A

PROJECT REPORT ON

DEEP LEARNING APPROACH FOR DETECTION AND CLASSIFICATION OF ALZHEIMER'S DISEASE

SUBMITTED TO SAVITRIBAI PHULE PUNE UNIVERSITY
FOR PARTIAL FULFILLMENT
OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

BACHELOR OF ENGINEERING
In
Electronics and Telecommunication Engineering

By

ADITI BANKAR	**B400050028**
SHREYA BANSOD	**B400050029**
AAYUSH MOHOD	**B400050177**

GUIDE
Dr. M. P. TURUK

DEPARTMENT OF
ELECTRONICS AND TELECOMMUNICATION ENGINEERING
PUNE INSTITUTE OF COMPUTER TECHNOLOGY
PUNE – 43

2024-25

Department of Electronics and Telecommunication Engineering
Pune Institute of Computer Technology, Pune – 43

CERTIFICATE

This is to certify that the Project Report entitled

DEEP LEARNING APPROACH FOR DETECTION AND CLASSIFICATION OF ALZHEIMER'S DISEASE

has been successfully completed by

Aditi Bankar	B400050028
Shreya Bansod	B400050029
Aayush Mohod	B400050177

Is a bona fide work carried out by them under the guidance of Dr. M. P. Turuk and it is approved for the partial fulfillment of the requirement of the Savitribai Phule Pune University, Pune for the award of the degree of the Bachelor of Engineering (Electronics and Telecommunication Engineering). This project work has not been previously submitted to any other Institute or University for awarding any degree or diploma.

Dr. M. P. Turuk Guide	Dr. M. V. Munot HOD, E&TCE Dept.	Prof. Dr. S. T. Gandhe Principal, PICT

Place: Pune
Date :

ACKNOWLEDGEMENTS

We would like to express our heartfelt gratitude to Dr. M. P. Turuk for her constant encouragement, invaluable mentorship, and thoughtful guidance throughout the course of our project. Her expertise and unwavering support played a crucial role in its successful completion.

We are equally thankful to Dr. M. V. Munot, Head of the Department of Electronics and Telecommunication, for her consistent support, insightful suggestions, and motivation, which have been instrumental in shaping the direction of our work.

Our sincere appreciation also goes to Dr. R. C. Jaiswal and Mr. S. D. Hake for their valuable inputs and assistance, which significantly contributed to the development and refinement of our project.

Finally, we gratefully acknowledge the contributions of various authors and researchers whose work laid a strong academic foundation and greatly informed our study.

Aditi Bankar
Shreya Bansod
Aayush Mohod

CONTENTS

ABSTRACT

Alzheimer's Disease (AD) is an escalating global health concern, particularly affecting aging populations. Early and precise detection is essential for timely intervention and effective treatment. Magnetic Resonance Imaging (MRI) serves as a critical modality for identifying neurodegenerative changes, though manual interpretation is labor-intensive and susceptible to variability. To address this, an automated deep learning pipeline has been developed for the detection and classification of AD using MRI scans.

The pipeline is constructed using data sourced from the Harmonized Protocol (HARP) and employs advanced computational methods for image preprocessing, segmentation, and classification. Preprocessing of 3D NIFTI MRI volumes is performed using tools such as Antspy and Nibabel, involving denoising, skull stripping, intensity normalization, and spatial registration. The FMRIB Software Library (FSL) supports structural corrections and region-based analysis.

Segmentation of brain regions is performed using a UNet-based architecture, achieving a segmentation dice score of 89%. Additionally, Swin UNet—a vision transformer-based model utilizing hierarchical attention mechanisms and non-overlapping patch embeddings—has been integrated to improve spatial representation in volumetric data. The performance of Swin UNet currently reflects an dice score of 75%, with potential for further optimization.

A teacher-student knowledge distillation framework is incorporated to transfer learned representations from a high-capacity UNet (teacher) to a lightweight Swin UNet model (student). The student network is trained using a combined loss function, balancing binary cross-entropy loss with distillation loss derived from soft targets generated by the teacher model. This methodology enhances the generalizability and interpretability of the student model while maintaining computational efficiency.

The developed system showcases the potential of computational intelligence to automate and accelerate Alzheimer’s Disease diagnosis, thereby supporting more effective clinical decision-making through robust and reproducible deep learning solutions.

Abbreviations and Acronyms

AD	Alzheimer’s Disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
AIBL	Australian Imaging, Biomarker and Lifestyle
CN	Cognitively Normal
CNN	Convolutional Neural Network
FSL	FMRIB Software Library
HARP	Harmonized Protocol
MCI	Mild Cognitive Impairment.
MLP	Multilayer Perceptron
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
NACC	National Anti-Corruption Commission
NifTI	Neuroimaging Informatics Technology Initiative
ResNet	Residual Neural Network
SWIN	Shifted Window
UNet	U Network
ViT	Vision Transformer

List of Symbols

$\hat{y}$	Final segmentation prediction after activation.
P	Patch Size
W	Weight matrix of convolution layer
Wp	Weight matrix for patch embedding
b	Bias term
X	Input Shape
X_p	Patch split image
Z_0	Linear projection of flattened patches
Z	Feature representation through Swin blocks.
Q	Query matrix
K	Key matrix
V	Value matrix
W_Q	Linear projection matrices for Q
W_K	Linear projection matrices for K
W_V	Linear projection matrices for V
$+$	Addition
-	Subtraction
$\times$	Multiplication
$\cdot$	Matrix Multiplication
σ	Sigmoid Activation
$\mathscr{L}_{\mathscr{GT}}$	Ground Truth Loss
$\mathscr{L}_{\mathscr{KD}}$	Knowledge Distillation Loss
$\mathscr{L}_{total}$	Total Loss

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Background

Alzheimer's Disease (AD) is a prominent neurocognitive disorder, currently affecting over 55 million people worldwide, with approximately 10 million new cases emerging annually, as reported by the World Health Organization (WHO) in March 2023. As the seventh leading cause of death globally, AD imposes a growing burden on healthcare systems and is a major contributor to disability and dependency among the elderly. The disease progresses through seven clinical stages, yet diagnosis often occurs at stage 4, where moderate cognitive impairment begins to noticeably disrupt daily life.

One of the hallmark neuroanatomical features of AD is atrophy of the hippocampus, a brain region critical for memory consolidation and spatial navigation. As the disease advances, individuals exhibit progressive memory decline, difficulties in recognition, and increasing reliance on caregivers for routine tasks. Consequently, early and accurate detection of hippocampal changes is essential for timely intervention and effective management of the disease.

Traditional diagnostic methods rely on manual inspection and segmentation of brain MRIs, which is not only time-consuming but also prone to inter-observer variability. To address this, the application of automated deep learning models has gained momentum, offering the potential for fast, scalable, and highly consistent analysis of brain structures. In particular, hippocampal segmentation has emerged as a crucial task in the development of diagnostic pipelines for AD.

This project investigates a comprehensive deep learning-based pipeline for the segmentation of the hippocampus using 3D MRI scans, with data sourced from the ADNI and HarP datasets. Ground truth segmentations were created using FSL software, ensuring standardized and anatomically reliable labels. The study explores various deep learning models, including 2D and 3D U-Nets, Vision Transformers, and advanced training frameworks such as Diffusion Learning and Student-Teacher models, aiming to identify the most accurate and efficient approach for hippocampal segmentation in the context of

Alzheimer's detection.

1.2 Relevance

This project holds strong relevance both in real-world healthcare applications and within the scope of the Electronics and Telecommunication (EnTC) curriculum. By addressing the pressing global challenge of Alzheimer's disease detection, this work contributes to the advancement of technology-driven medical diagnostics, an area where innovation can directly impact patient outcomes and quality of life.

From an academic perspective, this project provides us with hands-on experience in developing and deploying deep learning models, including Vision Transformers, which represent the forefront of modern AI research. These skills are increasingly valuable as AI becomes embedded in all facets of engineering, from telecommunications to healthcare and beyond.

As students of Electronics and Telecommunication engineering, exploring the integration of deep learning techniques with medical imaging systems aligns closely with current industry trends toward AI-powered solutions. The work demonstrates how data processing, image analysis, and pattern recognition, core strengths of current discipline, can be leveraged to solve real-world problems in healthcare.

By developing expertise in these emerging technologies, the aim is to contribute meaningfully to the interdisciplinary convergence of engineering and medicine. This project not only enhances the technical proficiency but also prepares us to innovate in fields where engineering knowledge can be translated into impactful healthcare solutions.

1.3 Literature Survey

1.3.1 CNN Based Approaches

Convolutional Neural Networks (CNNs) have been widely applied in medical imaging, progressively extracting features through convolutional layers, pooling layers, and fully connected layers. While CNNs effectively recognize spatial patterns, they often lose important spatial details during down sampling, impacting precise feature localization.

For Alzheimer's disease (AD) detection, various studies have explored CNN-based methods. Bamber et al. [19] achieved 98% accuracy on the OASIS dataset but highlighted challenges such as long training times and preprocessing constraints. Carmo et al. [14] improved generalizability by varying training images across epochs. Liu et al. [10] combined CNNs with DenseNet for hippocampus segmentation and AD classification, identifying 62 × 48 × 58 as the optimal patch size for the ADNI dataset. Qiu et al. [12] introduced a 3D FCN+MLP model trained on ADNI and tested on AIBL, FHS, and NACC, incorporating non-imaging factors like age and gender to improve classification precision across datasets.

Beyond standard CNNs, U-Net has been widely used for medical image analysis, featuring an encoder-decoder structure with skip connections to retain spatial information lost during downsampling. Attention U-Net further improves this architecture by introducing attention gates in skip connections, refining features and allowing the decoder to focus on more critical spatial details, ultimately improving segmentation accuracy. The attention gate works by downsampling the skip connection, aligning it with the decoder input, and learning scaling factors to enhance key spatial features [4]. Several studies have built upon U-Net for hippocampus segmentation. Jiang et al. [30] developed CADyUNet, incorporating coordinate attention layers, which use larger convolution kernels to capture texture and background details, achieving an 83% segmentation accuracy. Helaly et al. [17] further optimized U-Net by introducing residual blocks and hyperparameter tuning, achieving 94% and 97% accuracy on a reduced OASIS dataset. ResNet, designed to address the vanishing gradient problem, introduces residual blocks for efficient training and better convergence. Odusami et al. [34] explored multimodal fusion, integrating PET and MRI scans with a modified ResNet-18, achieving 73.90% accuracy on ADNI. Similarly, Liu et al. [10] incorporated ResNet blocks within their CNN framework to accelerate training and improve robustness.

While CNN-based models have demonstrated strong performance, they still face limitations such as spatial information loss and computational inefficiency. These challenges have led to increasing interest in transformer-based architectures, which offer a promising alternative for improving segmentation accuracy and addressing biases in 3D neuroimaging.

1.3.2 Transformer Based Approaches

Vision Transformers (ViTs) have emerged as a powerful alternative to CNNs for medical imaging, leveraging self-attention mechanisms to model global dependencies. Unlike CNNs, which extract features hierarchically, ViTs treat image patches as tokens, passing them through transformer encoders to generate embeddings. Positional embeddings are added to these tokens to retain spatial relationships, as transformers lack inherent spatial inductive biases. The architecture also incorporates multi-head self-attention for better feature representation, layer normalization for stability, and residual connections to address vanishing gradients [13].

Hybrid approaches combining ViTs with CNNs enhance segmentation performance by leveraging the local feature extraction of CNNs with the global attention capabilities of transformers. Lei et al. [33] introduced a morphological ViT that refines segmentation through hierarchical contextual learning. Similarly, 3D-EffiViTCaps integrates capsule networks with EfficientViT blocks, preserving spatial relationships through dynamic routing. It incorporates 3D Patch Merging to balance resolution and channel dimensions while utilizing a cascaded group attention (CGA) mechanism to refine multi-scale features [36].

Other architectures explore lightweight transformer designs for hippocampus segmentation. Light3DHS [37] combines a simple CNN encoder for local feature extraction with a lightweight ViT to model global dependencies, improving segmentation resolution while maintaining efficiency. The UNet with Transformer introduces attention layers in both encoder and decoder, refining spatial feature extraction. Shah et al. [31] proposed Cascaded Modality Transformers (RMT), integrating multiple transformers for categorical, ordinal, and imaging data, achieving robust results even with missing data.

Among vision transformers, the Swin Transformer has gained prominence due to its hierarchical structure and shifted window attention mechanism (SW-MSA), enabling efficient multi-scale feature extraction. The architecture partitions an image into non-overlapping patches, which are progressively merged across stages while increasing feature dimensions. Unlike standard ViTs, Swin Transformers scale linearly with image size, making them ideal for high-resolution neuroimaging [15].

Swin Transformers have also been integrated into hybrid architectures. ST-MUNet fuses Swin Transformer with U-Net, enhancing segmentation precision. Its GAN-based pipeline consists of a generator using a Swin Transformer encoder and a discriminator evaluating segmentation authenticity. The approach incorporates preprocessing techniques such as Hybrid Kuan and Improved Frost (HKIF) filtering for noise removal, Geodesic Active Contour (GAC) skull stripping, and Bias Field Correction using Expectation-Maximization (EM)[23]. Another efficient approach is the Resizer-Swin model, which mitigates ViT's computational complexity by preprocessing grayscale images through a CNN-based resizer, converting single-channel inputs into multi-channel feature representations. This enables better collaboration between CNN feature extraction and Swin Transformer classification, enhancing segmentation performance [32].

As transformer-based models continue to evolve, their ability to reduce spatial biases and adapt to 3D neuroimaging data positions them as promising alternatives to traditional CNN architectures, particularly in hippocampus segmentation and other structural brain analyses.

1.4 Motivation

Research in Alzheimer's disease detection using deep learning has made significant progress, particularly in automating brain MRI segmentation and classification. However, real-world adoption remains limited due to critical challenges such as overfitting, insufficient dataset diversity, lack of interpretability, and high computational costs. Most existing approaches for 3D medical image segmentation in Alzheimer's disease detection rely heavily on convolutional neural networks (CNNs), which are effective at capturing local features but often fall short in modeling global context—crucial for understanding complex anatomical structures in brain MRIs. To address this, transformer-based models like SwinUNet3D are gaining attention for their ability to extract both local and long-range dependencies using hierarchical attention mechanisms. However, these models typically come with high computational costs, making them impractical for deployment in real-world, resource-constrained clinical environments This project proposes a hybrid solution by utilizing SwinUNet3D to ensure strong global contextual awareness, while addressing its limitations through a knowledge distillation framework in which a powerful UNet

teacher guides a lightweight SimpleSwinUNet3D student. By combining the representational power of transformers with the efficiency of distillation, and incorporating TPU acceleration, dynamic loss adaptation, and Grad-CAM-based interpretability, the system achieves a balanced trade-off between accuracy, efficiency, and transparency. The goal is to develop a model that not only performs well in research settings but is also practical and reliable for real-time clinical diagnostics in Alzheimer's care.

1.5 Aim of the Project

To develop a deep learning-based system using Vision Transformers (ViT) for segmentation and classification of MRI brain scans, with a focus on minimizing errors in overlapping segmented regions due to the critical nature of medical data. ViTs are preferred over traditional CNNs as they can capture global context more effectively, which is particularly important for 3D medical images that exhibit sequential spatial structure across slices. The project emphasizes interpretability and explainability to address the black-box nature of deep learning models, ensuring transparency in predictions. The goal is to assist medical professionals, not replace them, by providing reliable insights that enhance diagnostic accuracy and efficiency.

1.6 Scope and Objectives

This project focuses on designing and implementing a deep learning-based segmentation of hippocampus and classification of Alzheimer's disease through MRI analysis. The key objectives and scope for this include:

- Effective Pre-processing and Standardization: Developing a deep learning model that can automatically implement the entire set of accurate pre-processing to analyse MRI scans. This includes standardizing image intensities, spatial alignment, and artifact removal to ensure uniformity and quality across all input data.
- Segmentation of Hippocampus: Achieve high accuracy in segmentation of certain regions of the brain that are required for detection of neurodegenerative diseases using Deep Learning Techniques. Medical-grade precision is required, with

evaluation of overlap between model output and expert-annotated ground truth to assess and analyze errors.

- Classification of MRIs: Classify obtained MRIs into types as Cognitively Normal and Alzheimer's Disease Patient. In the future scope, the aim is to implement multi-class classification based on progression of the disease
- Simplicity and Efficiency: Design a streamlined pipeline that eliminates the need for additional tests and assessments, reducing the time and resources required for diagnosis.
- Scalability and Generalizability: Ensuring that the developed pipeline can be applied to diverse MRI datasets and is scalable for potential integration into clinical settings.
- Interpretability and Explainability in Clinical Workflows: Deep learning models often act as black boxes. Introducing interpretability along with explainability is essential for clinical adoption, allowing practitioners to understand and trust model decisions.

1.7 Technical Approach

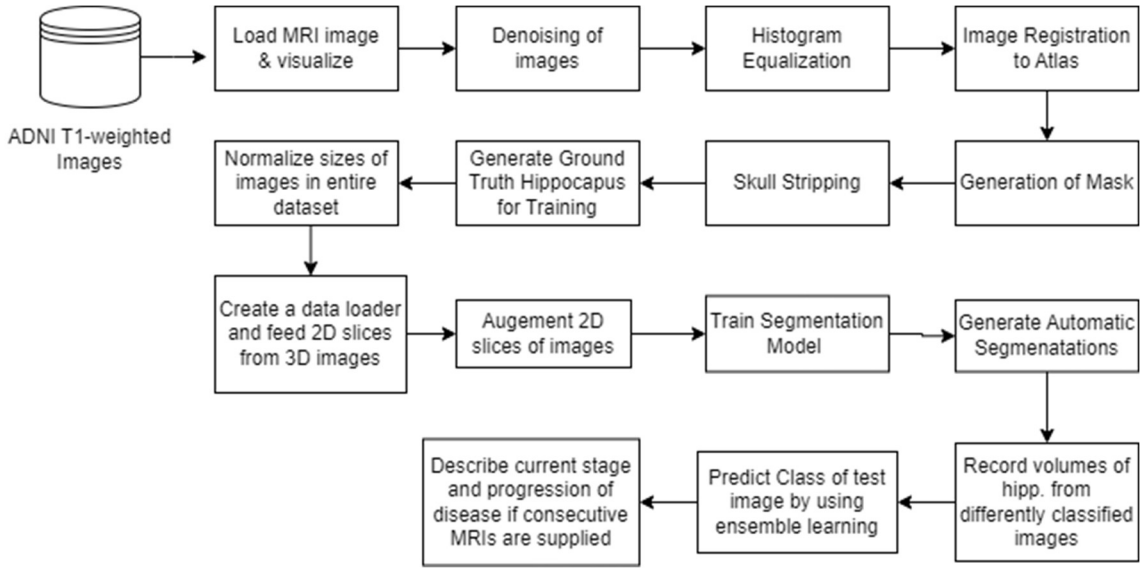


Fig 1.1 Block Diagram

Fig 1.1 demonstrates the overall clinical workflow which was implemented, encompassing MRI acquisition, preprocessing, hippocampal segmentation using deep learning models, volume extraction, and classification of Alzheimer's disease status through integration of anatomical features with clinical metadata

1.7.1 WorkFlow

Data Preprocessing

- Loading and visualization of 3D brain MRIs in NIfTI format
- Denoising to reduce scanner-induced artifacts
- Histogram equalization for consistent intensity scaling
- Alignment to standard brain atlases for spatial normalization
- Skull stripping to isolate brain regions
- Generation of ground truth hippocampal segmentations (FSL-FIRST + manual annotations)
- Normalization and resampling to consistent shape (256×256×166)
- Custom PyTorch data loaders for 2D slices and 3D volumes
- Data augmentation (flipping, rotation, intensity scaling)

Model Architecture Exploration

- UNet-2D: Fast but fragmented segmentations
- UNet-3D: Better spatial relationships but higher computational cost
- Attention UNet: Enhanced boundary precision with increased complexity
- MONAI's UNet: Optimized for medical imaging with strong baseline performance
- SWIN-Unet: Promising theory but difficult implementation with hardware constraints

Advanced Training Techniques

- Knowledge distillation: 3D UNet teacher guiding smaller student model
- Reverse knowledge distillation: Bidirectional learning between CNN and Transformer models
- TPU utilization for efficient complex model training

Classification Methodology

- Extraction of segmented hippocampal volumes
- Integration with demographic metadata (age, gender)
- Machine learning classification (Random Forests, shallow neural networks)
- Fusion of anatomical and clinical information

Performance Observations

- UNet-2D: Lightweight but structurally limited
- UNet-3D: Strong accuracy with higher compute demands
- Attention UNet: Precise boundaries with added complexity
- MONAI UNet: Balanced performance with clinical workflow integration
- SWIN Unet: Theoretical power limited by computational demands
- Reverse knowledge distillation: Best overall performance leveraging complementary architectures.

This comprehensive workflow illustrates a methodical and technically sound approach to automated segmentation and classification of brain MRIs for Alzheimer's Disease research. From robust preprocessing—ensuring data consistency and anatomical relevance—to exploring a spectrum of deep learning architectures (ranging from classical UNet variants to transformer-based models like SWIN-Unet), each step is carefully designed to improve model performance and clinical applicability.

The integration of advanced training strategies such as knowledge distillation, including its reverse variant, demonstrates a deep understanding of leveraging complementary strengths across models. Additionally, combining volumetric segmentation data with demographic metadata in the classification stage highlights a holistic approach that goes beyond imaging alone.

Overall, this pipeline reflects a strong alignment with both computational rigor and clinical relevance, making it a promising framework for real-world neuroimaging applications.

CHAPTER 2

Methodology and Resources

2.1 Introduction

This section outlines the methodologies and resources used to develop the deep learning-based system for MRI brain scan segmentation and classification. The process involves several key stages, including the acquisition and preprocessing of biomedical images, the application of segmentation techniques, and the use of knowledge distillation to optimize computational efficiency. Additionally, the resources, such as datasets and tools, that support the development and evaluation of the system are detailed. The integration of Vision Transformers allows for improved handling of the 3D nature of MRI scans, providing more accurate segmentation and classification results.

2.2 Bio-Medical Images

Various sequences of MRI images from the ADNI and HarP datasets were utilized to enhance generalizability across multiple datasets. Specifically, selecting grad-wrapped, B1 non-uniformity corrected, and N3 non-uniformity corrected MRI scans acquired at 1.5T field strength. Choosing these modalities minimized additional preprocessing requirements while preserving data quality. This approach improved segmentation consistency and classification accuracy, with models trained on ADNI demonstrating strong generalization when validated on external datasets.

2.3 Pre-processing

MRI image preprocessing is essential for optimizing data quality, ensuring consistency, and improving model performance in hippocampus segmentation and disease classification. The process starts with dataset acquisition from sources like ADNI and HARP, followed by denoising techniques such as Gaussian filtering or wavelet denoising to preserve anatomical details. Images are aligned to a standardized coordinate system using the MNI atlas, and skull stripping removes non-relevant structures. Additional steps include intensity normalization, resizing, histogram equalization, and slice selection to

enhance anatomical visibility. This preprocessing pipeline refines MRI images, enabling precise segmentation and accurate disease classification.

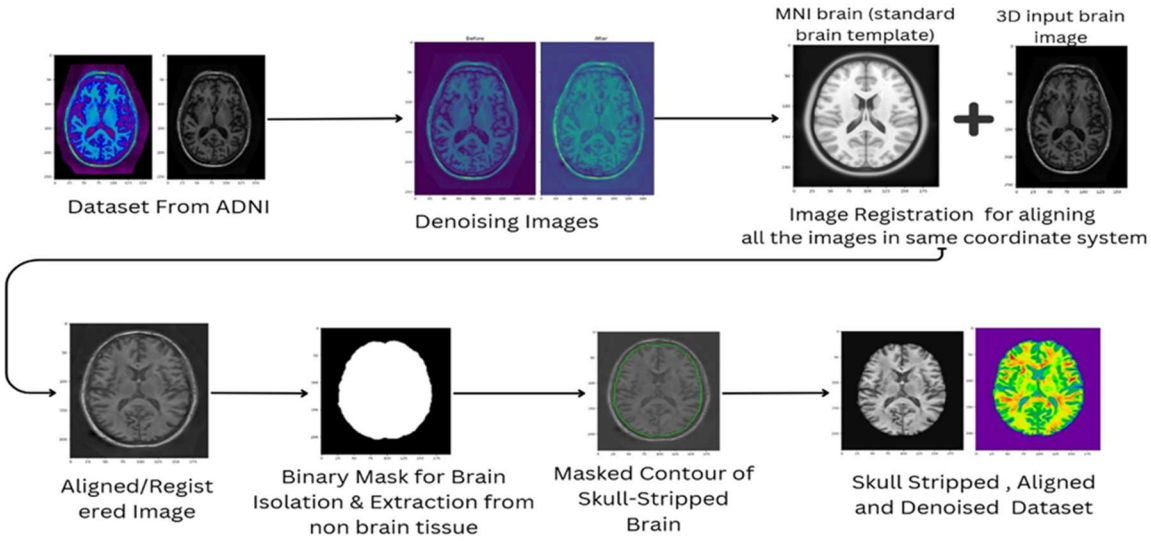


Fig 2.1 .(A) Pre-Processing Pipeline

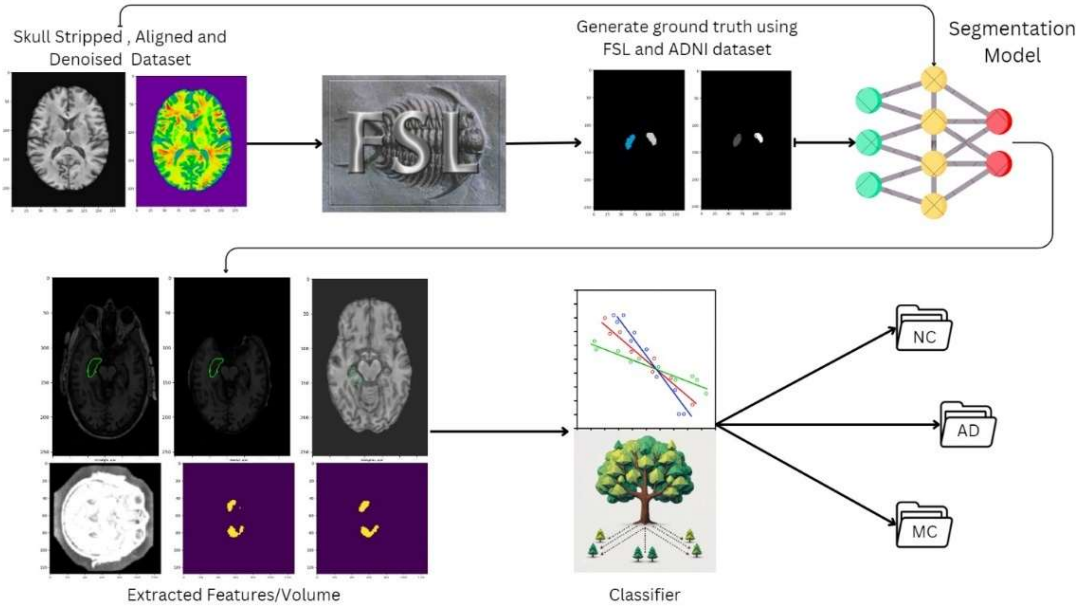


Fig 2.1 .(B) Generic pipeline for AI based Approach towards Alzheimer's Disease detection and classification

This figure shows the generic pipeline used for Alzheimer's Disease classification using an AI based approach. Initially, the image undergoes skull stripping, aligning with a

standard template and denoising. Then, it is passed through the FSL software to generate Ground Truth labels of the hippocampus. The model is trained based on these preprocessed images and their respective ground truth labels, The classifier is then trained based on the generated labels or hippocampus volumes, to decide the presence and stage of the disease in the patient.

2.3.1 Denoising

Histogram equalization was employed as a key denoising technique to improve image contrast and normalize intensity distribution across MRI scans. This method enhances structural visibility by adjusting pixel intensities, making features such as the hippocampus and amygdala more distinguishable. By applying histogram equalization before skull stripping, better separation between brain tissue and non-brain structures was ensured, leading to more accurate segmentation.

2.3.2 Image Registration and Alignment

Image registration over the MNI atlas ensures consistency by aligning MRI scans to a common coordinate system, facilitating generalization across datasets. This step standardizes spatial orientation and anatomical structures, making comparisons across subjects more reliable. This alignment is crucial for hippocampus segmentation, as it ensures that the region of interest is consistently positioned across all images, enhancing the accuracy of deep learning models.

2.3.3 Skull Stripping

Skull stripping is a critical preprocessing step that enhances the accuracy of segmentation and classification tasks by isolating brain tissue and removing non-relevant structures such as the skull, scalp, and dura. To improve contrast and tissue differentiation, histogram equalization was first applied, enhancing the visibility of anatomical structures. A binary mask was then utilized to extract only the brain region, ensuring that unwanted structures were eliminated. The final output was a skull-stripped, aligned, and denoised MRI dataset, optimized for hippocampus segmentation and classification, ultimately enhancing the performance and generalizability of deep learning models.

2.4 Segmentation

Segmentation of the hippocampus is a key task in detecting neurodegenerative diseases. Deep learning techniques (e.g., Vision Transformers for their ability to capture global context in 3D images) are employed to segment the hippocampus. This method improves the accuracy of segmentation compared to traditional methods, particularly in complex 3D medical images. The segmentation pipeline is designed to handle overlapping areas and minimize errors, with accuracy validated against expert-annotated ground truth to ensure precision.

2.4.1 Swin Transformer

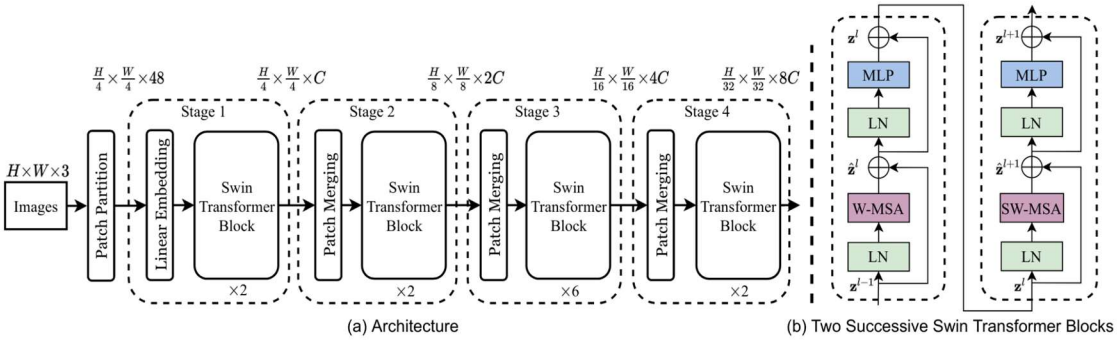


Fig 2.2. SWIN Transformer

The Swin Transformer is a hierarchical vision transformer that uses shifted windows to capture local and global features for image processing. Its structure captures the Multi-Scale features efficiently with linear computational complexity, enabling scalable and high resolution image processing. Basic architectural blocks include a patch partitioning layer which splits 3D image into non-overlapping patches (e.g., 4×4 pixels) ,each serving as a 'token' for the model. These patches are passed through the linear embedding layer to project them into an embedding space of dimension C. In Stage 1, multiple Swin Transformer block process the patches at a resolution of H/4×W/4 , using Window Multi-head self attention (W-MSA) to compute key-query weights for patches
within local windows. Global context limitations are compensated by Shifted W-MSA (SW-MSA), which connects patches across windows in consecutive layers. Patch merging layers in stage 2, 3 and 4 progressively reduce resolution (e.g., H/8 , H/16 , H/32
) while increasing feature dimensions (2C , 4C , 8C) .

Swin Transformer can also be integrated into hybrid architectures as a backbone, combining its hierarchical feature extraction with other methods (e.g., CNN) for enhanced performance, as detailed below.

2.4.2 Swin UNet

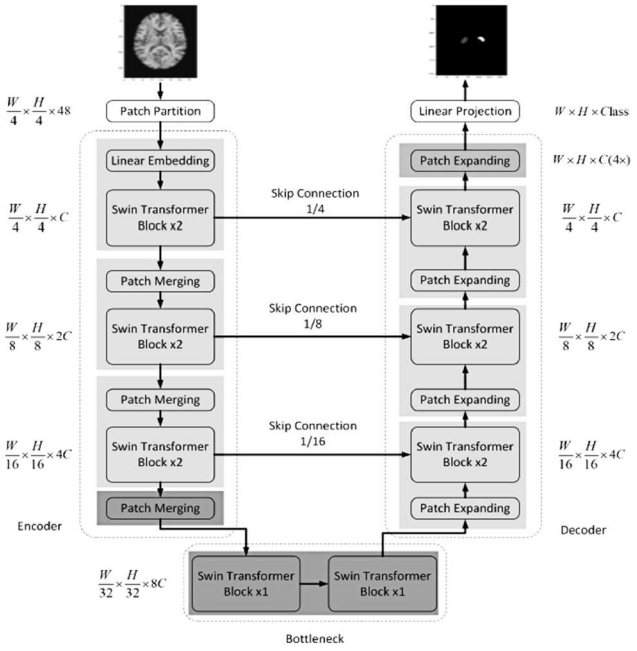


Fig 2.3. SWIN-UNet

The Swin-UNet architecture follows an encoder-decoder structure using Swin Transformer blocks. The encoder begins with an input image of size 224×224 and a patch size of 4. The image is tokenized into C-dimensional patches at a resolution of *H/4×W/4*, which are then processed through two Swin Transformer blocks without altering feature dimensions or resolution. A patch merging layer then reduces the number of tokens by half (2× downsampling) while doubling the feature dimension, and this process is repeated three times. Each merging layer splits the input into four parts, concatenates them, and applies a linear layer to unify dimensions. At the bottleneck, two additional Swin Transformer blocks are used to capture deep representations while keeping the dimensions unchanged. The decoder mirrors the encoder, using patch expanding layers to perform 2× upsampling while halving the feature dimensions. Each expanding layer first doubles the input feature

dimensions through a linear projection, then rearranges them to increase resolution and reduce feature dimensions accordingly. To preserve spatial information, skip connections concatenate shallow features from the encoder with the corresponding upsampled decoder features, followed by a linear layer to fuse multi-scale information effectively.

2.5 Computation Reduction via Knowledge Distillation

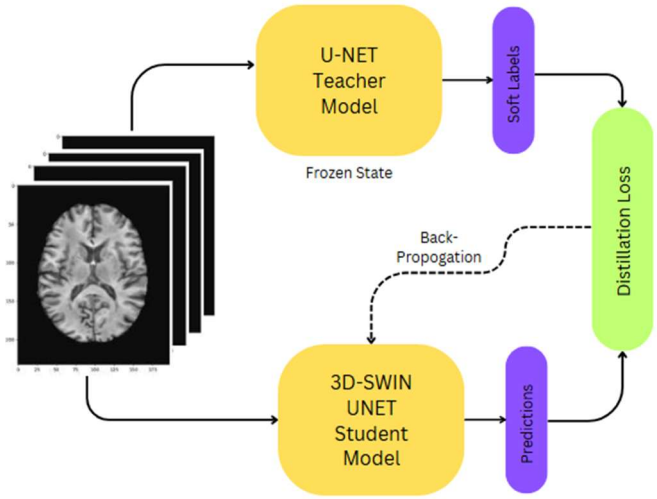


Fig 2.4. Knowledge Distillation with UNET-3D as teacher and SWIN-UNet-3D as student.

Knowledge Distillation (KD) is a model compression technique where a compact student model is trained to replicate the behavior of a high-capacity teacher model. The goal is to preserve the predictive performance of the larger model while significantly reducing computational demands during inference, which is particularly important in resource-constrained environments or real-time medical imaging applications.

In this setup, the teacher model is a 3D UNet trained on high-resolution brain MRI data, optimized for precise segmentation of structures such as the hippocampus. The student model is a computationally efficient, custom-designed Swin UNet 3D, based on a hierarchical Vision Transformer architecture that introduces spatial self-attention and patch merging to process volumetric data effectively.

Unlike traditional supervised learning where the model learns directly from hard ground truth labels (e.g., binary masks), knowledge distillation leverages soft targets — the output

probability distributions of the teacher. These soft labels contain inter-class similarities and uncertainties that encode how confident the teacher is in its predictions. This richer information acts as a form of regularization, enabling the student to learn smoother decision boundaries and generalize better.

2.5.1 Implementation Steps

1. Teacher Model Pretraining

The teacher model is trained on the complete training set using standard supervised learning with segmentation ground truth masks. Its performance is validated to ensure it acts as a reliable source of knowledge for distillation.

2. Freezing the Teacher

Once trained, the teacher model's weights are frozen. This ensures that during distillation, the teacher's knowledge remains static and is not influenced by updates to the student. This also reduces memory usage and computation during training.

3. Matching Student and Teacher Outputs

The student is trained not on the ground truth directly but to minimize the divergence between its predictions and the teacher's predictions. The approach uses:

i. Binary Cross-Entropy (BCE) loss when the teacher outputs are probabilities (e.g., post-sigmoid).

ii. Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss) when the teacher outputs logits (pre-sigmoid values).

This choice ensures numerical stability and preserves the probabilistic behavior of the model depending on the teacher's output type.

4. Activation Compatibility

If the teacher produces probabilities, a sigmoid activation is applied to the student's output before computing the loss, to ensure that both outputs lie in the same range. This alignment is crucial for meaningful distillation.

5. Efficient Prototyping and Validation

To iterate rapidly, the student model is initially trained on a smaller subset of the training data. This approach allows for testing configurations (e.g., learning rates, initialization

schemes, architectural tweaks) and validate the effectiveness of distillation strategies without incurring full computational costs.

6. Monitoring and Evaluation

Throughout training, segmentation accuracy (based on thresholded predictions vs ground truth) and the distillation loss are logged. This provides insight into how well the student is mimicking the teacher, and whether further improvements (e.g., better initialization or architectural modifications) are necessary.

2.5.2 Benefits Achieved

One of the most significant advantages of employing Knowledge Distillation (KD) in the segmentation framework is the substantial gain in computational efficiency. The student model, being architecturally lightweight and significantly smaller in parameter count compared to the teacher model, enables faster inference and drastically lower memory usage, making it highly suitable for real-time clinical applications or deployment on edge devices with limited hardware capabilities. Despite its compactness, the student model achieves high accuracy due to the rich supervision provided by the soft targets of the teacher. These soft outputs embed nuanced inter-class relationships and uncertainty estimates that go far beyond the binary information in traditional ground truth labels. As a result, they act as a powerful form of implicit regularization, guiding the student toward more generalized decision boundaries and reducing the risk of overfitting—particularly important in medical imaging where data can be noisy, imbalanced, or sparse. Furthermore, this framework has the potential to lower the annotation burden in future iterations, as the student can effectively learn from teacher predictions even when explicit manual labels are limited or partially unavailable. From a training perspective, the benefits are equally compelling: due to its reduced complexity, the student model can be trained in a highly time-efficient manner, taking only 60 to 90 minutes over 150 epochs when executed on a Google Colab TPU. This remarkable speed enables rapid experimentation and prototyping, facilitating an agile development cycle for deploying high-performance segmentation models in clinical workflows.

2.6 Classification

To classify the stages of cognitive impairment based on a combination of demographic and structural brain features, a robust supervised machine learning pipeline was followed. The dataset initially included important features such as physical_volume_mm3 (brain volume), Age, and Gender, among others, with the target variable being stage (labelled as CN, MCI, or AD). Label encoding was applied to transform the categorical class labels into numeric form suitable for model training.

In preprocessing, polynomial feature transformation was introduced to capture interaction terms and non-linear relationships between the existing features. Using `PolynomialFeatures` of degree 2, the feature space was expanded while excluding the bias term, allowing the model to capture higher-order patterns that might be predictive of disease stages.

After preprocessing, the dataset was split into training and testing sets using a stratified train-test split to maintain class balance across splits. A powerful ensemble classification system was then constructed using the `StackingClassifier` from scikit-learn. This ensemble combined three diverse base learners: a Random Forest classifier, a Gradient Boosting classifier, and an XGBoost classifier. These base learners were selected to capture different kinds of patterns in the data – randomness, boosting gradients of errors, and optimized decision trees respectively.

A logistic regression model was used as the meta-learner in the stacking ensemble, taking the output probabilities from the base learners and combining them to make a final prediction. The ensemble model was trained using 5-fold cross-validation to ensure generalizability and avoid overfitting. Model performance was then evaluated using the `classification_report` metric, which provided precision, recall, F1-score, and support for each class.

To further improve model generalization, class imbalance and low data volume were addressed by applying controlled data augmentation. Synthetic data points were generated by adding Gaussian noise to numeric features such as brain volume and age. This approach expanded the dataset from under 1000 samples to around 2000, strengthening model learning and reducing overfitting, without compromising the correlation between

volume and stage.

Lastly, to ensure the model can be queried efficiently in real-world scenarios, the trained model was saved using `joblib`, and a small utility function was developed to load it and make predictions. This function enables stage prediction based on just three inputs: brain volume, age, and gender – the minimal information likely available in real-world medical assessments.

CHAPTER 3

Implementation, Testing & Debugging

3.1 Implementation

1. Model Integration

- The segmentation task was handled using CNN based and Transformer based core architectures: UNet , SwinUNet3D and their varients , implemented using PyTorch.
- To optimize computation, lightweight architectures such as SimpleSwinUNet3D were adopted and further enhanced through a knowledge distillation approach, where a more complex UNet teacher guided the learning process.
- The classification module was built separately using scikit-learn and XGBoost, forming a hybrid deep learning and machine learning pipeline.

2. Frameworks and Tools

- PyTorch: Backbone framework for designing deep learning models.
- MONAI: Specialized toolkit for medical imaging, used for preprocessing and model utilities.
- scikit-learn & XGBoost: Used for structured data classification. Scikit-learn was also utilized for generating visualizations to evaluate model performance and interpret classification results.
- Nibabel: Utilized for loading and managing 3D medical imaging data in NIfTI format.
- NumPy and Pandas: Used in data manipulation, statistical preprocessing, and analysis.
- Joblib: For model serialization and deployment support.

3. Preprocessing Pipeline

- Selecting MRI images of T1-weighted modality from Harp dataset. The dataset has been reviewed by various research papers and thereby selected due to its generalizability.

- Images used in the pipeline were GradWarp, N3 bias field corrected, and B1 intensity corrected, ensuring standardized and high-quality input.
- Applied histogram equalization to normalize intensity distributions across MRI volumes.
- Performed skull stripping and spatial standardization via registration to a standardized space like MNI format for spatial alignment.
- Generation of ground truth masks via the FSL software.
- In case of classification , used label encoding for categorical features (e.g., gender, diagnosis stage) in the pipeline.
- Applied Polynomial Features (degree 2) to enhance feature representation by capturing non-linear relationships.

4. Training Hyperparameters and Optimization Techniques

- Optimizer: The Adam optimizer was employed with a learning rate of 1e-3, weight decay of 1e-5, and AMSGrad enabled (Adam(model.parameters(), lr=1e-3, weight_decay=1e-5, amsgrad=True)) to ensure efficient and stable model training.
- Batch Size: 1, to handle the high memory demands of 3D volumes.
- Epochs: Configurable based on model complexity and early stopping conditions.
- Loss Function: Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss) for segmentation.
- Stratified Train-Test Split: Maintained label distribution across training and testing sets for classification.

5. Computation Optimization Strategy

- The SimpleSwinUNet3D student model was trained not only on ground truth labels but also to minimize divergence from the UNet teacher's soft predictions.
- A sigmoid activation was applied to student outputs when the teacher produced probability maps, ensuring consistent output scaling.
- This setup reduced computational costs and memory consumption while retaining segmentation quality close to the teacher's.
- The SimpleSwinUNet3D student model was trained not only on ground truth labels but also to minimize divergence from the UNet teacher's soft predictions, enabling efficient knowledge transfer.

- A sigmoid activation was applied to student outputs when the teacher produced probability maps, ensuring consistent output scaling and compatibility between both models.
- This setup significantly reduced computational costs and memory consumption while retaining segmentation quality close to the teacher's.
- Unlike the original SwinUNet, which is computationally expensive and time-consuming to train, the SimpleSwinUNet3D model benefited from reduced training time due to the knowledge distillation approach.
- The SwinUNet model captured global context, while the UNet teacher focused on local features, providing complementary strengths for the student model's learning process.

3.2 Algorithm

3.2.1 UNet

1. Input images of size (H, W, C) (3.1)
2. Apply convolution followed by ReLU activation for downsampling

   $Y_{down} = \text{ReLU}(\text{Conv}(X, W) + b)$ (3.2)
3. Compress information at the bottleneck using convolution layers.

   $Y_{bottleneck} = \text{ReLU}(\text{Conv}(Y_{down}, W) + b)$ (3.3)
4. Upsample and concatenate feature maps from the contracting path.

   $Y_{up} = \text{Conv}\big(\text{Concat}\big(Y_{up}, Y_{down}\big), W\big) + b$ (3.4)
5. Apply a final convolution to generate the segmentation map.

   $$\hat{Y} = \text{Sigmoid}\big(Y_{final}\big)$$ (3.5)

3.2.2 SWIN Transformer

1. Input the image data with dimensions

   $$((H, W, C)).\ X \in R^{H \times W \times C}$$ (3.6)
2. Divide the input image into non-overlapping patches

   $$X_p = \text{PatchSplit}(X, P)$$ (3.7)
3. Flatten the patches and linearly embed them into a lower-dimensional space .

$$Z_0 = X_p W_p \tag{3.8}$$

4. Apply a series of Swin Transformer blocks

$$Z = \text{SwinBlock}(Z_{i-1}) \qquad \text{for } i = 1, \dots, B \tag{3.9}$$

5. Compute self-attention within local windows.

$$\text{Attention}(Z_i) = \text{softmax}\left(\frac{Z_i Z_i^T}{\sqrt{D}}\right) Z_i \tag{3.10}$$

6. Merge neighboring patches to reduce the spatial dimensions while increasing the channel depth

. $$Z_{i+1} = \text{PatchMerge}(Z_i) \tag{3.11}$$

7. Apply a fully connected feedforward network to each output of the Swin block.

$$Y = \text{FFN}(Z_i) \tag{3.12}$$

8. Use the output from the last block for classification or regression tasks.

$$\hat{Y} = \text{Softmax}(Y) \tag{3.13}$$

3.2.3 SWIN UNet

1. Patch Splitting and Embedding: Divide the input image $X \in R^{H \times W \times C}$ into non-overlapping patches of size $P \times P$, then flatten and project:

$$X_p = \text{PatchSplit}(X, P) \tag{3.14}$$

$$Z_0 = X_p \cdot W_p \tag{3.15}$$

2. Transformer Block : For each block in the encoder/decoder.

$$Z = \text{SwinBlock}(Z_{i-1}) \qquad \text{for } i = 1, \dots, B \tag{3.16}$$

Within each block :

$$\text{Attention}(Z) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right) V \tag{3.17}$$

Where

$$Q = ZW_Q, \quad K = ZW_{K,} \quad V = ZW_V$$

Followed by :

$$Z = Z + \text{Attention}(Z) \tag{3.18}$$

$$Z = Z + \text{FFN}(Z) \tag{3.19}$$

3. Patch Merging Layer (Encoder Downsampling) : Split each patch into 4 sub-patches, concatenate and apply linear projection

$$Z_{\text{cat}} = \text{Concat}(Z_1, Z_2, Z_3, Z_4) \tag{3.20}$$

$$Z_{\text{merged}} = Z_{\text{cat}} \cdot W_{\text{merge}} \tag{3.21}$$

4. Bottleneck : At the deepest level

$$Z_b = \text{SwinBlock}(Z_{\text{enc3}}) \tag{3.22}$$

5. Patch Expanding Layer (Decoder Upsampling) : Linear expand channel dimensions and reshape.

$$Z_{\text{expand}} = Z \cdot W_{\text{expand}} \tag{3.23}$$

$$Z_{\text{upsampled}} = \text{Rearrange}(Z_{\text{expand}}) \tag{3.24}$$

6. Skip Connections : Fuse features from encoder and decoder

$$Z_{\text{skip}} = \text{Concat}(Z_{\text{encoder}},\ Z_{\text{decoder}}) \tag{3.25}$$

$$Z_{\text{fused}} = Z_{\text{skip}} \cdot W_{\text{fuse}} \tag{3.26}$$

3.2.4 Teacher Student Knowledge Distillation

1. The UNet model is an encoder-decoder architecture commonly used in segmentation task.

2. Encoder contracts the spatial dimensions while increasing channels.

$$E_{i+1} = \text{DownConv}_i(E_i), \quad E_0 = x \tag{3.27}$$

3. Decoder expands the features while fusing with skip connections.

$$D_{i-1} = \text{UpConv}_i(D_i) + E_{i-1} \tag{3.28}$$

4. The final segmentation output can be represented as :

$$\widehat{y_T} = \sigma\big(f_T(D_0)\big) \tag{3.29}$$

5. The Swin UNet leverages shifted window attention in 3D space, replacing convolutions with self-attention blocks.

$$z_0 = \text{PatchEmbed}(x) \tag{3.30}$$

6. Swin Transformer block with shift :

$$z_l = \text{SwinBlock}(z_{l-1}), \quad l = 1, ..., \tag{3.31}$$

7. Each SwinBlock consists of Multi-head self-attention within windows and Shifted windowing between layers

8. Decoder Reconstruction :

$$\widehat{y_S} = \sigma\big(f_S(z_L)\big) \tag{3.32}$$

9. Knowledge Distillation Loss:

$$\mathcal{L}_{\mathcal{KD}} = \text{BCEWithLogits}(f_S(x), \widehat{y_T}) \tag{3.33}$$

10. If using labels between 0 and 1 , use BCE loss between student output and ground truth:

$$\mathcal{L}_{\mathcal{GT}} = -[y \log(\widehat{y_S}) + (1 - y) \log(1 - \widehat{y_S})] \tag{3.34}$$

11. Combined loss between KD and ground truth supervision:

$$\mathcal{L}_{total} = \lambda \cdot \mathcal{L}_{\mathcal{KD}} + (1 - \lambda) \cdot \mathcal{L}_{\mathcal{GT}} \tag{3.35}$$

12. Used to evaluate segmentation performance voxel-wise

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left[\left(\widehat{y_S^{(i)}} > 0.5\right) = y^{(i)}\right] \tag{3.36}$$

3.3 Testing

- Dataset and Generalization: All models were tested on a reserved test set, separate from training and validation data, to ensure unbiased evaluation and check generalization performance, especially given the limited dataset size.
- Evaluation Metrics: Segmentation performance was assessed using Dice Coefficient, IoU, Precision, and Recall, while classification accuracy was measured using precision, recall, F1-score, and overall accuracy via classification_report.
- Thresholding and Output Processing: Probability outputs from segmentation models were thresholded and binarized to obtain final masks for accurate metric computation and comparison against ground truth.
- Quantitative and Visual Validation: Model predictions were validated through both statistical metrics and qualitative inspection, with segmentation overlays on MRI slices reviewed for anatomical correctness and consistency.
- UI Testing: The interface was tested for responsiveness and functionality, with deployment handled via ngrok for secure public access during development.

3.4 Performance Parameters

3.4.1 Accuracy

Accuracy reflects the number of correct predictions made by the model out of all predictions. It provides an overall view of the model's performance across all classes.

$$Accuracy = \frac{Correct\ Positive\ Predictions + Correct\ Negative\ Predictions}{Total\ Predictions} \tag{3.37}$$

3.4.2 Precision

Precision measures how accurately the model identifies positive cases among all predicted positives.It calculates the number of true positive results out of everything the model classified as positive.

$$Precision = \frac{True\ Positive\ Predictions}{True\ Positive\ Predictions + Fa \quad Positive\ Predictions} \tag{3.38}$$

3.4.3 Recall

Recall evaluates the model's ability to capture all relevant positive instances. It calculates the ratio of correctly identified positive instances to the actual total of positive instances. This metric provides insight into how thoroughly the model detects positive cases.

$$Recall = \frac{True\ Positive\ Predictions}{True\ Positive\ Predictions+ \quad Negative\ Predictions} \tag{3.39}$$

3.4.4 F1-Score

F1-Score combines Precision and Recall into a single metric, balancing both aspects of performance .It offers a more comprehensive measure of accuracy, especially in situations with imbalanced data. This metric ensures that both false positives and false negatives are considered when evaluating the model's performance.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recal} \tag{3.40}$$

3.5 Debugging and Refinement

- Error Tracing: Model development involved systematic debugging using detailed logging and inspection of tensor shapes to resolve runtime errors and shape mismatches, ensuring consistency across 3D data operations.
- Model Tuning: Hyperparameters such as learning rate, batch size, and epochs were fine-tuned iteratively. Optimization was guided by validation loss trends and performance metrics, enhancing both convergence and generalization.
- Loss Monitoring: Training and validation losses were continuously monitored to detect signs of overfitting or underfitting. Early stopping mechanisms were employed to maintain optimal training durations.
- Visualization Tools: Libraries like matplotlib were used to visualize loss curves, segmentation overlays, and classification outputs to guide refinement.
- Interactive Widgets: Used in the web interface and Jupyter notebooks for real-time input selection, parameter tuning, and dynamic visualization of segmentation and classification results.

Each stage of the system , from data preprocessing and model training to deployment was iteratively debugged and tested to identify bottlenecks, resolve inconsistencies, and ensure optimal functionality. This rigorous process led to significant improvements in model performance, stability, and overall user experience, making the pipeline more robust and reliable for real-world application.

3.6 Deployment and Integration

- Model Deployment: The trained models were integrated into a lightweight Streamlit-based application, creating an intuitive web interface that allows users to upload MRI scans and receive automated segmentation and classification results.
- Preprocessing Integration: Automated preprocessing steps, such as normalization, resizing, and intensity standardization, were incorporated within the application to ensure consistency with the model's training pipeline, reducing user burden and improving inference accuracy.
- Server Exposure: To facilitate remote testing and real-time demonstrations, Ngrok was employed to securely tunnel and expose the local Streamlit server. This enabled external access without complex hosting infrastructure during the development phase.
- Web Interface: Streamlit provided a responsive and interactive front-end framework. The interface supports file uploads, model inference visualization, and real-time display of segmentation overlays and diagnostic classifications.
- Integration Testing: Comprehensive testing was conducted to ensure the end-to-end pipeline, from image input to result visualization functioned smoothly. Performance, responsiveness, and reliability were validated under typical usage conditions.

These deployment strategies collectively ensured a seamless user experience by bridging the gap between complex deep learning models and end-users through an accessible, real-time interface. The integration of tools like Streamlit and Ngrok not only facilitated rapid prototyping and demonstration but also laid the groundwork for scalable deployment in future clinical or research settings.

3.7 Website Development for User Interaction

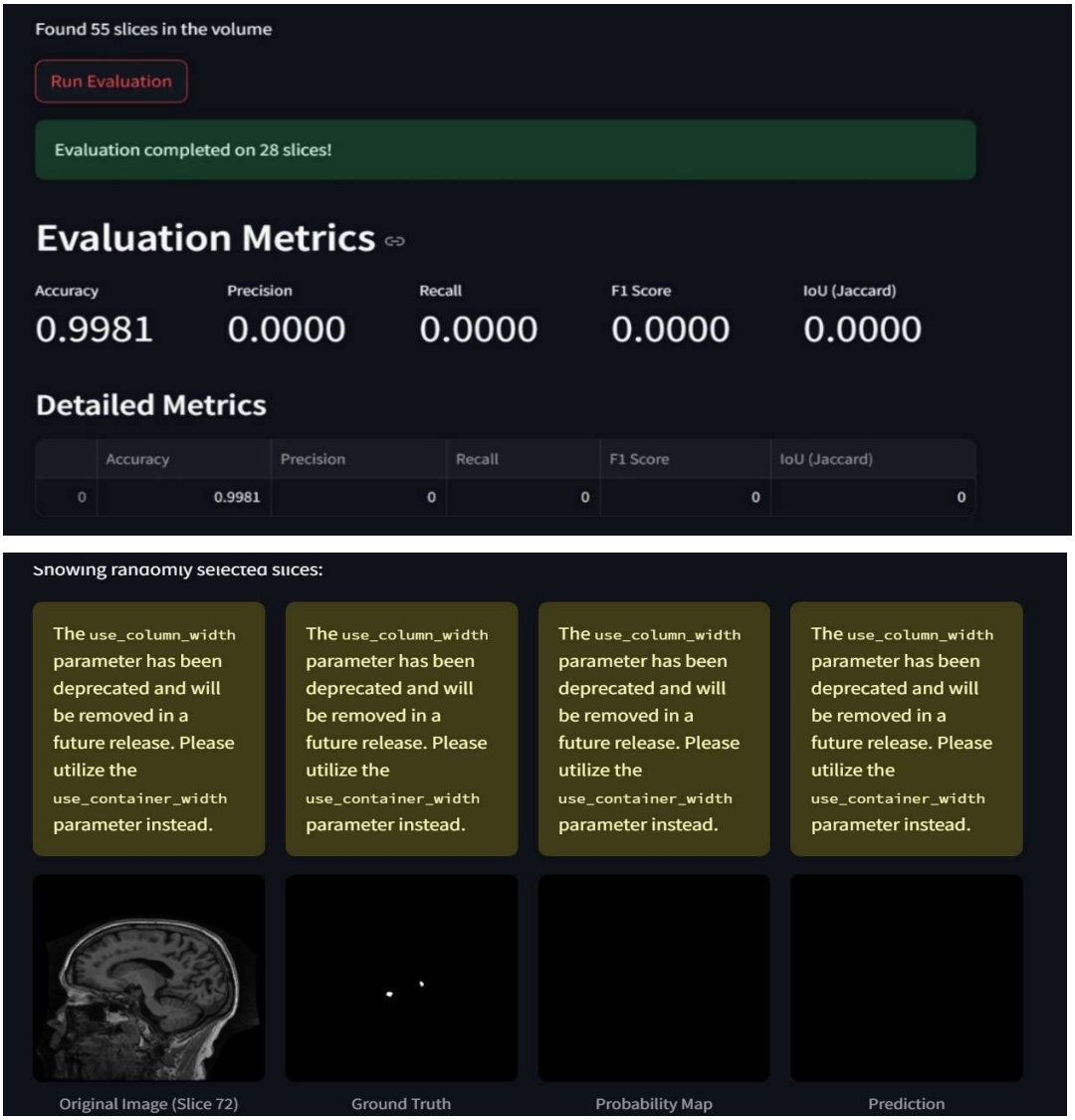


Fig 3.1. Web Interface

The web interface, built using Streamlit, provides an intuitive platform for users to interact with the model.

- Evaluation and Metrics: The UI allows users to trigger model evaluation on MRI slices by clicking a simple button ("Run Evaluation"). After processing, the interface displays key performance metrics in a clean, organized dashboard format with both summary and detailed views, making complex data easily digestible.

- Visualization of Results: The interface presents multiple visualization panels that display the original images alongside various analytical outputs. This side-by-side arrangement enables immediate visual comparison, enhancing the interpretability of the model's performance without requiring users to navigate between different screens.
- Interactivity and Usability: Interactive elements like buttons and selection tools enable users to control the analysis process directly from the interface. The responsive design provides immediate feedback as users interact with different components, creating a seamless experience for both technical and non-technical users alike.
- Streamlit Advantages: Streamlit's Python-based framework enables rapid development of interactive data applications with minimal frontend coding required. Its widget system allows for quick implementation of controls like sliders, dropdowns, and file uploaders that can be directly tied to the underlying machine learning model's parameters.
- Deployment with ngrok: For sharing and collaboration, ngrok provides a straightforward deployment solution by creating secure tunnels to the locally hosted Streamlit application. This approach enables immediate sharing of the interactive interface with stakeholders or team members anywhere in the world, without requiring complex cloud infrastructure setup during development and testing phases.

CHAPTER 4

Results and Discussions

4.1 Segmentation Results

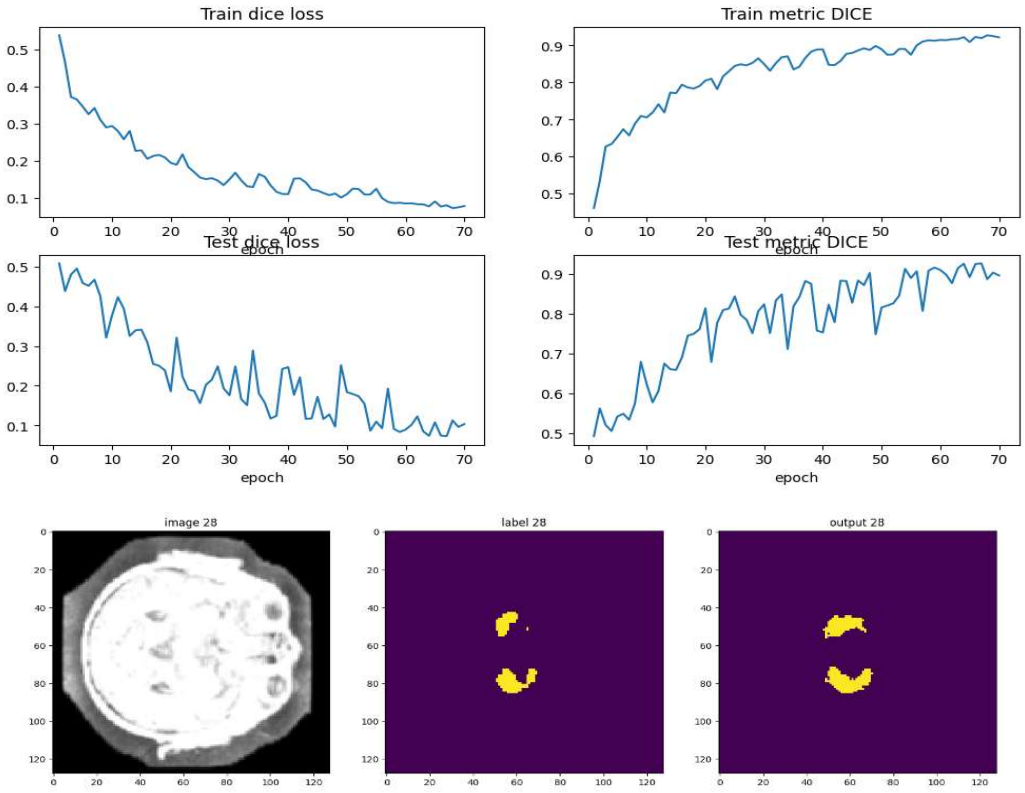


Fig 4.1. 3D-UNET metrics performance and prediction output

Fig 4.1 shows U-Net evaluation graphs with training and testing Dice loss and coefficient over epochs, along with ground truth (GT) and predicted segmentation (Seg) outputs. The steady drop in loss and rise in Dice score (~0.9 train, ~0.85 test) indicate accurate segmentation and good generalization.

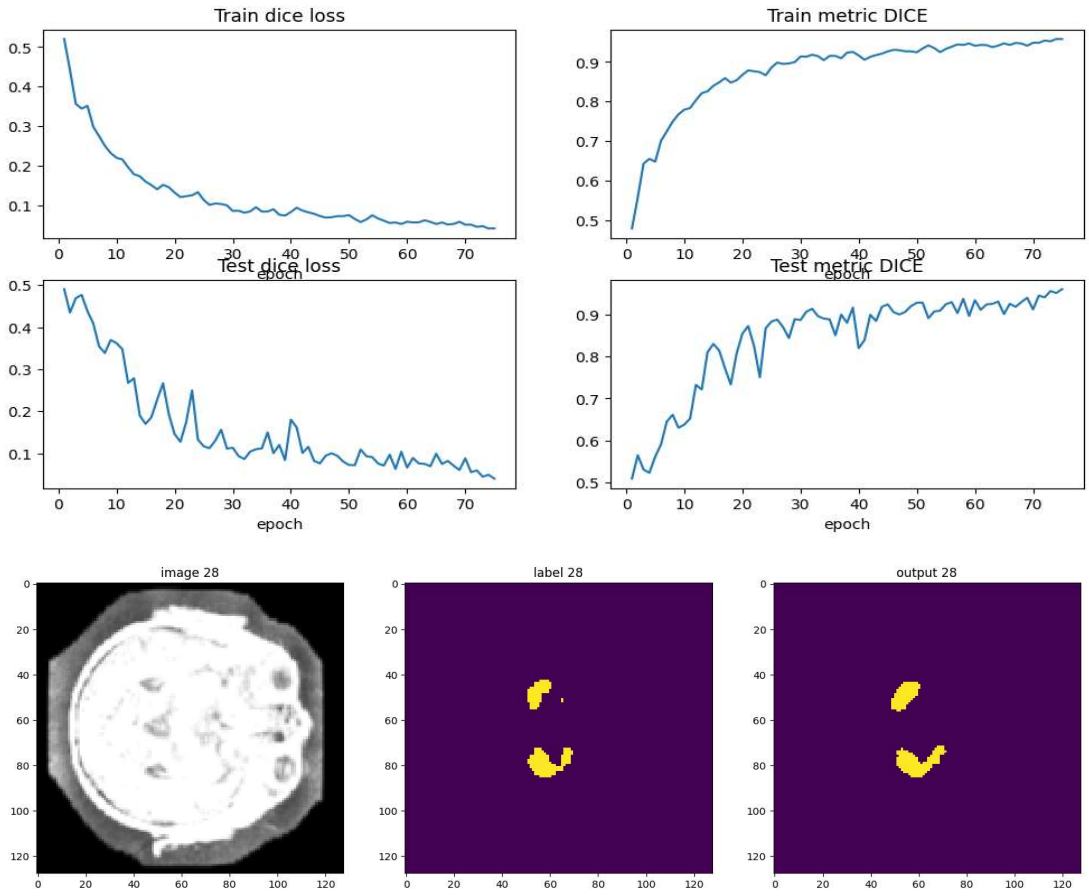


Fig 4.2. MONAI-UNET metrics performance and prediction output

Fig 4.2 shows MONAI U-Net evaluation with ground truth (GT) and predicted outputs, where Dice loss drops and Dice score rises (~0.92 train, ~0.88 test), confirming accurate and generalized segmentation. This indicates successful model convergence with minimal overfitting.

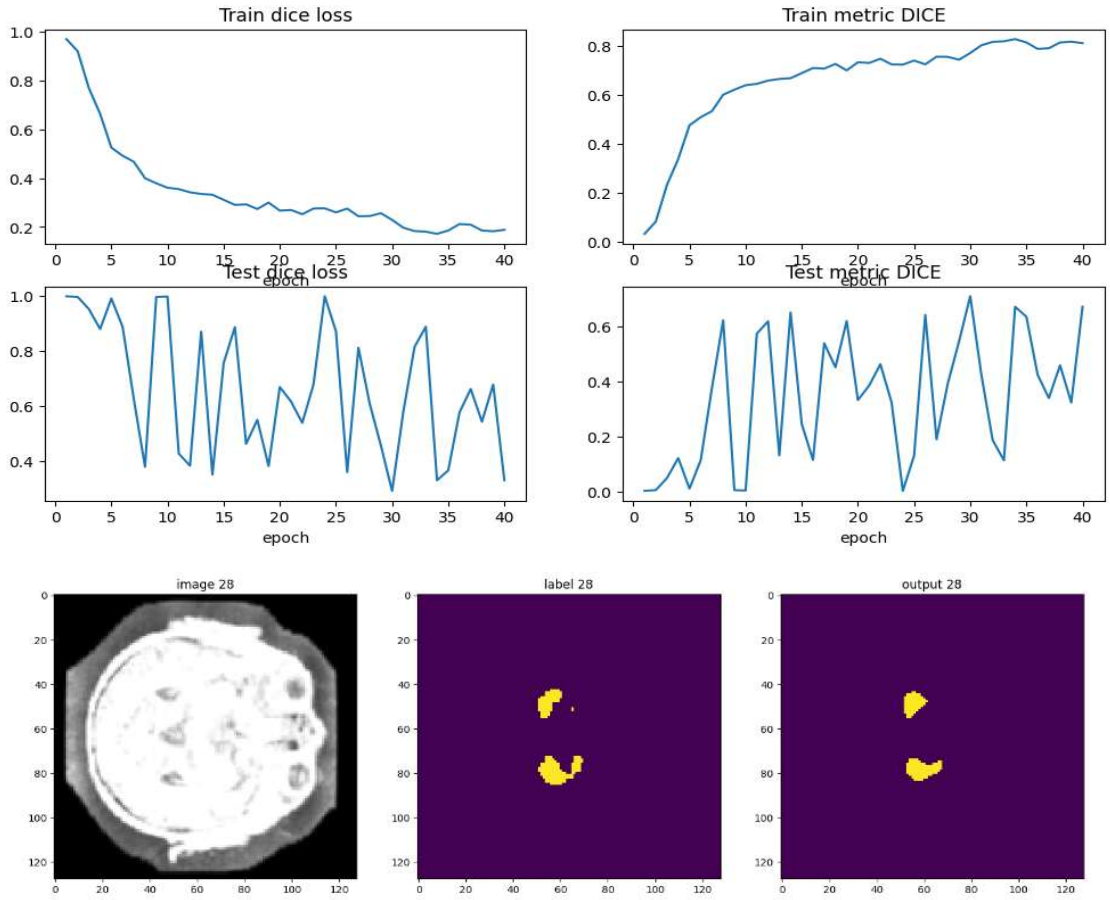


Fig 4.3. Attention-UNET metrics performance and prediction output

Fig 4.3 shows Attention U-Net evaluation with Ground Truth (GT) and predicted outputs, where training Dice loss decreases and Dice score improves (~0.8), but test curves are highly unstable, indicating poor generalization and inconsistent segmentation on unseen data.

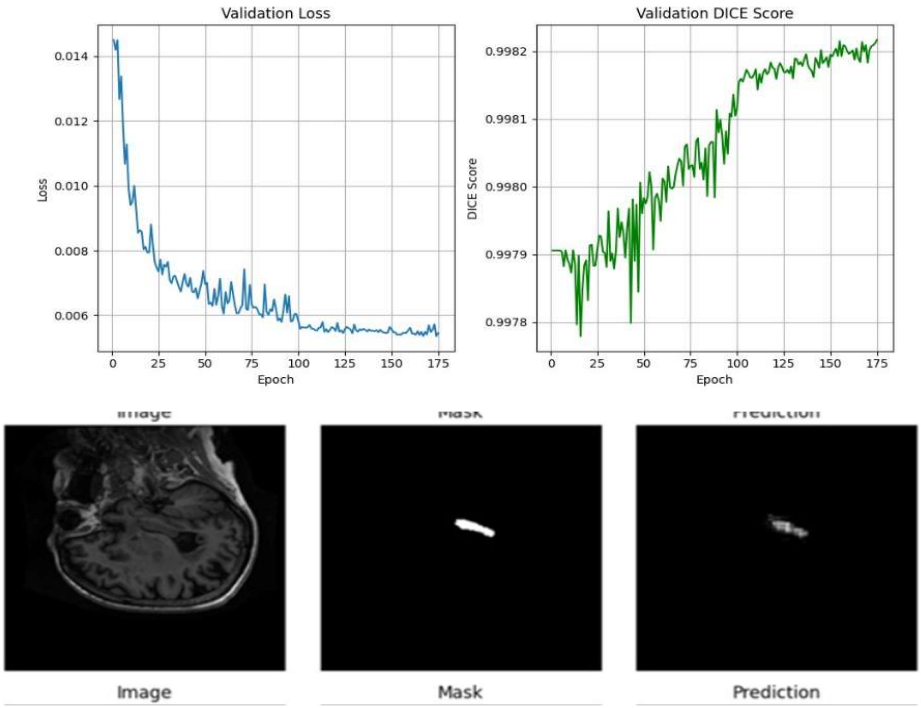
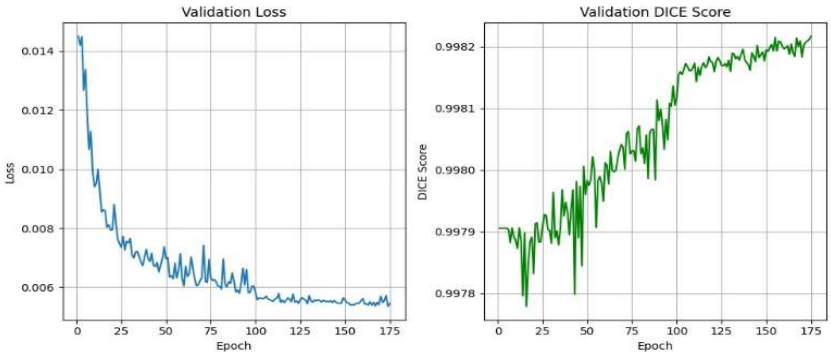


Fig 4.4. SWIN-UNET metrics performance and prediction output

Fig 4.6 shows validation performance of Swin U-Net along with its ground truth(GT) and predictions, which, despite its transformer-based design, shows comparatively lower performance with a Dice score of ~0.75 and higher Dice loss (~0.24), indicating weaker segmentation accuracy and less confident predictions.



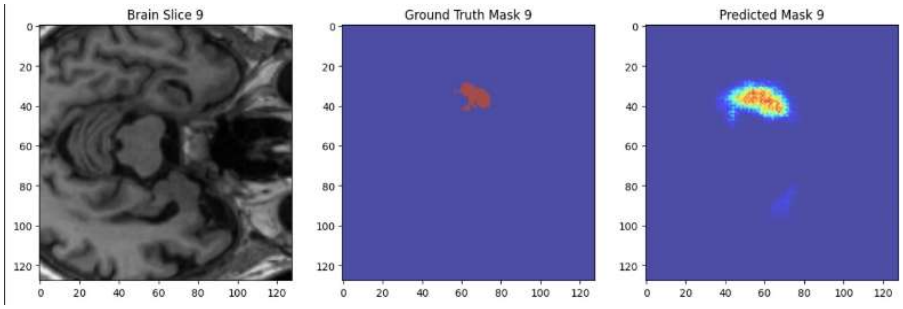


Fig 4.5. KD with UNET3D as student and teacher (different layers and channels) metrics performance and prediction output

Fig 4.5 shows validation results along with Ground Truth (GT) and predicted outputs, for knowledge distillation with U-Net (teacher) and a lighter U-Net (student). The sharp loss drop and high Dice (~0.9942) indicate efficient learning, strong generalization, and accurate segmentation with reduced model complexity.

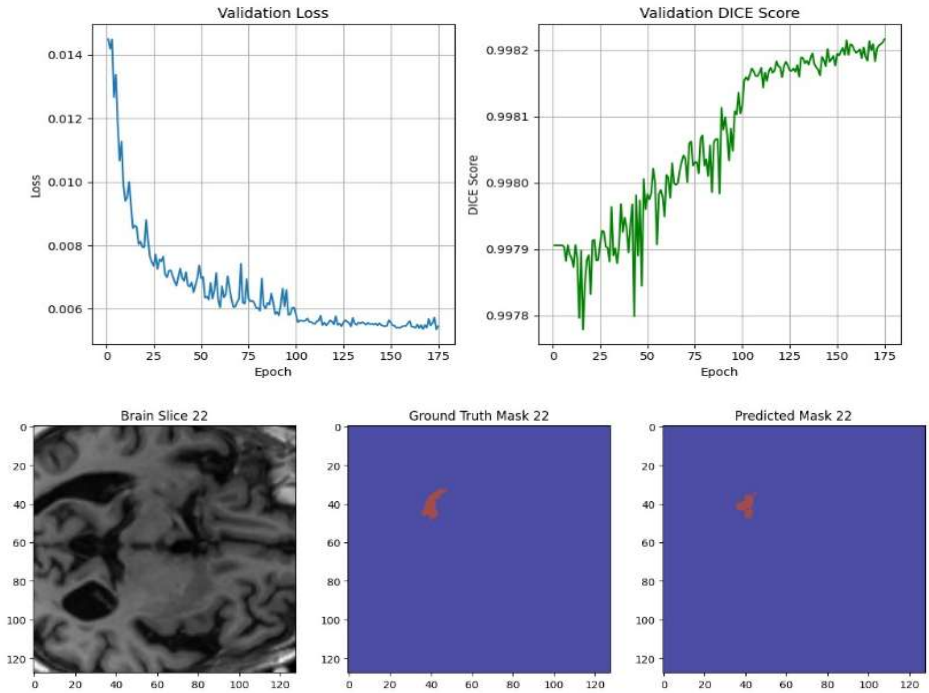


Fig :4.6. RKD with SWIN-UNET3D as student and UNET 3D as teacher metrics performance and prediction output

Fig 4.6 shows the validation performance of the knowledge distillation setup along with Ground Truth (GT) and predicted outputs, where U-Net (frozen) , which is a relatively

smaller model acts as the teacher and Swin U-Net 3Dt as the student. The sharp decline in loss and consistent rise in Dice score (~0.9982) indicate that the student model effectively learns from the teacher's soft predictions, achieving excellent segmentation accuracy, strong generalization, and stable training.

Table 4.1: Segmentation Evaluation Metrics

Model	Epochs	Optimizer	Dice-Coefficient	Dice-Loss
3D-UNET	70	Adam	0.89	0.10
MONAI-UNET	70	Adam	0.96	0.03
Attention-UNET	40	Adam	0.67	0.32
Swin UNET	70	Adam	0.75	0.24
KD 3D-UNET	70	Adam	0.90	0.005
RKD (3D-UNET as teacher SWIN-UNET-3D as student	150	Adam	0.96	0.006

The above table 4.1 summarizes the performance of various segmentation models on the ADNI dataset. It can be observed that traditional methods such as the 3D-UNet and the MONAI-UNet (which employs a transfer learning approach using the MONAI library) achieve high segmentation performance, with Dice coefficients of 0.89 and 0.96 respectively. These models demonstrate strong generalization capabilities on a single dataset. In contrast, the Swin UNet architecture, despite its transformer-based design, exhibits relatively lower performance with a Dice coefficient of 0.75 and a higher Dice loss of 0.24, indicating less confidence in its predictions. Interestingly, the performance of the Swin UNet-3D improves significantly when used as a student model in a reverse knowledge distillation (RKD) framework with 3D-UNet as a teacher, achieving a Dice coefficient of 0.96. Furthermore, knowledge distillation from a 3D-UNet teacher to a student model also yields improvements, as seen with the KD 3D-UNet model, which reaches a Dice coefficient of 0.90 and a Dice loss as low as 0.005. These results highlight the potential of knowledge distillation techniques in enhancing the predictive capabilities of models like Swin UNet that may otherwise underperform.

4.2 Classification Results :

Table 4.2: Random Forest Classifier

Class	Precision	Recall	F1-score	Support
AD	0.88	0.89	0.89	137
CN	0.93	0.93	0.93	110
MCI	0.86	0.86	0.86	153
Accuracy	-	-	0.89	400
Macro Avg	0.89	0.89	0.89	400
Weighted Avg	0.89	0.89	0.89	400

Table 4.3: Gradient Boosting Classifier

Class	Precision	Recall	F1-score	Support
AD	0.87	0.92	0.89	137
CN	0.94	0.95	0.95	110
MCI	0.92	0.86	0.89	153
Accuracy	-	-	0.91	400
Macro Avg	0.91	0.91	0.91	400
Weighted Avg	0.91	0.91	0.91	400

Table 4.4: XGBoost Classifier

Class	Precision	Recall	F1-score	Support
AD	0.86	0.93	0.89	137
CN	0.95	0.94	0.94	110
MCI	0.91	0.86	0.89	153
Accuracy	-	-	0.91	400
Macro Avg	0.91	0.91	0.91	400
Weighted Avg	0.91	0.91	0.91	400

Tables 4.2, 4.3, and 4.4 present a comprehensive comparison of machine learning classifiers for Alzheimer's disease classification using the ADNI dataset. The Random Forest Classifier (Table 4.2) demonstrates solid performance with an overall accuracy of 0.89, showing balanced classification across Alzheimer's Disease (AD), Cognitively Normal (CN), and Mild Cognitive Impairment (MCI) classes. The Gradient Boosting Classifier (Table 4.3) shows improvement with an overall accuracy of 0.91, with particularly strong performance on CN samples (F1-score 0.95). The XGBoost Classifier (Table 4.4) maintains the 0.91 accuracy while showing the highest recall for AD patients (0.93), suggesting enhanced sensitivity for detecting Alzheimer's cases. Notably, both boosting algorithms outperform Random Forest, with CN classification consistently achieving the highest metrics among all three classes, while MCI classification presents the greatest challenge. These results highlight the potential of ensemble-based methods, particularly gradient boosting techniques, in accurately differentiating between the three cognitive states based on neuroimaging features.

CHAPTER 5

Conclusions

This project presents a comprehensive machine learning pipeline for both hippocampus segmentation and Alzheimer's Disease (AD) classification. The pipeline is structured around two primary tasks: first, the segmentation of the hippocampus from 3D brain MRI scans; and second, the classification of individuals into cognitive states—Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), or Cognitively Normal (CN)—based on a combination of clinical and neuroimaging features such as hippocampal volume, age, and gender.

For the segmentation task, several deep learning models were explored, including UNet-2D, UNet-3D, Attention UNet, Swin UNet, and the 3D Swin Transformer. Among these, the most accurate segmentation results were achieved using a reverse knowledge distillation strategy between 3D UNet and 3D Swin UNet. This hybrid method leveraged the spatial learning capabilities of the UNet alongside the global contextual understanding provided by the Swin Transformer, effectively capturing both fine anatomical structures and long-range dependencies within the data.

For classification, hippocampal volume was utilized as a key neuroimaging biomarker, complemented by demographic features such as age and gender. Models including Random Forest, Gradient Boosting, and XGBoost were implemented, with a final Stacking Ensemble model delivering the highest classification accuracy of 93%. This ensemble demonstrated strong performance across all cognitive classes, with high precision, recall, and F1-scores.

The integration of segmentation and classification within a single pipeline highlights the potential of combining state-of-the-art deep learning techniques with traditional machine learning methods to enhance early detection of Alzheimer's Disease. The use of Google Colab TPUs significantly accelerated the training of computationally intensive 3D models, contributing to an efficient and scalable solution. Overall, the pipeline supports a clinically viable approach for reliable early screening and diagnosis of cognitive decline.

CHAPTER 6

Future Scope

In terms of segmentation, there are several areas for potential improvement. Although the reverse knowledge distillation approach between 3D UNet and 3D SWIN Unet achieved the highest accuracy, further experimentation could be conducted with hybrid models combining CNNs and transformers in different configurations, such as multi-scale or multi-resolution approaches, to better capture complex anatomical structures and variations in the hippocampus. Additionally, the SWIN-Unet 3D model, while theoretically promising, faced significant challenges due to its large size and slow training times. Future iterations could involve model compression techniques, such as knowledge distillation or pruning, to improve the feasibility of using such models on limited hardware resources like Google Colab TPUs.

For classification, future work could integrate additional biomarkers, such as genetic factors (e.g., APOE status), functional MRI data, or other regions of interest in the brain (e.g., cortical thickness, amygdala volume), which could provide more nuanced insights into Alzheimer's progression and improve the model's predictive accuracy. Incorporating longitudinal data—which tracks changes over time—could also help shift the focus from static classification to dynamic progression prediction, which is more clinically relevant for monitoring Alzheimer's Disease.

Furthermore, improving model interpretability will be crucial for gaining trust in clinical applications. While tree-based models like Random Forests and XGBoost are somewhat interpretable, advanced interpretability methods such as SHAP or LIME could be employed to explain the decisions made by the deep learning models, ensuring that clinicians can understand the reasoning behind predictions, especially in critical healthcare settings.

Another avenue for future work would be the development of a real-time, web-based diagnostic tool that leverages both segmentation and classification pipelines. Such a tool could allow healthcare professionals to input MRI scans and clinical data and receive predictions about both hippocampal segmentation and cognitive state in real time.

In conclusion, this project lays a solid foundation for integrating AI with healthcare diagnostics, particularly in the early detection and monitoring of Alzheimer's Disease. The integration of advanced segmentation techniques with predictive models offers a comprehensive solution, and future work could further improve these methods, both in terms of accuracy and clinical applicability. This work contributes significantly to the understanding of how AI can assist in the early diagnosis of cognitive diseases and paves the way for the development of practical, deployable diagnostic tools in the healthcare industry.

References

[1] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs.CV] (2015).

[2] He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs.CV] (2015).

[3] Gunawardena, K.A.N.N.P., Rajapakse, R.N., Kodikara, N.D.: Applying convolutional neural networks for pre-detection of Alzheimer's disease from structural MRI data. In: 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), pp. 1–7. IEEE, Auckland, New Zealand (2017).

[4] Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas. arXiv:1804.03999 [cs.CV] (2018).

[5] Pellegrini, E., Ballerini, L., Hernandez, M.D.C.V., Chappell, F.M., González-Castro, V., Anblagan, D., Danso, S., Muñoz-Maniega, S., Job, D., Pernet, C., Mair, G., MacGillivray, T.J., Trucco, E., Wardlaw, J.M.: Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 10, 519–535 (2018).

[6] Bidani, A., Gouider, M.S., Travieso-González, C.M.: Dementia detection and classification from MRI images using deep neural networks and transfer learning. In: Rojas, I., Joya, G., Catala, A. (eds.) Advances in Computational Intelligence, IWANN 2019, LNCS, vol. 11506, pp. 925–933. Springer, Cham (2019).

[7] Khan, N.M., Abraham, N., Hon, M.: Transfer learning with intelligent training data selection for prediction of Alzheimer's disease. IEEE Access 7, 72726–72735 (2019).

[8] Battineni, G., Chintalapudi, N., Amenta, F.: Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). Informatics in Medicine Unlocked 16, 100200 (2019).

[9] Li, H., Habes, M., Wolk, D.A., Fan, Y.; Alzheimer's Disease Neuroimaging Initiative and the Australian Imaging Biomarkers and Lifestyle Study of Aging: A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal

magnetic resonance imaging data. Alzheimer's & Dementia 15(8), 1059–1070 (2019)

[10] Liu, M., Li, F., Yan, H., Wang, K., Ma, Y., Shen, L., Xu, M.; Alzheimer's Disease Neuroimaging Initiative: A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. NeuroImage 208, 116459 (2020).

[11] Wang, Z., Xin, J., Wang, Z., Gu, H., Zhao, Y., Qian, W.: Computer-aided dementia diagnosis based on hierarchical extreme learning machine. Cognitive Computation 13, 34–48 (2021).

[12] Qiu, S., Joshi, P.S., Miller, M.I., Xue, C., Zhou, X., Karjadi, C., Chang, G.H., Joshi, A.S., Dwyer, B., Zhu, S., Kaku, M., Zhou, Y., Alderazi, Y.J., Swaminathan, A., Kedar, S., Saint-Hilaire, M.-H., Auerbach, S.H., Yuan, J., Sartor, E.A., Au, R., Kolachalama, V.B.: Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain 143(6), 1920–1933 (2020).

[13] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929 [cs.CV] (2020).

[14] Carmo, D., Silva, B., Yasuda, C., Rittner, L., Lotufo, R.: Hippocampus segmentation on epilepsy and Alzheimer's disease studies with multiple convolutional neural networks. Heliyon 7(2), e06226 (2021).

[15] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. arXiv:2103.14030 [cs.CV] (2021).

[16] Antor, B., Jamil, A.H.M.S., Mamtaz, M., Khan, M.M., Aljahdali, S., Kaur, M., Singh, P., Masud, M.: A comparative analysis of machine learning algorithms to predict Alzheimer's disease. Journal of Healthcare Engineering 9917919, 12 pages (2021).

[17] Helaly, H.A., Badawy, M., Haikal, A.Y.: Toward deep MRI segmentation for Alzheimer's disease detection. Neural Computing and Applications 34, 1047–1063 (2022).

[18] Helaly, H.A., Badawy, M., Haikal, A.Y.: Deep learning approach for early detection of Alzheimer's disease. Cognitive Computation 14, 1711–1727 (2022).

[19] El-Geneedy, M., Moustafa, H.E.-D., Khalifa, F., Khater, H., AbdElhalim, E.: An

MRI-based deep learning approach for accurate detection of Alzheimer's disease. Alexandria Engineering Journal 63, 211–221 (2023).

[20] Hu, Z., Wang, Z., Jin, Y., Hou, W.: VGG-TSwinformer: Transformer-based deep learning model for early Alzheimer's disease prediction. Computer Methods and Programs in Biomedicine 229, 107291 (2023).

[21] Balasundaram, A., Srinivasan, S., Prasad, A., Malik, J., Kumar, A.: Hippocampus segmentation-based Alzheimer's disease diagnosis and classification of MRI images. Arabian Journal for Science and Engineering 48(8), 10249–10265 (2023).

[22] Rajab, M.D., Jammeh, E., Taketa, T., Brayne, C., Matthews, F.E., Su, L., Ince, P.G., Wharton, S.B., Wang, D.: Assessment of Alzheimer-related pathologies of dementia using machine learning feature selection. Alz Res Therapy 15, 47 (2023).

[23] Gharaibeh, N., Abu-Ein, A.A., Al-hazaimeh, O.M., Nahar, K.M.O., Abu-Ain, W.A., Al-Nawashi, M.M.: Swin Transformer-Based Segmentation and Multi-Scale Feature Pyramid Fusion Module for Alzheimer's Disease with Machine Learning. International Journal of Online and Biomedical Engineering (iJOE) 19(04), 22–50 (2023).

[24] Arafa, D.A., Moustafa, H.ED., Ali, H.A., Ali-Eldin, A.M.T., Saraya, S.F.: A deep learning framework for early diagnosis of Alzheimer's disease on MRI images. Multimed Tools Appl 83, 3767–3799 (2024).

[25] Bamber, S.S., Vishvakarma, T.: Medical image classification for Alzheimer's using a deep learning approach. J. Eng. Appl. Sci. 70, 54 (2023).

[26] Odusami, M., Maskeliūnas, R., Damaševičius, R., Misra, S.: Explainable Deep-Learning-Based Diagnosis of Alzheimer's Disease Using Multimodal Input Fusion of PET and MRI Images. J. Med. Biol. Eng. 43, 291–302 (2023).

[27] Asgharzadeh-Bonab, A., Kalbkhani, H., Azarfardian, S.: An Alzheimer's disease classification method using fusion of features from brain Magnetic Resonance Image transforms and deep convolutional networks. Healthcare Analytics 4, 100223 (2023).

[28] Al Olaimat, M., Martinez, J., Saeed, F., Bozdag, S., Alzheimer's Disease Neuroimaging Initiative: PPAD: a deep learning architecture to predict progression of Alzheimer's disease. Bioinformatics 39(Supplement_1), i149–i157 (2023).

[29] Xin, J., Wang, A., Guo, R., Liu, W., Tang, X.: CNN and swin-transformer-based efficient model for Alzheimer's disease diagnosis with MRI. Biomed. Signal Process.

Control 86(Part B), 105189 (2023).

[30] Jiang, J., Liu, H., Yu, X., Zhang, J., Xiong, B., Kuang, L.: Hippocampus Segmentation Method Applying Coordinate Attention Mechanism and Dynamic Convolution Network. Appl. Sci. 13(13), 7921 (2023).

[31] Liu, L., Liu, S., Zhang, L., To, X.V., Nasrallah, F., Chandra, S.S.: Cascaded multi-modal mixing transformers for Alzheimer's disease classification with incomplete data. NeuroImage 277, 120267 (2023).

[32] Huang, Y., Li, W.: Resizer Swin Transformer-Based Classification Using sMRI for Alzheimer's Disease. Appl. Sci. 13(16), 9310 (2023).

[33] Lei, Y., Ding, Y., Qiu, R.L.J., Wang, T., Roper, J., Fu, Y., Shu, H.K., Mao, H., Yang, X.: Hippocampus substructure segmentation using morphological vision transformer learning. Phys. Med. Biol. 68(23), 235013 (2023).

[34] Fathi, S., Ahmadi, A., Dehnad, A., Almasi-Dooghaee, M., Sadegh, M.: A Deep Learning-Based Ensemble Method for Early Diagnosis of Alzheimer's Disease using MRI Images. Neuroinform 22, 89–105 (2024).

[35] Gan, D., Chang, M., Chen, J.: 3D-EffiViTCaps: 3D efficient vision transformer with capsule for medical image segmentation. arXiv (2024).

[36] Xiao, Z., Zhang, Y., Deng, Z., Liu, F.: Light3DHS: A lightweight 3D hippocampus segmentation method using multiscale convolution attention and vision transformer. NeuroImage 292, 120608 (2024).

[37] Alp, S., Akan, T., Bhuiyan, M.S., Disbrow, E.A., Conrad, S.A., Vanchiere, J.A., Kevil, C.G., Bhuiyan, M.A.N.: Joint transformer architecture in brain 3D MRI classification: its application in Alzheimer's disease classification. Sci Rep 14, 8996 (2024).

[38] Suchitra, S., Krishnasamy, L., Poovaraghan, R.J.: A deep learning-based early Alzheimer's disease detection using magnetic resonance images. Multimed Tools Appl (2024).

[39] Shah, S.M.A.H., Khan, M.Q., Rizwan, A., Jan, S.U., Samee, N.A., Jamjoom, M.M.: Computer-aided diagnosis of Alzheimer's disease and neurocognitive disorders with multimodal Bi-Vision Transformer (BiViT). Pattern Anal Applic 27, 76 (2024).