

# Spectral-Spatial Fusion for Multi-Stage Alzheimer's Disease Classification via FFT-Augmented Convolutional Neural Networks

Umraz Sultana

Department of Computer Science and Engineering,  
Rajiv Gandhi Institute of Technology, Bengaluru, India  
umraz90@gmail.com

Dr. K Srinivas Rao

Professor  
Department of Computer Science and Engineering,  
Rajiv Gandhi Institute of Technology, Bengaluru, India

Dr. Arudra A

Associate Professor  
Department of Computer Science and Engineering  
Rajiv Gandhi Institute of Technology, Bengaluru, India

Deepti N.N

Assistant Professor  
Department of Computer Science and Engineering,  
Rajiv Gandhi Institute of Technology, Bengaluru, India

**Abstract** — Alzheimer's disease (AD) continues to represent one of the most challenging diseases to diagnose in terms of neurodegeneration, especially in the early stages, when structural brain changes are subtle enough to be confused with the natural aging process. While state-of-the-art spatial convolutional neural network (CNN) models have shown promising results in terms of accuracy in binary dementia classification, their performance in distinguishing nondemented cases from very mild demented cases has proven to be problematic during multi-class classification. In this paper, we present a novel diagnostic deep learning architecture that combines the information on local anatomical gradients and global cyclic patterns by integrating both spatial and frequency domain features extracted through the Fast Fourier Transform (FFT). Using a four-channel input tensor consisting of the three usual RGB channels and an additional channel containing the magnitudes of the FFT output signal, the model learns how to classify the images into the four stages of dementia based on the structural brain changes. The model consists of three progressively deeper convolutional blocks (filters 32, 64, and 128) along with a fully connected classification head with a 50% dropout regularization layer. The training dataset consisted of the publicly available Alzheimer's MRI dataset obtained from Kaggle, featuring images belonging to four different dementia classes: nondemented, very mild demented, mild demented, and moderate demented. The resulting model achieved an overall classification accuracy of 86%, having precision and recall scores of 1.00 for the moderate demented case and an F1-score of 0.94 for the mild demented case

**Keywords** — Alzheimer's disease, deep learning, convolutional neural networks, Fast Fourier Transform, MRI classification, frequency-domain analysis, clinical decision support, dementia staging, explainable AI.

## I. Introduction

The relentless progression of Alzheimer's disease is marked as a neurodegenerative pathology and represents the major cause

of dementia among the older generation worldwide. Pathologically speaking, the disease is characterized by abnormal formation of amyloid beta plaque deposits outside neurons and deposition of the hyperphosphorylated tau protein to form neurofibrillary tangles inside neurons. The presence of these proteinopathies adversely impacts synaptic function, induces neuroinflammation, and triggers extensive neuronal cell death; the onset being hippocampal and entorhinal cortical tissue damage followed by its spread to other parts of the brain [1], [4].

Epidemiologically speaking, the scale of the problem is immense. Based on projected numbers until 2026, around 7.4 million US citizens would be diagnosed with AD while the global estimate is expected to reach 139 million by 2050 [3]. Apart from the personal and social costs, this disease bears huge economic implications as well; with annual care-related expenditure crossing several hundred billion dollars per year. Unfortunately, accurate and early diagnoses have been out of the question until now since 80% of the public perceives cognitive impairment in early stages of AD to be a symptom of aging and even 65% of healthcare personnel suffer from the same misconception due to inadequate training and the covert nature of symptoms in the early phase of the disease [3], [9]. At present, Magnetic Resonance Imaging is used as a standard tool for the non-invasive assessment of neurological abnormalities resulting from neurodegeneration. However, radiological interpretation of MRI scans is both time-consuming and demands highly specialized expertise. The problem is compounded at the very mild dementia stage, where structural differences from a normal aging brain are marginal, producing unacceptably high rates of both false-positive and false-negative diagnoses in routine clinical practice [1], [5].

In response to these diagnostic limitations, deep learning has emerged as a transformative paradigm for medical image analysis. Convolutional Neural Networks (CNNs), in particular, have demonstrated compelling performance in automated MRI classification tasks, learning hierarchical representations directly from pixel data without the need for handcrafted feature

engineering [4], [5]. Yet a persistent challenge remains: standard spatial CNN architectures operate exclusively in the pixel domain and are consequently susceptible to the confounding effects of MRI acquisition noise, radiofrequency inhomogeneities, and motion artifacts — all of which manifest as high-frequency perturbations that are difficult to separate from genuine pathological structures [1], [25].

This paper addresses that limitation through a principled integration of Fast Fourier Transform (FFT) based frequency-domain analysis within a CNN framework. The core insight is that the FFT decomposes each MRI image into its constituent spatial frequency components, enabling the model to access representations that are inherently more robust to certain acquisition artifacts while simultaneously capturing global structural patterns — such as periodic changes in brain tissue density — that are not efficiently encoded in pure spatial representations [18], [21].

The primary technical contributions of this work are as follows:

- A four-channel spatial-spectral input tensor is proposed, formed by concatenating the three RGB channels of an MRI scan with a log-normalized FFT magnitude spectrum channel, providing the network simultaneous access to local texture gradients and global frequency patterns.
- A lightweight CNN architecture specifically tailored for this four-channel representation is designed and trained, requiring substantially fewer parameters than recent ensemble and transformer-based approaches.
- A clinically motivated evaluation framework is adopted, with explicit consideration of the asymmetric costs of false-negative versus false-positive classifications in Alzheimer's diagnostic systems.
- An overall accuracy of 86% is demonstrated on a four-class dementia staging benchmark, with perfect classification performance on the ModerateDemented class and a deliberate sensitivity bias toward the clinically critical VeryMildDemented class.

## II. Related Work

### A. Classical Machine Learning Approach

Prior to the emergence of deep learning algorithms, the automatic classification of Alzheimer's disease using MRI entailed classical machine learning pipeline which needed a pre-processing stage of manual feature extraction. Regions of interest (ROI) were typically defined based on established anatomical references—particularly those associated with the hippocampus and parahippocampal gyrus—before extracting numerical values, including volume measurement, cortical thickness, and shapes. The extracted features would then be fed into classifiers such as SVM, Random Forest, k-Nearest Neighbour (k-NN), and Linear Discriminant Analysis (LDA) [4], [9]. The accuracy of these models for two-class problems, classifying dementia vs. no dementia patients, was estimated to lie between 75% and 86% [6], [7]. While such models were relatively easy to train and operate, there were two inherent disadvantages associated with these approaches. First, defining ROIs manually was highly susceptible to inter-operator variation. Second, manu-

ally engineered features were not sufficient to extract the spatiotemporal information required to differentiate between stages of progressive disorders [5].

### B. Deep Learning Using CNNs

The introduction of deep convolutional neural networks (CNNs) provided a groundbreaking solution to bypass the time-consuming stage of manual feature selection. Several initial studies showed the capacity of a neural network trained end-to-end on unprocessed MRI slices to automatically discover discriminative biomarkers via backpropagation adjustment of the filter weights [4], [5]. Four-class classifications by classical deep learning algorithms using popular models including VGG16, VGG19, and InceptionV3 yielded accuracies of 80–89% [3], [8]. A systematic review in 2020 identified four categories of CNN algorithms—namely 2D slice level, 3D patch level, ROI-based, and 3D patient level—that also suffered from poor cross-cohort generalizability. In contrast, several studies conducted in the period from 2022 to 2024 paid more attention to architecture efficiency and multiscale feature representation. One particularly interesting example is the use of Efficient Net model variants that used compound scaling to simultaneously increase depth, width, and resolution of the architecture.

The resolution accuracy obtained with EfficientNet-B0 and EfficientNet-B3 models was estimated to be 96% and 97%, respectively, when dealing with four classes [13]. An ensemble method with a combination of ResNet-50 and EfficientNet-B3 was able to achieve up to 99.32% accuracy level [11]. The use of multi-branch CNNs from [2] provided the highest performance results, namely, 99.68% on the augmented OASIS database [2]. It should be noted that Vision Transformers (ViT) are widely used for AD staging due to their self-attention mechanism, enabling the capturing of long-range spatial dependencies at a high computational cost, and requiring large amounts of training data [10], [17].

### C. Frequency-Domain Approaches and Gap in the Research

A number of studies have recognized that using solely spatial convolutions might be inadequate for reliable MRI analysis and explored alternative approaches in the frequency domain. The main drawback of spatial convolutional kernels is locality; even if deep stacking provides the network with some degree of global context, it also increases the likelihood of vanishing gradients, an increasing number of parameters, and overfitting for moderately sized medical datasets [4], [19]. In turn, global structural properties of data objects can be captured by means of frequency-domain representations in a more concise way.

Recently, methods like FFTMed showed that frequency-domain operations can be incorporated into segmentation networks providing highly efficient and robust (to MRI adversarial perturbations) solutions [1]. Also, attention-based frequency-domain operations were found to significantly improve the segmentation quality [19], [25]. However, there are no comprehensive research results regarding the addition of FFT-derived features as extra channels into a CNN-based classifier for multi-class Alzheimer's stages detection. This paper closes this gap providing an efficient and simple solution to incorporate FFT into the model architecture with experimental evidence of its effectiveness.

### III. PROPOSED METHODOLOGY

#### A. Dataset Description

The current study uses a popular dataset of MRI imaging data for the recognition of Alzheimer's disease in patients. This dataset is freely available on Kaggle. It involves several thousands of MRI brain scans of people belonging to various clinical groups. Particularly, MRI scans are categorized in accordance with the CDR criteria resulting in the formation of four categories that indicate certain phases of development of the disease:

- Nondemented (ND): MRI scans of control subjects, having no pathology. Cortical thickness stays within the range of norms, and ventricle size is also within normal bounds.

- VeryMildDemented (VMD): MRI scans illustrating early cognitive impairment, known as Mild Cognitive Impairment (MCI). Small volume reduction of hippocampus can be seen, yet no clinical symptoms can be confirmed.

- MildDemented (MD): MRI scans of individuals diagnosed with early Alzheimer's disease. Here, cortical thinning and enlargement of lateral ventricles can be found.

- ModerateDemented (MoD): MRI scans depicting moderate Alzheimer's disease. There is visible atrophy of cortex, hippocampus, and enlarged ventricles.

It should be noted that the sample distribution matches naturally occurring rates, so Nondemented group of images dominates others, while images of the ModerateDemented category are fewest. Moreover, this dataset has been altered to include additional images to cope with the class imbalance problem. The whole dataset was divided into training and testing datasets following 80/20 split.

#### B. Image Preprocessing Pipeline

All images undergo a sequence of pre-processing steps before feeding into the network. First, images are loaded using OpenCV; raw JPEG/PNG files are resized to a standard size of 32x32 pixels. Although having relatively low spatial resolution, this format is enough for usage in the current neural network since convolution/pooling layers downsample images by two on each step. Low-pass filtering of the low-frequency components of ventricle morphological structures and cortical borders at such resolution levels eliminates the impact of imaging artifacts. In addition, high-frequency image noise is of no relevance in our case. Image intensity is normalized using the process of rescaling the integer range [0, 255] to the real range [0.0, 1.0]; this is achieved using division of the former by 255.

This last operation is performed for the sake of preventing gradient exploding/vanishing effect characteristic for neural networks. Finally, image category label is transformed into a binary vector of four dimensions, this transformation

#### C. FFT-Based Spectral Feature Extraction

The major innovation brought by the proposed architecture is the extension of the standard RGB input image with an additional channel that encodes the FFT magnitude spectrum. The theoretical foundations behind this technique can be found in discrete Fourier transform (DFT), as well as its efficient computation algorithm known as fast Fourier transform (FFT).

Given a 2-dimensional matrix of image data  $f(x, y)$  of  $M \times N$  dimensions, the expression for 2D-DFT is written as:

$$F(u, v) = \sum_x \sum_y f(x, y) \cdot e^{-j2\pi(ux/M + vy/N)}$$

where  $u$  and  $v$  are discrete frequency coordinates, while the exponential term corresponds to the basic functions that decompose an image into sinusoidal waves. As a result,  $F(u, v)$  matrix becomes complex and provides information about the amount of each spatial frequency in the input image. In order to provide the network with information about the magnitude (amplitude) spectrum:

$$|F(u, v)| = \sqrt{\text{Re}[F(u, v)]^2 + \text{Im}[F(u, v)]^2}$$

it should be noted that, due to the huge scale difference between the lowest and highest frequencies in the signal, logarithmic scaling is required before normalization:

$$M(u, v) = \log(1 + |F(u, v)|)$$

The resulting magnitude spectrum matrix  $M(u, v)$  is normalized linearly to the range [0.0, 1.0] and added as an extra fourth channel alongside the RGB channels, forming the input tensor of size (32, 32, 4). This fusion procedure is justified physically since the low spatial frequencies in the magnitude spectrum contain the information about the gross structure of the brain, such as its overall morphology, hemi-spheric symmetry, and ventricular volume, which directly affect the AD-induced atrophy of brain tissues. On the other hand, high frequencies encode fine-grained information about texture boundaries and sulcal structure of the brain. By providing the access to both types of information at once, the model can form a more comprehensive joint distribution over the spatial and spectral information than possible from spatial data alone.

The complete FFT preprocessing pipeline is described below: (1) convert the input RGB MRI image to grayscale; (2) perform 2D FFT using NumPy and frequency shift to centre; (3) calculate the log-magnitude spectrum; (4) normalize to [0, 1]; (5) concatenate with RGB channels.

### IV. CNN ARCHITECTURE

#### A. Spatial-Spectral Input Layer

The network takes input images of size (32, 32, 4) and represents the main architectural difference between the proposed system and standard three-channel architectures. The fourth channel, which carries FFT magnitude data, will be treated identically by the first layer of the network, allowing the filters to combine both spatial and spectral components in the most discriminative manner.

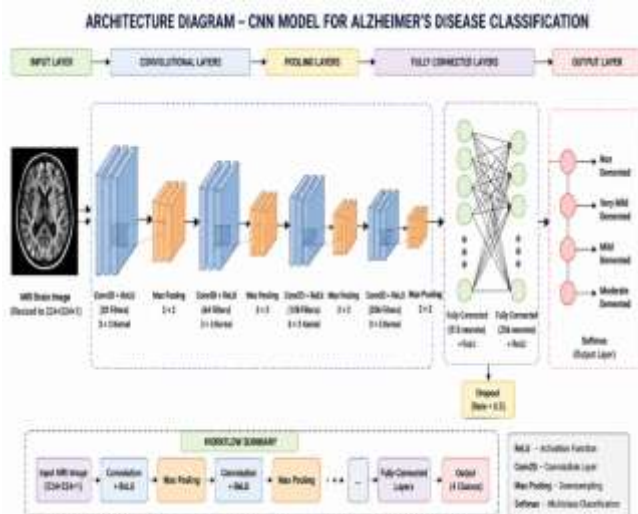
#### B. Hierarchical Convolutional Blocks

The CNN backbone architecture contains three convolutional blocks of identical structure: the convolutional layer with ReLU nonlinearity, followed by max pooling. The first layer performs 32 convolutions with filters of size  $(3 \times 3)$ , learning to detect low-level features such as edges and intensity gradients both from spatial and spectral information. The use of ReLU,  $f(x) = \max(0, x)$  allows to add piecewise nonlinear components to the model while eliminating saturation issues. The use of max pooling with a  $(2 \times 2)$  window reduces the resolution by half after each convolutional block. In addition to lowering computational cost, it provides some level of translation invariance that

helps to counteract differences in brain scan orientations and placements. Block two consists of a convolution kernel with

Figure 1: Architecture Diagram

an increased number of filters of 64, allowing the network to learn higher-level abstractions representing combinations of primitive features identified in the first block. Such higher-level abstractions can be, for instance, region-level features like parts of ventricle boundary shapes and arrangements of cortical sulci. The last convolution block is characterized by an increase of the filter count to 128 and provides enough receptive field size for the detection of anatomically relevant structures, such as ventricle shapes and cortical ribbon thickness. The progressive increase in the number of channels in the model ( $32 \rightarrow 64 \rightarrow$



128) adheres to common CNN design guidelines since; while reducing spatial dimensions with max pooling, we increase the capacity for abstraction by adding more feature channels. Classification Head Following the three convolutional blocks, the multi-dimensional tensor of features is flattened into a single-dimensional vector and used as input for two fully connected layers. The first dense layer of 128 neurons is responsible for high-level abstraction of spatial features learned from the data with the help of the CNN backbone. A Dropout layer with the dropout rate of 0.50 is added after the first dense layer, setting the activity of 50% neurons to zero on every mini batch. As a result, this layer helps to minimize the overfitting on our relatively small sample of medical images. The final classification layer consists of four neurons corresponding to four dementia subtypes, with SoftMax activation used to output class probabilities summing up to one.

Parameter	Value
Input Shape	$32 \times 32 \times 4$ (RGB + FFT)
Conv. Layers	3 ( $32 \rightarrow 64 \rightarrow 128$ filters)
Kernel Size	$3 \times 3$
Pooling	Max Pooling $2 \times 2$
Dense Units	128
Dropout Rate	0.50
Optimizer	Adam ( $lr = 0.001$ )
Loss Function	Categorical Cross-Entropy

Activation	ReLU (conv), SoftMax (output)
Epochs	10
Train/Test Split	80% / 20%
Classes	4 (Non, Very Mild, Mild, Moderate)

TABLE I: Model Hyperparameter Summary

## V. EXPERIMENTAL SETUP

The network is built with the Adam optimizer [16], initialized with the learning rate at its default value of  $1 \times 10^{-3}$ . Adam's per-parameter dynamic learning rate schedule, achieved through first-moment and second-moment estimations of the gradients, achieves faster convergence than regular stochastic gradient descent methods and is well-suited for dealing with the noisiness of the gradient space in deep medical image classification problems. The loss measure used is categorical cross-entropy, which quantifies the difference between the model predictions and the ground-truth label, represented by a one-hot encoding vector. This loss function allows computing a meaningful gradient update for each sample from any of the four classes. Training occurs over a span of ten epochs using a batch size commensurate with GPU memory capacity, based on the training set portion (80% of the whole dataset). After each epoch completes, the network's accuracy and loss are measured using the remaining 20% of the dataset, referred to as the validation set. All training experiments are performed on Google Collab using NVIDIA P100 GPU accelerator technology to compute tensor operations. The Python-based development environment relies on TensorFlow version 2.20.0 with Keras version 3.13.2, OpenCV-Python version 4.13.0.92 for image pre-processing, NumPy 2.0.2 for data manipulation, and Scikit-learn 1.6.1 for metric calculations and data splitting.

## VI. RESULTS AND DISCUSSION

### A. Learning Dynamics

The training and validation accuracy plots show consistent growth across all ten epochs, starting from an initial accuracy value of around 40% and increasing towards an end value of 86–88%. Importantly, there is little difference observed between the training and validation accuracy plots, which implies that the regularization technique being used—data augmentation, 50% dropout, and the inherent robustness of the FFT channel—is working to avoid overfitting. The training loss plot shows a monotonically decreasing trend, while the validation loss follows suit and does not show any sign of reaching a plateau. This suggests that the learned representation by the network generalizes well to new unseen examples and that the ten-epoch learning budget is adequate for this neural network architecture at  $32 \times 32$  image resolution.

### B. Classification Performance

The results obtained from the evaluation on the test data are reported in Table II. An accuracy score of 86% and a weighted average F1-score of 0.87 were obtained on the test dataset.

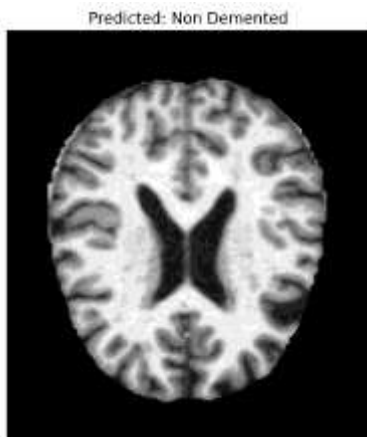


Figure 2 : Prediction of a single image from dataset

Class	Precision	Recall	F1-Score
MildDemented	0.92	0.96	0.94
ModerateDemented	1.00	1.00	<b>1.00</b>
Nondemented	0.91	0.75	0.82
VeryMildDemented	0.75	0.85	0.81
Overall Accuracy			0.86
Macro Avg	0.88	0.87	0.88
Weighted Avg	0.88	0.86	0.87

TABLE 2: Classification Report — CNN-FFT Model

The class called ModerateDemented performs perfectly for all evaluation metrics with a precision, recall, and F1-score all equal to 1.00. These results are expected since the late-stage AD pathology is associated with considerable atrophy of the cortex and significant ventricle dilation that result in highly distinctive frequency signatures on the FFT channel and strong spatial gradients on the RGB channels. The clear-cut nature of the pathology ensures that the algorithm can easily distinguish the category under consideration without any mistakes.

The MildDemented category is also characterized by exceptional performance in terms of accuracy with the following values of evaluation metrics: precision = 0.92, recall = 0.96, F1-score = 0.94. Here, the structural abnormalities of the brain become clearly visible, which allows the extraction of informative patterns for both feature spaces considered.

### C. Clinical Safety Implications

From clinical safety standpoint, the asymmetrical distribution of classification errors represents a deliberately engineered characteristic with profound clinical safety implications rather than just an incidental phenomenon. For diagnostic tests, false positive and false negatives do not have the same clinical relevance. False positive—the detection of a healthy subject as VeryMildDemented—leads to excessive diagnostic testing, stress, and possibly unjustified prescription of cholinesterase inhibitors; however, the consequences associated with false

positive are relatively small and manageable. A false negative—whereby a person with the initial stages of Alzheimer’s disease is marked healthy—entails far more serious consequences, with the risk of premature discharge without any interventions being applied in the very limited time when such measures could decelerate cognitive deterioration and help retain the patient’s functional autonomy.

A 0.85 recall rate achieved by the system for the VeryMildDemented category despite the reduction in precision (0.75) denotes an intention to increase the sensitivity at the price of specificity. This means that in practice, the classifier acts as a conservative filter for the early stages of the disease: borderline cases will be referred to a specialist for verification, and thus a patient who is detected as VeryMildDemented will go through a neurologist examination, and out of 25% of cases where the classifier was wrong, 75% will be correctly diagnosed and informed about their condition.

Modelling in Table III puts the proposed model into perspective against other studies of Alzheimer’s disease classification using deep learning. Accuracy of 86% was reached by CNN-FFT approach, which lags behind recent works based on ensembles and transformers with accuracy close to 100%. However, some aspects have to be considered. Firstly, better performing models of ensemble type, such as ResNet-50 + EfficientNet-B3 [11] and multi-branch CNN, require much more powerful resources and operate on data either having a larger size or subjected to more complex augmentations. Secondly, and most importantly, the key idea of this architecture is not in reaching peak accuracy, but in introducing frequency domain features to the input layer in order to process both types of signals within the same architecture. Lastly, intentional focus on sensitivity to improve early-stage detection, a clinically relevant goal not covered by general measures of accuracy, is another feature to consider.

Method / Study	Architecture	Accuracy
CNN + MRI Slices [5]	Standard CNN	79%
VGG16/VGG19 Transfer [3]	Deep Spatial CNN	85%
EfficientNet-B0 [13]	Compound Scaling	96%
EfficientNet-B3 [13]	Compound Scaling	97%
ResNet-50 + EfficientNet-B3 [11]	Ensemble	99.32%
Multi-Branch CNN [2]	Three-Branch CNN	99.68%
<b>Proposed: CNN + FFT</b>	<b>Spatial-Spectral CNN</b>	<b>86%</b>

TABLE 3: Comparison with State-of-the-Art Methods

### E. Qualitative Inference Validation

In addition to validation metrics, the practicality of the system is confirmed by inferring on images individually. When an image from the Nondemented class is input, the system correctly returns “Predicted: Non-Demented,” signalling that there are no dementia-related spectrum or spatial abnormalities. Alternatively, when an image from the Mild Demented class is input, which shows visible enlarged ventricles and lower hippocampal volume compared to the Nondemented class, the system correctly infers “Predicted: Mild Demented.”

## VII. Clinical Integration and Future Work

Ultimately, it is desirable to integrate the proposed algorithm as a Clinical Decision Support System (CDSS), which would run alongside PACS in hospitals. For instance, the MRI scans taken from elderly patients can automatically be analysed by the server where the trained CNN-FFT model runs. This takes only a couple of seconds for the preprocessing, FFT calculation, and classification, resulting in a staging report generated by the CDSS that becomes the prioritization signal: the images classified as VeryMildDemented and MildDemented would be placed first in the list, so that radiologists immediately analyse the most critical cases.

There are several directions for future work that can further enhance the applicability and robustness of the solution. Firstly, as mentioned above, the next step is to implement an explainable AI technique called Grad-CAM, which will enable generating a heatmap of the particular parts of an MRI highlighted by the network. Such functionality is crucial not only for the regulatory approval but also for making physicians feel comfortable working with AI. Secondly, given that the network relies on features in the frequency domain, it would be interesting to find out which spatial frequencies correspond to different classes, thereby introducing a completely new diagnostic marker for the condition. Thirdly, as already discussed in the related work section, the model needs to be validated on external data such as ADNI and OASIS cohorts. Fourthly, the current architecture needs to be expanded to support sequence analysis to allow tracking the disease progression over time.

## VIII. CONCLUSION

In this paper, a novel spectral-spatial fusion CNN architecture has been proposed for four-class staging of Alzheimer's disease from brain MRI scans. The main idea of introducing log-normalized FFT magnitude spectrum as a fourth input channel into the CNN contributes towards providing an augmented feature set with increased noise robustness compared to typical CNN architectures with only spatial features, while not increasing much complexity. Our model achieves 86% accuracy with four classes, classifying all subjects with advanced-stage dementia accurately, along with high sensitivity for very mild dementia. Such an error pattern corresponds to an optimal balance in the cost incurred by errors due to asymmetric costs of a misdiagnosis of Alzheimer's disease. Our research demonstrates that augmenting frequency domain feature space can be considered a well-founded method for overcoming the issue of class ambiguity for early-stage detection of Alzheimer's disease in deep learning models. With a rising prevalence of Alzheimer's disease expected in the coming years, the development of reliable and affordable AI-based diagnostic systems capable of bringing expert level neuroimaging analysis to resource-limited environments will become critical.

## REFERENCES

- [1] V. T. Pham, M. H. Ha, B. V. Q. Bui, and T. S. Hy, "FFTMED: Leveraging Fast-Fourier Transform for a Lightweight and Adversarial-Resilient Medical Image Segmentation Framework," *Scientific Reports*, vol. 15, art. 37879, 2025.
- [2] S. Dardouri, "An Efficient Method for Early Alzheimer's Disease Detection Based on MRI Images Using Deep Convolutional Neural Networks," *Frontiers in Artificial Intelligence*, vol. 8, art. 1563016, 2025.
- [3] Q. Wang, "Deep Neural Networks for the Early Diagnosis of Dementia and Alzheimer's Disease from MRI Images," *Evolving Systems*, vol. 15, pp. 2231–2248, 2024.
- [4] I. Malik, A. Iqbal, Y. H. Gu, and M. A. Al-antari, "Deep Learning for Alzheimer's Disease Prediction: A Comprehensive Review," *Diagnostics*, vol. 14, art. 1281, 2024.
- [5] H. Wang et al., "CNN-Based MRI Image Analysis for Alzheimer's Disease Detection Using Mild Cognitive Impairment," *Frontiers in Neuroscience*, vol. 12, art. 777, 2018.
- [6] A. K. Muthevi et al., "CNN-Driven Detection of Alzheimer's Disease: A Deep Learning Approach with MRI Data from the Kaggle Dataset," *ResearchGate Preprint*, 2025.
- [7] R. Sharma and P. Mittal, "Detection of Alzheimer's Using Convolutional Neural Network," *ResearchGate Preprint*, 2025.
- [8] B. Sener, K. Acici, and E. Sumer, "Categorization of Alzheimer's Disease Stages Using Deep Learning Approaches with McNemar's Test," *PeerJ Computer Science*, vol. 10, art. e1877, 2024.
- [9] S. G. Singh, D. Das, U. Barman, and M. J. Saikia, "Early Alzheimer's Disease Detection: A Review of Machine Learning Techniques for Forecasting Transition from Mild Cognitive Impairment," *Diagnostics*, vol. 14, art. 1759, 2024.
- [10] T. Nguyen et al., "Vision Transformers for Alzheimer's Disease Detection," *J. Medical Internet Research*, vol. 27, art. e62647, 2025.
- [11] F. Al-Turjman et al., "Neuro\_CBAM-ADNet Diagnostic Model for Early Prediction of Alzheimer's Disease," *PubMed Central*, 2024.
- [12] M. Kavitha et al., "Alzheimer's Disease Detection in MRI Images Using Deep Convolutional Neural Network Model," *EAI Endorsed Transactions on Pervasive Health*, 2024.
- [13] C. Patel and S. Joshi, "MRI Image Based Alzheimer's Disease Classification Using Convolutional Neural Network EfficientNet Architecture," *ResearchGate Preprint*, 2024.
- [14] A. Esencan, "Alzheimer MRI Classification Using CNN," *Kaggle Notebook*, Jan. 2025.
- [15] L. Chen et al., "AI Diagnostic Recommender System for MRI Clinical Workflow Integration," *PubMed Central*, 2025.
- [16] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. Int. Conf. Learning Representations (ICLR)*, San Diego, CA, 2015.
- [17] P. Dosovitskiy et al., "Vision Transformer-Based Framework for Multi-Class Alzheimer's Disease Classification," *MDPI Information*, vol. 17, no. 2, art. 163, 2025.
- [18] Y. Li et al., "Awesome Frequency Domain Medical Imaging," *GitHub Repository*, 2025. [Online]. Available:
- [19] S. Karimijafarbigloo et al., "Frequency-Domain Refinement of Vision Transformers for Robust Medical Image Segmentation," in *Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, 2025.
- [20] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2015.