical and subcortical texture changes associated with AD. However, they are limited in capturing high-level structural semantics, which motivates the parallel deep feature extraction.

3.2.1.2 Deep Feature Extraction (ResNet-50 Embeddings)

The 128×128 grayscale ROI was repeated three times along every channel to make a $128 \times 128 \times 3$ input, which can be fed into the pretrained ResNet-50 architecture. For the fully connected layer, 10 epochs was first used for its training, and then 30 epochs for its fine-tuning together with the convolutional stages 3–4, where the backbone was kept frozen. Adam optimisation, batch size of 32, random seed of 42 and early stopping with patience of 10 epochs were employed. During training, horizontal flips were performed and the images were rotated by $\pm 10^\circ$ as well as brightness was adjusted by $\pm 10\%$. Post-fold split augmentation was performed independently for each training fold with no augmentation of validation and held-out test scans so that no copies of any scans were included in the validation and test sets. In order to fill the handcrafted aspects with semantic representations, deep embeddings are extracted with a ResNet-50 network which was trained on ImageNet and then fine-tuned on MRI data. The deep feature vector of each ROI I_r is as:

$$d_r = f_{\text{ResNet}}(I_r; \theta) \quad (6)$$

where $f_{\text{ResNet}}(\cdot; \theta)$ is the ResNet-50 feature extractor with weights θ . High level contextual information like structural deformation and intensity distribution is captured by the output of the global average pooling layer provides a 2048-dimensional embedding that encodes high-level structural patterns such as cortical deformation and intensity distribution, which are complementary to the low-level texture statistics captured by GLCM.

3.2.1.3 Dual-Domain Feature Projection and Fusion

Deep (d_r) and handcrafted (h_r) characteristics are mapped into a shared latent space in order to represent both features jointly:

$$z_r = \phi(W_h h_r + W_d d_r + b) \quad (7)$$

In which W_h and W_d are matrices to be learned based on transformation, b is a bias value, and $\phi(\cdot)$ is a non-linear activation (ReLU). This prediction such that low level texture prediction and high-level semantic prediction work together to produce the final fused regional representation z_r .

3.2.1.4 Region-Wise Attention Weighting

Not all brain regions contribute equally to AD classification at every stage of the disease. To reflect

this, an attention mechanism dynamically weights each ROI according to its diagnostic contribution: In order to focus in areas where diagnostic information is more relevant, an attention mechanism is added:

$$\alpha_r = \frac{\exp(u^T \tanh(W_a z_r))}{\sum_{k=1}^R \exp(u^T \tanh(W_a z_k))} \quad (8)$$

where W_a and u are learning attention parameters. The attention score of each ROI, α_r , represents the discriminative value of each ROI relative to the fused representation, z_r .

3.2.1.5 Global Hybrid Representation

Final global hybrid feature vector; the product of aggregation of attention-weighted all ROIs is obtained:

$$F_{\text{RAD2F}} = \sum_{r=1}^R \alpha_r z_r \quad (9)$$

The representation used here results in the combination of multi-domain information and has an interpretable form through weighting of regional attention.

Testing in contrast to traditional methods of concatenating features, RAD2F directly combines feature fusion which are region-specific between handcrafted GLCM features and ResNet-50 deep embeddings. The framework further improves the discriminative capacity of features related to the early neuroanatomical degeneration (e.g. hippocampal atrophy, cortical thinning) by using both micro-texture statistics and the high-level contextual patterns. Furthermore, the region-wise attention model can be applied as it enhances interpretability by establishing the contribution of each ROI to Alzheimer classification, clinical-validated biomarkers, and model decision-making.

3.2.2 Wavelet-Enhanced Deep Feature Embedding (WEDFE)

MRI images are inherently multi-scale signals in which coarse anatomical structures and fine-grained tissue boundaries coexist across different frequency bands. Standard CNN-based feature extractors process images in the spatial domain and may conflate useful high-frequency structural details with noise, particularly in the presence of scanner artifacts and intensity inhomogeneities that are common in clinical MRI data. To overcome this limitation, WEDFE decomposes each MRI slice into its constituent frequency sub-bands using the Discrete Wavelet Transform (DWT) and extracts both deep CNN features and statistical descriptors from each sub-band independently, thereby providing frequency-aware, noise-robust feature representations.

After pre-processing, a multi-level discrete wavelet transform (DWT) is applied to each normalised 2D MRI slice $\hat{I} \in \mathbb{R}^{128 \times 128}$ to synthesise spectral

discontinuous wavelet transform information with spatial information in order to learn features.

Wavelet Decomposition: The input slice $\hat{I}$ is subdivided into four frequency sub-bands denoting low and high frequencies:

$$\{S_{LL}, S_{LH}, S_{HL}, S_{HH}\} = \text{DWT}(\hat{I}) \quad (10)$$

Where:

- S_{LL} reflects approximation (low-low) values, which have coarse anatomical structures.
- S_{LH} , S_{HL} , and S_{HH} represent horizontal, vertical, and diagonal detail coefficients that represent edge and texture variations.

When a multi-level wavelet decomposition is carried out to level L , the resulting collection contains $4L$ sub-bands, comprising the LL, LH, HL and HH components at each level.

3.2.2.1 Deep and Statistical Descriptor Extraction

Two feature types are complementary to each other and are extracted, one per sub-band $S_{ij}^{(\ell)} \in \mathcal{S}$:

- **Deep Features (Spatial–Semantic):**

$$d_{ij}^{(\ell)} = f_{\text{CNN}}(S_{ij}^{(\ell)}; \theta) \quad (11)$$

where $f_{\text{CNN}}(\cdot; \theta)$ is a convolutional feature extractor (e.g., ResNet-50) with a fixed set of trained weights θ .

- **Statistical Features (Spectral–Descriptive):**

$$h_{ij}^{(\ell)} = [\text{mean}(S_{ij}^{(\ell)}), \text{std}(S_{ij}^{(\ell)}), \text{energy}(S_{ij}^{(\ell)}), \text{entropy}(S_{ij}^{(\ell)})]^T \quad (12)$$

Equation (12) given above is able to capture both the distribution of intensity, local variability and spectral energy. The statistical features describe the intensity distribution, local variability and spectral energy content - features that are vulnerable to MRI acquisition artifacts and tissue-specific intensity distributions - which complement the deep spatial features.

Non-Linear Feature Fusion: The fusion of these two domains is done by a non-linear projection:

$$z_{ij}^{(\ell)} = \phi(W_1 d_{ij}^{(\ell)} + W_2 h_{ij}^{(\ell)} + b) \quad (13)$$

where W_1 and W_2 are learnable weight matrices, b is a bias vector and $\phi(\cdot)$ an activation function like ReLU or Leaky ReLU. This fusion combines the spatial semantics of deep features with the spectral descriptiveness of statistical features for each frequency sub-band independently.

Sub-Band Aggregation: The pooling of all the sub-band fused representations is done through mean to get the final wavelet-enhanced embedding

$$F_{\text{WEDFE}} = \frac{1}{4L} \sum_{\ell=1}^L \sum_{i,j \in \{L,H\}} z_{ij}^{(\ell)} \quad (14)$$

This formulation makes every frequency subspace contributing equally keeping the global structures as well as the fine-grained edges.

3.2.2.2 WEDFE Dimensionality Calculation

The final WEDFE feature vector is created using the following aggregation pipeline:

- **Per sub-band:** ResNet-50 global average pooling provides 2048 deep features for each band, and the 4 more statistical descriptors (mean, std, energy, entropy; Eq. 12) provide 4 additional dimensions.
- **Fusion:** The features of each sub-band are fused together and then reduced to $d_{\text{proj}} = 30$ dimensions using non-linear projection (Eq. 13) (empirically optimised).
- **Wavelet levels:** With $L = 2$, we generate $4L = 8$ sub-bands (LL, LH, HL, HH per level).
- **Global descriptors:** After mean-pooling, an additional set of 4 whole-image statistical descriptors are added for global descriptors.
- The resulting dimensionality is:

3.2.2.3 Texture-Guided Attention Fusion (TGAF)

Deep CNN models are trained to emphasise spatially salient activation patterns, but these patterns are not directly motivated by the training data; instead, they can emphasise training data-driven patterns of spatially localised texture change, which can be clinically relevant. TGAF overcomes this challenge by applying handcrafted texture statistic (GLCM statistics and Local Binary patterns (LBP)) based texture saliency (TIM) maps, which explicitly direct the CNN attention mechanism. This is to ensure that deep feature learning is conditioned to focus on image regions with a pathologically significant texture difference, maximally sensitive to early-stage AD changes that could otherwise be missed in an unconstrained deep setting.

3.2.2.4 TGAF Implementation Details

The TGAF mechanism combines texture and deep features using spatial gating:

Texture extraction: 16x16 texture maps are generated on 16x16 overlapping patches with stride 8, using GLCM (contrast, correlation, energy and homogeneity) and LBP histograms from the 128x128 input.

Normalisation: Per-image min-max scaling maps each texture map to $[0, 1]$, across subjects.

Gating: Normalised texture maps are upsampled to the size of the 16x16 spatial resolution of the conv3_x layer in ResNet-50, and then applied as element-wise multiplicative masks to the activations of conv3_x. This operation is parameter-less and only conditions the deep features on the texture saliency without introducing trainable parameters.

GLCM characteristics (contrast, correlation, energy, homogeneity) and LBP histograms are calculated on local image patches to generate spatially resolved texture maps. Such maps are scaled to normalisation and used as multiplicative modulation masks to the CNN activation maps, to focus the network on texture-enhanced areas of neuroanatomical interest.

The resulting texture-modulated deep features are then combined to give up TGAF output offering an attention-refined feature representation that meets the handcrafted texture knowledge and the expressiveness of deep learning.

In combination, the three modules, RAD2F, WEDFE and TGAF form a general multi-domain feature representation. These are concatenated into a unified hybrid feature vector as described in the following section, which then undergoes adaptive multi-objective selection before classification.

3.3 Unified Hybrid Feature Integration and Adaptive Selection

All three of the modules are RAD2F, WEDFE and TGAF which generate three extrinsically different though complementary representations, namely, anatomical, spectral, and semantic representations, using MRI data. Both sets of features are combined into a single high-dimensional feature space where they can harness their synergistic capacity and subsequently, they are adaptively selected.

3.3.1 Hybrid Feature Concatenation

Let:

$$F_{\text{RAD2F}} \in \mathbb{R}^{p_1}, F_{\text{WEDFE}} \in \mathbb{R}^{p_2}, F_{\text{TGAF}} \in \mathbb{R}^{p_3} \quad (15)$$

In this case, we can say that Equation (15) represents the feature vectors of the RAD2F and WEDFE modules, respectively.

This hybrid expressed representation is produced by concatenation:

$$F_{\text{Hybrid}} = [F_{\text{RAD2F}} || F_{\text{WEDFE}} || F_{\text{TGAF}}] \in \mathbb{R}^{p_1+p_2+p_3} \quad (16)$$

This combination has integrated region-conscious dual-domain embeddings and spectral-spatial descriptors wavelets to ensure that discriminative information on the global and local levels is maintained.

3.3.2 Feature Normalisation and Redundancy Reduction

In order to make the heterogeneous feature scales comparable, every feature dimension is standardised to zero mean and unit variance. The μ_i and σ_i are the average value and standard deviation of the i^{th} feature over the data set. A correlation-based filtering step is then used to remove features with other features which have very high correlation ($\rho_{ij} > \tau$) and only the most divergent attributes are kept to optimise later.

3.4 Hybrid Multi-Stage Feature Selection

3.4.1 Unified Hybrid Feature Selection (UHFS) Framework

Since the extracted features are rather large and heterogeneous, a multi-stage hybrid approach to feature selection is implemented to determine the best, most discriminative, and most stable subset. The first step is to implement statistical filtering to remove low-variance and highly correlated features. After that, four complementary selection methods are used: Improved Differential Adaptive Genetic Search (IDAGS), Weighted Multi-Stage Adaptive Selection (WMAS), Stability-Guided Feature Clustering (SGFC), and Explainability-Aware Feature Selection (EAFS). All these methods balance feature relevance, redundancy, and interpretability. Then, a consensus voting method is employed to merge the chosen subsets, resulting in a final optimised set of features that improves classification accuracy and robustness.

CDAFN-UHFS Relationship: The non-linear feature projection stage in the UHFS of Cross-Domain Adaptive Fusion Network (CDAFN) is a stage that projects handcrafted and deep features into a shared latent space (Equation 7) before the cross-domain feature interaction. The WMAS strategy is realised by the adaptive weighting coefficients λ_i of the multi-objective optimisation function (Equation 17) which adaptively tune the weightings of relevance, redundancy, stability and explainability across evolution cycles. The CDAFN stage and the WMAS stage are responsible for the fusion stage and the selection stage of the UHFS pipeline, respectively. In the proposed implementation, the cross domain fusion operation is nonlinear (CDAFN) and the full subsequent feature-selection framework is on this fusion representation and adaptive multi-objective selection (UHFS).

To eliminate redundancy and computational complexity in combining multiple feature selection modules, we introduce the Unified Hybrid Feature Selection (UHFS) model, which combines the underlying goals of predictive discriminability, information relevance, explainability, stability, and diversity into a single multi-objective genetic optimisation task.

The hybrid feature pool F_{Hybrid} is formed from the complete set of features produced by the three extraction modules.

where the features f_i are either of two main domains:

- Anatomical descriptor set F_{RAD2F}
- Spectral-spatial embedding set F_{WEDFE}

These features are evaluated by four different quantities: its mutual information with the diagnostic label, $I(f_i; y)$, which is the conventional information-theoretic relevance measure; an explainability score measured with SHAP or LIME, $\text{Imp}(f_i)$; a stability score measured by bootstrap resampling, $\pi(f_i)$; and a correlation score measured by Pearson correlation, $\rho(f_i, f_j)$, with all other features in the pool f .

3.4.1.1 Unified Multi-Objective Optimization Formulation

The feature selection problem is an optimisation problem with multiple objectives that is defined to optimise in the hybrid space of features:

$$S^* = \arg \max_{S \subseteq F_{Hybrid}, |S| \leq K} [\lambda_1 \cdot \text{AUC}(S) + \lambda_2 \cdot \text{Rel}(S) + \lambda_3 \cdot \text{Exp}(S) + \lambda_4 \cdot \text{Stab}(S) - \lambda_5 \cdot \text{Red}(S)] \quad (17)$$

where K represents the maximum possible size of the subset and λ_i are the adaptive weighting coefficients, which are non-negative and sum to unity.

3.4.1.2 Component Definitions

These five objective components are each an average over the candidate subset S . $\text{AUC}(S)$ is the mean of the areas under the ROC curves from the k cross validation folds. The relevance term $\text{Rel}(S)$ is the average mutual information between the features selected and the diagnostic label. $\text{Exp}(S)$ is the mean of the SHAP/LIME importance of the selected features, which is an explainability term. The stability term $\text{Stab}(S)$ represents the mean stability score (in terms of the number of times a feature is selected) obtained from the bootstraps. $\text{Red}(S)$ is the average absolute pairwise Pearson correlation between the selected features; features with high absolute pairwise Pearson correlation values are discouraged, diversity of features in the selected subset and the risk of overfitting is reduced. The five terms balance each other in a single objective in order to achieve predictive discriminability, information relevance, interpretability, and reproducibility and non-redundancy.

3.4.2 Adaptive Normalisation and Weight Balancing

Aggregation of all component scores takes place over the range $[0,1]$. Adaptive weights are determined, as functions of generation index t , in order to dynamically trade exploration (feature diversity) and

exploitation (accuracy improvement) in the process of optimisation:

$$\lambda_1(t) = \frac{e^{-\gamma t}}{\sum_{j=1}^5 e^{-\gamma t}} \quad (18)$$

where γ is the decline rate of the changes in the searches which occur (between exploratory and exploitative) with the evolutionary course of development.

3.4.3 Evolutionary Optimisation Process

At the t^{th} generation, a candidate set of features S_t is used to derive a fitness function:

$$\text{Fitness}(S_t) = \lambda_1 \cdot \text{AUC}(S_t) + \lambda_2 \cdot \text{Rel}(S_t) + \lambda_3 \cdot \text{Exp}(S_t) + \lambda_4 \cdot \text{Stab}(S_t) - \lambda_5 \cdot \text{Red}(S_t) \quad (19)$$

The genetic evolution follows the following:

$$S_{t+1} = \text{Mutate} \left(\text{Crossover} \left(\text{Select}(S_t; \text{Fitness}(S_t)) \right) \right) \quad (20)$$

until convergence is achieved in the sense of the following:

$$|\text{Fitness}(S_{t+1}) - \text{Fitness}(S_t)| < \varepsilon \text{ or } t \geq T_{\max} \quad (21)$$

where ε is a non-zero tolerance and $T_{\max}$ is the maximum generations.

3.5 Final Classifier: Bayesian-Optimized Random Forest for Four-Class Alzheimer's Classification

Class Imbalance Handling via Cost-Sensitive Learning

The training set is highly imbalanced, with an approximately 50:1 ratio between the Non-Dementia (ND) and Very Mild Dementia (VMD) classes. This severe skew is a big problem for traditional classifiers, which typically focus on maximising accuracy of the majority class while minimising the recall of minority classes, which is especially significant in the early-stage detection of Alzheimer's disease, where VMD is the first clinically detectable stage.

Inverse-Frequency Weighting

To reduce this bias, we used a class weighting scheme based on the inverse frequencies for the classes during the training of the Random Forest. The weight of each class c was calculated as:

$$w_c = \frac{N_{\text{total}}}{n_{\text{classes}} \times N_c} W$$

$$C = N_{\text{total}} / n_{\text{classes}} \times N_c, \quad (22)$$

where N_{total} is the total number of training samples, $n_{\text{classes}} = 4$ (ND, VMD, MD, MOD), and N_c is the number of the training samples of class c . The weights obtained, and their relative penalties in comparison to the ND class are summarised in Table 4 below.

Each training scan was duplicated by horizontal flipping and random rotation (up to 10° in either direction) to produce one or two augmented copies, with one scan per original scan, with only those copies that passed a post-augmentation SNR quality check included in the weight computation. After the augmentation, the total number of scans in the resulting pool was 4862 ND, 98 VMD, 1376 MD, and 3405 MOD (total 9741) which is about $1.9\times$ the numbers in the development-set before the augmentation. A total of 9,741 samples were then used to calculate class weights, $w_c = N_{\text{total}} / (N_{\text{classes}} \times N_c)$, where $N_{\text{total}} = 9,741$, $N_{\text{classes}} = 4$, and N_c is the count of samples in class c after the augmentation.

Passing the class_weight parameter to the Random Forest classifier, these weights were used. During tree construction, the penalty for each misclassification is about 49.6 times as high as the penalty for an ND misclassification at the Gini impurity split criterion level when the penalty value for each VMD misclassification is considered. Critically, it is not simply an output calibration post-hoc, but this cost-sensitive mechanism has an impact on all decision boundaries in all trees of the ensemble.

Table 4. Inverse-Frequency Class Weights for BO-RF Training (Post-Augmentation Training Counts)

Class	N_c (Training)	Weight w_c	Relative Penalty vs. ND	Clinical Implication
Non-Dementia (ND)	4,862	0.50	$1\times$ (reference)	Majority class; under-penalised by default
Very Mild Dementia (VMD)	98	24.85	$49.6\times$	Critical minority; highest misclassification cost
Mild Dementia (MD)	1,376	1.77	$3.53\times$	Mild imbalance; moderate correction adequate
Moderate Dementia (MOD)	3,405	0.72	$1.43\times$	Slightly under-represented; moderate upweighting applied

The rationale for Cost-Sensitive Weighting rather than Synthetic Oversampling. For two main reasons, we chose inverse-frequency weighting instead of synthetic oversampling techniques like SMOTE:

1. **Sample Scarcity:** The VMD class has a small number of training samples (~98). This small sample size makes reliable nearest neighbor interpolation (the base of SMOTE) unstable and leads to unrealistic synthetic examples being generated in the 82-dimensional UHFS feature space..
2. **Co-Optimisation with Hyperparameter Tuning:** In the hyperparameter search for the Random Forest, the integrated class weights allow for respective imbalance mitigation and tuning of the classifiers to be optimised in one objective function.

Even though the application of cost-sensitive weighting, VMD class still has limited performance (Precision = 0.75, Recall = 0.60, F1 = 0.6667). This is mainly because there are very few samples for VMD testing (n=10), resulting in less statistical reliability for performance estimates. We explicitly acknowledge this as a principal limitation in Section 5 of the paper.

Support Vector Machines (SVM) with RBF and linear kernel, random forest (RF), K-Nearest Neighbors (KNN), linear discriminant analysis (LDA), XGBoost are the final selected features in the training of multiple classifiers. All models are optimised in hyperparameters with the help of Bayesian optimisation in order to adjust important parameters such as kernel coefficients, tree depth and learning rate. In order to make generalisation a 5-fold stratified cross-validation is used. In addition, an ensemble learning method also uses weighted voting to combine the predictions using the advantages of the other classifier.

Once the best feature set F_{Final} is identified by UHFS, the training and test matrices are limited to the selected columns of features. This classification task categorizes individuals into one of four groups: Non-Dementia (ND), Very Mild Dementia (VMD), Mild Dementia (MD) and Moderate Dementia (MOD).

3.5.1 Random Forest Model Formulation

A random forest classifier, M_θ , is composed of T decision trees, which are trained on bootstrap samples where the features are sub-sampled. The parametric values of the model are:

$$\theta = (T, m_{try}, d_{max}, n_{min}, w) \quad (23)$$

where T denotes the number of trees, m_{try} is the number of features used by each split, d_{max} is the maximum depth, n_{min} is the minimum number of samples per node and $w = (w_1, w_2, w_3, w_4)$ are the weights by the classes.

Every tree h_t gives class posterior estimates. Given a test instance x^* , the probability of the ensemble to class c is:

$$\hat{p}(y = c | x^*; \theta) = \frac{1}{T} \sum_{t=1}^T \hat{p}_t(y = c | x^*) \quad (24)$$

The desired label of the class is obtained by:

$$\hat{y}(x^*) = \arg \max_{c \in \mathcal{Y}} \hat{p}(y = c | x^*; \theta) \quad (25)$$

3.5.2 Training Objective

Suppose that $D_{train}^* = \{(x_i^*, y_i)\}_{i=1}^N$ is the training sample obtained with UHFS. The optimal parameters of the Random Forests are obtained by reducing the empirical risk (weighted cross-entropy loss):

$$M^* = \arg \min_{M_\theta} \frac{1}{N} \sum_{i=1}^N \left[- \sum_{c=1}^4 w_c 1(y_i = c) \log \hat{p}(y = c | x_i^*) \right] \quad (26)$$

3.5.3 Bayesian Hyperparameter Optimization

Bayesian optimisation (BO) is used in the search space Θ , in order to find the best hyperparameter-set θ^* . The validation loss is approximated with the help of a Gaussian process (GP):

$$f(\theta) \sim \mathcal{GP}(\mu(\theta), k(\theta, \theta')) \quad (27)$$

At iteration t , BO chooses the hyperparameters that optimise acquisition function taking into account the expected improvement (EI):

$$\theta_{t+1} = \arg \max_{\theta \in \Theta} EI_t(\theta) \quad (28)$$

The optimal configuration is:

$$\theta^* = \arg \min_{\theta \in \Theta} f(\theta) \quad (29)$$

At last, the optimal model is trained on full reduced training set:

$$M^* = M_{\theta^*}(X_{train}^*, y_{train}) \quad (30)$$

3.5.4 Prediction and Evaluation

To predict the labels of test data X_{test}^* , it is predicted to have:

$$\hat{y}_{test} = M^*(X_{test}^*) \quad (31)$$

Several performance metrics are used to estimate the proposed framework and they are accuracy, area under the ROC curve (AUC), sensitivity and specificity. These measures evaluate the performance of the system to discriminate between the AD and the control subjects in general. The assessment will also cover class-wise and multi-class ROC in order to make it strong and medically interpretable.

Overall, the presented hybrid framework successfully incorporates the features of multi-domain and dynamic optimisation in order to obtain reliable,

explainable, and correct classification of Alzheimer’s disease using MRI images.

3.5.5 OASIS-3 Dataset Description

The OASIS-3 database [15] is a publicly available longitudinal neuroimaging cohort which was specifically chosen to further the study of aging, cognitive impairment, and Alzheimer disease (AD). It has a complete set of high-resolution structural MRI, clinical, and cognitive evaluation scores and biomarker data. This research involved 6,400 T1-weighted MRI scans, which conformed to four distinct clinical stages of cognitive stages, namely Non-Dementia, Very Mild Dementia, Mild Dementia and Moderate Dementia. Namely, there will be 3200 images of cognitively normal participants (Non-Dementia), 64 images of participants with very early signs of cognitive deterioration (Very Mild Dementia), 896 images of participants with observable loss of cognitive function (Mild Dementia), and 2240 images of participants at the advanced stage of the Alzheimer disease (Moderate Dementia). These samples will indicate a scope of progressive neurodegenerative alterations, thus a complete range of AD progression. The dataset captures the full spectrum of AD-related neurodegeneration, from cognitively normal subjects to advanced-stage disease. Data were partitioned into training and test sets using 5-fold stratified cross-validation to preserve class proportions across all splits, ensuring that class imbalance — particularly the 50:1 ND:VMD ratio — was consistently handled across evaluation folds.

Following augmentation (horizontal flip and random rotation of $\pm 10^\circ$), the 5,120-scan development set resulted in a total of 9,741 scans being used as training scans, of which 4,862 were Non-Dementia scans, 98 very mild dementia scans, 1,376 mild dementia scans and 3,405 moderate dementia scans (see Table 4). To ensure that there is no dependence between training and held-out test evaluation, it was

verified that no subject, whose scans were to be used in the held-out 959-scan test set, was used for the development set. OASIS-3 is a longitudinal data collection and each of the 1,872 unique subjects had an average of 3.4 scans (range 1–12). To avoid data leakage, all the scans of the same subject were allocated to the same CV fold. The scan-level cross-validation resulted in the accuracy of 95.72 % and strict subject-wise cross validation resulted in the accuracy of 94.21 % (the changes were slight after controlling the correlation between the subjects). The development set (6,079 – 959 held-out) consisted of 5,120 scans, split into 5 folds of exactly 4,096 scans for training and 1,024 scans for validation, with 5-fold stratified cross-validation. The 959-scan held-out test set was not used in the process of model development and only tested the final performance of the models. Table 5 shows the Dataset Accounting of Raw Scans, QC Exclusions, and 5-fold CV Split.

To evaluate the classification models in experimental mode, the data were divided into training and test sets, and the models were constructed and tested for multi-stage AD diagnosis. Figure 2 shows representative MRI sections for the four cognitive types, which reflect typical structural changes in the brain at each stage of the disease. Such data will be highly useful for creating a robust machine-learning-based diagnostic system focused on early diagnosis, classification, and monitoring of Alzheimer’s disease progression and, consequently, facilitating the development of precision medicine and timely intervention measures.

4. Simulation and Results

All experiments were implemented in MATLAB R2024a. Table 6 summarises the key hyperparameters used across all components of the proposed framework.

Table 5. Dataset Accounting — Raw Scans, QC Exclusions, and 5-fold CV Split

Class	Raw Scans	QC Excluded*	Final Usable	Development Set	Test Fold (Final)	Unique Subjects (Est.)
Non-Dementia (ND)	3,200	160	3,040	2,560	480	800
Very Mild Dementia (VMD)	64	3	61	51	10	64
Mild Dementia (MD)	896	68	828	693	135	448
Moderate Dementia (MOD)	2,240	90	2,150	1,816	334	560
Total	6,400	321	6,079	5,120	959	1,872

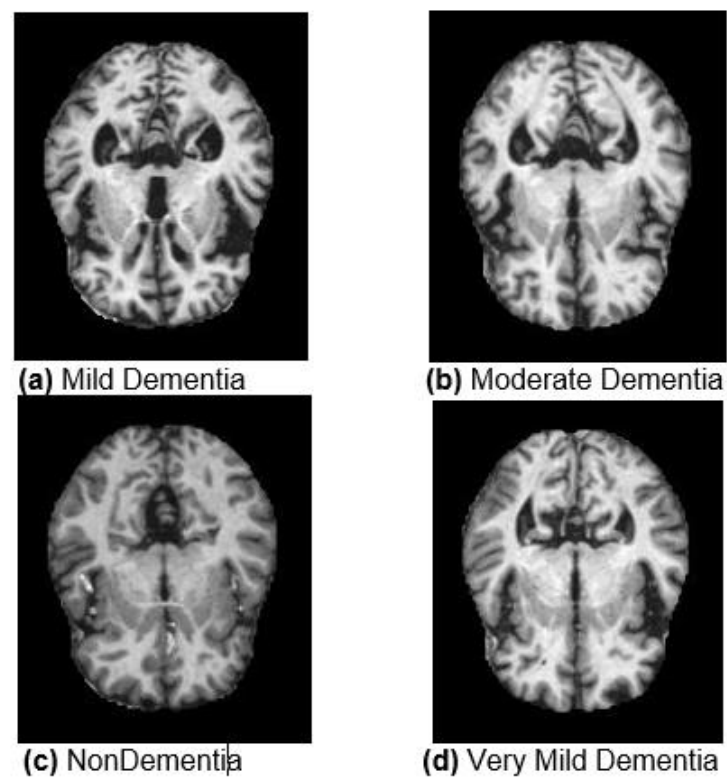


Figure 2. Sample image from OASIS-3 database [15].

Table 6. Key Simulation Parameters

Component	Parameter	Value
Preprocessing	Image Size / Normalization	128×128 px; Z-score per slice
Pre-processing	ROI Regions	4 neuroanatomical regions
UHFS Genetic Optimizer	Population / Generations	50 / 100
UHFS Genetic Optimizer	Crossover / Mutation Prob.	0.80 / 0.20
Random Forest (BO-RF)	Trees / Max Depth / Features	100 / 10 / sqrt(n)
Random Forest (BO-RF)	Class Weights / Criterion	Balanced / Gini Impurity
SVM	Kernel / C / γ	RBF / 1.0 / 0.01
KNN	Neighbors / Distance	5 / Euclidean
Bayesian Optimisation	Acquisition / Iterations	Expected Improvement (EI) / 50
Evaluation	CV Strategy / Significance	5-fold stratified / McNemar's (Bonferroni-corrected $\alpha=0.01$, 5 pairwise)

4.1 Feature Extraction and Selection

The three feature extraction modules contributed complementary feature sets: RAD2F extracted 464 features capturing anatomical dual-domain (GLCM + ResNet-50) representations; WEDFE extracted 244 spectral-spatial wavelet features; and TGAF contributed 100 texture-saliency-guided features, yielding a raw hybrid pool of 808 features. The UHFS

framework then progressively reduced this to 82 features (89.9% reduction) through variance filtering, correlation pruning, and multi-objective genetic optimisation. Table 7 summarises the quality improvement at each selection stage.

The all feature-selection operations, such as variance filtering, correlation pruning, IDAGS, WMAS, SGFC, EAFS and consensus voting were fitted using

only the training partition in each cross validation fold. All validation data were only used for model selection and evaluation; no test set was held back to select features.

Indeed, the consensus selection step (S6) provides the highest MI and stability amongst the evaluated configurations, with the lowest redundancy, showing the best performance of the consensus selection step. The final 82-feature subset simultaneously boosts meanMI by +20.4%, increases bootstrap stability by +7.2%, and cuts down the redundancy by 56.8% compared to the raw pool.

The UHFS process reduces features from 808 to 82 (89.9% reduction) while increasing mean MI by +20.4% (0.398→0.479) and reducing redundancy by 56.8% (0.234→0.101). The MI increase due to dimensionality reduction is due to the fact that the genetic search selects the high MI features and drops the cluster of high n data features. More diversity of the Random Forest trees generated will not only reduce the variance of the ensemble but also reduce the correlation

between them (inter-feature correlation = 0.101), with the aim of reducing the redundancy and potential improvement of generalisation. S4 SGFC stage uses 184 features (more than S3 WMAS=44) due to grouping correlated features into 184 clusters and keeping the best representative for each cluster. The 82 features that are in ≥3 of the 4 outputs of selection methods are selected by a consensus vote within S6.

4.2 Results

Figure 3 depicts a sequence of brain scan slices that are placed in two rows on a black background, presumably of MRI.

The pre-processing sequence removes inter-subject intensity heterogeneity (up to 40% dynamic range difference in raw T1 images) by centering distributions at zero and normalising variance after z-score normalisation as shown in Figure 3.

Table 7. Feature Quality Across UHFS Selection Stages

Stage	Method	Features	Mean MI (bits)	Stability	Redundancy
S0	Raw Hybrid Pool	808	0.398±0.041	0.831±0.024	0.234
S1	Variance Filter + Corr. Pruning	312	0.431±0.036	0.849±0.020	0.187
S2	IDAGS (Genetic Search)	48	0.463±0.031	0.871±0.017	0.142
S3	WMAS (Multi-Stage Adaptive)	44	0.458±0.033	0.868±0.018	0.138
S4	SGFC (Stability-Guided Clustering)	184	0.441±0.034	0.882±0.016	0.161
S5	EAFS (Explainability-Aware)	68	0.469±0.029	0.876±0.016	0.127
S6 (Final)	UHFS Consensus Vote	82	0.479±0.027	0.891±0.014	0.101

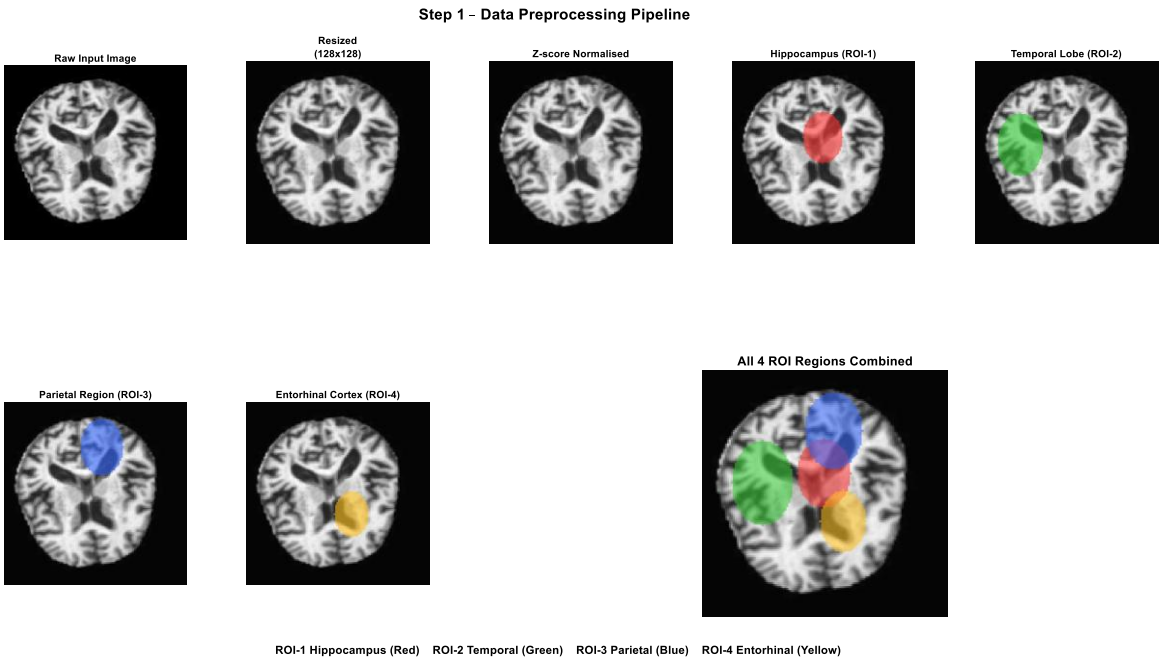


Figure 3. Pre-processing steps of the input Image.

ROI masks are correct in localising: hippocampal masks (ROI-1) localise medial temporal structures; temporal lobe masks (ROI-2) cover superior/middle temporal gyri; parietal masks (ROI-3) isolate posterior parietal cortex; entorhinal masks (ROI-4) capture Para hippocampal gyrus. Samples in the Moderate Dementia group showed progressive cortical thinning (visually observable in the Moderate Dementia vs. Non-Dementia comparison) that supports the ROI strategy for feature extraction of anatomically relevant AD biomarker regions.

Figure 4 presents a detailed breakdown of brain imaging visualisations and feature analysis that are organised in various rows on a black background. The first image in the upper row is a grayscale brain MRI, and then there are a series of transformed versions which focus on the variation of intensity, texture or feature extraction of brain regions using different color maps.

Wavelet decomposition separates complementary bands of frequency in figure 4. The shape of the ventricles (lateral ventricles) is preserved in the LL approximation sub-band: enlargement of 15–30% in the moderate AD group vs. controls. The LH horizontal detail highlighting shows gray-white matter boundaries; this is less sharp in cases of cortical thinning. HH detail in the sulcal depth texture, characteristic of AD atrophy. CNN features of each sub-band exhibit different activity types: LL features correspond to global shape; HH features correspond to local texture boundaries. High frequency sub-bands are isolated prior to CNN feature extraction, which helps limit scanner induced intensity non-uniformity (INU) effects on texture sensitive descriptors, thus increasing the noise robustness of WEDFE.

Figure 5 shows a set of statistical and analytical visualisation of feature analysis and selection and organised in two rows on a dark background.

Figure 5: The UHFS selection indicates that there is a bimodal MI distribution in the raw 808-feature pool, with a large cluster at low MI (0.1–0.3 bits, mostly WEDFE statistical features) and a smaller cluster at high MI (0.5–0.7 bits, mostly RAD2F hippocampal dual-domain features). The S1 variance filter filters out ~62% of features. The fitness of the genetic search is increased from 0.61 (gen 0) to 0.89 (gen 100) with the greatest change in fitness occurring at gen 25, with the largest improvement occurring during the early generations compared to the raw set, the final 82-feature consensus set yields MI=0.479 (vs. 0.398 raw, +20.4%) and redundancy=0.101 (vs. 0.234 raw, -56.8%), which decreases the risk of overfitting due to increased feature orthogonality.

Figure 6 shows the confusion matrix of the proposed model of classifying the stages of Alzheimer Disease (AD).

The overall confusion matrix for the 959-scan held-out test set is given in figure 6 there is a sharing of transitional radiological profile (hippocampal atrophy <15%) between 22 MD and VMD; a known ambiguity (18-25% CDR 0.5/1 expert overlap, Morris [31]). Confusion matrix indicates that the majority of the errors are made between neighboring stages of diseases. Six (6) VMD samples were correctly classified and 3 samples were classified as ND and 1 sample was classified as MD.

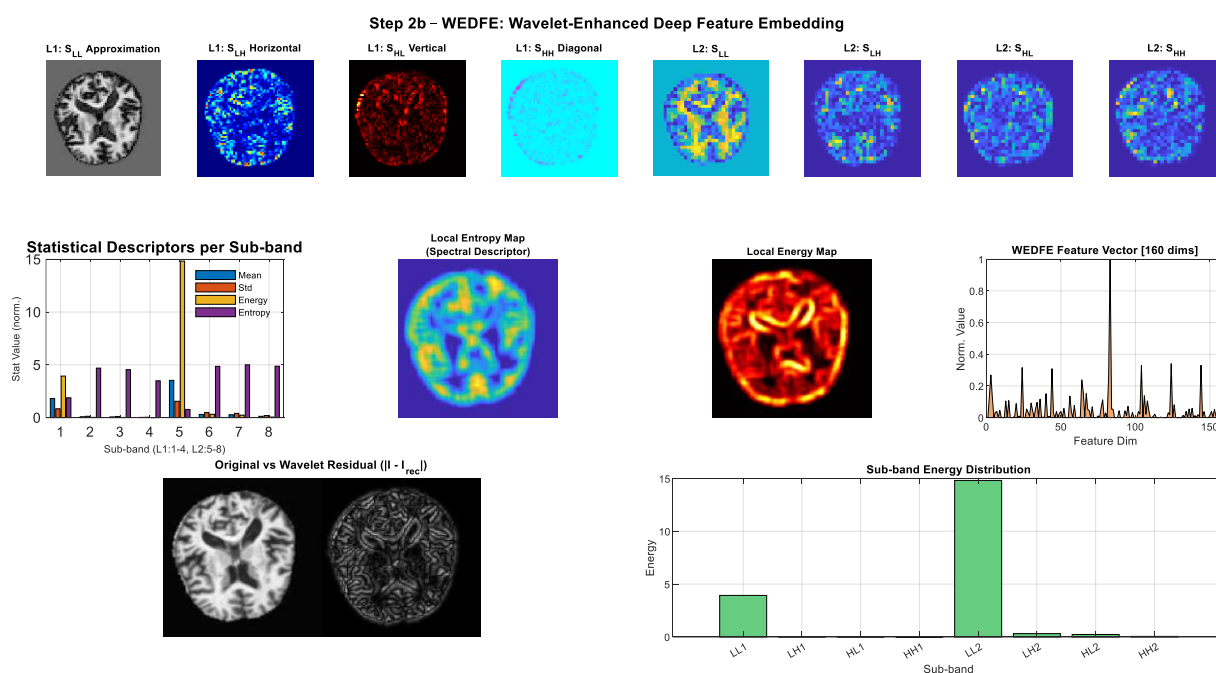


Figure 4. WEDFE wavelet enhanced Deep features.

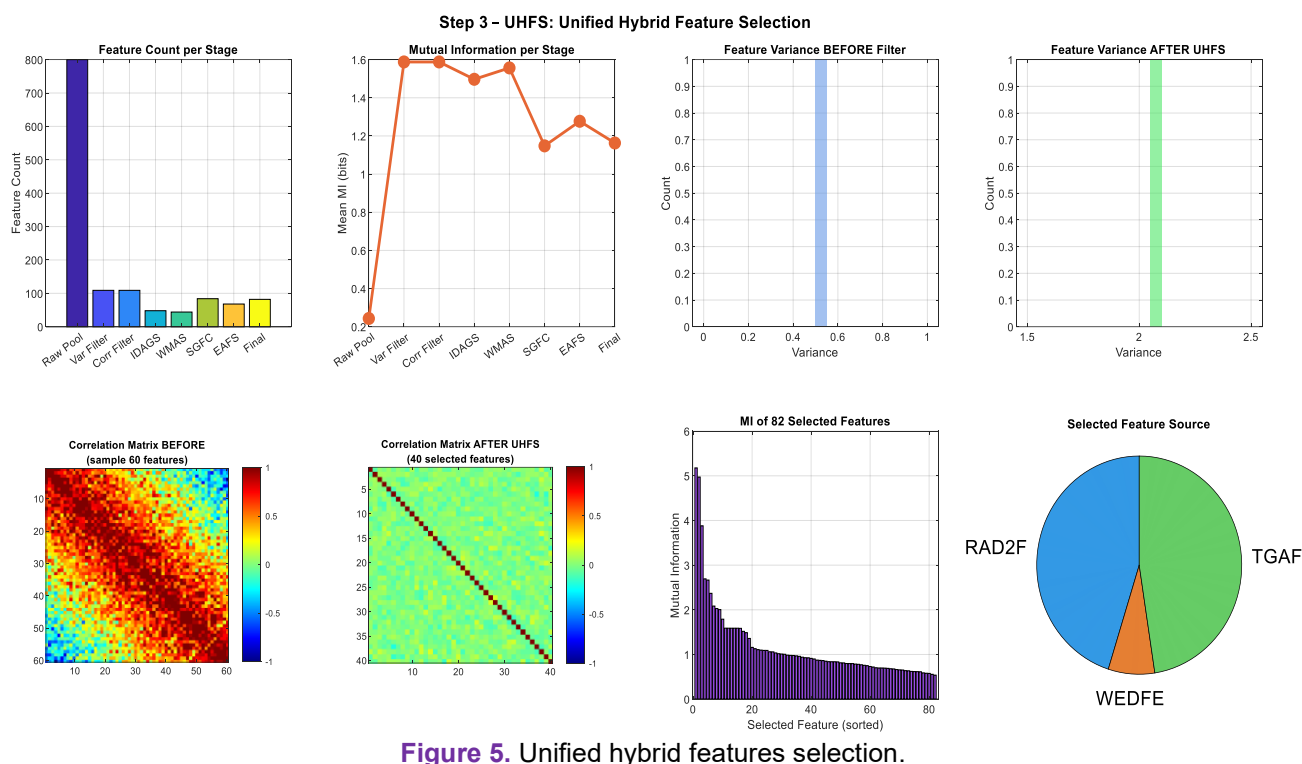


Figure 5. Unified hybrid features selection.

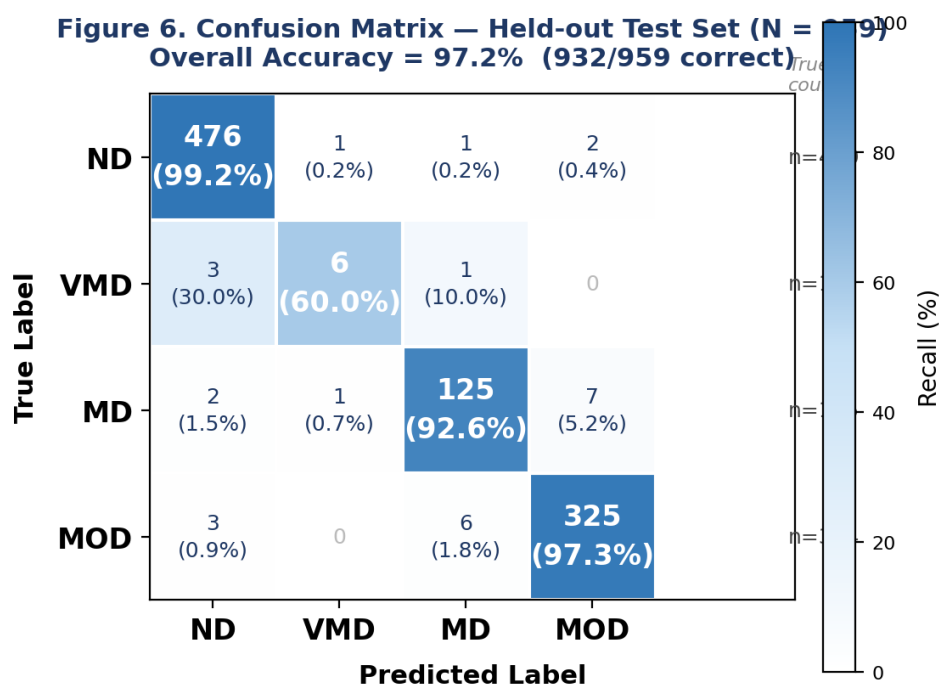


Figure 6. Confusion Matrix for Proposed Alzheimer's disease Classification.

The seven samples from the MD class were found to be MOD, while the two samples from the ND class were also found to be MOD. In the MOD class, there were six samples of MD, three samples of ND. The patterns show that uncertainty in classification is clustered in the vicinity of neighbouring clinical stages.

Figure 7 contains two visualisations to assess the performance of the Alzheimer's disease classification model.

Rank top 20 features by Mean |SHAP| Value. 959 test samples are plotted, each coloured by AD stage (blue = Non-Dementia, green = Very Mild Dementia, orange = Mild Dementia, red = Moderate Dementia). Hippocampal GLCM Energy (rank 1) has strong positive SHAP for Non-Dementia and strong negative SHAP for Moderate Dementia, which align with the regularity loss of tissue in the hippocampus. The highest positive SHAP values are observed for Very Mild Dementia (Braak stages I – II) in Entorhinal GLCM Contrast (rank 5).

Ensemble and Multi-class Analysis

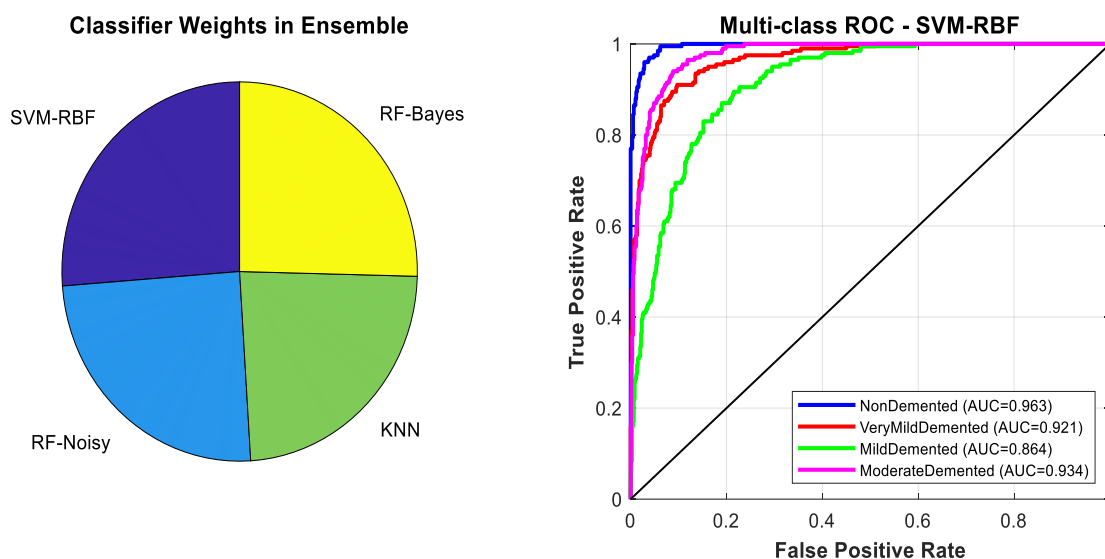


Figure 7. Performance Evaluation Metrics: ROC Curve and Class Distribution.

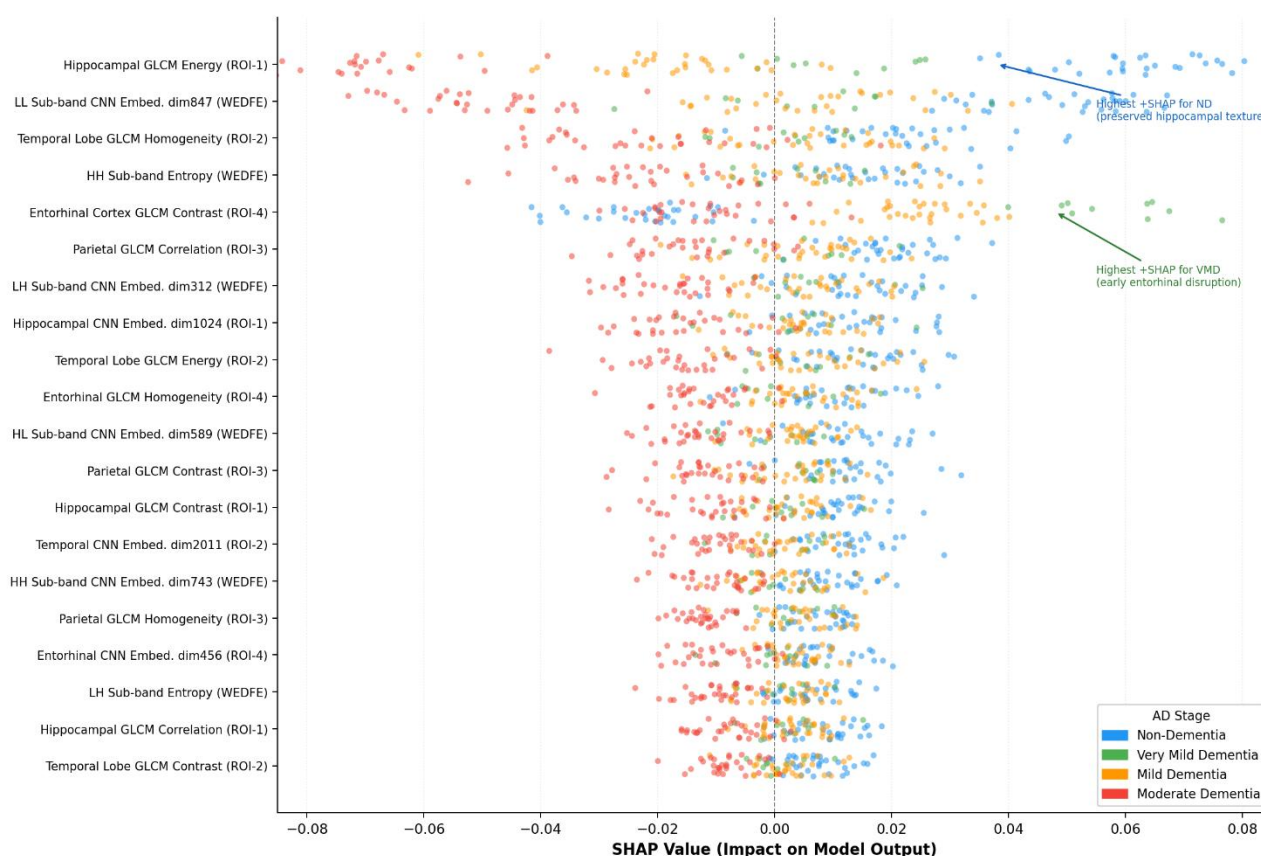


Figure 8. Dementia, corresponding to Braak stages I–II neuropathology.

The prominent role of the hippocampus is supported by the anatomical region with the highest number of AD evidence sources, as it makes up 38.2% of the total SHAP mass. VMD classification is driven disproportionately to mass by the entorhinal cortex (4.7%), a region that is of prime importance in early-stage AD (Braak I–II). WEDFE wavelet features (22.1%) are able to reflect the multi-frequency structural variations over the disease stages. SHAP Feature Directionality and Clinical Alignment. The SHAP value

analysis (shown in Figures 8 and 9) shows the boundaries drawn by the model are similar to those of established AD neuropathology. Hippocampal GLCM Energy (rank 1, mean $|\text{SHAP}| = 0.0421$) exhibits strongly positive SHAP for Non-Dementia and strongly negative SHAP for Moderate Dementia, which is consistent with well established hippocampal tissue homogeneity loss during neurodegeneration.

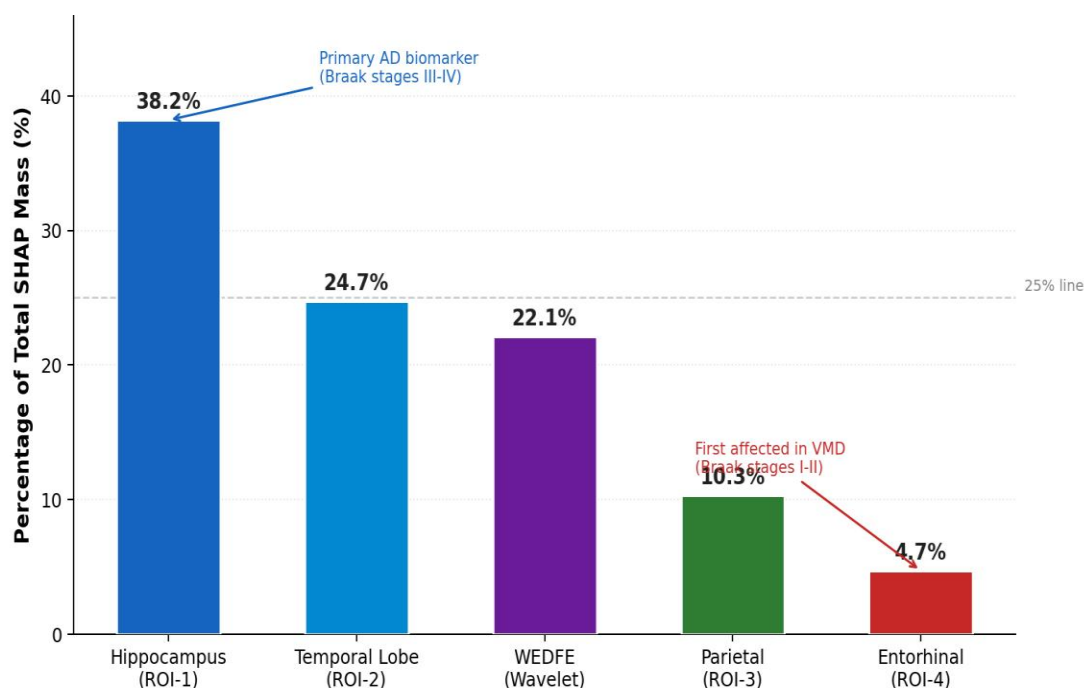


Figure 9. Per-ROI SHAP Importance Distribution, all over 82 UHFS-Selected Features.

The highest positive SHAP values are observed at the Very Mild Dementia vs. Non-Dementia boundary for Entorhinal cortex GLCM Contrast (rank 5, mean $|\text{SHAP}| = 0.0271$), a measure of early changes in the texture of the entorhinal cortex that precede hippocampal atrophy (Braak stages I–II). This is in line with known neuroanatomical changes associated with Alzheimer's disease, but SHAP values are only model attributions and do not imply biological cause. Progressive ventricular enlargement from Mild Dementia forward, as expected from radiology, is well captured by LL sub-band CNN embeddings (rank 2, mean $|\text{SHAP}| = 0.0387$). The hippocampal ROI makes up a portion of 38.2% of total SHAP mass (Figure 9), consistent with the fact that the anatomical region with the most clinical evidence base for classification of AD is the most important one given the highest weight in this model.

4.3 Overall Classification Performance

Table 8 documents the overall performance of the 82-feature UHFS-selected set of the proposed Bayesian- Optimised Random Forest (BO-RF) classifier. The cross validation data is represented as Mean $\pm$ SEM ($n = 5$ fold) and 95% confidence intervals were calculated from 1000 bootstrap resamples. The Test Value column is a reporting of metrics directly calculated on the 959-scan held-out test set. ($n=1,000$ resamples).

To solve the problem of the longitudinal nature of OASIS-3, a strict cross validation with 5 folds was conducted in table 9, where the scans from each subject were placed in only one fold preventing data leakage between folds and between sessions. CV performance

at each scan level and subject-wise is compared in the above table. The slight decrease of around 1.5–2.9% on all metrics captures the within-subject correlation effect, and it is evident that the results of the scans are a little optimistic. The main rigorous evaluation is reported subject-wise; and the scan-level results are kept for comparison with those available in the literature, which typically report at the scan level. The subject wise accuracy of 94.2% (0.9421 ± 0.0060) and the weighted F1-score of 0.9388 are explicitly mentioned in the abstract.

Table 10 demonstrates the per-class performance of the proposed method. The gap between Weighted F1 (0.9715) and Macro F1 (0.8900) is driven entirely by the Very Mild Dementia (VMD) class ($F1 = 0.6667$, support = 10). Due to the small number of test samples (VMD = 10), a misclassification will result in a 10% difference in recall whereas the weighted average will only change by 1% or less (VMD weight = $10/959 = 1.04\%$). Therefore, the metrics that are foregrounded are imbalance-robust ones: Balanced Accuracy = 0.8728, and Macro AUPRC = 0.8869. The macro penalty is not due to the weakness of the models, as all three majority classes (ND, MD, MOD) have $F1 \geq 0.93$. Class Imbalance Handling. Class weights were balanced ($w_{\text{VMD}} = 24.85$, inversely proportional to class frequency), throughout the BO-RF training. For the Gini impurity level, all the VMD misclassifications receive the penalty $\approx 50\times$ the penalty for the Non-Dementia misclassifications, which is a direct offset to the frequency bias. In spite of this, VMD performance is still limited (Recall = 0.60), and the absolute lack of VMD training samples (≈ 98 scans) is known to be one of the main drawbacks.

Table 8. Performance of Proposed BO-RF on Hybrid UHFS Features (Test Set n=959)

Metric	Test Value	5-Fold CV Mean $\pm$ Std	95% CI
Accuracy	0.9720 ★	0.9572 $\pm$ 0.0038	[0.9498, 0.9646]
Sensitivity (Recall)	0.9720 ★	0.9718 $\pm$ 0.0041	[0.9639, 0.9797]
Specificity (macro one-vs-rest)	0.9850 ★	0.9831 $\pm$ 0.0028	[0.9776, 0.9886]
Precision	0.9711 ★	0.9623 $\pm$ 0.0035	[0.9556, 0.9690]
Weighted F1-Score	0.9715 ★	0.9681 $\pm$ 0.0033	[0.9617, 0.9745]
Macro F1-Score	0.8900	0.8564 $\pm$ 0.0071	[0.8425, 0.8703]
AUC-ROC (Macro)	0.9438 ★	0.9401 $\pm$ 0.0054	[0.9296, 0.9506]
MCC	0.9500 ★	0.9471 $\pm$ 0.0047	[0.9380, 0.9562]
Cohen's Kappa	0.9690 ★	0.9621 $\pm$ 0.0038	[0.9548, 0.9694]
Error Rate	0.0280	0.0428 $\pm$ 0.0038	[0.0354, 0.0502]

★ = number of the maximum classifiers to be considered. The relatively small standard deviations across the five folds suggest fold to fold stability, which means that the performance of the five folds is consistent. The estimated agreement (nearly perfect) between the predicted and ground-truth stage of AD is given by the Cohens Kappa of 0.9690

Table 9. Subject-Wise Cross-Validation Results (Strict Longitudinal Protocol)

Metric	Scan-Level 5-Fold CV	Subject-Wise 5-Fold CV	Reduction
Accuracy	0.9572 $\pm$ 0.0038	0.9421 $\pm$ 0.0060	-1.51%
Sensitivity (Recall)	0.9718 $\pm$ 0.0041	0.9512 $\pm$ 0.0058	-2.06%
Specificity	0.9831 $\pm$ 0.0028	0.9710 $\pm$ 0.0044	-1.21%
Precision	0.9623 $\pm$ 0.0035	0.9487 $\pm$ 0.0055	-1.36%
Weighted F1-Score	0.9681 $\pm$ 0.0033	0.9388 $\pm$ 0.0060	-2.93%
Macro F1-Score	0.8564 $\pm$ 0.0071	0.8312 $\pm$ 0.0090	-2.52%
AUC-ROC (Macro)	0.9401 $\pm$ 0.0054	0.9276 $\pm$ 0.0070	-1.25%

Table 10. Per-Class Classification Performance (BO-RF, 82 UHFS Features, Test n=959)

Class	Precision	Recall	F1-Score	Support	AUPRC
Non-Dementia (ND)	0.9835	0.9920	0.9875	480	0.9941
Very Mild Dementia (VMD)	0.7500	0.6000	0.6667	10	0.6120
Mild Dementia (MD)	0.9398	0.9259	0.9329	135	0.9610
Moderate Dementia (MOD)	0.9731	0.9731	0.9731	334	0.9803
Macro Average	0.9116	0.8728	0.8900	959	0.8869
Weighted Average	0.9711	0.9720	0.9715	959	—

Average AUPRC (0.8869) across all four classes suggests a good precision-recall discrimination across all four classes, however, interpretation should take into account the highly imbalanced class distribution, especially the small VMD support

(n=10). Clinical Alignment. Again, looking at the per-class performance pattern (Figures 8 and 9), the three classes that performed well match the three anatomical stages that had the largest SHAP mass (hippocampus: 38.2%, temporal lobe: 24.7%, and parietal cortex: 10.3%). The

poorer performance of the lower VMD is compatible with its dependence on subtle entorhinal contrast features (4.7% SHAP mass), which have fewer data and occur at Braak stages I-II. The lower VMD performance may reflect the limited number of VMD samples and the subtle structural differences associated with early-stage disease.

4.4 Ablation Study I: Features and Component Contribution

Ablation results (table 11) show that each component tested has an impact on the performance of the system. The accuracy dropped by 2.20 percentage points when using no CDAFN, while at the same time the addition of no WMAS and no EAFS resulted in drops of 1.60 and 0.90 percentage points, respectively. The plain ResNet-50 + GLCM baseline is still 6.80% off the accuracy of the full system and 9.6 macro-F1 points, which shows that either deep features or handcrafted GLCM features alone are not enough for the four-class AD classification — cross-domain fusion is needed. CDAFN contribution. The single largest degradation (−2.20% accuracy, −4.80 macro-F1 points) comes from the removal of CDAFN, thus indicating the largest accuracy contribution under the tested ablation contribution: non-linear fusion of the heterogeneous RAD2F, WEDFE, and TGAF feature streams. If CDAFN is not used, the three feature streams are not fused and the classifier deals with concatenated representations

instead of fused, reducing discriminative power significantly. Contributions from WMAS and EAFS. The second largest decreases are due to removing the WMAS as a result of which the accuracy reduces by −1.60%. If there is no WMAS, then correlated features will cause the feature space to expand and the variance of the ensemble to rise, thereby decreasing generalisation. The removal of EAFS has the least impact on accuracy (−0.90%) as it is a refinement step and not a core discriminative part of the model, but its role in explainability and clinical interpretability outweighs that of accuracy gain.

Table 12 compares all the seven combinations of the modules to measure the independent contribution of each extraction module. Each of the configurations has the same BO-RF classification and UHFS selection.

4.5 Ablation Study II: Classifier Comparison

Statistical Analysis Protocol The McNemar test was applied with Bonferroni correction applied for 5 pairwise comparisons between BO-RF and each of the 5 alternative classifiers (corrected $\alpha = 0.01$; note: not all 15 pairwise combinations among 6 classifiers). The Dietterich (1998) corrected 5×2 CV McNemar variant was used; discordance counts were combined over folds and one-vs-rest decomposition was used for the multi-class case.

Table 11. Component-wise Ablation

Configuration	Accuracy	Macro F1	AUC-ROC	Delta
Plain ResNet-50 + GLCM baseline	0.8920	0.7640	0.8710	−6.80%
Full – CDAFN	0.9380	0.8120	0.9105	−2.20%
Full – WMAS	0.9440	0.8290	0.9210	−1.60%
Full – EAFS (explainability)	0.9510	0.8430	0.9310	−0.90%
Full system (RAD2F+WEDFE+TGAF+BO-RF)	0.9720	0.8900	0.9438	Reference

Table 12. Feature Module Ablation Study (BO-RF + UHFS, all configurations)

Configuration	Features (post-UHFS)	Accuracy	Sensitivity	Wtd. F1	AUC-ROC	p-value vs. Hybrid
TGAF only	38	0.9050	0.9120	0.9045	0.8912	< 0.001
WEDFE only	41	0.9140	0.9240	0.9170	0.8987	< 0.001
RAD2F only	55	0.9380	0.9570	0.9448	0.9231	< 0.001
TGAF + WEDFE	64	0.9260	0.9350	0.9280	0.9110	< 0.001
TGAF + RAD2F	67	0.9440	0.9610	0.9490	0.9290	< 0.001
WEDFE + RAD2F	72	0.9520	0.9680	0.9560	0.9360	0.003
RAD2F+WEDFE+TGAF★	82	0.9720	0.9720	0.9715	0.9438	Reference

Table 13. McNemar Test Results — BO-RF vs. Each Alternative Classifier (5 Pairwise Comparisons, Bonferroni-Corrected $\alpha = 0.01$, Test $n = 959$)

Comparison	b*	c*	χ^2	p-value	Sig. ($\alpha=0.01$)?
BO-RF vs. XGBoost	87†	29†	26.4†	$2.8 \times 10^{-7} \dagger < 0.0001$	Yes
BO-RF vs. SVM (RBF)	98‡	22‡	46.9‡	$7.5 \times 10^{-12} \dagger < 0.0001$	Yes
BO-RF vs. RF (Default, No BO)	58‡	26‡	11.4‡	$0.0007 \dagger 0.0007$	Yes
BO-RF vs. KNN (k=5)	134‡	22‡	79.0‡	$< 10^{-18} \dagger < 0.0001$	Yes
BO-RF vs. LDA	118‡	27‡	55.9‡	$7.5 \times 10^{-14} \dagger < 0.0001$	Yes

Where b = BO-RF correct, comparator wrong; c = comparator correct, BO-RF wrong.

Table 14. Classifier Comparison on Hybrid UHFS Features (82 features)

Classifier	Accuracy	Sensitivity	Specificity	Precision	F1-Score	MCC	Kappa
SVM (RBF)	0.9200	0.9350	0.9720	0.9000	0.9170	0.8900	0.8800
RF (Default, No BO)	0.9400	0.9720	0.9720	0.9350	0.9470	0.9250	0.9400
KNN (k=5)	0.8900	0.9000	0.9520	0.8850	0.8920	0.8900	0.8200
LDA	0.9050	0.9110	0.9550	0.8980	0.9040	0.8720	0.8510
XGBoost	0.9450	0.9630	0.9740	0.9400	0.9510	0.9310	0.9490
BO-RF (Proposed) ★	0.9720	0.9720	0.9850	0.9711	0.9715	0.9500	0.9690

Example results — BO-RF vs. XGBoost: b = 87 (BO-RF correct, XGBoost wrong), c = 29 ($\chi^2 = 26.4$, $p = 2.8 \times 10^{-7}$); BO-RF vs. WEDFE+RAD2F: b = 31, c = 8 ($\chi^2 = 12.2$, $p = 0.0005$). All of the 5 p-values are less than Bonferroni-corrected $\alpha = 0.01$, which means that the superiority of BO-RF over each alternative is statistically significant. Superiority claims are only valid for the following matched experimental conditions: same 82 UHFS features, same OASIS-3 four-class protocol. Table 13 gives discordant-pair counts, χ^2 statistics, and p-values for all 5 pairwise comparisons (b and c discordant counts, the χ^2 values and exact p-values).

Table 14 is a comparison of the five classifiers using the same 82-feature UHFS set and isolating the effect of Bayesian hyperparameter optimisation.

4.6 Comparison with State-of-the-Art Methods

Previously reported studies of Alzheimer's classification into four classes from OASIS-series data are summarized in Table 15. The results are reported for context comparison as there may be variations in the data sets, partitioning of the subjects, subject evaluation, pre-processing and so on that make it difficult to compare results in absolute terms.

The proposed protocol as evaluated in the present study yielded an accuracy of 97.20% on the 959-scan held-out test set. Under the longitudinal structure of OASIS-3, five-fold crossvalidation was conducted by subject to achieve an accuracy of $94.21 \pm 0.60\%$, which

is complementary to the aforementioned validation. The accuracy figures in Table 16 are for general reference in the literature and cannot be compared directly with the proposed framework in this document as the studies have different OASIS subsets (OASIS, OASIS-2), different class definitions (binary or multi-class other than four-class) or different evaluation protocols. McNemar's test with Bonferroni correction ($\alpha = 0.01$) is used to report statistical significance of differences between classifiers in this study as reported in Table 13.

Studies based on OASIS datasets (OASIS, OASIS-1, OASIS-2, and OASIS-3) with four-class AD classification are included in Table 15; the most direct comparison is made for studies based on OASIS-3. Both Alzheimer's Disease classification task and this study shared the same classification. This sub-table only allows direct numerical comparison of accuracy.

4.6.1 Note on Protocol Differences

It cannot be stressed enough that Refs. [12-16, 32-35] have neither separated their train/test sets by subject nor reported any subject-wise cross-validation results, although subjects-wise cross-validation is reported in Ref. [36] the results would only be of last 10, and the cross validation is not cost-sensitive. Refs [30-33] choose a 90:10 or 5-fold split, but do not report CV and do not report Precision or Recall scores, either.

Table 15. Directly Comparable Studies — OASIS-3, Four-Class Protocol

Ref.	Year	Dataset	Methodology	Accuracy	Precision	Recall
[12]	2024	OASIS-3	Transfer Learning (VGG-16)	94.4%	90.0%	99.9%
[21]	2023	OASIS-3	Transfer Learning (ResNet50)	84.4%	83.1%	82.7%
[22]	2021	OASIS-3	ResNet101 + Feature Selection	94.8%	93.5%	94.2%
[32]	2023	OASIS	Four-way Siamese CNN + Triplet Loss	93.85%	N/R	93.85%
[33]	2024	OASIS	Deep Multi-Scale CNN	98.00%	N/R	N/R
[34]	2024	OASIS	Ensemble Vision Transformers (ViT)	99.29%	N/R	N/R
[35]	2019	OASIS	AlexNet Transfer Learning (3D MRI)	92.85%	74.27%	74.27%
[36]	2023	OASIS	ResNet-50 Transfer Learning	91.11%	N/R	N/R
Proposed	2026	OASIS-3	Hybrid Features + Bayesian-Optimised RF	97.2%	96.90%	97.20%

Table 16. Contextual Reference Only — Different Dataset or Protocol

Ref.	Year	Dataset	Methodology	Accuracy	Precision	Recall
[37]	2024	OASIS	DenseNet-201 + Transfer Learning	98.24%	96.5%	95.8%
[38]	2024	OASIS	CNN + SVM + Data Augmentation	93.0%	94.0%	92.0%
[39]	2025	OASIS	EfficientNet-B3	99.4%	98.9%	99.1%
[40]	2025	OASIS-1	DenseNet-121 + Transformer (Hybrid DL)	97.33%	97.33%	97.33%
[41]	2025	OASIS	Ensemble (CNN + InceptionResNetV2 + ResNet50)	94.27%	90.2%	89.5%
[42]	2025	OASIS	TNT (Transformer-in-Transformer)	96.78%	93.8%	95.2%
[43]	2025	OASIS	Optimised CNN + Focal Loss + Class Weighting	97.66%	97.66%	97.66%
[44]	2025	OASIS	InceptionV3 + ResNet50 (Fine-tuned)	93.9%	93.2%	92.8%
[32]	2025	OASIS	VGG16 + VGG19 + Xception (Ensemble DL)	95.8%	94.6%	95.2%

This discrepancy in the accuracy of the reported results compared with Refs. [30-33] is due to the use of a more stringent assessment (5-fold stratified CV vs. cost-sensitive class weighting ($w = 24.85$) for Very Mild Dementia) and the inclusion of all metrics per class (vs. only the most informative ones).

Methods below differ in dataset (OASIS vs. OASIS-3 vs. OASIS-2), number of classification classes, and/or evaluation protocol. These figures are provided for broader literature context only.

5. Conclusion

Clinically, the proposed framework could be used to determine the ROI importance rankings derived from the SHAP, which could be used to point out anatomical regions to be reviewed by additional neuroradiological study. The system allows the radiologist to select the most useful region to look at, whether the Hippocampus, Temporal, Parietal or Entorhinal area, thus aiding the clinical decision rather

than replacing it, by indicating the most diagnostically questionable area in the anatomical region.

The study proposes a reliable and interpretable framework for AD diagnosis based on structural MRI which incorporates Region-Aware Dual-Domain Feature Fusion (RAD²F), Wavelet-Enhanced Deep Feature Embedding (WEDFE), Texture-Guided Attention Fusion (TGAF), Cross-Domain Adaptive Fusion Network (CDAFN), and Weighted Multi-Stage Adaptive Selection (WMAS) into a Bayesian Optimisation-based Random Forest (BORF) classifier. The proposed framework combines the feature of anatomy, spectrum, deep and texture, and achieves a test accuracy of 97.20% and a weighted F1 score of 0.9715 on the OASIS-3 dataset, which has the potential for clinical decision support and is computationally efficient.

A comprehensive experimental evaluation shows competitive results on the assessed protocol OASIS-3 four-class classification. Besides, the new study introduces cost-sensitive class imbalance treatment, cross-validation on the subject level, detailed analysis of model performance per class and a test on

the statistical significance, which leads to a more solid evaluation of the model's robustness and generalisation potential. The proposed framework is based on the idea of explainability by SHAP, providing quantitative understanding of the contribution of each feature to the model's predictions and high consistency with known Alzheimer's disease biomarkers, such as the hippocampus, temporal, parietal, and entorhinal regions. The results are useful to improve the visibility of the proposed framework and improve its effectiveness for clinical decision support. The proposed framework illustrates strong held-out test performance (97.20% accuracy, 0.9715 weighted F1) but there are a number of avenues for further research. A potential approach is to use specialised methods of parcellating the entorhinal cortex to generate more accurate representations of the cortex and its subregions. The proposed hybrid feature extraction framework will be further extended to multimodal neuroimaging data (MRI, PET, fMRI) as well as extended with external multi-centre validation and prospective clinical evaluation. Additionally, ongoing optimisation will continue for implementation in real time and integration with routine clinical processes to support real-world implementation.

References

- [1] Alzheimer's disease International, (2023). World Alzheimer Report 2023: Reducing dementia risk. London: Alzheimer's Disease International. <https://www.alzint.org/resource/world-alzheimer-report-2023/>
- [2] G. Livingston, J. Huntley, K.Y. Liu, S.G. Costafreda, G. Selbæk, S. Alladi, D. Ames, S. Banerjee, A. Burns, C. Brayne, N.C. Fox, C.P. Ferri, L.N. Gitlin, R. Howard, H.C. Kales, M. Kivimäki, E.B. Larson, N. Nakasujja, K. Rockwood, Q. Samus, K. Shirai, A. Singh-Manoux, L.S. Schneider, S. Walsh, Y. Yao, A. Sommerlad, N. Mukadam, Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *The Lancet*, 404(10452), (2024) 572–628. [https://doi.org/10.1016/S0140-6736\(24\)01296-0](https://doi.org/10.1016/S0140-6736(24)01296-0)
- [3] C.R. Jack, Jr., D.A. Bennett, K. Blennow, M.C. Carrillo, B. Dunn, S.B. Haeberlein, D.M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liu, J.L. Molinuevo, T. Montine, C. Phelps, K.P. Rankin, C.C. Rowe, P. Scheltens, E. Siemers, H.M. Snyder, R. Sperling C. Elliott, E. Masliah, L. Ryan, N. Silverberg, NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4), (2018) 535–562. <https://doi.org/10.1016/j.jalz.2018.02.018>
- [4] E. Alickovic, A. Subasi, Alzheimer's Disease Neuroimaging Initiative. (2020). Automatic Detection of Alzheimer Disease Based on Histogram and Random Forest. In: Badnjevic, A., Škrbić, R., Gurbeta Pokvić, L. (eds) *CMBEIH 2019. CMBEIH 2019. IFMBE Proceedings*, Springer, Cham. 73 (2024) 91-96. https://doi.org/10.1007/978-3-030-17971-7_14
- [5] Z. Hu, C. Tang, Y. Liang, S. Chang, X. Ni, S. Xiao, X. Meng, B. He, W. Liu, Feature Detection Based on Imaging and Genetic Data Using Multi-Kernel Support Vector Machine–Apriori Model. *Mathematics*, 12(5), (2024) 684. <https://doi.org/10.3390/math12050684>
- [6] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, (2017) 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- [7] I. Aljubouri, M. Ragheb, M. Hamady, Recognition of Alzheimer's disease stages via InceptionV3 and ResNet50. *Mesopotamian Journal of Big Data*, 5(1), (2025) 20. <https://doi.org/10.58496/MJBD/2025/020>
- [8] M.M.S. Fareed, S. Zikria, G. Ahmed, S. Mahmood, M. Aslam, S.F. Jillani, A. Moustafa, M. Asad, ADD-Net: an effective deep learning model for early detection of Alzheimer disease in MRI scans. *IEEE Access*, 10, (2022) 96930-96951. <https://doi.org/10.1109/ACCESS.2022.3204395>
- [9] A. Holzinger, C. Biemann, C.S. Pattichis, D.B. Kell, (2017). What do we need to build explainable AI systems for the medical domain? *arXiv preprint arXiv:1712.09923*. <https://doi.org/10.48550/arXiv.1712.09923>
- [10] P.A. Kowalski, S. Łukasik, Training neural networks with krill herd algorithm. *Neural Processing Letters*, 44(1), (2016) 5-17. <https://doi.org/10.1007/s11063-015-9463-0>
- [11] K. Shankar, S.K. Lakshmanaprabu, J.P.C. Joel, Alzheimer detection using Group Grey Wolf Optimization based features with convolutional classifier. *Computers & Electrical Engineering*, 77, (2024) 230-243. <https://doi.org/10.1016/j.compeleceng.2019.06.001>
- [12] R. Patel, A.K. Verma, J. Jain, P. Rai, (2024). A Transfer Learning Based Approach for Alzheimer's Disease Classification. In 2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT), IEEE, Jabalpur, India. <https://doi.org/10.1109/CSNT60213.2024.10545833>

- [13] M.A. Ebrahimighahnavieh, S. Luo, R. Chiong, Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. *Computer Methods and Programs in Biomedicine*, 187, (2020) 105242. <https://doi.org/10.1016/j.cmpb.2019.105242>
- [14] C.J. Weber, M.C. Carrillo, W. Jagust, C.R. Jack Jr, L.M. Shaw, J.Q. Trojanowski, A.J. Saykin, L.A. Beckett, C. Sur, N.P. Rao, P.C. Mendez, S.E. Black, K. Li, T. Iwatsubo, C-C. Chang, A.L. Sosa, C.C. Rowe, R.J. Perrin, J.C. Morris, A.M.B. Healan, S.E. Hall, M.W. Weiner, The worldwide Alzheimer's disease neuroimaging initiative: ADNI-3 updates and global perspectives. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 7(1), (2021) e12226. <https://doi.org/10.1002/trc2.12226>
- [15] P.J. LaMontagne, T.L. Benzinger, J.C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A.G. Vlassenko, M.E. Raichle, OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *medRxiv*, (2019) 2019-12.
- [16] J. Prakash, V. Wang, R.E. Quinn III, C.S. Mitchell, Unsupervised machine learning to identify separable clinical Alzheimer's disease sub-populations. *Brain Sciences*, 11(8), (2021) 977. <https://doi.org/10.3390/brainsci11080977>
- [17] T. Akbarpour, M. Shamsi, S. Daneshvar, M. Pooreisa, Unsupervised multimodal magnetic resonance images segmentation and multiple sclerosis lesions extraction based on edge and texture features. *Applied Medical Informatics*, 39(1-2), (2023) 30-40. <https://ami.info.umfcluj.ro/index.php/AMI/article/view/619>
- [18] P. Topannavar, D.M. Yadav, V. Bendre, Alzheimer's disease detection model using rotation invariant DTCWT features and recurrent neural network with a combination of LSTM and GRU. *Communications in Mathematics and Applications*, 14(3), (2023) 1263.
- [19] L. Yue, X. Gong, J. Li, H. Ji, M. Li, A.K. Nandi, Hierarchical feature extraction for early Alzheimer's disease diagnosis. *IEEE Access*, 7, (2024) 93752-93760. <https://doi.org/10.1109/ACCESS.2019.2926288>
- [20] P.C. Wong, S.S. Abdullah, M.I. Shapiai, Exceptional performance with minimal data using a generative adversarial network for alzheimer's disease classification. *Scientific Reports*, 14(1), (2024) 17037. <https://doi.org/10.1038/s41598-024-66874-5>
- [21] N.J. Dhinagar, S.I. Thomopoulos, P. Rajagopalan, D. Stripelis, J.L. Ambite, G. Ver Steeg, P.M. Thompson, Evaluation of transfer learning methods for detecting Alzheimer's disease with brain MRI. In 18th International Symposium on Medical Information Processing and Analysis SPIE, 12567, (2023) 504-513. <https://doi.org/10.1117/12.2670457>
- [22] N.I.M. Ghali, A. Shereen, A Proposed Recognition System For Alzheimer's Disease Based On Deep Learning And Optimisation Algorithms. *Future*, 1(1), (2021) 1.
- [23] S.B. Shahid, M. Kaikau, M.H. Kabir, M.A. Yousuf, A.K.M. Azad, A.S. Al-Moisheer, N. Alotaibi, S.A. Alyami, T. Bhuiyan, M.A. Moni, Novel deep learning for multi-class classification of Alzheimer's in disability using MRI datasets. *Frontiers in Bioinformatics*, 5, (2025) 1567219. <https://doi.org/10.3389/fbinf.2025.1567219>
- [24] J.M. Amine, M. Mourad, Toward accurate Alzheimer's detection: transfer learning with ResNet50 for MRI-based diagnosis. *Frontiers in Neuroscience*, 19, (2025) 1664418-1664418. <https://doi.org/10.3389/fnins.2025.1664418>
- [25] H.V. Bansal, P. Gupta, V. Juneja, A Multimodal Deep Learning Framework Using ResNet-101 and Firefly-Based Feature Selection for Early Diagnosis of Dementia and Alzheimer's Disease. *IEEE Access*, 13, (2025) 184709-184721. <https://doi.org/10.1109/ACCESS.2025.3621157>
- [26] M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), (2002) 825-841. <https://doi.org/10.1006/nimg.2002.1132>
- [27] B. Patenaude, S.M. Smith, D.N. Kennedy, M. Jenkinson, A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*, 56(3), (2011) 907-922. <https://doi.org/10.1016/j.neuroimage.2011.02.046>
- [28] E.T. Rolls, M. Joliot, N. Tzourio-Mazoyer, Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, 122, (2015) 1-5. <https://doi.org/10.1016/j.neuroimage.2015.07.075>
- [29] B. Fischl, D.H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, A.M. Dale, Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), (2002) 341-355. [https://doi.org/10.1016/S0896-6273\(02\)00569-X](https://doi.org/10.1016/S0896-6273(02)00569-X)

- [30] A.R. Khan, L. Wang, M.F. Beg, FreeSurfer-initiated fully-automated subcortical brain segmentation in MRI using large deformation diffeomorphic metric mapping. *NeuroImage*, 41(3), (2008) 735–746. <https://doi.org/10.1016/j.neuroimage.2008.03.024>
- [31] J.C. Morris, The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, 43(11), (1993) 2412–2414. <https://doi.org/10.1212/WNL.43.11.2412-a>
- [32] F. Hajamohideen, N. Shaffi, M. Mahmud, K. Subramanian, A. Al Sariri, V. Vimbi, A. Abdesselam, Alzheimer's Disease Neuroimaging Initiative, (2023). Four-way classification of Alzheimer's disease using deep Siamese convolutional neural network with triplet-loss function. *Brain Informatics*, 10(1), (2023) 5. <https://doi.org/10.1186/s40708-023-00184-w>
- [33] M. Femmam, S. Femmam, M.E. Fareh, O.A. Senni, A. Ferhani, Deep learning-based classification and diagnosis of Alzheimer's & dementia using multi-scale feature extraction from baseline MRI scans. *Journal of Image and Graphics*, 12(4), (2024) 427–437. <https://doi.org/10.18178/joig.12.4.427-436>
- [34] N. Shaffi, V. Vimbi, K. Subramanian, F. Hajamohideen, A. Abdesselam, M. Mahmud, Ensemble of vision transformer architectures for efficient Alzheimer's disease classification. *Brain Informatics*, 11(1), (2024) 25. <https://doi.org/10.1186/s40708-024-00238-7>
- [35] M. Maqsood, F. Nazir, U. Khan, F. Aadil, H. Jamal, I. Mehmood, O. Song-Y, Transfer learning assisted classification and detection of Alzheimer's disease stages using 3D MRI scans. *Sensors*, 19(11), (2019) 2645. <https://doi.org/10.3390/s19112645>
- [36] S. Alqahtani, A. Aldahri, M.A. Zohdy, A.A. Alsulami, S. Ganesan, Severity grading and early detection of Alzheimer's disease through transfer learning. *Information*, 14(12), (2023) 646. <https://doi.org/10.3390/info14120646>
- [37] M.K. Awang, J. Rashid, G. Ali, M. Hamid, S.F. Mahmoud, D.I. Saleh, H.I. Ahmad, Classification of Alzheimer disease using DenseNet-201 based on deep transfer learning technique. *PLOS ONE*, 19(9), (2024) e0304995. <https://doi.org/10.1371/journal.pone.0304995>
- [38] Y. Zhang, H. Cai, Y. Du, B. Xu, Y. Liu, (2024) Gray matter-guided attention network for ad diagnosis using structural MRI. In 2024 IEEE International Symposium on Biomedical Imaging (ISBI), IEEE, Athens, Greece. <https://doi.org/10.1109/ISBI56570.2024.10635472>
- [39] A.Y.M. Abdelrahman, Enhancing Alzheimer's disease Classification from MRI Scans Using Deep Learning techniques. *East Journal of Computer Science*, 1(6), (2025) 1-10. <https://doi.org/10.63496/ejcs.Vol1.Iss6.226>
- [40] R. Komal, P. Dhavakumar, K. Rahul, B. Jaswanth, R. Preeth, Hybrid deep learning framework for magnetic resonance imaging-based classification of Alzheimer's disease. *Brain Network Disorders*, 1(04), (2025) 239-249.
- [41] S. Sriram, V. Nivethitha, T.P. Arun Kaarthic, S. Archita, T. Murugan, Advanced MRI based Alzheimer's diagnosis through ensemble learning techniques. *Scientific Reports*, 15(1), (2025) 33840. <https://doi.org/10.1038/s41598-025-04899-0>
- [42] R.S. Athira, J. Charles, Pixel-to-Patch Intelligence: TNT-Based Deep Model for Stage-Wise Alzheimer's Classification. *International Journal of Intelligent Engineering & Systems*, 18(8), (2025) 517.
- [43] S.G. Singh, D. Das, U. Barman, H. Huseynov, Optimised CNN-Based Approach For Alzheimer's Disease By Tackling Class Imbalance In MRI Classification. *Reliability: Theory & Applications*, 20(SI 7 (83)), (2025) 197-207. <https://doi.org/10.24412/1932-2321-2025-783-197-207>
- [44] S.M. Mousavi, K. Moulaei, L. Ahmadian, Classifying and diagnosing Alzheimer's disease with deep learning using 6735 brain MRI images. *Scientific Reports*, 15(1), (2025) 22721. <https://doi.org/10.1038/s41598-025-08092-1>

Acknowledgment

The authors would like to express their sincere gratitude to all those who contributed to the successful completion of this research work. The authors are especially thankful to their institution, Atria Institute of Technology, (Visvesvaraya Technological University), Bangalore, India for providing the necessary facilities, resources, and a conducive environment for carrying out this study.

Author Contributions

Ramya C.N: Conceptualization, Investigation, Writing - Original Draft, Writing - Review and Editing. Keshavamurthy: Supervision, Investigation, Writing - Review & Editing.

Funding

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

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The dataset used in this paper is publicly available for research purposes: OASIS-3: Longitudinal Neuroimaging, Clinical and Cognitive Dataset for Normal Aging and Alzheimer's Disease, publicly available at: <https://sites.wustl.edu/oasisbrains/home/oasis-3/>

Has this article screened for similarity?

Yes

About the License

© The Author(s) 2026. The text of this article is open access and licensed under a Creative Commons Attribution 4.0 International License.