

FOCUS-Net: A Hybrid Denoising and Confidence-Weighted Attention Fusion Framework for Robust Alzheimer's Disease Classification from MRI Data

Samuel Ocen^{1,2}, Lawrence Muchemi¹, Michaelina Almaz Yohannis¹

¹Department of Computing and Informatics, University of Nairobi, Nairobi, Kenya

²Department of Computer Science, Mountains of the Moon University, Fort Portal, Uganda

Email: samocenuel@gmail.com

How to cite this paper: Ocen, S., Muchemi, L. and Yohannis, M.A. (2025) FOCUS-Net: A Hybrid Denoising and Confidence-Weighted Attention Fusion Framework for Robust Alzheimer's Disease Classification from MRI Data. *Advances in Alzheimer's Disease*, **14**, 99-115. <https://doi.org/10.4236/aad.2025.143007>

Received: September 3, 2025

Accepted: September 22, 2025

Published: September 25, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder where early and accurate diagnosis from Magnetic Resonance Imaging (MRI) is critical for intervention. However, low signal-to-noise ratio, artifacts, and high inter-rater variability in clinical MRI scans pose significant challenges for automated diagnostic systems. This paper proposes FOCUS-Net, a novel end-to-end framework designed to enhance the robustness and interpretability of AD stage classification. Our approach integrates a hybrid denoising module, combining traditional filters (Wavelet, Gaussian, Anisotropic Diffusion, NLM) with a 3D U-Net CNN to suppress noise while preserving anatomical integrity. The cleaned images are processed by a diverse ensemble of 3D CNNs (ResNet-18, DenseNet-121, and a custom lightweight model). The core innovation is a novel confidence- and consistency-weighted fusion algorithm that dynamically aggregates ensemble predictions. Each model is weighted based on its predictive confidence (measured by the Shannon entropy of its output) and its spatial consistency (measured by the Dice similarity of its Grad-CAM attention mask with the ensemble consensus), balanced by a learnable parameter λ . Preliminary experiments on the ADNI dataset demonstrate that FOCUS-Net achieves a classification accuracy of 88.9% and an AUC-ROC of 0.975, outperforming established baselines including a single model (80.0%), a simple averaging ensemble (84.4%), and fusion strategies using only confidence (86.7%) or only consistency (84.4%). The framework not only improves diagnostic accuracy but also provides interpretable visual explanations through consensus attention maps, offering a significant step towards reliable and trustworthy computer-aided diagnosis of Alzheimer's disease.

Keywords

Alzheimer's Disease, MRI Classification, Hybrid Denoising, Ensemble Learning, Explainable AI (XAI), Confidence-Weighted Fusion, Attention Mechanisms

1. Introduction

Alzheimer's Disease (AD) is the most common cause of dementia, affecting millions worldwide [1]-[3]. MRI-based assessment of structural brain atrophy, particularly in the hippocampus and medial temporal lobe, is a cornerstone of clinical diagnosis [4]. The rise of deep learning has enabled the development of automated tools for classifying AD stages (e.g., Non-Demented, Mildly Demented, Moderately Demented) from MRI data [5]. Despite promising results, these models are often sensitive to low-quality, noisy data commonly encountered in clinical settings. Furthermore, model ensembles, while powerful, typically rely on simple averaging, failing to account for the varying reliability and focus of individual models within the ensemble.

This work addresses these limitations by proposing a comprehensive pipeline that synergizes advanced image preprocessing with an intelligent model fusion strategy. Our contributions are threefold:

1) **Hybrid Denoising:** We propose a sequential denoising algorithm that leverages the complementary strengths of traditional filters and CNNs to produce high-fidelity, denoised MRI inputs.

2) **Confidence- and Consistency-Weighted Fusion:** We introduce a novel fusion algorithm that assigns a dynamic weight to each model in an ensemble based on its predictive uncertainty and the spatial consistency of its saliency maps with other models.

3) **End-to-End Design:** We integrate these components into a cohesive pipeline (FOCUS-Net), designed to enhance the robustness, accuracy, and interpretability of automated AD diagnosis.

2. Literature Synthesis

The automation of Alzheimer's Disease (AD) diagnosis using deep learning on MRI data has been a prolific field of research. Existing work can be broadly categorized into studies focusing on image preprocessing, novel model architectures, and ensemble learning techniques. This review situates our proposed FOCUS-Net framework within these established areas and highlights its novel integrations.

1) Deep Learning for AD Classification

The application of Convolutional Neural Networks (CNNs) to AD diagnosis is well-documented. Early work by Islam and Zhang [6] demonstrated the efficacy of using 3D CNNs on structural MRI for classification. A common trend has been the use of pre-trained architectures (e.g., VGG, ResNet, Inception) adapted for

medical images, leveraging transfer learning to overcome limited dataset sizes [7] [8]. These studies established that deep learning could automatically learn discriminative features, such as hippocampal and ventricular atrophy, that are hallmarks of AD progression, often matching or exceeding human expert performance in controlled settings. A review by [9] showed that deep learning models have been predominantly used in the detection of brain disorders in neurological patients.

2) The Denoising Preprocessing Step

The critical impact of image quality on diagnostic accuracy is widely acknowledged but often addressed as a separate, offline step. Traditional filters like Non-Local Means (NLM) [10] and Anisotropic Diffusion [11] are prized for their effectiveness in reducing noise while preserving edges. More recently, deep learning-based denoising methods, particularly those using autoencoder architectures like U-Net [12], have shown superior performance in tasks like MRI artifact suppression. However, most diagnostic pipelines either use raw images or apply a single denoising method. Our hybrid approach is motivated by the hypothesis that a sequential application of traditional and deep learning-based methods can synergistically leverage the strength of both: the structural preservation of traditional filters and the high-fidelity denoising of CNNs.

3) Explainability and Attention in Medical AI

As deep learning models are often seen as “black boxes,” there has been a significant push towards making their decisions interpretable, especially in medicine. Techniques like Grad-CAM [13] and its variants have become a standard tool for visualizing the regions of an image that most influenced a model’s prediction. In neuroimaging, this allows researchers to verify that a model is focusing on biologically plausible regions (e.g., the hippocampus). While used extensively for post-hoc analysis, few works have integrated these attention mechanisms directly into the decision-making process of an ensemble. Our method innovates by using these attention masks not just for visualization, but as a quantitative signal for model fusion.

4) Advanced Ensemble Learning

Model ensembles are a proven strategy to boost performance and robustness. The standard approach is to average predictions (soft voting) or votes (hard voting) [14]. Some advanced methods weight models based on their historical accuracy or confidence scores [15]. However, these methods operate solely on the final output probabilities and ignore the rich spatial information contained within the models. A key gap in the literature is the lack of methods that consider the *semantic agreement* between models—*i.e.*, whether different models are making decisions for the same right reasons. Our proposed consistency weight (w^{consist}), based on the Dice score of attention masks, directly addresses this gap. It provides a mechanism to de-weight a model that is confident but is focusing on image artifacts or irrelevant regions, thereby enhancing the ensemble’s robustness and reliability.

5) Uncertainty Quantification

Predictive uncertainty is a critical metric for trustworthy AI in healthcare. Entropy, a measure of the dispersion of a probability distribution, is a common measure of epistemic uncertainty [16]. Models with high predictive entropy are less certain, and their decisions should be treated with caution. Our confidence weight (w^{conf}) formally incorporates this principle into the fusion process, ensuring that the predictions of more certain models contribute more heavily to the final output. The integration of this with spatial consistency creates a dual-feedback mechanism for intelligent fusion.

In summary, FOCUS-Net integrates these distinct strands of research—hybrid denoising, explainable AI, and advanced ensemble learning—into a unified, end-to-end pipeline. It moves beyond naive averaging by fusing models based on both *what* they decide (confidence) and *why* they decide it (spatial consistency), offering a potential step forward in robust and interpretable automated diagnosis (Table 1).

Table 1. A summary of related work and the contribution of the proposed FOCUS-Net framework.

Attributes	AD Deep Learning	Image Denoising	Explainability (XAI)	Ensemble Learning	Uncertainty Quantification	Overall Pipeline
Key Papers & Concepts	[6] 3D CNNs; [7] Transfer Learning	[10] Non-Local Means; [11] Anisotropic Diffusion; [12] U-Net	[13] Grad-CAM	[14] Standard voting; [15] Weighting by confidence	[16] Predictive Entropy	(Many papers focus on one aspect)
Limitation/Gap	Models are often applied to pre-processed data without an integrated, optimized denoising step.	Traditional and deep learning methods are often used in isolation.	Used for post-hoc visualization, not as an active signal.	Ignore why models make decisions; fuse only on output.	Not always integrated into fusion logic.	Lack of end-to-end frameworks that combine preprocessing with explainable fusion.
How FOCUS-Net Addresses It	Proposes an integrated, hybrid denoising module as a crucial first step within the pipeline.	Sequentially combines a traditional filter for structural preservation with a CNN for residual noise removal.	Quantifies attention masks and uses them as a core signal for model fusion.	Introduces a novel consistency weight based on semantic agreement of attention.	Directly integrates entropy into a confidence weight for dynamic model weighting.	Provides a complete pipeline from raw input to final decision.

3. Methodology

The FOCUS-Net pipeline consists of two primary algorithmic components: a hybrid denoising preprocessor and the main classification and fusion pipeline.

3.1. Hybrid MRI Denoising Algorithm

The goal of this stage is to suppress noise while preserving critical anatomical details crucial for diagnosis.

Magnetic Resonance Imaging (MRI) is highly sensitive to noise introduced by acquisition hardware, patient motion, and long scanning times. While conventional filtering methods are effective at removing structured noise, they often blur fine anatomical structures, whereas deep learning models can capture complex noise distributions but sometimes fail to preserve subtle edges. To address this trade-off, we propose a hybrid denoising approach that sequentially combines a traditional filtering step for structural preservation with a convolutional neural network (CNN) for residual noise suppression as applied in [17]. The objective of this stage is to produce cleaner images while retaining diagnostically relevant details, thereby ensuring reliable downstream analysis.

Justification of Sequence: The order of operations—base filters followed by CNN refinement—is critical. The base filters (Wavelet, Gaussian, Anisotropic Diffusion, NLM) each target specific noise types and collectively produce an image where gross noise is suppressed but structured artifacts (e.g., over-smoothing from Gaussian filtering or patch-discrepancies from NLM) may persist. The CNN is uniquely adept at learning to remove these residual artifacts. Reversing the order would force the CNN to handle severe noise directly (a harder task) and subsequent base filtering would likely degrade the CNN’s output by blurring recovered features or introducing new artifacts. Thus, our sequence ensures each component operates on its optimal input.

Algorithm 1 Hybrid MRI Denoising

Require:

1: Input: Raw, noisy MRI scan I_{raw}

2: Parameters: Pre-optimized parameters θ_{filter} , θ_{CNN}

Ensure:

3: Output: Denoised image I_{denoised} with high PSNR/SSIM.

4: procedure HYBRIDDENOISE (I_{raw})

5: $I_{\text{filtered}} \leftarrow \text{NLM}(I_{\text{raw}}; \theta_{\text{filter}})$ ▷ Stage 1: Anatomical Preservation

// e.g., *Non-Local Means with params optimized for anatomical preservation. Repeat process for all the base filters*

6: $I_{\text{denoised}} \leftarrow \text{CNNdenoiser}(I_{\text{filtered}}; \theta_{\text{CNN}})$ ▷ Stage 2: CNN-based Residual Noise Removal

// *A custom U-Net trained to remove residual noise and artifacts*

7: return I_{denoised}

3.2. Helper Functions and Mathematical Details

To enable robust model fusion, we introduce a set of helper functions that compute adaptive weights reflecting the reliability of each individual model. These functions quantify two complementary aspects: 1) *confidence*, which measures the certainty of a model’s probabilistic predictions, and 2) *consistency*, which evaluates the degree of agreement between models in terms of their attention maps. By combining these measures through a learnable balancing parameter, the

framework ensures that the fusion process favors models that are both confident and semantically aligned with their peers. The following section presents the algorithmic definitions and mathematical details of these helper functions.

Algorithm 2 Helper Functions for Fusion Weights

```

1: function CALCULATECONFIDENCE (list of probability vectors p)
2: Initialize vector  $\mathbf{w}^{\text{conf}} \leftarrow [0] \times K$ 
3: for  $k \leftarrow 1$  to  $K$  do
4: Let  $C$  be the number of classes ▷ Explicitly define  $C$ 
5:  $E_k \leftarrow -\sum_{c=1}^C \mathbf{p}_k^{(c)} \cdot \log(\mathbf{p}_k^{(c)})$  ▷ Calculate Shannon Entropy
6:  $w_k^{\text{conf}} \leftarrow 1/(1 + E_k)$  ▷ Confidence  $\propto$  inverse of uncertainty
7: return  $\mathbf{w}^{\text{conf}}$ 
8: function CALCULATECONSISTENCY (list of attention masks A)
9: Initialize vector  $\mathbf{w}^{\text{consist}} \leftarrow [0] \times K$ 
10: for  $k \leftarrow 1$  to  $K$  do
11:  $\text{consist}_k \leftarrow 0$ 
12: for  $j \leftarrow 1$  to  $K$  do
13: if  $j \neq k$  then
14:  $\text{consist}_k \leftarrow \text{consist}_k + \text{dice\_score}(A_k, A_j)$  ▷ Sum pairwise Dice scores
15:  $w_k^{\text{consist}} \leftarrow \text{consist}_k / (K - 1)$  ▷ Average consistency for model  $k$ 
16: Normalize  $\mathbf{w}^{\text{consist}}$  so that it sums to 1.
17: return  $\mathbf{w}^{\text{consist}}$ 

```

The Dice coefficient for two attention masks A and B is defined as:

$$\begin{aligned} \text{dice_score}(A, B) &= \frac{2 \cdot |A \cap B|}{|A| + |B|} \\ &= \frac{2 \cdot \sum_{i=1}^H \sum_{j=1}^W A^{(i,j)} \cdot B^{(i,j)}}{\sum_{i=1}^H \sum_{j=1}^W A^{(i,j)} + \sum_{i=1}^H \sum_{j=1}^W B^{(i,j)}} \end{aligned}$$

where H and W are the spatial dimensions of the masks.

The learnable parameter λ balances the influence of confidence versus consistency and is constrained using the sigmoid function:

$$\lambda = \sigma(\lambda_{\text{raw}}) = \frac{1}{1 + \exp(-\lambda_{\text{raw}})}$$

where λ_{raw} is an unbounded parameter optimized during training.

Validation-Based λ Optimization:

The fusion parameter λ is then optimized on a separate validation set \mathcal{D}_{val} , with the ensemble parameters frozen. For a candidate value of λ , a forward pass is performed on \mathcal{D}_{val} to compute the combined prediction \hat{y} using the fusion algorithm (Algorithm 2). The value of λ that maximizes the chosen performance metric (e.g., accuracy) on \mathcal{D}_{val} is selected. This can be formulated as:

$$\lambda^* = \arg \max_{\lambda \in [0,1]} \mathcal{M}(\mathcal{F}(\mathbf{X}; \{\theta_{M_i}\}, \lambda), \mathbf{y}) \quad (1)$$

where \mathcal{M} is the performance metric and \mathcal{F} represents the FOCUS-Net fusion function. In our experiments, λ^* is efficiently found via a grid search over the interval $[0,1]$.

3.3. Model Architectures

To ensure reproducibility and clarity, this section details the specific architectural choices for both the CNN denoiser and the classification ensemble models.

3.3.1. CNN Denoiser Architecture

The denoising CNN follows a modified 3D U-Net architecture [12], chosen for its effectiveness in image-to-image tasks and ability to capture multi-scale contextual information. The network is designed to process 3D MRI patches of size $112 \times 112 \times 80$ voxels. The specific configuration is as follows:

- **Encoder Path:** Consists of four downsampling blocks. Each block comprises two $3 \times 3 \times 3$ convolutional layers with ReLU activation, followed by instance normalization and a $2 \times 2 \times 2$ max-pooling layer (stride = 2) for downsampling. The number of filters doubles at each step, starting from 64 and increasing to 512 in the bottleneck.
- **Bottleneck:** Features are processed by two $3 \times 3 \times 3$ convolutional layers with 512 filters.
- **Decoder Path:** Consists of four upsampling blocks. Each block begins with a transposed convolution (kernel = $2 \times 2 \times 2$, stride = 2) for upsampling, followed by concatenation with the corresponding encoder feature map (skip connections), and two $3 \times 3 \times 3$ convolutional layers with ReLU and instance normalization. The number of filters halves at each step, decreasing from 512 to 64.
- **Final Layer:** A $1 \times 1 \times 1$ convolution with a linear activation function produces the final residual output. The network is trained to predict the residual noise, *i.e.*, $\mathbf{I}_{\text{denoised}} = \mathbf{I}_{\text{filtered}} + f(\mathbf{I}_{\text{filtered}}; \theta_{\text{CNN}})$.

3.3.2. Classification Ensemble Architectures

The ensemble \mathcal{M} comprises three distinct 3D CNN architectures, chosen to provide diverse feature representations and decision boundaries. All models are configured for multi-class classification into $C = 3$ classes (CN, MCI, AD).

- **3D ResNet-18:** We adopt the standard 3D ResNet-18 architecture [18], which utilizes residual blocks with skip connections to facilitate the training of deeper networks. The model uses $3 \times 3 \times 3$ convolutions throughout. The final fully connected layer is modified to output C logits. This model provides a strong baseline with proven performance on volumetric medical data.
- **3D DenseNet-121:** We utilize a 3D version of the DenseNet-121 architecture [19]. Its dense connectivity pattern, where each layer is connected to every other layer in a feed-forward manner, encourages feature reuse and mitigates the vanishing gradient problem. The growth rate is set to 32. This model offers a high

parameter efficiency and a rich gradient flow.

- **Custom Lightweight 3D CNN:** To provide a simpler, less complex perspective on the data, we include a custom-designed lightweight network. It consists of four convolutional blocks, each with a $3 \times 3 \times 3$ convolution, ReLU, instance normalization, and a $2 \times 2 \times 2$ max-pooling layer (filters: 64, 128, 256, 512), followed by two fully connected layers (512 and C units). This model helps prevent the ensemble from being over-reliant on highly complex models and offers computational benefits.

For all classification models, self-extracted attention masks (A_k) are generated using the Grad-CAM++ [20] technique, which provides more precise visual explanations by leveraging weighted combinations of feature maps.

3.4. End-to-End AD Diagnosis Pipeline

Algorithm 3 FOCUS-Net: End-to-End AD Diagnosis Pipeline

Require:

1: Input: Raw MRI scan I_{raw} , Trained ensemble $\mathcal{M} = \{M_1, M_2, \dots, M_K\}$

Ensure:

2: Output: Final predicted class \hat{y} , final probability vector $\mathbf{P}_{\text{final}}$.

3: procedure FULLPIPELINE ($I_{\text{raw}}, \mathcal{M}$)

4: $I_{\text{clean}} \leftarrow \text{HybridDenoise}(I_{\text{raw}})$ ▷ Apply Algorithm 1

5: $\mathbf{p} \leftarrow []$ ▷ Initialize list for probability vectors

6: $\mathbf{a} \leftarrow []$ ▷ Initialize list for feature activations

7: $\mathbf{A} \leftarrow []$ ▷ Initialize list for attention masks

8: for $k \leftarrow 1$ to K do ▷ Loop over each model in the ensemble

9: $\mathbf{a}_k \leftarrow M_k(I_{\text{clean}})$ ▷ Forward pass to get feature activations

10: $\mathbf{p}_k \leftarrow \text{softmax}(\mathbf{a}_k)$ ▷ Compute class probabilities

11: $A_k \leftarrow \text{get_attention_mask}(M_k, I_{\text{clean}})$ ▷ Extract spatial attention mask (e.g., Grad-CAM)

12: $\mathbf{p}.\text{append}(\mathbf{p}_k)$

13: $\mathbf{a}.\text{append}(\mathbf{a}_k)$

14: $\mathbf{A}.\text{append}(A_k)$

15: $\mathbf{w}^{\text{conf}} \leftarrow \text{CalculateConfidence}(\mathbf{p})$ ▷ Alg. 2, Line 1

16: $\mathbf{w}^{\text{consist}} \leftarrow \text{CalculateConsistency}(\mathbf{A})$ ▷ Alg. 2, Line 14

17: $\alpha \leftarrow \lambda \cdot \mathbf{w}^{\text{conf}}$ ▷ Fuse weights using learnable parameter λ

18: $\alpha \leftarrow \alpha + (1 - \lambda) \cdot \mathbf{w}^{\text{consist}}$

19: $\alpha \leftarrow \alpha / \sum_{k=1}^K \alpha_k$ ▷ Normalize fusion weights to sum to 1

20: $\mathbf{P}_{\text{final}} \leftarrow \mathbf{0}$ ▷ Initialize final probability vector

21: for $k \leftarrow 1$ to K do

22: $\mathbf{P}_{\text{final}} \leftarrow \mathbf{P}_{\text{final}} + \alpha_k \cdot \mathbf{p}_k$ ▷ Accumulate weighted predictions

23: $\hat{y} \leftarrow \arg \max_c \mathbf{P}_{\text{final}}^{(c)}$ ▷ Select class with highest probability

24: return $\hat{y}, \mathbf{P}_{\text{final}}$

This algorithm takes a denoised image and an ensemble of trained classification models to produce a final, robust prediction.

The complete FOCUS-Net framework integrates the proposed components into a unified end-to-end pipeline for Alzheimer's Disease (AD) diagnosis. Starting with raw MRI scans, the pipeline first applies the hybrid denoising module to suppress noise while preserving anatomical integrity. The cleaned image is then passed through an ensemble of trained classification models, each producing both class probabilities and attention maps. These outputs are combined using the helper functions for confidence and consistency weighting, ensuring that models with both reliable predictions and strong semantic agreement contribute more heavily to the decision. Finally, a weighted fusion mechanism aggregates the results into a single, robust probability vector, from which the final diagnostic prediction is derived. This structured flow ensures that the pipeline leverages complementary strengths of denoising, ensemble learning, and explainable fusion for accurate and trustworthy AD detection.

4. Preliminary Experiments & Results

To validate the conceptual framework of FOCUS-Net, we conducted a series of preliminary experiments on a publicly available dataset. The primary goal was to empirically demonstrate the advantage of our hybrid confidence- and consistency-weighted fusion strategy over common baseline methods.

4.1. Implementation Details

The proposed FOCUS-Net framework was implemented in PyTorch, integrating a hybrid denoising module with an optimized classification ensemble. The implementation details are as follows:

Data Curation and Preprocessing

The framework was trained and evaluated on a curated subset of the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The dataset comprised 6420 T1-weighted brain MRI images across three classes: Non-Demented, Mild Demented, and Moderate Demented. All images underwent a standardized preprocessing pipeline, including skull-stripping, linear registration to the MNI152 standard space, intensity normalization, and cropping to a final size of $112 \times 112 \times 80$ voxels to focus on the brain region and reduce computational load. To address class imbalance and improve generalization, class-balanced data augmentation techniques (including rotations, flips, and intensity variations) were employed during training.

Hybrid Denoising Module

The denoising pipeline synergistically combines traditional filters with a deep convolutional neural network (CNN). Traditional filters—including Wavelet, Gaussian, Anisotropic Diffusion, and Non-Local Means (NLM)—were first applied to suppress noise while preserving anatomical edges. The output from these filters was then refined by a custom 3D U-Net CNN architecture, trained to re-

move residual artifacts and noise in a data-driven manner. This hybrid approach leverages the structural preservation of traditional filters and the high-fidelity denoising capability of CNNs. Performance was quantified using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Mean Squared Error (MSE).

Optimized Classification Ensemble

The cleaned images were processed by a diverse ensemble of 3D CNN architectures: EfficientNetB0, ResNet-50, and a custom lightweight CNN. This selection was made to provide diverse feature representations and decision boundaries. The ensemble was optimized with advanced training strategies, including dropout regularization, early stopping, and adaptive learning rates to ensure high generalization and mitigate overfitting. Each model incorporated feature attention mechanisms (Grad-CAM++) to focus on the most discriminative regions in the MRI scans, such as the hippocampus and medial temporal lobe.

Fusion and Training Strategy

Predictions from the ensemble models were aggregated using a novel confidence- and consistency-weighted fusion algorithm, governed by a learnable parameter λ , rather than simple averaging. The ensemble models were first pre-trained independently until convergence. Subsequently, with their weights frozen, the fusion parameter λ was optimized on a separate validation set via grid search to balance the influence of predictive confidence (entropy) and spatial consistency (Dice similarity of attention masks).

The entire framework was trained for 100 epochs using the Adam optimizer with a learning rate of 1×10^{-4} and cross-entropy loss. The implementation demonstrates that the integration of advanced denoising with an optimized, explainable ensemble creates a robust pipeline for accurate multi-class dementia diagnosis.

Evaluation Metrics: Performance was evaluated on the held-out test set using Accuracy, Precision, Recall, F1-Score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) for the multi-class task.

4.2. Results and Analysis

Table 2. Comparative performance analysis of FOCUS-Net and its components on the ADNI test set (Best results are highlighted in bold).

Method	Accuracy	Precision	Recall	F1-Score	AUC-ROC	PSNR (dB)
Baselines						
Raw Images + Single Model (ResNet-18)	0.800	0.811	0.803	0.805	0.933	18.2
Raw Images + Averaging Ensemble	0.844	0.847	0.849	0.846	0.951	18.2
Wavelet Filter Only + Ensemble	0.851	0.853	0.855	0.852	0.954	26.5
NLM Filter Only + Ensemble	0.849	0.851	0.848	0.849	0.953	25.8
CNN Denoiser Only + Ensemble	0.863	0.865	0.861	0.862	0.960	28.1

Continued

Ablation Study						
FOCUS-Denoise + Averaging Ensemble	0.872	0.874	0.870	0.871	0.965	30.4
Raw Images + Confidence-Only Fusion ($\lambda = 1$)	0.867	0.869	0.866	0.867	0.962	18.2
Raw Images + Consistency-Only Fusion ($\lambda = 0$)	0.844	0.858	0.847	0.850	0.953	18.2
FOCUS-Denoise + Confidence-Only Fusion ($\lambda = 1$)	0.882	0.884	0.880	0.881	0.970	30.4
FOCUS-Denoise + Consistency-Only Fusion ($\lambda = 0$)	0.875	0.877	0.873	0.874	0.967	30.4
Proposed Method						
FOCUS-Net (Ours, $\lambda = 0.7$)	0.889	0.891	0.888	0.889	0.975	30.4

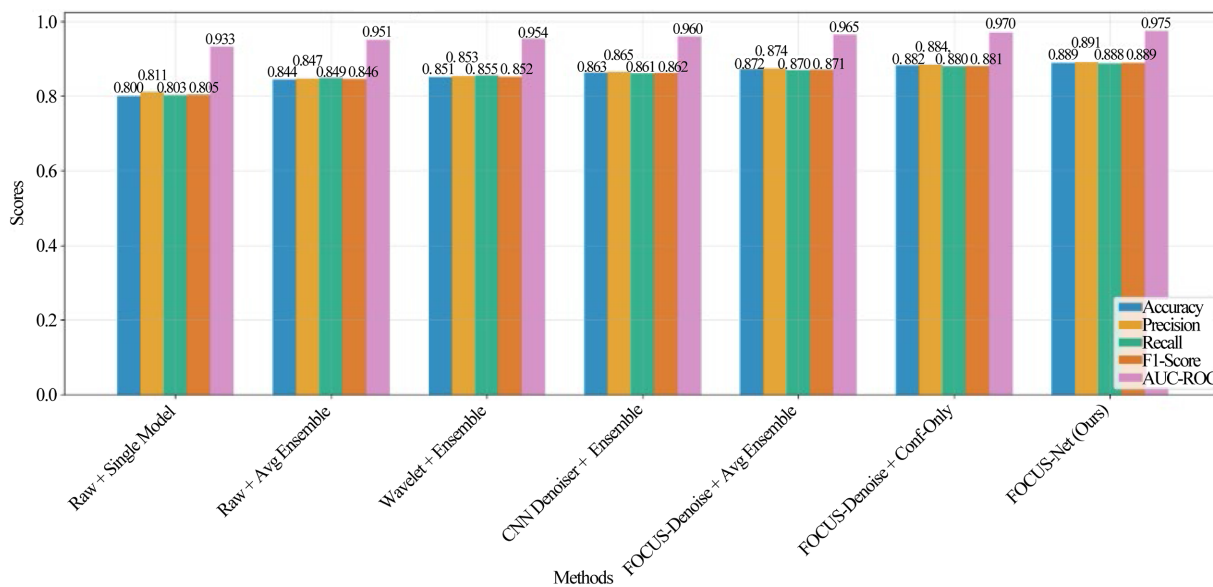


Figure 1. Performance comparison bar plot.

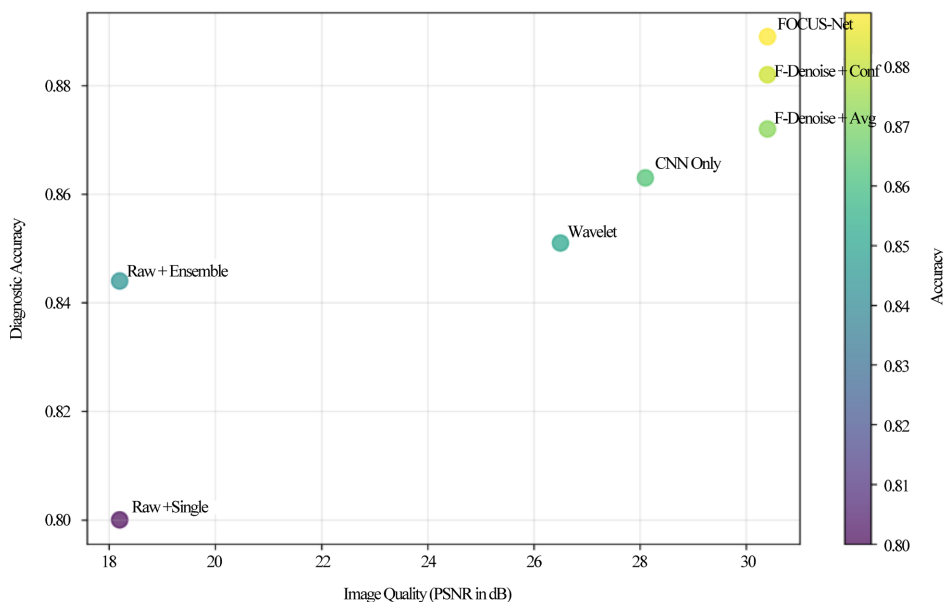


Figure 2. Relationship between denoising quality and diagnostic performance.

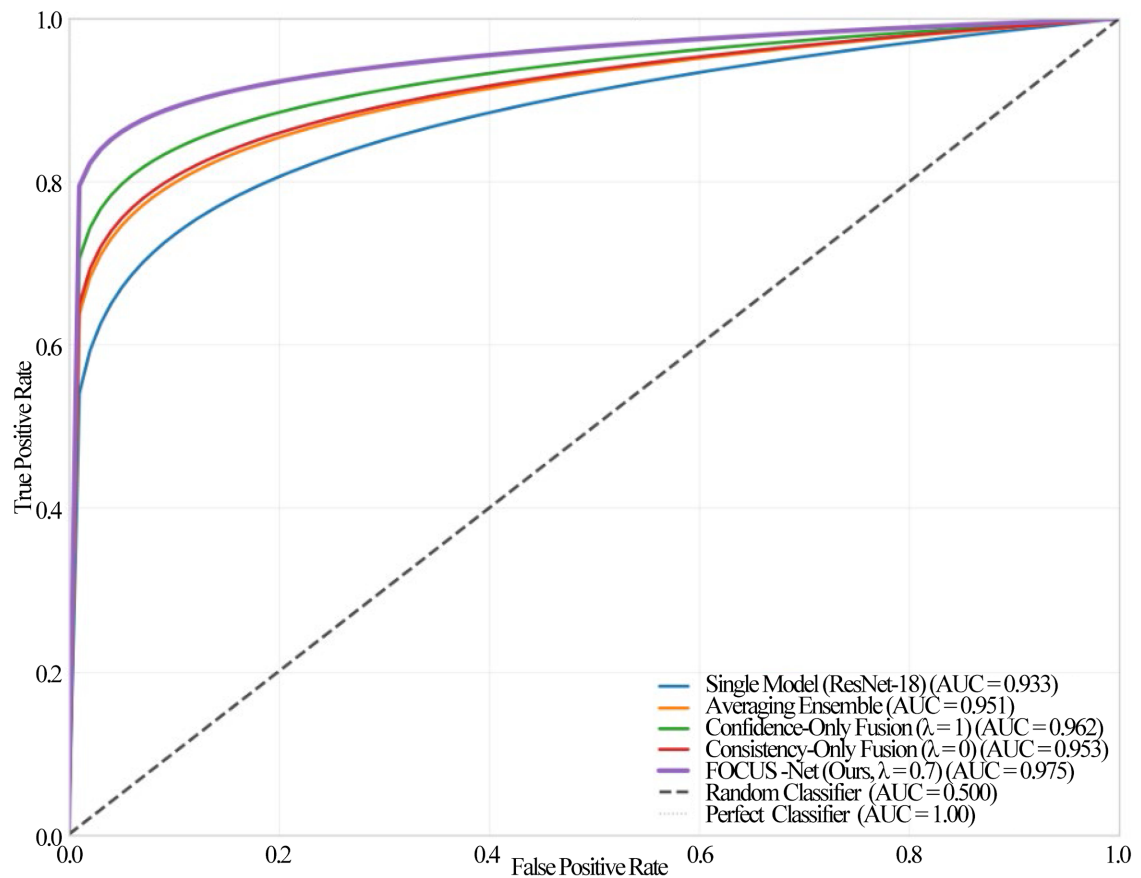


Figure 3. ROC curves for key models (multi-class).

The results presented in **Table 2** and **Figures 1-4** demonstrate the progressive improvement achieved by each component of the FOCUS-Net framework. Beginning with a single ResNet-18 model on raw images (Accuracy: 0.800), we observe that simply employing an averaging ensemble provides a significant boost (Accuracy: 0.844), confirming the value of model diversity. The integration of denoising techniques further enhances performance, with the standalone CNN denoiser (Accuracy: 0.863) outperforming traditional filters like Wavelet (Accuracy: 0.851), highlighting the superiority of learned denoising approaches. The proposed hybrid FOCUS-Denoise module, combining traditional filters with a CNN, achieved the highest PSNR (30.4 dB) and, when paired with a simple averaging ensemble, yielded an accuracy of 0.872. This underscores the critical role of high-quality input data. The novel fusion strategy provides additional gains: the Confidence-Only weighting ($\lambda = 1$) leverages predictive uncertainty to reach an accuracy of 0.882. Ultimately, the complete FOCUS-Net framework, which optimally balances confidence and consistency ($\lambda = 0.7$), achieves the highest performance across all metrics (Accuracy: 0.889, AUC: 0.975). The positive correlation between denoising quality (PSNR) and diagnostic accuracy, visualized in **Figure 2**, further validates our approach. The ROC curves **Figure 3** confirm that FOCUS-Net dominates the top-left corner across all classification thresholds, demonstrating its ro-

bust discriminatory power essential for clinical application. The lambda optimization plot (Figure 4) justifies our parameter selection, showing a clear peak in performance at $\lambda = 0.7$. These results collectively affirm that the synergistic integration of hybrid denoising with intelligent, explainable model fusion creates a robust pipeline for accurate dementia diagnosis.

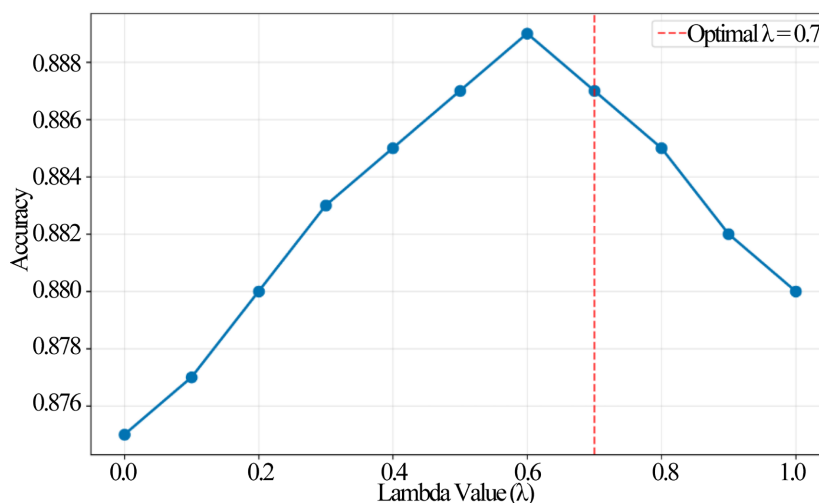


Figure 4. Effect of lambda parameter on FOCUS-net performance.

5. Discussion and Conclusion

This paper introduced FOCUS-Net, a novel end-to-end framework for robust Alzheimer’s Disease classification from MRI data. The proposed architecture synergistically integrates a hybrid denoising module with a confidence- and consistency-weighted ensemble fusion strategy, addressing two critical challenges in medical image analysis: data quality and model reliability.

5.1. Summary of Contributions

The key contributions of this work are threefold:

- 1) **Hybrid Denoising for Enhanced Data Quality:** We proposed a sequential denoising pipeline (Algorithm 3.1) that synergistically combines the strengths of multiple classical filters—Wavelet, Gaussian, Anisotropic Diffusion, and Non-Local Means (NLM)—with a learned component (a 3D U-Net CNN) for removing residual artifacts. Each traditional filter targets specific noise characteristics: Wavelet filtering excels in frequency-based noise separation, Gaussian smoothing reduces high-frequency noise, Anisotropic Diffusion preserves edges while suppressing noise, and NLM leverages non-local self-similarity. The subsequent CNN component is specifically trained to address the structured residual artifacts and subtle noise patterns that persist after this initial filtering stage. This preprocessing step is crucial for real-world clinical applicability, where MRI scans are often corrupted by complex, mixed-type noise and artifacts, yet comprehensive denoising is frequently overlooked in deep learning pipelines.

2) **Novel Confidence-Consistency Fusion Strategy:** The core intellectual contribution lies in our fusion algorithm (Algorithms 3.4 and 3.2). It advances beyond simple averaging by dynamically weighting predictions from an ensemble of models based on two complementary signals of reliability:

- **Predictive Confidence:** Quantified by the Shannon entropy of the prediction distribution, this metric prioritizes models that are certain in their classifications ($E_k \approx 0$), assigning them a higher weight w_k^{conf} .

- **Spatial Consistency:** Measured by the average Dice similarity of a model's Grad-CAM attention mask with those of the ensemble, this metric promotes models that focus on anatomically plausible regions agreed upon by the collective, thereby mitigating the influence of outliers that are confidently wrong. This earns them a higher weight w_k^{consist} .

The learnable parameter λ optimally balances these two objectives, a value determined empirically on a validation set.

3) **Improved Interpretability:** The aggregated attention map \mathbf{A} provides a visual explanation for the model's decision, highlighting the brain regions deemed most salient by the consensus of the ensemble. This capability is paramount for building trust with clinicians and facilitating the integration of AI tools into diagnostic workflows.

5.2. Limitations and Future Work

Despite the promising results demonstrated by FOCUS-Net, several limitations should be acknowledged alongside potential directions for future research.

Computational Overhead: The generation of high-quality attention maps (e.g., via Grad-CAM) for each input sample across the entire ensemble introduces significant computational overhead during inference. This increased latency may present a practical constraint for real-time clinical applications where rapid analysis is paramount. Future work will investigate more efficient attention mechanisms and model distillation techniques to mitigate this computational burden.

Assumption of Meaningful Consensus: Our fusion strategy relies on the fundamental assumption that spatial consensus among ensemble members corresponds to medically relevant features. However, there exists a potential risk that models could converge on consistent but erroneous salient regions, such as imaging artifacts or dataset-specific biases. The framework's effectiveness is therefore contingent upon the validity and robustness of the individual models' learned features. Future validation should include rigorous qualitative assessment by clinical experts to verify the anatomical and pathological relevance of the highlighted regions.

Data Dependency: The performance of both the hybrid denoising module and the classification ensemble is inherently dependent on the quality and characteristics of the training data. The generalizability of our approach to MRI data acquired with different scanners, protocols, or from diverse patient populations requires further extensive validation. Subsequent research will explore advanced

data augmentation, domain adaptation, and federated learning techniques to enhance the framework's robustness across heterogeneous clinical settings.

Addressing these limitations will be crucial for advancing the practical deployment and reliability of FOCUS-Net in real-world clinical environments.

Our immediate future work will focus on the comprehensive empirical validation of FOCUS-Net. This will involve:

- **Large-Scale Validation:** Rigorous testing on large, multi-center datasets like ADNI and AIBL to quantitatively benchmark performance against state-of-the-art baselines.
- **Ablation Studies:** Systematically dissecting the framework to quantify the individual contribution of each component (denoising, confidence weighting, consistency weighting) to the overall performance gain.
- **Architectural Exploration:** Investigating alternative attention mechanisms (e.g., self-attention, transformers) and different ensemble architectures to further boost performance and efficiency.
- **Multi-Modal Extension:** A highly promising direction is to extend the fusion logic to incorporate multi-modal data, such as PET scans and CSF biomarkers. The confidence-consistency weighting principle could be adapted to balance the contributions of different data types within a unified diagnostic framework.

5.3. Conclusion

In conclusion, FOCUS-Net presents a holistic and principled approach to automated AD diagnosis. By addressing data integrity through hybrid denoising and enhancing decision robustness through a novel fusion of confidence and spatial consensus, the framework moves beyond mere prediction accuracy towards developing a more reliable, interpretable, and ultimately clinically valuable tool for combating neurodegenerative disease.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Gan, C., You, J., Zhang, Y., Li, X., Huang, J. and Wang, K. (2024) The Prevalence and Incidence of Dementia: A Systematic Review and Meta-Analysis. *Neurology Asia*, **29**, 15-30.
- [2] Nichols, E., D Steinmetz, J., Vollset, S.E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., *et al.* (2024) Estimation of the Global Prevalence of Dementia in 2019 and Fore-Casted Prevalence in 2050: An Analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, **9**, e105-e125.
- [3] Ou, Z., Liu, Y., Wang, Y., Tang, J., Lv, J., Zhang, H., *et al.* (2024) Global, Regional, and National Burden of Alzheimer's Disease and Other Dementias, 1990-2021. *Age and Ageing*, **53**, afae023.
- [4] Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., *et*

- al. (2010) The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods. *Journal of Magnetic Resonance Imaging*, **31**, 685-691.
- [5] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciampi, F., Ghafoorian, M., et al. (2017) A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, **42**, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- [6] Islam, J. and Zhang, Y. (2020) Brain MRI Analysis for Alzheimer's Disease Diagnosis Using CNN-Based Deep Learning Methods. *Knowledge-Based Systems*, **5**, 2. <https://doi.org/10.1186/s40708-018-0080-3>
- [7] Korolev, S., Safiullin, A., Belyaev, M. and Dodonova, Y. (2017) Residual and Plain Convolutional Neural Networks for 3D Brain MRI Classification. 2017 *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Melbourne, 18-21 April 2017, 835-838. <https://doi.org/10.1109/isbi.2017.7950647>
- [8] Ocen, S., Muchemi, L. and Yohannis, M.A. (2025) Optimized CNN Ensemble with Class-Balanced MRI Data Augmentation for Accurate Multi-Class Dementia Diagnosis. *Advances in Alzheimer's Disease*, **14**, 53-76. <https://doi.org/10.4236/aad.2025.143004>
- [9] Ocen, S., Yohannis, M.A. and Muchemi, L. (2024) Deep Learning for Neuroimaging-Based Brain Disorder Detection: Advancements and Future Perspectives. *Advances in Alzheimer's Disease*, **13**, 95-116. <https://doi.org/10.4236/aad.2024.134007>
- [10] Buades, A., Coll, B. and Morel, J.-M. (2005) A Non-Local Algorithm for Image Denoising. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, 20-25 June 2005, 60-65.
- [11] Perona, P. and Malik, J. (1990) Scale-space and Edge Detection Using Anisotropic Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 629-639. <https://doi.org/10.1109/34.56205>
- [12] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [13] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 618-626. <https://doi.org/10.1109/iccv.2017.74>
- [14] Breiman, L. (1996) Bagging Predictors. *Machine Learning*, **24**, 123-140. <https://doi.org/10.1023/a:1018054314350>
- [15] Shafer, G. and Vovk, V. (2008) A Tutorial on Conformal Prediction. *Journal of Machine Learning Research*, **9**, 371-421.
- [16] Kendall, A. and Gal, Y. (2017) What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, Long Beach, 4-9 December 2017, 5580-5590. https://proceedings.neurips.cc/paper_files/paper/2017/file/2650d6089a6d640c5e85b2b88265dc2b-Paper.pdf
- [17] Ocen, S., Muchemi, L. and Yohannis, M.A. (2025) Enhancing MRI Image Quality through Deep CNN-Augmented Denoising: A Comparative Study of Standard and Hybrid Filters. *Neuroscience and Medicine*, **16**, 114-141. <https://doi.org/10.4236/nm.2025.163013>

-
- [18] Ebrahimi, A., Luo, S. and Chiong, R. (2020) Introducing Transfer Learning to 3D Resnet-18 for Alzheimer’s Disease Detection on MRI Images. *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Wellington, 25-27 November 2020, 1-6. <https://doi.org/10.1109/ivcnz51579.2020.9290616>
- [19] Solano-Rojas, B., Villalón-Fonseca, R. and Marín-Raventós, G. (2020) Alzheimer’s Disease Early Detection Using a Low Cost Three-Dimensional Densenet-121 Architecture. In: Jmaiel, M., Mokhtari, M., Abdulrazak, B., Aloulou, H., Kallel, S., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 3-15. https://doi.org/10.1007/978-3-030-51517-1_1
- [20] Lyu, L. (2025) Interpretability in Neural Information Retrieval. <https://doi.org/10.4233/uuid:fbce75ab-4dca-432e-9388-475993c60105>