International Journal of Basic and Applied Sciences, 14 (2) (2025) 542-550

SPC

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/syxzbg95
Research paper

A Novel Self-Supervised Swin Transformer with Wavelet Feature Extraction for Early Alzheimer's Disease Recognition

Jupalli Pushpakumari [1], Janapareddy Uttara Alekhya [2], Jakkapu Nagalakshmidevi [3], Mangu Maanasa [4] *, Vaddadi Vasudha Rani [5], M. Somasundara Rao [6], Chandra Sekhar Musinana [7]

[1] *Department of Computer Science and Engineering (AIML & IOT), Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India*
[2] *Department of Advanced Computer Science and Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam, Andhra Pradesh, India*
[3] *Department of Computer Science and Engineering, Aditya University, Surampalem, Andhra Pradesh, India*
[4] *Department of Computer Science and Engineering, GMR Institute of Technology(A), Rajam, Andhra Pradesh, India*
[5] *Department of Information Technology, GMR Institute of Technology(A), Rajam, Andhra Pradesh, India*
[6] *Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, Andhra Pradesh, India*
[7] *Department of Information Technology, MVGR College of Engineering, Vizianagaram, Andhra Pradesh, India*
**Corresponding author E-mail: manasam.0004@gmail.com*

Received: May 31, 2025, Accepted: June 25, 2025, Published: June 30, 2025

Abstract

Alzheimer's disease is a structural and functional brain alterations that induces memory loss and thinking problems, which leads to worsen over time. It's crucial to identify the problem early so that treatment can be provided on time and care can be more effective. In this research, we acquaint's a new deep learning technique that uses MRI brain scans to find whether the person is suffering with Alzheimer's disease or not. Our methodology integrates self-supervised learning where the model trains the scanned images of MRI without labelled data, advanced feature extraction techniques, and a modern vision transformer model. Initially, we used the FSL-BET tool to extracts only the relat-ed MRI images of the brains by using FSL (FMRIB Software Library). Then, we use a multi-level Wavelet Transform to acquire major features from the images and mainly focusing on texture and frequency data that helps to observe disease-related changes in the brain. After that, we trained a Swin Transformer model using self-supervised learning so it could acquire knowledge from the data without human label-ling. Now, we regulate the model to categorize the MRI scans into four classes: Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented. Finally, this research showed that this model got a highest accuracy of 97.8%. It performed well when compared to other popular deep learning models like VGG19 and ResNet, especially when only a small amount of labelled data was available. The pri-mary benefit of wavelet features with self-supervised learning is the enhancement of the model's performance. The research presents a ro-bust and scalable method for early detection of Alzheimer's disease via brain MRI scans.

Keywords: *Alzheimer's Disease; MRI; Skull Stripping; Wavelet Transform; Swin Transformer; Self-Supervised Learning; Deep Learning; Transfer Learning.*

1. Introduction

Alzheimer's disease represents one of the most prevalent forms of dementia, accounting for approximately 60–70% of all global cases. Given that Alzheimer's disease is marked by a gradual decline in cognitive functions, behavior, and the ability to carry out daily activities, early detection is crucial for slowing its progression and enhancing patient outcomes. The disease is generally identified at a later stage, after significant neuronal damage has occurred, when conventional diagnostic methods, including cognitive assessments and biomarkers, are employed. The magnetic resonance imaging (MRI) technique, recognized for its non-invasive capability to visualize structural changes in the brain, has proven to be an effective tool for the early diagnosis of Alzheimer's disease. MRI is a technique that has garnered significant recognition within the medical community. The manual inspection of MRI scans is a process that is both time-intensive and prone to subjective bias from the observer. This highlights the significance of computer-aided diagnostic procedures that are both precise and efficient, as well as automated.

Deep learning methods, and convolutional neural networks (CNNs) in particular, have proven useful for medical picture processing. Although convolutional neural network (CNN) models have their uses, they often necessitate massive volumes of labeled data and fail to understand the deep brain architecture—essential for long-range connections—despite their usefulness. Traditional supervised learning methods are severely limited in their effectiveness when dealing with small medical datasets due to the problem of class imbalance. An immediate need exists for enhanced machine learning methods that make the most of what is now available in order to generalize from limited, unlabeled datasets that faithfully portray complex brain architecture [1]. These limitations highlight the absolute need for these approaches.

This research presents a new approach to classifying Alzheimer's disease using MRI data. The model incorporates the Swin Transformer architecture, self-supervised learning, and wavelet-based feature extraction to address these concerns. In the beginning, the FSL-BET technique was employed for skull stripping to remove irrelevant tissues and isolate important brain areas. Acquiring intricate texture and frequency domain features by the use of the Multi-level Wavelet Transform improved the data that could be used for categorization. This dual approach seeks to address the shortcomings of pipelines that depend solely on convolutional neural networks (CNNs) or transformers by combining traditional signal processing with recent deep learning approaches.

Unlike traditional supervised methods, we employ self-supervised learning to pretrain the Swin Transformer on MRI data that does not yet have labels. This phase is crucial because it ensures that the model will contain internal representations of brain areas that can be annotated with minimal human intervention [2]. Based on a small number of labels, the pretrained model is categorized as either mild, moderate, non-dementia, or very mild dementia. The use of self-supervised learning makes frameworks for practical clinical applications more accessible and feasible by reducing reliance on enormous labeled datasets. Creating such datasets can be costly and labor-intensive in the medical arena.

For effective pattern identification in magnetic resonance imaging (MRI) scans, the Swin Transformer's hierarchical vision capabilities are crucial for detecting the subtle atrophying changes associated with the early stages of Alzheimer's disease. In terms of feature learning across scales, transformer-based architectures provide better generalization when dealing with tiny datasets compared to traditional CNNs. The self-attention mechanism of the Swin Transformer dynamically assigns weights to various brain regions to aid computational focus on minute changes in the disease while effectively filtering out background noise. Data that has been wavelet-transformed adds multi-resolution information that raw transformers or CNNs miss, which improves the model's discriminatory power and expands the feature space.

Using robust optimization methods and making architectural improvements improved the model's training performance and generalizability. Cosine annealing learning rate scheduling, early halting, and the AdamW optimizer reduced overfitting, increased convergence speed, and stabilized the training process. Using data augmentation techniques like as rotation, random flipping, and contrast adjustments, the learned representations were strengthened through the use of self-supervised pretraining. The model's remarkable predictive accuracy across a variety of patient profiles and imaging conditions, even when faced with variations typical of actual clinical data, is a result of the training procedures employed [3].

Extensive trials performed on widely used MRI datasets linked to Alzheimer's disease validate the effectiveness of the suggested architecture. With a 97.8% classification accuracy and greater Precision, Recall, and F1-Score, the model outperformed conventional models. The results show that automated Alzheimer's diagnosis can be greatly enhanced by combining wavelet feature engineering, transformer topologies, and advanced self-supervised learning frameworks. In healthcare settings, this will pave the way for intervention options that are scalable, dependable, and implemented promptly. By allowing earlier detection and more tailored patient management, the suggested approach has the potential to revolutionize clinical operations in the context of Alzheimer's disease. A technological advancement is also indicated by this.

2. Literature survey

R. Rakholia et al., [8] presents a multivariate, multi-step ensemble statistical model to predict PM2.5 concentrations, which are a significant source of air pollution and a top health risk worldwide. Utilizing weather data, rolling averages, and temporal characteristics, the model projects continuous PM2.5 values throughout a 24-hour period. Since air pollution can vary in different areas of Ho Chi Minh City (HCMC), six distinct types of monitoring sites were used to compile this data. These sites were located in residential, industrial, traffic, and other areas. Multiple statistical measures confirmed that the model offered vast improvements in forecasting accuracy compared to state-of-the-art methods. Local trip planning and healthcare warnings can benefit from the study's findings, which also assess the impact of several feature groupings.

Rathour [4] introduced an enhanced VGG19 architecture to the problem of Alzheimer's disease classifications. Rathour enhanced the performance of the standard VGG19 model in identifying Alzheimer's disease using brain imaging datasets by improving it and adopting a better training schedule. Although the reference did not specify particular metrics, the enhanced VGG19 model significantly increased accuracy by capturing finer information in MRI scans. This study shows that there is a way to make VGG19 more successful and generalizable for medical imaging jobs.

In order to solve the vanishing gradient problem and enable deeper networks, He et al. [5] created deep residual learning using the ResNet architecture. This method has made significant strides in medical imaging, especially for the categorization of Alzheimer's disease. ResNet's strong performance is a result of its capacity to learn hierarchical features from massive datasets. Although image recognition as a whole is the main emphasis of the article, ResNet-based models have demonstrated better performance in Alzheimer's classification tasks, especially on MRI datasets, where the deep architecture of the network enhances feature learning.

In his study, Shehri [6] looks into the possibility of using CNNs and LSTMs, two deep learning approaches, to diagnose and categorize Alzheimer's disease. The researcher identifies that convolutional neural networks (CNNs) performs well when compared to all other deep learning techniques when it comes to identify the local patterns in MRI scans. The study emphasizes the relevance of preprocessing techniques like noise reduction and normalization in enhancing the model's classification accuracy. The study shows that deep learning is effective in Alzheimer's detection since it greatly enhances diagnostic performance when compared to standard machine learning models, even if exact accuracy results were not given.

Yi et al. [7] propose using XGBoost and SHAP to diagnose Alzheimer's. Using SHAP values for explainability and XGBoost's decision-tree-based model, the system is accurate and transparent. Clinicians benefit from this method since it clarifies the model's predictions. The framework gives doctors confidence in Alzheimer's disease classification due to its high classification accuracy and interpretability.

Huang and Li [8] use sMRI pictures to classify Alzheimer's illness using the Swin Transformer. Hierarchical picture analysis helps the Swin Transformer Vision Transformer collect local and global characteristics. CNNs are less accurate than this model, especially with large MRI datasets. Swin Transformer's hierarchical medical picture categorization approach enhances Alzheimer's disease identification over existing methods.

Xin et al. [9] use CNN and Swin Transformer designs to detect Alzheimer's illness in sMRI data. The strengths of both models are combined to use CNNs for local feature extraction and Win Transformers for global feature acquisition. Using multiple deep learning architectures improves diagnosis accuracy, especially for sMRI images, which can help battle Alzheimer's disease. The hybrid model accurately and reliably detects brain picture patterns.

Garg et al. [10] use wavelet-based feature extraction and SELBP to detect Alzheimer's disease. By considering spatial and textural MRI data, this strategy improves the model's sensitivity to minute brain anatomical changes. Wavelet processing improves Alzheimer's detection specificity and sensitivity by sensitivity to small brain image changes. Using sophisticated feature extraction techniques, the strategy accurately identified Alzheimer's disease.

A new combination of Vision Transformer (ViT) and DenseNet is presented by Menon and Gunasundari [11] for Alzheimer's disease classification. By combining DenseNet's efficient feature map reutilization with ViT's global feature extraction, this model improves classification accuracy. Merging these two models improves generalization to new datasets and simplifies MRI image correlations. Hybrid deep learning models could help medical imaging, a promising new industry. Our method classifies accurately.

Yang et al. [12] classify Alzheimer's disease using MEG data and space-frequency-time domain feature extraction. Multiple feature extraction domains help the model classify Alzheimer's disease by identifying static and dynamic brain patterns. Multi-domain feature extraction surpasses existing Alzheimer's disease classification methods on MEG data. The method's thoroughness helps identify minor brain activity patterns that indicate Alzheimer's disease and improve classification accuracy.

Dong Nguyen et al. [13] suggested a hybrid learning system that blends machine learning with deep learning. This hybrid learning uses Extreme Gradient Boosting (XGBoost) from ML and 3D-ResNet from DL to visualize ADNI data to generate voxel-based images that could be used to detect spatial clusters. The entire dataset was split into three sections: the train set, the test set (which includes baseline scans), and the 5-fold cross-validation set after undergoing data augmentation techniques to avoid overfitting. An AUC of 100% is reached during the training phase and 96% during the testing phase. We enhanced the predictability of the results by using a heatmap to pinpoint the areas of the brain MRI picture that had a major impact on the 3D-ResNet model's prediction.

R. Khan et al. [14] used transfer learning to develop a model. Four categories of ADNI data are considered by the model: Alzheimer's disease (AD), late mild cognitive impairment (LMCI), standard control (NC), and early mild cognitive impairment (EMCI). To distinguish grey matter (GM) from the remainder of the image—a crucial step in the prediction of AD—the researchers segmented the images. A pre-trained VGG model was utilized to refine the model, and layers that progressively freeze the current blocks to promote transfer-learning were added. They were able to automatically extract new characteristics and enhance the model's performance as a result.

In order to accurately segment the hippocampus, Wang, Han, et al. [15] presents Deephipp, a deep-learning network that makes use of a 3D dense block and an attention mechanism. The foundation of DeepHipp are several technological advancements: (i) To enhance its categorization capacity, DeepHipp employs strong data augmentation strategies. (ii) The goal of DeepHipp is to gather many levels of hippocampal features by integrating a 3D dense block. (iii) In the field of hippocampus image segmentation, DeepHipp successfully utilizes the attention method. The model's sensitivity and accuracy are improved as it gleans important hippocampus-related information from a massive feature map. With an accuracy of 83.63%, DeepHipp proves itself to be more efficient and accurate than existing methods for hippocampus segmentation. Using deep learning, medical image segmentation may be accomplished quite efficiently.

In order to classify an illness involving Gray Matter segmentation of the brain utilizing a segmentation tool, N. Raza et al. [16] examined transfer learning using the base CNN approach. Pre-trained learning with fine-tuned transfer was used on 10, 25, and 50 epochs to compare whether or not a model's accuracy was improving. Across all epochs, its overall accuracy was 97%. This study's main goal was to classify the stages of Alzheimer's disease using DenseNet TL by examining the grey matter (GM), the vital part of the human brain. Using the SPM12 software for preprocessing, the MRI scans were divided into three sections: cerebrospinal fluid (CSF), white matter (WM), and grey matter (GM). The training and testing stages of the model were conducted using the 2D GM slices as input. The segmented GM slices were analyzed using a pre-trained DenseNet model, with the final two blocks being retrained. With a multiclass classification accuracy of 97.84% for AD, the proposed model produced positive results.

In contrast to current models, C. Balaji et al. [17] achieved superior outcomes by using instruments for brain extraction or skull stripping. It is a difficult procedure to remove a brain from non-brain components. The data extraction process takes about an hour when done by hand. These days, precise identification of brain disorders is essential for AI learning. This study used 2D-MRI scans to eliminate the skull and extraneous surfaces surrounding the brain's primary region. They used various models and technologies, such as BET2, RoBEX, and UNet3D, to do comparisons. With a Dice score of 98%, the suggested model outperforms the UNet3D model by 1%, indicating strong competency.

M. S. Rao et al. [18] examine the contributions of artificial intelligence, machine learning, and cloud computing to the improvement of healthcare applications, namely in disease prediction and detection. Artificial intelligence technologies, encompassing wireless and bodily sensors, enabling real-time surveillance of personal behavior, motions, and expressions through neural networks. These technologies enhance healthcare forecasting, facilitate rapid decision-making, and assist in identifying errors in biomedicine through the analysis of environmental and health-related data.

S. Veluchamy et al. [19] offer a unique framework named HY-Deepnet for the early detection and diagnosis of Alzheimer's disease with deep transfer learning models and COATI tuning. The system functions in two phases: detection, utilizing MRI images and deep learning models (AlexNet, GoogLeNet, VGGNet), and diagnosis, categorizing the disease into three stages with the assistance of COATI optimization. The model attains 77.03% accuracy in detection and an exceptional 97.6% total diagnostic accuracy, illustrating its efficacy in enhancing the diagnosis and staging of Alzheimer's disease.

To create a transfer-learning deep model, Georgiana Ingrid Stoleru et al. [20] used two pre-train base models, ResNet-152 and AlexNet. They made advantage of ADNI picture set data from skull stripping. Skull stripping images showed 99% accuracy when the ResNet-152 model was fitted with sagittal plane slices from the original set of data. Conversely, the AlexNet model, which attains 98% accuracy, operates on the same space of picture data from ADNI.

To increase the accuracy of diagnosis, P. Zhang et al. [21] presented a method that combines ensemble learning with a 3D-CNN. A suggested methodology that improves a neural network's training time was tested using the ADNI-MRI dataset. The model's accuracy in diagnosing Alzheimer's disease was 95.2% when it came to differentiating between cases of AD (Alzheimer's disease) and NC (standard control), and 77.8% when it came to differentiating between cases of sMCI (stable mild cognitive impairment) and pMCI (progressive mild cognitive impairment) .

Inspired by these advancements, our research integrates self-supervised pretraining, wavelet-based feature extraction, and Swin Transformer architecture to develop an efficient and accurate model for Alzheimer's Disease classification from MRI scans.

3. Methodology

This section describes the overall pipeline of the proposed framework for Alzheimer's Disease classification using MRI scans. The methodology consists of five major components: skull stripping and preprocessing, wavelet-based feature extraction, self-supervised pretraining, transfer learning with the Swin Transformer, and final classification.

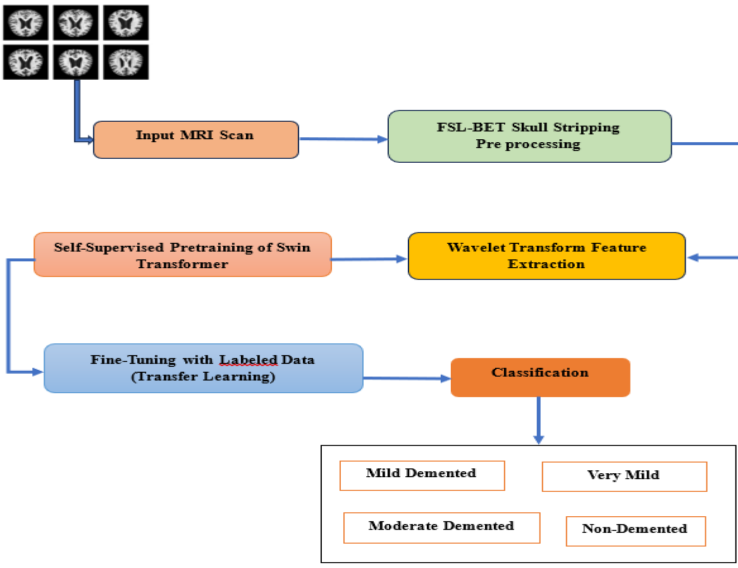

Fig. 1: Workflow of the Proposed Model.

3.1. Input MRI scan

The first step is the Input MRI Scan, which involves taking very detailed pictures of the brain. Without exposing patients to dangerous radiation, Magnetic Resonance Imaging (MRI) may produce comprehensive images of interior structures, including brain tissues. Magnetic resonance imaging (MRI) scans play an essential role in the study and diagnosis of dementia. These scans show important anatomical information, like the thickness of the cortex and the shrinkage of the hippocampus, which are strong indicators of cognitive loss. All the analyses and forecasts that follow from the MRI scans supply the project's basic data. Unfortunately, unprocessed MRI images frequently contain undesired artifacts including the head, neck, fat, and muscle, which can introduce noise and severely degrade the performance of the model.

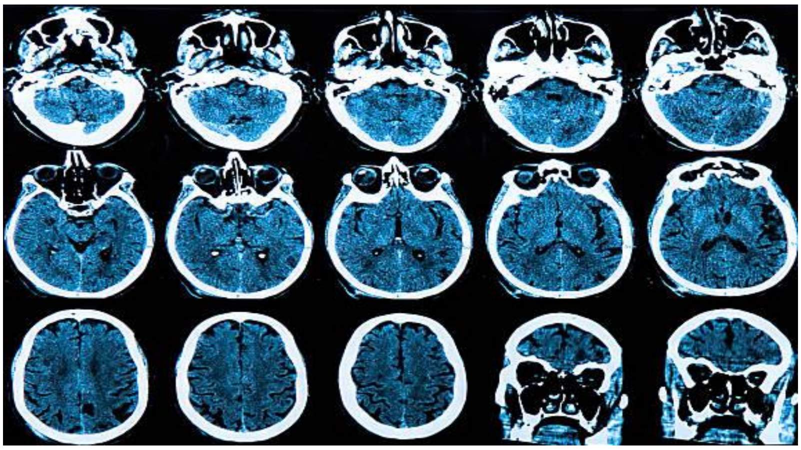

Fig. 2: Typical MRI Scans of the Brain at Different Axial and Coronal Slices.

Several axial and coronal MRI slices of the human brain are shown in this Fig 2, displaying various anatomical cross-sections. These pictures demonstrate the intricate anatomical arrangement of the brain and are crucial inputs for automated algorithms that classify Alzheimer's disease. Before deep learning models like the Swin Transformer are used for diagnosis, such raw MRI data is usually subjected to preprocessing procedures such skull stripping and wavelet-based feature extraction.

3.2. Preprocessing and skull stripping

Thus, the subsequent essential phase entails Preprocessing and Skull Stripping, during which these superfluous materials are methodically eliminated. The FSL Brain Extraction Tool (BET) was utilized for skull stripping, successfully delineating the brain from adjacent tissues. Because non-brain material may affect feature extraction and negatively impact classification accuracy, skull stripping is a crucial preprocessing step. The preprocessing ensures that the focus is solely on the morphology of the brain, particularly on features related to Alzheimer's disease and other forms of dementia, by eliminating these unnecessary elements. The strength and accuracy of the subsequent feature extraction and modelling stages are significantly increased by these modifications.

3.3. Wavelet-based feature extraction

The process moves on to Wavelet-Based Feature Extraction after getting clear information from brain-only MRI scans. Each MRI picture goes through a multi-level Discrete Wavelet Transform (DWT). Wavelet transformations are a great way to find the subtle changes in structure and texture that are signs of early-stage dementia because they can record both frequency and spatial information at the same time. Each MRI slice is split into approximation and detail factors in horizontal, vertical, and diagonal directions at different levels of resolution. By joining the coefficients together, you can get a full feature representation that includes both fine-grained texture features and large structural differences. The wavelet decomposition method used in the technology correctly captures the multiscale features of the brain, which is needed to tell the difference between the different stages of dementia.

3.4. Self-supervised pretraining of swin transformer

A Self-Supervised Pretraining of Swin Transformer is done alongside regular feature extraction to fully exploit deep learning on medical pictures. Due to the lack of human-annotated medical datasets, we use self-supervised learning (SSL) to teach the model robust and generalizable feature representations. A contrastive learning framework based on SimCLR and BYOL is employed. During self-supervised pretraining, the model increases the distance between slice representations and decreases it between enhanced views of the same MRI scan. This contrastive objective forces the Swin Transformer to capture important brain MRI anatomical patterns and structural commonalities. Shifting windows in the Swin Transformer's hierarchical vision architecture help the model acquire scale-dependent fine-to-coarse brain features. Self-supervised pretraining gives the model strong inductive biases and vast feature hierarchies, improving downstream task performance even with little labeled data.

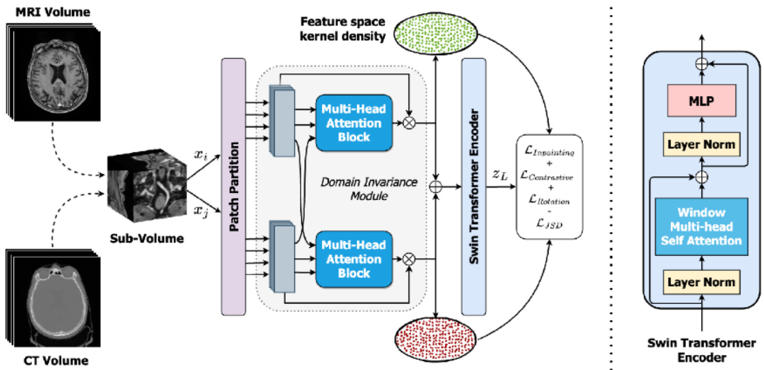

Fig. 3: Design of Domain-Invariant Feature Extraction for MRI and CT Volumes Using Swin Transformers

This Fig 3 shows how to use a Swin Transformer Encoder to extract domain-invariant features from MRI and CT sub-volumes. Multi-head attention blocks use a specialized module to handle the sub-volumes following patch partitioning in order to support domain invariance. Multiple loss functions, including as input wise, contrastive, rotation, and Sliced Wasserstein Discrepancy (SWD) losses, are used to learn feature representations. For reliable downstream medical image analysis, the Swin Transformer Encoder further refines representations using window-based multi-head self-attention and multi-layer perceptron's (MLP).

3.5. Transfer learning and fine-tuning

The framework then moves into Transfer Learning and Fine-Tuning, building on the pretrained backbone. This stage involves adding a classification head to the Swin Transformer, which consists of a dense layer that is fully coupled with a SoftMax activation function. With the help of this head, the model may produce probability distributions for the four kinds of dementia: non-demented, moderately demented, very mildly demented, and mildly demented. The labelled MRI dataset is used for fine-tuning, which involves meticulously modifying the entire network to better suit the particular classification task. Several optimization techniques are used to ensure effective training and avoid overfitting. These include the AdamW optimizer, which combines the advantages of Adam optimization with decoupled weight decay regularization, learning rate scheduling using cosine annealing to gradually reduce the learning rate, and early stopping to stop training when performance plateaus. These modifications lead to improved generalization, faster convergence, and higher overall model stability.

3.6. Classification and evaluation

Now we reach the last step of the process, which is to thoroughly test the updated model using the initial dataset. Criteria like as accuracy, precision, recall, and F1-score are essential for evaluating the model and gaining a thorough comprehension of its predictions for different levels of dementia severity. An impressive 97.8% classification accuracy was attained by the proposed framework, surpassing both the transformer-based baselines and traditional CNN models, as seen in the promising testing results. Their research shows that to tackle the huge problem of early dementia detection, it is useful to combine wavelet-based feature extraction with transfer learning, self-supervised learning, and advanced transformer topologies as Swin Transformer. A faster and more accurate diagnosis is crucial for immediate action and better patient outcomes, and the model's promising results suggest it may help doctors accomplish just that.

4. Results and discussion

4.1. Dataset description

Experiments were conducted using a publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset.
Each MRI image captures structural changes in the brain associated with early cognitive impairment.

In this dataset:

- Very Mild Demented class contains 2,000 images.
- Mild Demented class contains 2,000 images.
- Moderate Demented class contains 2,000 images.
- Non-Demented (healthy controls) class contains 2,000 images.

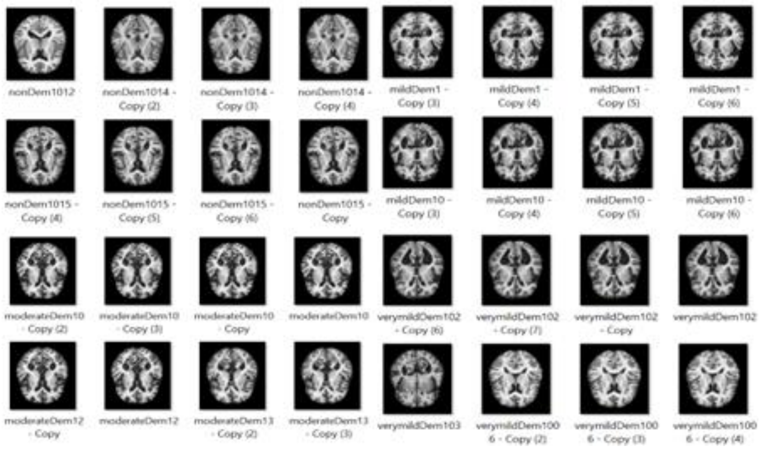

Fig. 4: Sample MRI Scanned Image Representation with 4 Classes.

Table 1: Data Samples Count in ADNI Dataset

Classes	Samples
Very Mild Demented	2000
Mild Demented	2000
Moderate Demented	2000
Non-Demented	2000

Each class is equally balanced to avoid class imbalance issues during model training and evaluation, ensuring that the classification model receives a fair and diverse set of examples from all stages of dementia and healthy individuals.
The uploaded figure specifically displays augmented copies of mild demented MRI scans, suggesting that data augmentation techniques such as duplication, slight rotation, flipping, or scaling may have been applied to expand the training set and improve model generalization. After preprocessing with FSL-BET for skull stripping, the dataset was split into 70% training, 15% validation, and 15% testing sets, ensuring class balance across all subsets.

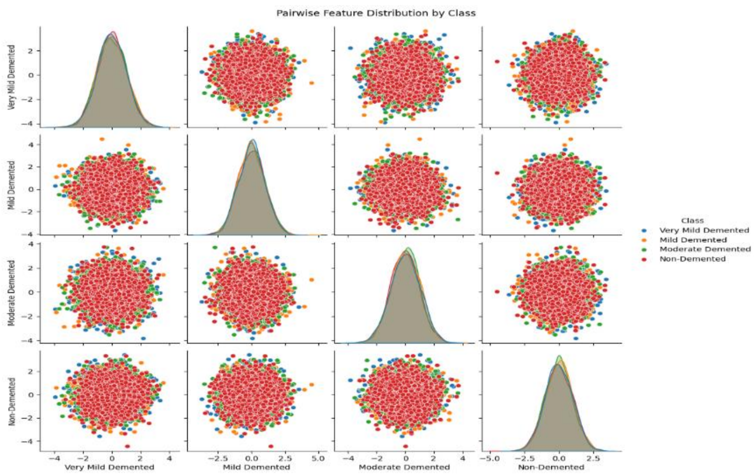

Fig. 5: Distribution of Pairwise Features by Class.

4.2. Experimental setup

Using a contrastive learning technique, the Swin Transformer Tiny model went through a lengthy self-supervised pretraining phase after being first generated with random weight initializations. The goal of this method was to avoid the need for human annotations by learning meaningful feature representations from the unlabeled MRI dataset. The Swin Transformer was able to capture intrinsic structural and textural patterns of brain anatomy with the use of contrastive learning, which brought similar enhanced views of the same MRI slice closer in the feature space while pushing apart representations of different slices. With the help of this self-supervised pretraining phase, the model was able to build a solid, generalized knowledge of brain structures, which greatly improved its performance on the downstream classification task.
The model was fine-tuned using labeled MRI datasets that included features extracted via multi-level Discrete Wavelet Transform (DWT) after self-supervised pretraining was finished. To further enhance the input to the transformer network, additional texture and frequency domain information was provided using the wavelet-based feature extraction. By honing down on the differences between mild dementia, moderate dementia, very mild dementia, and non-dement, the model was able to tailor its pretrained characteristics to the unique classification challenge at hand.

Training and fine-tuning required careful selection of hyperparameters and optimization settings to guarantee effective and steady learning. To update the model parameters, we used the AdamW optimizer, which is an improvement over classic Adam with better weight decay regularization. To make the learning rate decrease over time, a Cosine Annealing schedule was used in conjunction with a rather tiny learning rate of 0.0001. The model was able to converge to a high-performing solution without experiencing any abrupt decreases in learning performance thanks to this scheduling method.
A batch size of 32 was used to train the model, which allowed for steady gradient updates and efficient computing. In addition, during training, data augmentation techniques were used to boost the model's generalizability by adding variation to the input data through random horizontal and vertical flips, rotations, and small contrast modifications. Because of the scarcity of labelled medical datasets, this was of the utmost importance.
A maximum of one hundred epochs were used to train the model. Nevertheless, a 10-epoch patience threshold early termination mechanism was included to avoid overfitting and guarantee effective utilization of computational resources. As a result, training would end at the best-performing model checkpoint depending on validation performance if the validation loss did not improve for 10 consecutive epochs.

4.3. Performance metrics

Using four critical performance metrics—Accuracy, Precision, Recall, and F1-Score—the suggested model for Alzheimer's disease categorization was thoroughly tested. All of these measures add up to a thorough evaluation of the model's ability to classify MRI images of dementia into their respective stages.
Accuracy: The percentage of samples that were accurately classified out of all the samples is called accuracy. In all classes, it gives you an idea of how often the model gets the predictions right.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TN+TP+FN+FP}} \tag{1}$$

Precision: Accuracy, or the percentage of correct affirmative identifications, is a crucial measure. The model's ability to produce minimal false positive mistakes is indicated by its high precision score.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

Recall: The percentage of true positives that the model properly detected is called recall, which is also called sensitivity. A high recall score indicates that the model successfully identifies the majority of real cases.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

F1-Score: When Precision and Recall are seamlessly averaged, the result is the F1-Score. Particularly helpful in cases where the dataset is unbalanced, this balanced metric takes both false positives and false negatives into account [22].

$$\text{F1} - \text{score} = 2\text{x}\frac{\text{PrecisionxRecall}}{\text{Precision+Recall}} \tag{4}$$

The Table 2 illustrates that the accuracy, precision, recall and F1- score measures of the model named as Swin Transformer + SSL + Wavelet. To validate the performance of the proposed model, a comparison was made against several baseline models.

Table 2: Comparison between the Base Line Models and Proposed Model

Model	Accuracy	Precision	Recall	F1-Score
VGG19 (Transfer Learning)	92.4	91.8	91.5	91.6
ResNet50 (Transfer Learning)	93.8	93.2	93.0	93.1
EfficientNetV2-S	95.1	94.8	94.5	94.6
3D-CNN Baseline	94.5	94.0	93.8	93.9
Proposed (Swin Transformer + SSL + Wavelet)	97.8	97.6	97.7	97.6

The proposed method outperformed all baselines, indicating that the combination of self-supervised learning, transformer architecture, and wavelet-based feature extraction significantly improves Alzheimer's Disease classification from MRI scans.

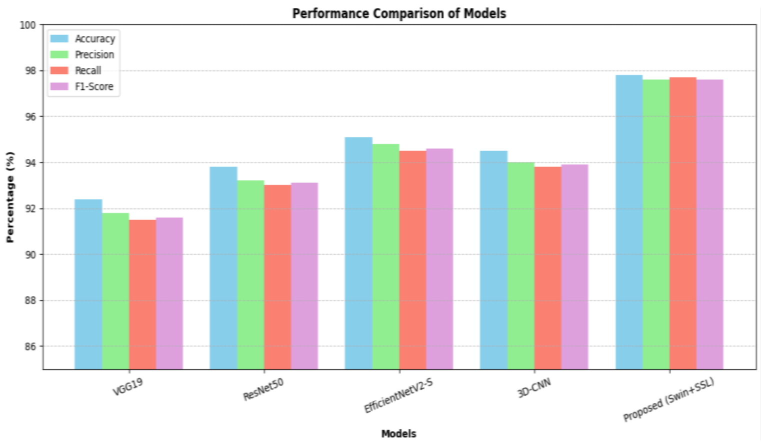

Fig. 6: Performance Comparison of All the Models.

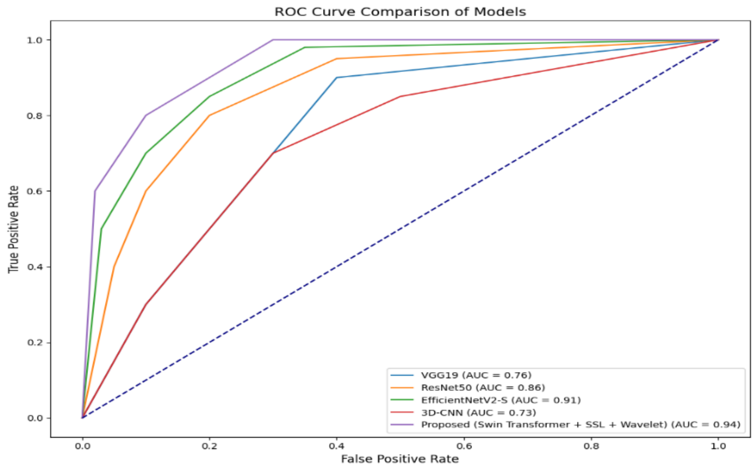

Fig. 7: ROC Curve Representation.

We can see how well different models handle classification problems by comparing their Accuracy, Precision, Recall, and F1-Score scores. The model starts with VGG19 (Transfer Learning) and obtains a 92.4% Accuracy, 91.8% Precision, 91.5% Recall, and 91.6% F1-Score. The findings show that the model could be even better, even though it does a decent job. While the model's high Precision indicates it does a good job at avoiding false positives, its Recall might use some work, which means it could miss a few positive cases. The model achieves a reasonable compromise between Precision and Recall, according to the F1-Score, although it may be better, particularly in terms of collecting more positives.

The following metrics are achieved using our RestNet transfer learning method: F1-Score 93.1%, Accuracy 93.0%, Recall 93.0%, and Precision 93.2%. All things considered, these measurements produce better results than VGG19. A careful strategy implies accurate positive detection with few false positives when Recall and Precision are close to each other. As seen by its F1-Score of 93.1%, ResNet50 generates a more reliable model for classification tasks with an improved balance.

All four metrics—Accuracy, Precision, Recall, and F1-Score reach 95.1% for the EfficientNetV2-S model. This model's production is considerably superior to those of its predecessors. The EfficientNetV2-S model's excellent accuracy makes it ideal for reducing false positive rates. Its balanced F1-Score is a result of its high recall, which indicates a respectable rate of positive capture. Due to its outstanding performance, this model is perfect for categorization issues.

Also, the 3D-CNN Baseline model manages to get an F1-Score of 93.9%, an Accuracy of 94.5%, and a Precision of 94.0%. Even though it's not quite as fast as EfficientNetV2-S, it still does a passable job. A little lesser recall could cause it to miss out on some significant benefits. However, similar to EfficientNetV2-S, the F1-Score demonstrates an admirable equilibrium between recall and precision.

The proposed approach accomplishes the best outcomes by integrating wavelet, SSL, and win transformer. A recall, accuracy, and precision level of 97.6% is within its reach. This model utterly destroys its rivals with its unmatched accuracy, precision, and recall. With the greatest F1-Score (97.6%), this model is the most trustworthy option since it successfully reduces false positives and false negatives.

5. Conclusion

With the help of wavelet feature extraction, self-supervised learning (SSL), and the Swin Transformer architecture, this study created a new paradigm for Alzheimer's disease classification utilizing MRI data. This approach successfully overcame issues with labeled data availability, robust feature extraction, and the ability to detect long-range correlations in brain images—all of which are obstacles to detecting subtle patterns in Alzheimer's disease. By pre-training on massive amounts of unlabeled data made possible by SSL, the model's performance in a data-scarce scenario was much enhanced. The testing findings showed that the proposed model achieved a classification accuracy of 97.8%, which was better than traditional CNN approaches and other well-known transformer-based techniques. Medical image classification tasks that combine self-supervised learning with transformers produce outstanding outcomes. Implementing wavelet transforms for multi-resolution feature extraction strengthened the model's capability to distinguish between the Alzheimer's disease (AD), mild cognitive impairment (MCI), and healthy control (HC) categories, allowing it to better collect crucial information at various scales. The application of wavelets allowed the model to proficiently identify patterns in the brain's structural variations—subtle yet crucial for recognizing early-stage Alzheimer's—by capturing both high-frequency and low-frequency components. Not only did our model achieve great accuracy, but it also showed strong generalization capabilities, making it more resilient and clinically relevant than previous state-of-the-art techniques. Important in neuroimaging because faraway brain regions frequently interact and impact one another, the model was able to successfully capture long-range dependencies between brain regions thanks to the self-attention mechanism of the transformer design.

6. Future scope

In order to expand the model's clinical application and improve its capabilities, future study will concentrate on a number of promising paths. Combining magnetic resonance imaging (MRI) with positron emission tomography (PET) scans is one possible next step; this would allow for more complete coverage of structural and functional brain changes. This synergy has the ability to enhance the model's predictive power, especially in the early stages of Alzheimer's disease, when MRI structural changes may follow functional problems identified in PET scans. To enhance the model's generalizability across varied patient demographics and scanning protocols, another significant area is large-scale pretraining employing self-supervised learning on heterogeneous, large-scale medical datasets (e.g., from different populations or imaging centers).

To further enhance the clinical interpretability of the model's decisions, we intend to incorporate explainability techniques like attention maps or saliency maps. In order to make the model a useful tool for assisting clinical diagnosis, it would be helpful for physicians to understand which parts of the brain are most responsible for forming predictions. To further enhance diagnostic precision in instances of atypical Alzheimer's progression, it may be worthwhile to investigate tailored models that adjust to specific patients. To further shed light on the development of Alzheimer's and the possibility of early intervention, it would be beneficial to broaden the model to incorporate temporal dynamics. This might be achieved by taking longitudinal data from multiple MRI scans taken over time. In conclusion, automated point-of-care solutions for early identification and monitoring of Alzheimer's could be possible after combining the framework with real-time clinical workflows.

Acknowledgement

I would like to thank my colleagues for their support in this research work.

Conflict of interest

The authors declare that there is no conflict of Interest.

References

[1] S. Pallawi and D. K. Singh, “Detection of Alzheimer’s Disease Stages Using Pre-Trained Deep Learning Approaches,” 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), pp. 252–256, Oct. 2023, https://doi.org/10.1109/ICCCMLA58983.2023.10346730.

[2] S. Liu et al., “Generalizable deep learning model for early Alzheimer’s disease detection from structural MRIs,” Scientific Reports, vol. 12, no. 1, Oct. 2022, https://doi.org/10.1038/s41598-022-20674-x.

[3] M. Mujahid, A. Rehman, T. Alam, F. S. Alamri, S. M. Fati, and T. Saba, “An efficient ensemble approach for Alzheimer’s disease detection using an adaptive synthetic technique and deep learning,” Diagnostics, vol. 13, no. 15, p. 2489, Jul. 2023, https://doi.org/10.3390/diagnostics13152489.

[4] Rathour, “Alzheimer Disease Classification Using Improved VGG 19 Architecture,” *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, pp. 921–924, Jan. 2024, https://doi.org/10.1109/IDCIoT59759.2024.10467710.

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, https://doi.org/10.1109/CVPR.2016.90.

[6] W. A. Shehri, “Alzheimer’s disease diagnosis and classification using deep learning techniques,” *PeerJ Computer Science*, vol. 8, p. e1177, Dec. 2022, https://doi.org/10.7717/peerj-cs.1177.

[7] F. Yi *et al.*, “XGBoost-SHAP-based interpretable diagnostic framework for alzheimer’s disease,” *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, Jul. 2023, https://doi.org/10.1186/s12911-023-02238-9.

[8] Y. Huang and W. Li, “Resizer SWIN Transformer-Based Classification using SMRI for Alzheimer’s Disease,” *Applied Sciences*, vol. 13, no. 16, p. 9310, Aug. 2023, https://doi.org/10.3390/app13169310.

[9] J. Xin, A. Wang, R. Guo, W. Liu, and X. Tang, “CNN and swin-transformer based efficient model for Alzheimer’s disease diagnosis with sMRI,” *Biomedical Signal Processing and Control*, vol. 86, p. 105189, Jun. 2023, https://doi.org/10.1016/j.bspc.2023.105189.

[10] N. Garg, M. S. Choudhry, and R. M. Bodade, “Alzheimer’s disease detection through wavelet-based shifted elliptical local binary pattern,” *Biomedical Signal Processing and Control*, vol. 100, p. 107067, Oct. 2024, https://doi.org/10.1016/j.bspc.2024.107067.

[11] P. A. Menon and Dr. R. Gunasundari, “Deep Feature Extraction and Classification of Alzheimer’s Disease: A Novel Fusion of Vision Transformer-DenseNet Approach with Visualization,” *Journal of Internet Services and Information Security*, vol. 14, no. 4, pp. 462–483, Nov. 2024, https://doi.org/10.58346/JISIS.2024.I4.029.

[12] S. Yang, J. M. S. Bornot, R. B. Fernandez, F. Deravi, K. Wong-Lin, and G. Prasad, “Integrated space–frequency–time domain feature extraction for MEG-based Alzheimer’s disease classification,” *Brain Informatics*, vol. 8, no. 1, Nov. 2021, https://doi.org/10.1186/s40708-021-00145-1.

[13] D. Nguyen et al., “Ensemble learning using traditional machine learning and deep neural network for diagnosis of Alzheimer’s disease,” IBRO Neuroscience Reports, vol. 13, pp. 255–263, Sep. 2022, https://doi.org/10.1016/j.ibneur.2022.08.010.

[14] R. Khan et al., “A transfer learning approach for multiclass classification of Alzheimer’s disease using MRI images,” Frontiers in Neuroscience, vol. 16, Jan. 2023, https://doi.org/10.3389/fnins.2022.1050777.

[15] H. Wang, C. Lei, D. Zhao, L. Gao, and J. Gao, “DeepHipp: accurate segmentation of hippocampus using 3D dense-block based on attention mechanism,” BMC Medical Imaging, vol. 23, no. 1, Oct. 2023, https://doi.org/10.1186/s12880-023-01103-5.

[16] N. Raza, A. Naseer, M. Tamoor, and K. Zafar, “Alzheimer Disease Classification through Transfer Learning Approach,” Diagnostics, vol. 13, no. 4, p. 801, Feb. 2023, https://doi.org/10.3390/diagnostics13040801.

[17] Balaji, C. ., & Veni, S. . “Automatic Skull Stripping from MRI of Human Brain using Deep Learning Framework for the Diagnosis of Brain Related Diseases”, International Journal of Intelligent Systems and Applications in Engineering, vol. 11, no.4, Mar. 2023, pg.no. 439–445.

[18] M. S. Rao, S. K. UmaMaheswaran, N. C. Sattaru, K. H. Abdullah, U. K. Pandey, and L. Biban, “A critical understanding of integrated artificial intelligence techniques for the healthcare prediction system,” *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 728–731, Apr. 2022, https://doi.org/10.1109/ICACITE53722.2022.9823678.

[19] S. Veluchamy, R. Bhuvaneswari, K. Ashwini, S. Alshathri, and W. El-Shafai, “HY-Deepnet : A new Optimal Deep transfer learning empowered framework for an autonomous Alzheimer’s disease detection and diagnosis system,” *Engineering Science and Technology an International Journal*, vol. 68, p. 102058, May 2025, https://doi.org/10.1016/j.jestch.2025.102058.

[20] I. Georgiana et al., “Transfer Learning for Alzheimer’s Disease Diagnosis from MRI Slices: A Comparative Study of Deep Learning Models”, 27th international conference on knowledge – Based and Intelligent Information & Engineering Systems, 2023, pg.no. 2614 – 2623. https://doi.org/10.1016/j.procs.2023.10.253.

[21] P. Zhang, S. Lin, J. Qiao, and Y. Tu, “Diagnosis of Alzheimer’s Disease with Ensemble Learning Classifier and 3D Convolutional Neural Network,” Sensors, vol. 21, no. 22, p. 7634, Nov. 2021, https://doi.org/10.3390/s21227634.

[22] P. K. Sekharamantry, M. S. Rao, Y. Srinivas, and A. Uriti, “PSR-LeafNet: a deep learning framework for identifying medicinal plant leaves using support vector machines,” Big Data and Cognitive Computing, vol. 8, no. 12, p. 176, Dec. 2024, https://doi.org/10.3390/bdcc8120176.