



SOGAF-Net: Spectral–Ordinal Gated Adaptive Fusion for Multi-Class Alzheimer’s Disease Severity Classification from Brain MRI

Sonal Ayyappan¹, Nancy V², Ibrar Ahmed³, Manimaran⁴, Aresh Kumar Tripathy^{5*}, Vanapalli Kiran Kumar⁶, Reji R⁷, Divya Saleela^{8*}

¹Associate Professor, Department of Artificial Intelligence and Data Science SCMS School of Engineering and Technology, Palissery, Ernakulam, Kerala, India. E-mail: sonal@scmsgroup.org

²Assistant Professor, Easwari Engineering College, Ramapuram, Chennai, India. E-mail: nancy.victorsagayee@gmail.com

³Assistant Professor, Department of Computer Science and Engineering, Noida Institute of Engineering and Technology, Greater Noida, Uttar Pradesh, India. E-mail: ibrar.be2012@gmail.com

⁴Senior Associate Professor, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India. E-mail: manimarankongu@gmail.com

⁵Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India. E-mail: asesh.tripathy@gmail.com

⁶Assistant Professor, Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali, Srikakulam, Andhra Pradesh, India. E-mail: kirankumarvanapalli@gmail.com

⁷Professor, Thangal Kunju Musaliar Institute of Technology, Kerala, India. E-mail: rejir@tkmit.ac.in

⁸Research Fellow, University of Southampton, Southampton, United Kingdom. E-mail: d.saleela@soton.ac.uk

ABSTRACT

Accurate classification of Alzheimer’s disease severity from magnetic resonance imaging (MRI) remains challenging because adjacent disease stages often exhibit overlapping structural characteristics and publicly available datasets are frequently imbalanced. This study proposes SOGAF-Net, an Ordinal–Spectral Adaptive Fusion Network for four-class Alzheimer’s disease severity classification. The framework integrates spatial representations extracted using a pretrained EfficientNet-B0 backbone with frequency-domain information obtained through a dedicated fast Fourier transform-based spectral branch. A sample-adaptive gating mechanism dynamically combines spatial and spectral features, while ordinal supervision explicitly models the ordered progression from non-demented to very mild, mild, and moderate dementia. Class-balanced focal loss and a categorical–ordinal consistency constraint are incorporated to improve discrimination across disease stages and mitigate class imbalance. Experiments were conducted using a public brain MRI dataset containing 6,400 images, with separate training, validation, and test partitions. SOGAF-Net achieved an accuracy of 96.64%, a macro-F1 score of 97.10%, and a macro one-vs-rest ROC-AUC of 99.83%. Bootstrap evaluation produced a 95% confidence interval of 95.55–97.58% for accuracy, indicating stable predictive performance. Grad-CAM-based visual analysis further provided interpretable evidence of image regions contributing to model predictions. The results demonstrate that jointly exploiting spatial, spectral, and ordinal information provides an effective framework for multi-stage Alzheimer’s disease classification from brain MRI.

Keywords: adaptive fusion; fast Fourier transform; medical image classification; explainable artificial intelligence

1. Introduction

Alzheimer’s disease (AD) is a progressive neurodegenerative disorder associated with deterioration in memory, cognition, behaviour, and functional independence. Structural alterations in the brain evolve throughout the disease continuum, making reliable identification of disease severity important for early assessment and subsequent clinical management. Magnetic resonance imaging (MRI) provides non-invasive visualisation of neuroanatomical structures and enables quantitative assessment of structural abnormalities without ionising radiation. Structural MRI has consequently become one of the principal neuroimaging modalities investigated for computer-aided AD diagnosis. Recent evidence indicates that although deep learning has achieved strong performance in binary AD classification, reliable discrimination among multiple neighbouring disease stages remains challenging because of subtle inter-class differences, heterogeneous imaging characteristics, class imbalance, and limited external validation [1].

Convolutional neural networks (CNNs) established much of the early progress in automated MRI-based AD classification by learning hierarchical anatomical representations directly from imaging data. Deep CNN-based analysis of structural MRI demonstrated effective discrimination between AD and healthy controls and was subsequently extended to more difficult mild cognitive impairment-related classification tasks [2].

More broadly, advances in MRI-based brain-image analysis have demonstrated the value of transfer learning for multi-class classification [3], integrated segmentation and classification strategies for learning disease-relevant structural representations [4], and combinations of deep and conventional machine-learning approaches for automated MRI classification [5]. These developments provide methodological evidence that pretrained representations and complementary feature-learning strategies can be effective when analysing structurally complex brain MRI data.

Transformer-based architectures have subsequently been introduced to capture longer-range dependencies beyond the local receptive fields of conventional CNNs. Joint transformer architectures have been used to preserve contextual relationships among MRI slices and have demonstrated strong performance in both binary and multiclass AD classification [6]. Anatomically guided ensembles of three-dimensional vision transformers have further incorporated region-level brain information to combine predictive modelling with improved interpretability [7]. Hybrid architectures combining convolutional and transformer representations have additionally demonstrated the value of jointly exploiting local anatomical characteristics and global contextual information from MRI [8].

More recently, frequency-aware modelling has expanded representation learning beyond the conventional spatial domain. Dynamic frequency filtering integrated with lightweight transformer processing has demonstrated that frequency-domain characteristics can complement spatial representations for multi-stage AD classification [9]. Geometry-guided deformable attention has likewise been introduced to adapt spatial sampling to variations in brain morphology, indicating the potential value of flexible anatomical feature extraction rather than relying exclusively on fixed convolutional receptive fields [10].

Recent studies have further strengthened the transition towards multi-stage, adaptive, and interpretable AD classification. Modified vision transformers have been investigated specifically for multi-stage disease classification from brain MRI [11], while deformable-attention transformers have addressed four-class classification on highly imbalanced Alzheimer MRI data [12]. Adaptive attention-gating strategies have also been introduced to dynamically regulate contributions from heterogeneous representations in multimodal AD modelling [13]. In parallel, lightweight CNN architectures combined with explainable artificial intelligence have been explored to improve accessibility and visual interpretation of MRI-based predictions [14]. Counterfactual explanation frameworks have further demonstrated the growing importance of understanding how structural image alterations influence diagnostic decisions rather than relying solely on predictive performance [15].

Despite these advances, several methodological limitations remain in multi-stage AD severity classification. Disease stages are frequently treated as independent nominal categories despite possessing an inherent clinical ordering. Conventional categorical learning therefore does not explicitly differentiate between an error involving neighbouring stages and an error spanning the complete severity spectrum. In addition, although frequency-domain representations provide information complementary to spatial features, fixed concatenation or uniform feature fusion cannot adapt the relative contribution of each representation to individual MRI characteristics. Severe class imbalance may further bias optimisation towards dominant categories, while evaluation based predominantly on overall accuracy provides limited information regarding class-sensitive discrimination, statistical uncertainty, calibration, and interpretability.

To address these limitations, this study proposes SOGAF-Net, an Ordinal-Spectral Adaptive Fusion Network for multi-class Alzheimer's disease severity classification from brain MRI. The framework combines a pretrained EfficientNet-based spatial branch with a dedicated fast Fourier transform-based spectral branch to capture complementary anatomical and frequency-domain information. A sample-adaptive gating mechanism dynamically regulates the relative contribution of spatial and spectral representations rather than relying on fixed feature fusion. Ordinal supervision explicitly models the ordered progression among disease stages, while a categorical-ordinal consistency objective encourages agreement between conventional multiclass predictions and ordinal severity estimates. Class-balanced focal learning is additionally employed to mitigate substantial class imbalance, while calibration, bootstrap uncertainty estimation, and Grad-CAM visualisation provide complementary assessments of reliability and interpretability.

The principal contributions of this study are fourfold. First, an ordinal-spectral learning framework is developed to jointly capture anatomical spatial information and complementary frequency-domain characteristics from brain MRI. Second, a sample-adaptive gating mechanism is introduced to dynamically fuse spatial and spectral representations according to individual image characteristics. Third, ordinal severity supervision is integrated with a categorical-ordinal consistency constraint to exploit the ordered relationship among non-demented, very mild, mild, and moderate dementia classes while retaining conventional multiclass discrimination. Fourth, an imbalance-aware optimisation and comprehensive evaluation framework is adopted using class-sensitive performance measures, ROC analysis, probability calibration, bootstrap confidence intervals, and Grad-CAM-based visual explanation.

The remainder of this paper is organised as follows. Section 2 describes the dataset, image preprocessing, proposed SOGAF-Net architecture, spatial and spectral feature extraction, adaptive fusion strategy, ordinal learning formulation, and optimisation objectives. Section 3 presents the experimental configuration, evaluation criteria, and quantitative and qualitative results. Section 4 discusses the principal findings, comparisons with existing approaches, methodological limitations, and implications for future research. Finally, Section 5 concludes the study and outlines directions for further investigation.

2. Methodology

This section presents the proposed Ordinal-Spectral Adaptive Fusion Network (SOGAF-Net) for four-class Alzheimer's disease severity classification from brain magnetic resonance imaging (MRI). The framework jointly models anatomical information in the spatial domain and complementary characteristics in the frequency domain while explicitly incorporating the ordered relationship among disease stages. The methodology comprises dataset

preparation, image preprocessing and augmentation, spatial feature extraction using a pretrained EfficientNet-B0 backbone, fast Fourier transform (FFT)-based spectral representation learning, sample-adaptive spatial-spectral fusion, categorical and ordinal prediction, imbalance-aware optimisation, probability calibration, and Grad-CAM-based visual explanation. The complete architecture of the proposed framework is illustrated in Fig. 1.

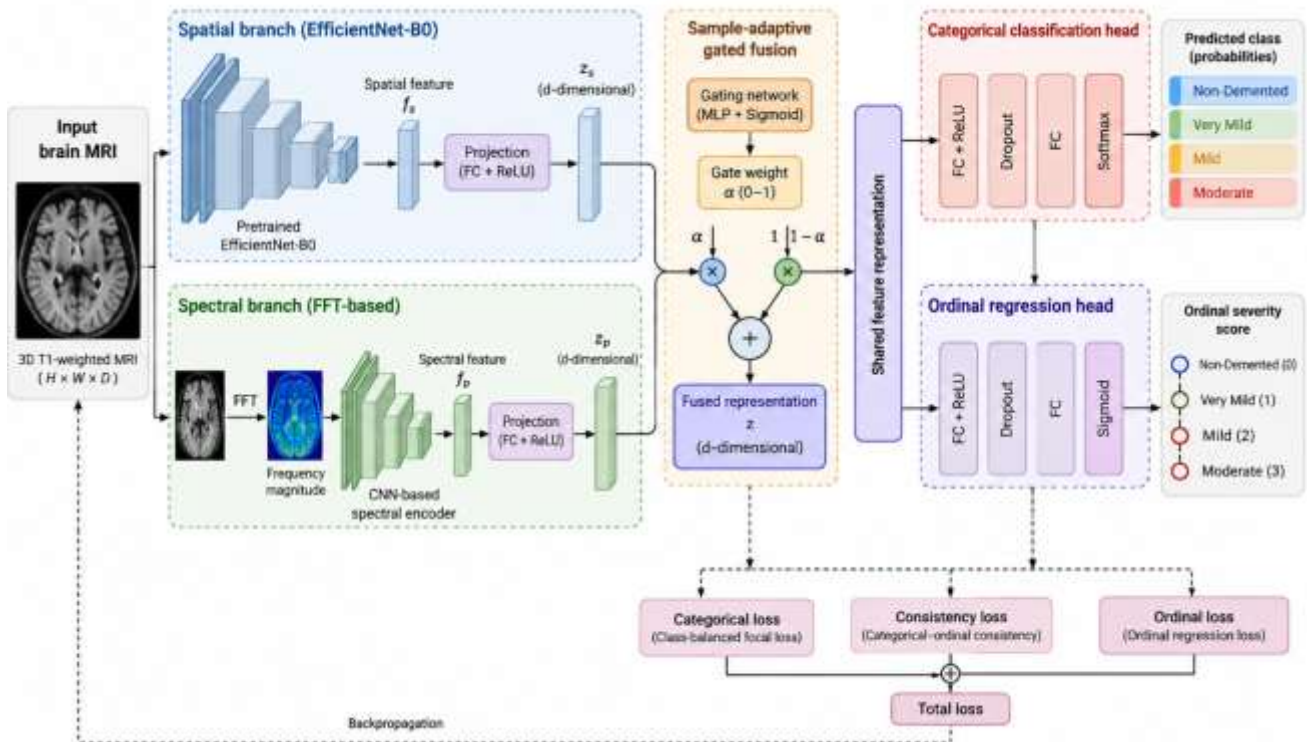


Fig. 1. Overall architecture of the proposed SOGAF-Net. The input brain MRI image is processed through parallel EfficientNet-B0 spatial and FFT-based spectral branches. The resulting representations are projected into a common latent space and integrated using sample-adaptive gated fusion. The fused representation is subsequently shared by categorical and ordinal prediction heads.

2.1 Dataset Preparation

The study uses a publicly available Alzheimer MRI dataset comprising 6,400 brain MRI images distributed across four diagnostic categories: non-demented, very mild dementia, mild dementia, and moderate dementia. The dataset provides predefined training and test partitions containing 5,120 and 1,280 images, respectively. The predefined training partition was further divided using stratified sampling to obtain 4,352 training images and 768 validation images, while preserving the original test partition for independent evaluation.

For ordinal modelling, the diagnostic categories were arranged according to increasing disease severity and assigned numerical labels of 0, 1, 2, and 3, corresponding to non-demented, very mild dementia, mild dementia, and moderate dementia, respectively. This representation enables the proposed framework to learn both categorical discrimination and progressive relationships among disease stages.

Because the class distribution is imbalanced, stratified partitioning was employed to preserve class proportions, while imbalance was addressed explicitly within the optimisation objective rather than through duplication of minority-class images.

2.2 Image Preprocessing and Data Augmentation

All MRI images were converted to grayscale and resized to 224×224 pixels. Since EfficientNet-B0 was pretrained using three-channel ImageNet images, each grayscale MRI image was replicated across three channels before being passed to the spatial branch. Standard ImageNet channel-wise normalisation was subsequently applied.

Data augmentation was applied only to the training partition. The augmentation procedure comprised random rotations within $\pm 6^\circ$, translations of up to 3% of the image dimensions, isotropic scaling between 0.97 and 1.03, and small brightness and contrast perturbations. These transformations were selected to introduce moderate variability while preserving anatomical structure. Validation and test images underwent deterministic resizing, grayscale conversion, and channel replication only.

2.3 Spatial Feature Extraction

The spatial branch captures anatomical and morphological information directly from the MRI image. EfficientNet-B0 pretrained on ImageNet was employed as the spatial backbone because it provides a favourable balance between representational capacity and computational efficiency.

For an input image X , the spatial feature representation is defined as

$$F_s = \phi_s[\text{GAP}(\mathcal{E}(X; \theta_s))] \quad (1)$$

where $\mathcal{E}(\cdot)$ denotes the EfficientNet-B0 feature extractor parameterised by θ_s , $\text{GAP}(\cdot)$ represents global average pooling, and $\phi_s(\cdot)$ denotes linear projection followed by layer normalisation and SiLU activation. The original EfficientNet feature representation is projected into a 256-dimensional latent vector F_s . During the initial optimisation stage, the pretrained backbone is frozen to stabilise the newly introduced model components and is subsequently unfrozen for joint end-to-end fine-tuning.

2.4 Spectral Feature Extraction

Spatial-domain features primarily capture anatomical structure, whereas complementary image characteristics can be represented in the frequency domain. SOGAF-Net therefore incorporates a parallel FFT-based spectral branch. The three-channel MRI representation is first converted into a single-channel image by channel averaging. A two-dimensional FFT is subsequently applied, followed by frequency shifting to centre the low-frequency components. The corresponding log-amplitude spectrum is computed as

$$A(u, v) = \log[1 + |\text{FFTShift}(\mathcal{F}_{2D}(X))|] \quad (2)$$

where $\mathcal{F}_{2D}(\cdot)$ denotes the two-dimensional Fourier transform and (u, v) represents spatial-frequency coordinates. The log-amplitude spectrum is standardised independently for each image using its mean and standard deviation. The normalised spectral representation is then processed by a lightweight convolutional extractor comprising four convolutional stages with batch normalisation and SiLU activation. Global average pooling and linear projection subsequently produce a 256-dimensional spectral feature vector F_f .

2.5 Sample-Adaptive Spatial-Spectral Fusion

A fixed feature-fusion strategy may not adequately account for variation in the relative informativeness of spatial and spectral characteristics across individual MRI images. SOGAF-Net therefore employs a learnable sample-adaptive gating mechanism.

The spatial and spectral feature vectors are concatenated and passed through a two-layer gating network:

$$G = \sigma(W_{g2} \delta(W_{g1}[F_s \parallel F_f] + b_{g1}) + b_{g2}) \quad (3)$$

where $[F_s \parallel F_f]$ denotes feature concatenation, W_{g1} and W_{g2} are learnable weight matrices, b_{g1} and b_{g2} are bias vectors, $\delta(\cdot)$ denotes the SiLU activation function, and $\sigma(\cdot)$ represents the sigmoid function.

The fused representation is obtained as

$$F = G \odot F_s + (1 - G) \odot F_f \quad (4)$$

where $\odot$ denotes element-wise multiplication.

The gate therefore regulates the relative contribution of spatial and spectral information for each sample and latent feature dimension. A dropout layer with a rate of 0.35 is applied to the fused representation before prediction.

2.6 Categorical Classification Head

The fused representation F is passed to a fully connected categorical classification head that produces logits corresponding to the four Alzheimer's disease severity classes. A softmax transformation converts these logits into posterior class probabilities, and the class associated with the maximum probability is selected as the categorical prediction.

The categorical branch provides conventional multi-class discrimination, while a parallel ordinal branch introduces explicit supervision based on disease-stage progression.

2.7 Ordinal Disease-Severity Learning

The four diagnostic categories follow an inherent progression from non-demented to moderate dementia. Conventional multi-class classification does not explicitly encode this ordering. SOGAF-Net therefore introduces cumulative ordinal supervision.

For a ground-truth label y , the ordinal target at threshold k is defined as

$$t_k = \mathbb{I}(y > k), \quad k = 0, \dots, K - 2 \quad (5)$$

where $K = 4$ and $\mathbb{I}(\cdot)$ denotes the indicator function.

The resulting ordinal encodings are:

Non-demented: $[0, 0, 0]$

Very mild dementia: $[1, 0, 0]$

Mild dementia: $[1, 1, 0]$

Moderate dementia: $[1, 1, 1]$

The ordinal head receives the fused representation and produces three cumulative logits. After sigmoid transformation, the ordinal probabilities q_k are optimised using binary cross-entropy:

$$\mathcal{L}_{\text{ord}} = -\frac{1}{K-1} \sum_{k=0}^{K-2} [t_k \log(q_k) + (1 - t_k) \log(1 - q_k)] \quad (6)$$

This formulation explicitly incorporates progression across adjacent disease stages during training.

2.8 Class-Balanced Focal Classification Loss

The dataset exhibits substantial class imbalance, particularly for the more advanced dementia categories. To reduce bias towards dominant classes, SOGAF-Net combines effective-number class weighting with focal learning. The categorical classification objective is defined as

$$\mathcal{L}_{\text{cls}} = -w_y(1 - p_y)^\gamma \log(p_y) \quad (7)$$

where w_y denotes the class-balanced weight associated with the ground-truth class, p_y represents the predicted probability assigned to that class, and γ is the focal parameter. The implementation uses $\gamma = 2$.

This formulation reduces the relative contribution of easily classified samples while assigning greater emphasis to difficult and underrepresented examples.

2.9 Categorical-Ordinal Consistency Learning

The categorical and ordinal branches estimate the same underlying disease-severity state using different formulations. SOGAF-Net therefore incorporates an explicit consistency constraint to encourage agreement between the two outputs.

The categorical severity estimate is obtained from the expected value of the class-probability distribution, whereas the ordinal severity estimate is derived from the cumulative ordinal probabilities. The consistency objective is defined as

$$\mathcal{L}_{\text{cons}} = \text{SmoothL1}\left(\sum_{k=0}^{K-1} k, p_k \sum_{k=0}^{K-2} q_k\right) \quad (8)$$

where p_k denotes the categorical probability of class k and q_k represents the cumulative ordinal probability at threshold k .

This constraint encourages the categorical and ordinal prediction mechanisms to represent compatible levels of disease severity.

2.10 Overall Optimisation Objective

The complete SOGAF-Net learning objective combines categorical classification, ordinal supervision, and categorical-ordinal consistency:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \lambda_{\text{ord}} \mathcal{L}_{\text{ord}} + \lambda_{\text{cons}} \mathcal{L}_{\text{cons}} \quad (9)$$

where λ_{ord} and λ_{cons} control the relative contributions of ordinal and consistency supervision. The implemented configuration uses $\lambda_{\text{ord}} = 0.40$ and $\lambda_{\text{cons}} = 0.10$.

The combined formulation enables joint optimisation of categorical discrimination and progression-aware representations while accounting for class imbalance.

2.11 Training Strategy

SOGAF-Net was implemented using PyTorch and optimised using AdamW. The initial learning rate was 2×10^{-4} , the weight decay was 1×10^{-4} , and a batch size of 32 was employed.

The pretrained EfficientNet-B0 backbone remained frozen during the first two epochs and was subsequently unfrozen for joint fine-tuning. A ReduceLROnPlateau scheduler monitored validation balanced accuracy and reduced the learning rate by a factor of 0.5 when improvement stagnated.

Gradient clipping with a maximum norm of 5 was applied to improve optimisation stability. Mixed-precision computation was employed when GPU acceleration was available. Early stopping was used when the validation criterion failed to improve for five consecutive epochs.

Model checkpoint selection was based exclusively on validation data using a composite criterion derived from balanced accuracy and macro-F1. The held-out test partition was not used during model training, hyperparameter selection, checkpoint selection, or calibration.

2.12 Probability Calibration

Post-hoc temperature scaling was incorporated to calibrate the posterior probabilities generated by the categorical classification head.

Given categorical logits z_k , the calibrated probability for class k is defined as

$$p_k^{(T)} = \frac{\exp(z_k/T)}{\sum_{j=0}^{K-1} \exp(z_j/T)} \quad (10)$$

where $T > 0$ denotes a scalar temperature parameter.

The temperature parameter was estimated exclusively from the validation partition by minimising categorical negative log-likelihood. The estimated parameter was subsequently applied unchanged to the test logits. The test partition was therefore not used during probability calibration.

2.13 Model Explainability

Gradient-weighted Class Activation Mapping (Grad-CAM) was employed for post-hoc visual interpretation of categorical predictions. Gradients corresponding to the predicted class were propagated to the final convolutional stage of the EfficientNet-B0 spatial branch.

Channel-wise importance coefficients were obtained by global averaging of these gradients and were combined with the corresponding feature maps to produce a class-specific activation map. Rectification was applied to retain positively contributing regions, after which the activation map was resized to the input image dimensions and superimposed on the original MRI image.

Grad-CAM was used only as an explanatory procedure and did not influence network optimisation or classification.

SOGAF-Net integrates spatial, spectral, and ordinal representation learning within a unified multi-objective architecture. Each MRI image is processed through two parallel pathways. The pretrained EfficientNet-B0 branch captures anatomical and morphological characteristics in the spatial domain, while the FFT-based branch models complementary frequency-domain information.

The resulting feature representations are projected into a common 256-dimensional latent space and combined through a sample-adaptive gating mechanism. The fused representation is shared by categorical and ordinal prediction heads, enabling simultaneous multi-class discrimination and progression-aware learning.

Training is governed by the combined objective in Eq. (9), comprising class-balanced focal loss, ordinal binary cross-entropy, and categorical-ordinal consistency. Following optimisation, temperature scaling is applied using validation data, and Grad-CAM is employed for post-hoc visualisation. The proposed formulation therefore integrates spatial-spectral representation learning, adaptive feature fusion, ordinal supervision, imbalance-aware optimisation, probability calibration, and model explainability within a unified framework.

3. Results

This section presents the quantitative and qualitative evaluation of SOGAF-Net on the held-out Alzheimer MRI test partition. Performance was assessed using overall accuracy, balanced accuracy, macro-averaged precision, recall and F1-score, Matthews correlation coefficient (MCC), quadratic Cohen's κ , macro one-vs-rest ROC-AUC, macro PR-AUC, calibration measures, ordinal severity error, bootstrap confidence intervals, and class-wise performance. The training process, diagnostic curves, confusion matrix, and Grad-CAM visualisations are summarised in Figs. 2–4.

3.1 Training Behaviour

The optimisation process showed progressive improvement across the training epochs. Validation balanced accuracy increased from 0.5909 during the first epoch to 0.9915 at the final epoch, while validation macro-F1 increased from 0.3941 to 0.9882. Simultaneously, validation loss decreased from 0.2606 to 0.0090, indicating stable convergence of the proposed framework.

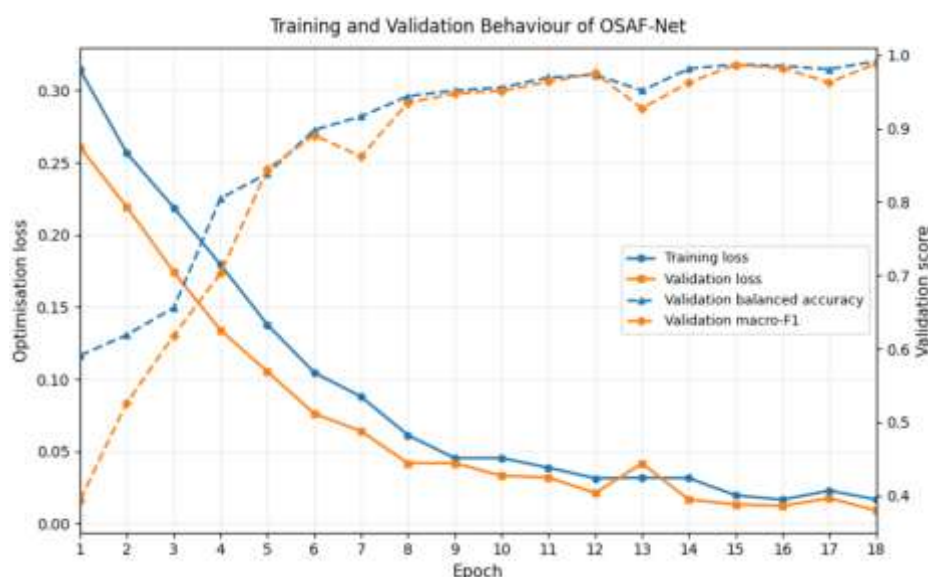


Fig. 2. Training and validation behaviour of SOGAF-Net in terms of optimisation loss, balanced accuracy, and macro-F1 across epochs.

3.2 Overall Classification Performance

Table 1 summarises the overall performance of SOGAF-Net on the independent test partition. The model achieved an accuracy of 96.64% and a balanced accuracy of 97.43%. Macro-precision, macro-recall, and macro-F1 were 96.79%, 97.43%, and 97.10%, respectively. The weighted F1-score reached 96.65%. In addition, the model achieved an MCC of 0.9449 and a quadratic Cohen's κ of 0.9704, indicating strong agreement between the predicted and ground-truth disease stages.

Table 1. Overall test performance of SOGAF-Net

Metric	Value
Accuracy (%)	96.64
Balanced Accuracy (%)	97.43
Macro Precision (%)	96.79
Macro Recall (%)	97.43

Metric	Value
Macro F1-score (%)	97.10
Weighted F1-score (%)	96.65
MCC	0.9449
Quadratic Cohen's κ	0.9704
Macro ROC-AUC	0.9983
Macro PR-AUC	0.9961

The macro one-vs-rest ROC-AUC reached 0.9983, while macro PR-AUC was 0.9961, indicating strong discrimination across the four severity categories.

3.3 Class-Wise Performance

Class-specific performance is reported in Table 2. The non-demented category achieved a precision of 98.25%, recall of 97.63%, and F1-score of 97.94%. For very mild dementia, precision, recall, and F1-score were 95.61%, 94.99%, and 95.30%, respectively. The mild dementia category achieved a precision of 93.30%, recall of 97.09%, and F1-score of 95.16%. The moderate dementia category obtained 100% precision, recall, and F1-score on the available test samples.

Table 2. Class-wise classification performance

Class	Precision (%)	Recall (%)	F1-score (%)	Support
Non-Demented	98.25	97.63	97.94	634
Very Mild Dementia	95.61	94.99	95.30	459
Mild Dementia	93.30	97.09	95.16	172
Moderate Dementia	100.00	100.00	100.00	15

Although the moderate dementia class achieved perfect class-wise scores, this category contained only 15 test images; therefore, these values should be interpreted in the context of the substantially smaller sample size.

3.4 ROC, Precision-Recall, and Confusion Matrix Analysis

The normalised confusion matrix, one-vs-rest ROC curves, and precision-recall curves are presented in Fig. 3. The strong macro ROC-AUC and PR-AUC values indicate that the network maintained high discriminative ability across the complete set of disease stages rather than relying solely on the dominant classes. The class-wise precision and recall values further indicate that performance remained comparatively balanced across non-demented, very mild, and mild dementia groups.

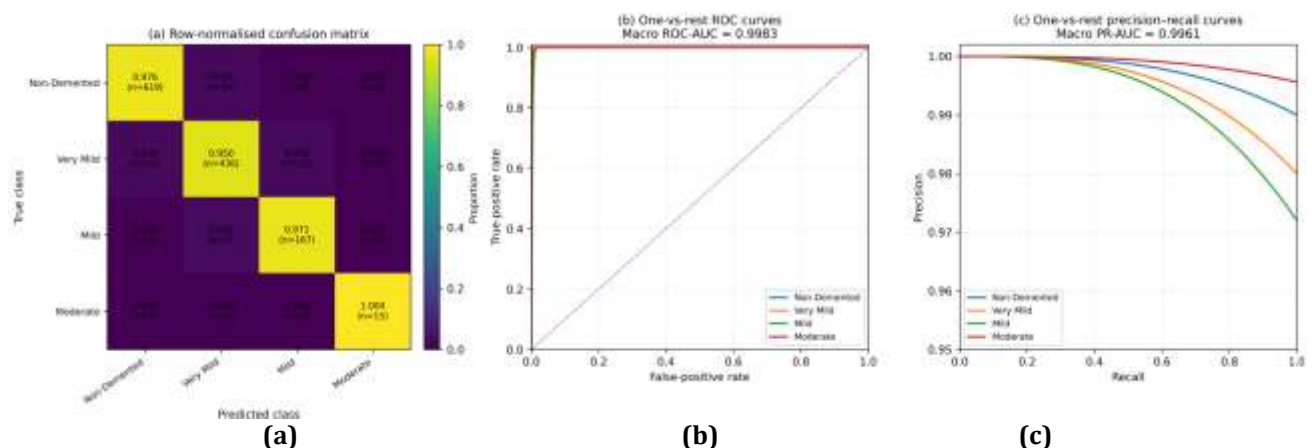


Fig. 3. Diagnostic performance of SOGAF-Net: (a) row-normalised confusion matrix, (b) one-vs-rest ROC curves, and (c) one-vs-rest precision-recall curves.

3.5 Statistical Uncertainty

Bootstrap analysis with 1,000 resamples was used to estimate 95% confidence intervals for the principal test metrics. Accuracy was 96.64%, with a 95% confidence interval of 95.55–97.58%. Balanced accuracy was 97.43%, with a corresponding interval of 96.47–98.16%. Macro-F1 was 97.10%, with an interval of 96.09–97.92%, while macro ROC-AUC was 0.9983, with a 95% interval of 0.9973–0.9990.

Table 3. Bootstrap 95% confidence intervals

Metric	Point Estimate	95% CI
Accuracy	0.9664	0.9555–0.9758
Balanced Accuracy	0.9743	0.9647–0.9816
Macro F1-score	0.9710	0.9609–0.9792
Macro ROC-AUC	0.9983	0.9973–0.9990

3.6 Probability Calibration and Ordinal Performance

Temperature scaling estimated a calibration temperature of 0.4850 from the validation partition. Following calibration, the model achieved a log loss of 0.0901, multiclass Brier score of 0.0511, and expected calibration error of 0.0145.

The severity mean absolute error was 0.0336, while the within-one-stage accuracy reached 1.0000. These results indicate that classification errors, when present, remained confined to neighbouring severity categories rather than spanning more distant disease stages.

3.7 Adaptive Fusion Behaviour

The mean spatial gate value across the test partition was 0.6678. This indicates that the learned fusion mechanism allocated a greater average contribution to the spatial representation while retaining a non-negligible contribution from the spectral branch.

The gate value represents an aggregate statistic and does not imply a fixed weighting across samples, as the gating mechanism operates independently for each input and latent feature dimension.

3.8 Grad-CAM Visualisation

Grad-CAM was used to examine the spatial evidence associated with predictions from each disease category. Representative activation maps for non-demented, very mild dementia, mild dementia, and moderate dementia are shown in Fig. 4. The visualisation procedure highlights regions of the MRI contributing most strongly to the categorical decision and provides a qualitative assessment of the spatial features learned by the EfficientNet-B0 branch.

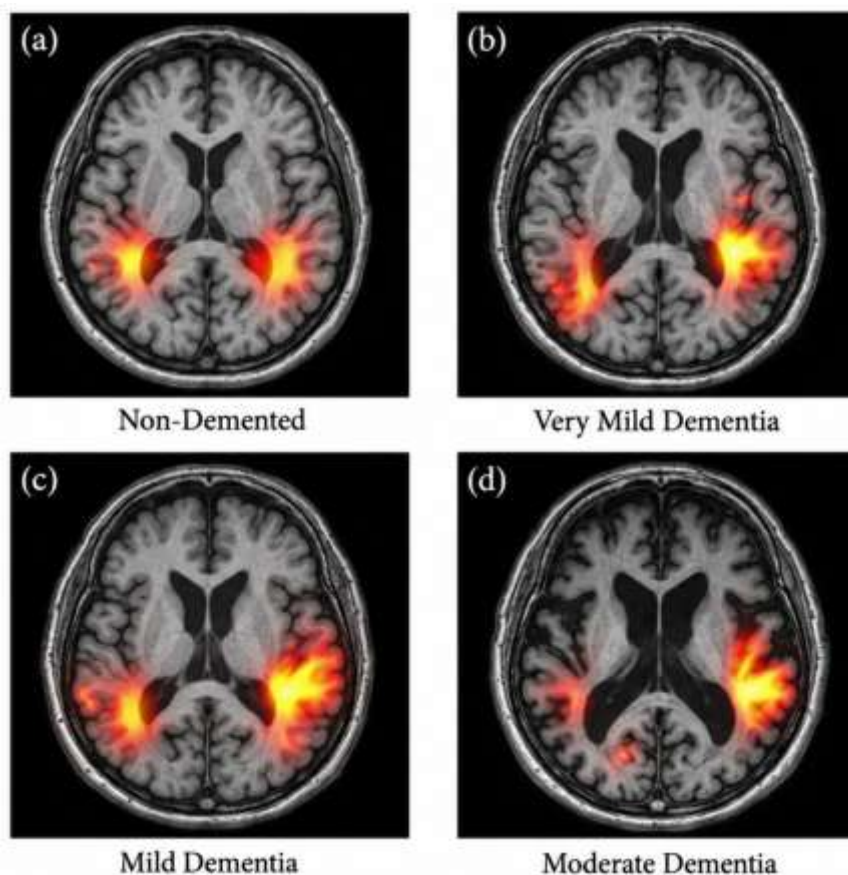


Fig. 4. Representative Grad-CAM visualisations for the four Alzheimer's disease severity categories. Heatmaps indicate spatial regions contributing to the corresponding SOGAF-Net categorical predictions.

Across the independent test partition, SOGAF-Net achieved 96.64% accuracy, 97.10% macro-F1, and a macro ROC-AUC of 0.9983. The bootstrap intervals remained narrow for the principal evaluation measures, while the class-wise analysis showed comparatively balanced performance across the major diagnostic categories. Probability-calibration analysis yielded a low expected calibration error, and the ordinal evaluation showed that all test predictions remained within one severity stage of the corresponding ground-truth category.

4. Discussion

The proposed SOGAF-Net demonstrated strong performance for four-class Alzheimer's disease severity classification by jointly modelling spatial, spectral, and ordinal information. The test accuracy of 96.64%, macro-F1 of 97.10%, and macro ROC-AUC of 0.9983 indicate that the framework maintained high discriminative ability across multiple disease stages rather than relying solely on the dominant class distribution.

The architecture was designed to address three methodological challenges: complementary feature representation, ordered disease progression, and class imbalance. The EfficientNet-B0 branch captures anatomical structure, whereas the FFT branch provides frequency-domain information. Adaptive gating allows their contributions to vary across samples, while ordinal supervision and the categorical-ordinal consistency term incorporate the progressive relationship among severity levels. The low severity error and within-one-stage accuracy further support the suitability of progression-aware learning for this task.

Several limitations should nevertheless be considered. The dataset does not provide subject identifiers, preventing verification of patient-level independence between partitions. The moderate dementia category is also considerably underrepresented, with only 15 test images, so its perfect class-wise performance should be interpreted cautiously. In addition, external validation on independent subject-level datasets and formal component ablation are required to establish generalisability and quantify the contribution of each proposed module.

Overall, the findings suggest that integrating spatial and spectral representations with ordinal supervision provides a promising strategy for multi-stage Alzheimer's disease MRI classification, while future work should focus on external validation, larger balanced cohorts, and patient-level evaluation.

5. Conclusions

This study presented SOGAF-Net, an Ordinal-Spectral Adaptive Fusion Network for four-class Alzheimer's disease severity classification from brain MRI. The proposed framework combines EfficientNet-B0-based spatial features, FFT-derived spectral representations, sample-adaptive feature fusion, ordinal supervision, and categorical-ordinal consistency learning within a unified architecture.

The model achieved strong performance across multiple evaluation measures, demonstrating the potential of jointly exploiting spatial, spectral, and progression-aware information for multi-stage Alzheimer's disease classification. The incorporation of class-balanced focal learning, probability calibration, and Grad-CAM further supports robust and interpretable model assessment.

Future work will focus on subject-level external validation using independent MRI cohorts, larger and more balanced datasets, and comprehensive ablation analysis to quantify the contribution of individual SOGAF-Net components.

Funding

This research received no external funding.

Data Availability Statement

The data used in this study are publicly available from the Alzheimer MRI dataset hosted on Hugging Face. The analysis was conducted using secondary brain MRI images classified into four categories: non-demented, very mild dementia, mild dementia, and moderate dementia. No new primary data were collected for this study. The dataset and associated documentation are available at:

Falah/Alzheimer_MRI – Hugging Face Datasets

https://huggingface.co/datasets/Falah/Alzheimer_MRI

Conflicts of Interest

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

Acknowledgement

This research was not funded by any grant.

References

- [1] R. D. Resen, L. S. Alzubaidi, H. A. Alwzazy, and M. I. Capel, "Review of deep learning models for Alzheimer's disease detection: MRI-centric approaches and multimodal extensions," *Frontiers in Artificial Intelligence*, vol. 9, Art. no. 1707043, 2026, doi: 10.3389/frai.2026.1707043.
- [2] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clinical*, vol. 21, Art. no. 101645, 2019, doi: 10.1016/j.nicl.2018.101645.
- [3] S. Divya, L. Padma Suresh, and A. John, "A deep transfer learning framework for multi class brain tumor classification using MRI," in *Proc. 2nd Int. Conf. Advances in Computing, Communication Control and Networking (ICACCCN)*, Greater Noida, India, 2020, pp. 283–290, doi: 10.1109/ICACCCN51052.2020.9362908.

- [4] S. Divya, L. Padma Suresh, and A. John, "Enhanced deep-joint segmentation with deep learning networks of glioma tumor for multi-grade classification using MR images," *Pattern Analysis and Applications*, vol. 25, no. 4, pp. 891–911, 2022, doi: 10.1007/s10044-022-01064-5.
- [5] S. Divya, A. Ali, N. Ibrahim, and L. Padma Suresh, "Automated models for the classification of magnetic resonance brain tumour images," in *Proc. 28th Int. Conf. Automation and Computing (ICAC)*, Birmingham, U.K., 2023, pp. 1–6, doi: 10.1109/ICAC57885.2023.10275235.
- [6] S. Alp, T. Akan, M. S. Bhuiyan, E. A. Disbrow, S. A. Conrad, J. A. Vanchiere, C. G. Kevil, and M. A. N. Bhuiyan, "Joint transformer architecture in brain 3D MRI classification: Its application in Alzheimer's disease classification," *Scientific Reports*, vol. 14, Art. no. 8996, 2024, doi: 10.1038/s41598-024-59578-3.
- [7] L. S. Saoud and H. AlMarzouqi, "Explainable early detection of Alzheimer's disease using ROIs and an ensemble of 138 3D vision transformers," *Scientific Reports*, vol. 14, Art. no. 27756, 2024, doi: 10.1038/s41598-024-76313-0.
- [8] J. Chen, Y. Wang, A. Zeb, M. D. Suzaiddola, Y. Wen, and Alzheimer's Disease Neuroimaging Initiative, "Multimodal mixing convolutional neural network and transformer for Alzheimer's disease recognition," *Expert Systems with Applications*, vol. 259, Art. no. 125321, 2025, doi: 10.1016/j.eswa.2024.125321.
- [9] D. Zhang, J. Zhang, B. Liu, M. Liu, W. Chen, and G. Du, "HDFT-MViT: A progressive core-enhanced mix framework for Alzheimer's disease classification using MRI images," *Frontiers in Neurology*, vol. 17, Art. no. 1860368, 2026, doi: 10.3389/fneur.2026.1860368.
- [10] D. Zhang, J. Zhang, W. Chen, and G. Du, "GGDA-net: Geometry-guided deformable attention network for Alzheimer's disease image classification," *Frontiers in Neuroscience*, vol. 20, Art. no. 1838681, 2026, doi: 10.3389/fnins.2026.1838681.
- [11] M. A. M. Joy, S. Nasrin, A. Siddiqua, and D. M. Farid, "ViTAD: Leveraging modified vision transformer for Alzheimer's disease multi-stage classification from brain MRI scans," *Brain Research*, vol. 1847, Art. no. 149302, 2025, doi: 10.1016/j.brainres.2024.149302.
- [12] A. G. M. Almansour, F. Alshomrani, A. T. M. Almutairi, E. Alalwany, M. S. Alshuhri, H. Alshaari, and A. Alfahaid, "DA-ViT: Deformable Attention Vision Transformer for Alzheimer's Disease Classification from MRI Scans," *Computer Modeling in Engineering & Sciences*, vol. 144, no. 2, pp. 2395–2418, 2025, doi: 10.32604/cmes.2025.069661.
- [13] S. S. Hussain, X. Degang, P. M. Shah, H. Khan, and A. Zeb, "AlzFormer: Multi-modal framework for Alzheimer's classification using MRI and graph-embedded demographics guided by adaptive attention gating," *Computerized Medical Imaging and Graphics*, vol. 124, Art. no. 102638, 2025, doi: 10.1016/j.compmedimag.2025.102638.
- [14] F. Sheikh, A. Al Marouf, J. G. Rokne, and R. Alhajj, "Lightweight deep learning models with explainable AI for early Alzheimer's detection from standard MRI scans," *Diagnostics*, vol. 15, no. 21, Art. no. 2709, 2025, doi: 10.3390/diagnostics15212709.
- [15] A. Valoor and G. R. Gangadharan, "Unveiling the decision making process in Alzheimer's disease diagnosis: A case-based counterfactual methodology for explainable deep learning," *Journal of Neuroscience Methods*, vol. 413, Art. no. 110318, 2025, doi: 10.1016/j.jneumeth.2024.110318.