

Comparative Analysis of Convolutional Neural Network and U-Net Segmentation Model for Alzheimer's Disease Detection

Jayashree Shetty, Manjula K Shenoy*, Sucheta V Kolekar,
M Mukhyaprabha Prabhu, Siddh Bhardwaj

Abstract—Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that demands the deployment of precise, automated, and reproducible diagnostic methodologies for effective clinical management and therapeutic decision-making. This study proposes and comparatively evaluates two computational frameworks for the detection of AD utilizing structural Magnetic Resonance Imaging (MRI) data. The first framework integrates overlay-based image segmentation techniques, employing intensity thresholding and pixel-wise differentiation, followed by classification of the extracted regions using a Convolutional Neural Network (CNN) architecture. The second framework incorporates a U-Net-based semantic segmentation model, coupled with ensemble classification schemes comprising Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbors (KNN) algorithms. A comprehensive quantitative analysis is performed to assess segmentation accuracy metrics and classification performance indices, including precision, recall, F1-score, and overall classification accuracy. Additionally, the influence of optimization algorithms—specifically Adam and RMSProp—on the convergence behavior and classification efficacy of the CNN model is systematically investigated. The proposed methodologies demonstrate classification accuracies within the range of 70% to 89%, providing comparative insights into the efficacy of conventional and deep learning-based segmentation-classification pipelines. The findings contribute to the advancement of neuroimaging-based diagnostic systems and offer critical guidance for researchers and clinicians in the selection of optimal computational approaches for medical image analysis applications.

Index Terms—Alzheimer's disease, overlaying analysis, thresholding, Brain MRI processing, Image Segmentation, U-Net, CNN, Ensemble Classification.

I. INTRODUCTION

Alzheimer's Disease (AD) is a progressive, debilitating neurodegenerative disorder that poses significant challenges in its diagnosis, classification, and management [1]. As reported by the World Health Organisation [2],

Manuscript received March 10, 2025; revised July 21, 2025.

Jayashree Shetty is an Assistant Professor of Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India (e-mail: jayashree.sshetty@manipal.edu).

Manjula K. Shenoy is a Professor of Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India (Corresponding author, e-mail: manju.shenoy@manipal.edu).

Sucheta V Kolekar is an Associate Professor of Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India (e-mail: sucheta.kolekar@manipal.edu).

M Mukhyaprabha Prabhu is a Professor of Department of Medicine, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, Karnataka, India (e-mail: mm.prabhu@manipal.edu).

Siddh Bhardwaj is a Cloud Engineer in Hewlett Packard Enterprise, Bengaluru, Karnataka, India (e-mail: siddhbhardwaj2002@gmail.com).

approximately 57 million people are living with dementia worldwide, over 60% of whom reside in low- and middle-income countries. Each year, nearly 10 million new cases of AD are diagnosed, accounting for 60–70% of dementia cases. Dementia is currently the seventh leading cause of death globally and one of the primary contributors to disability and dependency among the elderly. In 2019, dementia-related care cost economies approximately US \$1.3 trillion, with nearly 50% of this burden shouldered by informal caregivers. As the global population ages, the prevalence of AD continues to rise, making early and accurate detection increasingly vital for effective treatment planning and patient care.

However, traditional diagnostic approaches such as cognitive assessments and neurological examinations remain largely subjective, often resulting in inconsistent outcomes, particularly in the early stages of the disease. Clinical misdiagnosis rates for AD remain alarmingly high, reaching up to 30% in early stages due to the subtlety of initial symptoms and reliance on subjective clinical evaluations.

In contrast, advanced imaging techniques like Magnetic Resonance Imaging (MRI) offer objective, quantifiable data to support early detection and precise classification of AD. In recent years, developments in medical imaging, particularly MRI, have provided high-resolution insights into structural brain changes, offering new avenues for enhancing diagnostic accuracy. MRI biomarkers such as hippocampal atrophy and cortical thinning are among the earliest indicators of AD and are essential in differentiating between healthy and diseased brain tissues.

To harness this imaging data effectively, image segmentation and classification algorithms, especially deep learning-based methods, have become indispensable tools for extracting diagnostic features from neuroimaging data [3] [4]. Convolutional Neural Networks (CNNs) have demonstrated outstanding performance in various medical image analysis tasks, including AD classification. Additionally, ensemble classification approaches that combine predictions from multiple classifiers have shown improved robustness, reliability, and accuracy by mitigating individual model biases. This study introduces a comparative analysis of segmentation and classification methodologies for AD detection using MRI data sourced from a publicly available dataset. Specifically, it evaluates:

- A CNN-based classification framework incorporating overlay segmentation with thresholding.
- A U-Net-based segmentation pipeline combined with ensemble classifiers such as Random Forest (RF), De-

cision Tree (DT), and K-Nearest Neighbors (KNN).

The primary objectives of this research are:

- To evaluate the performance of CNN-based classification models with overlay segmentation in classifying publicly available AD-related MRI scans.
- To implement and assess a U-Net-based segmentation model integrated with ensemble classifiers (RF, DT, KNN) for improved AD classification accuracy.
- To conduct a comparative analysis of both segmentation-classification pipelines in terms of precision, sensitivity, specificity, F1-score, and overall reliability.
- To identify the strengths and limitations of each approach for potential application in clinical AD diagnosis workflows.
- To demonstrate the potential of ensemble classification strategies in enhancing the robustness and generalizability of AD detection models using open-source MRI data.

The findings of this comparative study have critical implications for refining diagnostic workflows and facilitating earlier, more accurate detection of AD, ultimately contributing to improved patient outcomes and better resource management [5] [6].

II. LITERATURE SURVEY

Hard and soft voting algorithms were used by Shah et al. [7] to categorize and pinpoint the early AD period. There are 437 patients in the data collection, ranging in age from 60 to 96. These are divided into 64 demented and 72 non-demented individuals. Of these, 70% are utilized to train the algorithm and 30% to test it. Decision trees, soft voting classifiers, and hard voting classifiers are classification algorithms. The vote classifier algorithm achieves an accuracy of 84%. Lazli et al. [8] created a technique that combines SVM-based classification with MRI data for tissue segmentation. To improve segmentation accuracy, their method uses fuzzy-possibilistic segmentation, which combines possibilistic clustering and fuzzy c-means clustering. A dataset of 60 MRI scans, comprising 30 of those with AD and 30 healthy controls, was used to test the approach. With a 93.33% accuracy rate, 96.67% sensitivity, 90% specificity, and an AUC of 0.983, the suggested procedure notably exceeded earlier approaches. In the study's conclusion, a CAD system for AD diagnosis that uses fuzzy-possibilistic segmentation and SVM classification is shown. The encouraging outcomes demonstrate its potential as a useful instrument for AD early diagnosis and detection. The use of unsupervised feature learning, which involves two processes, was emphasized by Razavi et al. [9]. The initial step is to take the raw data and extract its features. Uncontrolled neural layer networks and dispersed filtering are the techniques employed. Softmax is a technique for classifying healthy and ill people that combines regression with sparse filtering. To disseminate the gathered data, a few unsupervised learning methods are employed, including Boltzmann machines and scattered coding. ADNI with cerebrospinal fluids served as the data set for this methodology. There are 51 AD patients in total, and another 43 patients have only moderate AD symptoms. The 1.5T scanners were used to collect the MRI data. 98.3% accuracy is the best accuracy recorded while utilizing the softmax regression.

This early study by Su et al. [10] discovered that the gradiometer-based markers typically outperform their magnetometer-based counterparts. It's interesting to note that, out of the 10 regions of interest, the left frontal lobe performs roughly 8% better than the second-best region (left temporal lobe) for AD/MCI/HC classification in terms of mean recognition rate. Holilah et al. [3] presented a methodology for AD detection that involves the analysis of MRI brain images by employing watershed segmentation & K-means clustering techniques. The proposed approach follows three main stages. Initially, K-means clustering is applied to segment brain tissue from MRI scans. Next, the watershed segmentation technique is utilized to further divide the segmented tissue into distinct regions. Lastly, these segmented regions are analyzed to detect potential indicators of AD. The effectiveness of this method was assessed using an MRI dataset consisting of images from both AD patients and healthy individuals. The results demonstrated a high accuracy of 95.5% in identifying AD, outperforming existing state-of-the-art techniques. Pan D et al. [11] developed a 2D CNN model that processes inputs along sagittal, coronal, and transverse axes. The final classification is achieved through ensemble learning, incorporating a voting mechanism. The model identifies early signs of the disease by analyzing intersection points within the temporal lobe and other regions of the limbic system. However, its performance in early-stage detection was relatively lower. Similarly, Gupta Y et al. [12] introduced a classification framework to differentiate AD into three categories: AD, aAD (stable MCI), and mAD (MCI progressing to AD within 36 months). Their method utilized structural MRI scans to extract key features such as voxel-based morphometry (VBM), hippocampal volume (HV), and segmented cortical and subcortical regions.

Helaly, H. A et al. [13] conducted a comparative study on two different methods for classifying medical images and detecting AD. The first method employed a CNN for classification between two stages of AD, while the second method utilized a transfer learning approach. Beltrán et al. [14] proposed the use of Classification and Regression Trees (CART) for the classification of MRI images, and to address its limitations, a Random Forest classifier was also employed. However, this approach did not consider many imaging biomarkers. Liu et al. [15] conducted a study utilizing hippocampus MRI data, proposing an early detection model using deep learning techniques. The study utilized MRI scans to extract relevant features and employed a CNN to predict future clinical outcomes. By focusing on specific brain regions, the research highlighted the potential of deep learning for the early detection of AD.

To mitigate the challenge of limited data in machine learning applications, M. Orouskhani et al. [16] introduced a conditional deep Triplet network. Their approach integrated both the best and worst triplets into the conditional triplet loss function, facilitating AD diagnosis through a four-class classification framework based on brain MRI scans. Similarly, J. B. Bae et al. [17] developed a CNN-based model for AD classification using MRI scans of patients and age- and gender-matched healthy controls. The model was trained on coronal slices of T1-weighted images, particularly those encompassing the medial temporal lobe, and its performance was assessed on validation sets from both the same and

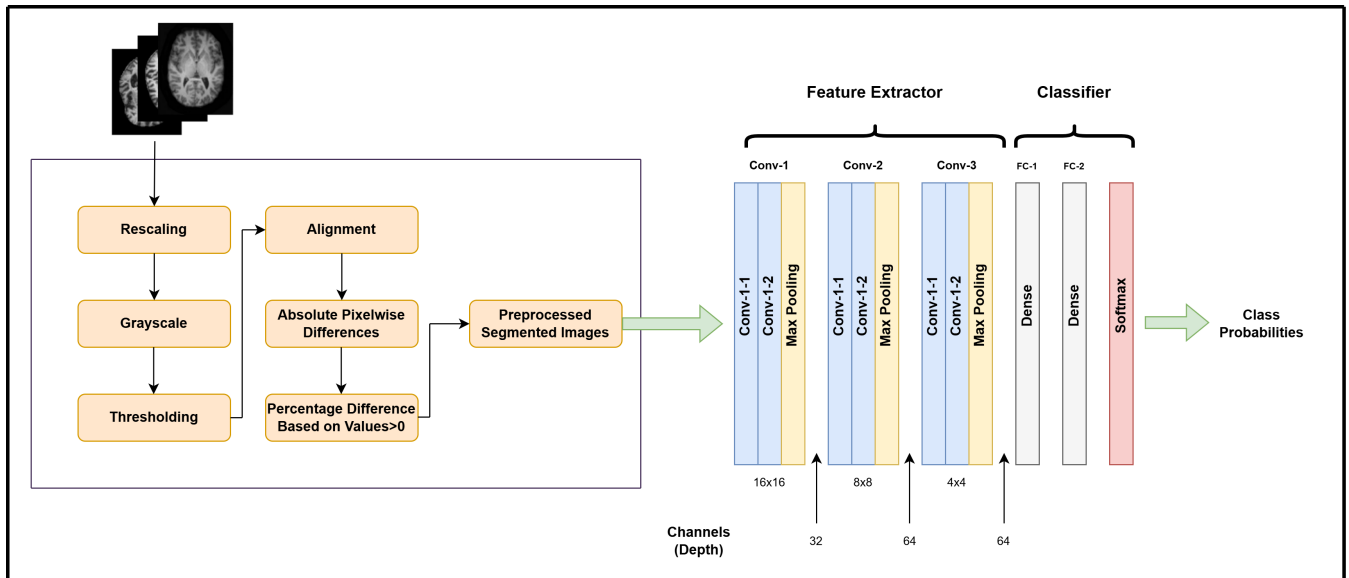


Fig. 1: Workflow of the proposed CNN-based classification model for AD detection, illustrating preprocessing, segmentation (overlying analysis), feature extraction, and classification modules

different populations.

MRI analysis is widely used in clinical research for AD diagnosis; however, distinguishing between AD and healthy aging MRI scans remains a challenge due to their similarities. Islam et al. [18] proposed a network designed to aid in the early detection of AD. Additionally, J. H. Cai et al. [19] provided an overview of texture analysis techniques, discussing their applications in MR imaging for AD diagnosis and summarizing the key factors contributing to AD-related cognitive decline.

III. MATERIALS AND METHODS

This section outlines the dataset employed and the methodology adopted for implementation and comparative evaluation of the segmentation-classification approaches explored in this study.

A. Dataset

The dataset utilized in this study is a publicly available secondary dataset comprising 6,400 T1-weighted MRI images of human brains. These images are systematically categorized into four clinically significant classes: *Mild Demented*, *Moderate Demented*, *Non-Demented*, and *Very Mild Demented*. Each category corresponds to a specific stage of dementia severity, ensuring a heterogeneous and comprehensive representation of the AD progression spectrum.

The availability of labelled images from each class facilitates the development and evaluation of machine learning models by providing a reliable ground truth for training and validation. For the purpose of this study, a carefully curated subset of 5,121 consistent images were selected from the original dataset. This selection process was aimed at maintaining uniformity across all four classes while eliminating images with anomalies or inconsistent labelling. The balanced and representative nature of the chosen images contributes to the reliability and robustness of the proposed segmentation and classification methodologies.

B. Image Segmentation Models

To facilitate a robust and comparative analysis, two distinct image segmentation techniques are implemented:

- Overlaying Analysis and Thresholding
- U-Net Segmentation

In the first method, a fixed thresholding value of 109, which is determined empirically, is applied to distinguish the region of interest within each image. This threshold value effectively captures minute variations and subtle differences in brain structure. In contrast, the second method employs U-Net segmentation—a deep learning-based technique that utilizes an encoder-decoder architecture with skip connections to enable precise delineation of relevant anatomical regions [20].

C. Overlaying Analysis and Thresholding

Prior to segmentation, a comprehensive preprocessing pipeline is applied to the MRI dataset to enhance image consistency, suppress noise artifacts, and improve segmentation accuracy. Given the inherently high dimensionality and varying signal intensities present in MRI acquisitions, preprocessing is a critical step to standardize the data and optimize its suitability for subsequent computational analysis. The overall workflow of the proposed CNN-based classification framework is depicted in Fig. 1. To ensure homogeneity in spatial resolution and maintain precise alignment during overlay operations, all MRI images are uniformly rescaled to a fixed resolution of 208×176 pixels, following established conventions in neuroimaging-based machine learning studies [21]. This rescaling operation minimizes geometric distortions while preserving critical anatomical structures across the image set. Additionally, images are converted from multi-channel to single-channel grayscale intensity images, thereby reducing computational complexity, memory overhead, and data dimensionality without compromising salient structural information required for AD detection [22].

A fixed-level thresholding technique is then applied to the grayscale images as an essential preprocessing step for

enhancing contrast between brain tissues and background regions. Thresholding serves to binarize the image by partitioning pixel intensities based on empirically selected threshold values. Specifically, a threshold value of 109 is applied for Alzheimer's-affected images and 75 for healthy brain scans, determined via iterative analysis of pixel intensity histograms and visual inspection of segmentation quality [23]. This operation facilitates reliable foreground-background separation, making disease-affected regions more discernible by suppressing non-informative background areas.

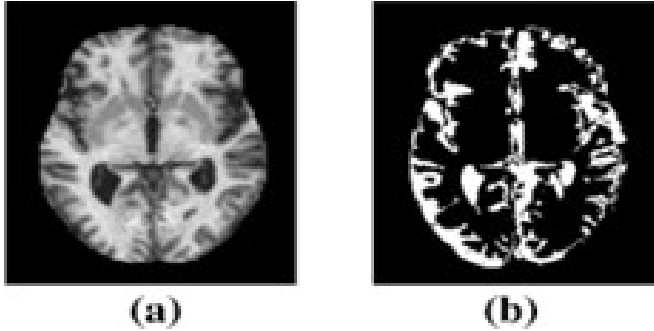


Fig. 2: (a) Original Axial T1-weighted MRI scan of an Alzheimer's-affected brain (b) Corresponding binary image after thresholding and overlay-based difference analysis

Following thresholding, an overlaying difference analysis is performed to identify morphologically significant alterations between Alzheimer's-affected and healthy brain scans [24], [25]. This is achieved using the absdiff function, which computes the per-pixel absolute difference between paired thresholded images. The resultant difference images effectively isolate structurally altered regions while suppressing minor variations and imaging noise. As the images have already been binarized, the difference maps yield a binary mask where morphologically abnormal regions are highlighted in white (pixel value = 255), as illustrated in Fig. 2.

The processed images, along with their corresponding difference masks, are subsequently organized into four distinct directories representing different clinical stages of AD for downstream classification. This preprocessing pipeline, comprising intensity normalization, spatial rescaling, fixed-threshold binarization, and overlay-based difference analysis, ensures enhanced segmentation quality, improved feature discriminability, and increased interpretability of the morphological patterns associated with AD pathology. It thus establishes a technically rigorous foundation for subsequent feature extraction and classification tasks within the proposed CNN-based framework.

D. U-Net architecture for segmentation

For the implementation of the U-Net segmentation model, a comprehensive preprocessing protocol is established to ensure that the input data conforms to the architectural and computational requirements of the network. This protocol involves a systematic sequence of operations including image resizing, grayscale conversion, intensity normalization, dimensional expansion, labeling of ground truth masks, and dataset shuffling. Collectively, these steps improve data uniformity, computational efficiency, and model generalization

TABLE I: Labels assigned to each class

Labels	Class
1	Mild Demented
2	Moderate Demented
3	Non-Demented
4	Very Mild Demented

performance. The workflow of the proposed U-Net-based segmentation model, incorporating these preprocessing procedures and model development phases, is depicted in Fig. 3.

Initially, all MRI images and their corresponding target segmentation masks are resized to a consistent spatial resolution. This resizing ensures dimensional alignment between images and masks, which is critical for maintaining accurate pixel-wise correspondence during training. Since the U-Net architecture involves multiple downsampling and upsampling operations through pooling and transposed convolution layers, ensuring consistent dimensions enables correct alignment of feature maps and skip connections throughout the network [26].

Following resizing, each image undergoes intensity normalization, where pixel values are scaled to a continuous range between 0 and 1. This step mitigates variations caused by differences in MRI acquisition settings, protocols, and equipment, thereby enhancing the stability and efficiency of the model during training. Normalization ensures numerical stability, accelerates convergence, and prevents issues such as vanishing or exploding gradients, which could otherwise hinder effective optimization [27].

Subsequently, each grayscale image is reshaped to include an explicit channel dimension, converting the data structure from a two-dimensional array of shape $(height, width)$ to a three-dimensional array of shape $(height, width, 1)$. This dimensional expansion ensures compatibility with the U-Net's convolutional layers, which expect input tensors with a defined channel depth.

In supervised segmentation tasks, it is essential to associate each image with an annotated ground truth mask containing labeled pixel values corresponding to the respective segmentation classes. In this study, the segmentation classes considered are *Mild Demented*, *Moderate Demented*, *Non-Demented*, and *Very Mild Demented*. These categorical class labels are mapped to numeric identifiers to enable efficient loss calculation and performance evaluation during model training. The mapping between numeric labels and class categories is summarized in Table I.

To further improve model generalization and minimize overfitting to specific image sequences or class patterns, the entire dataset is randomly shuffled prior to training. This randomization ensures that the model encounters image samples in varying sequences across epochs, promoting the extraction of more generalized features and reducing any potential learning bias related to sample ordering or class distribution. The incorporation of dataset shuffling enhances the robustness, reliability, and adaptability of the trained model when deployed on unseen test datasets.

E. Defining U-Net Architecture

The U-Net architecture adopted in this study is a fully convolutional, symmetric encoder-decoder network with skip

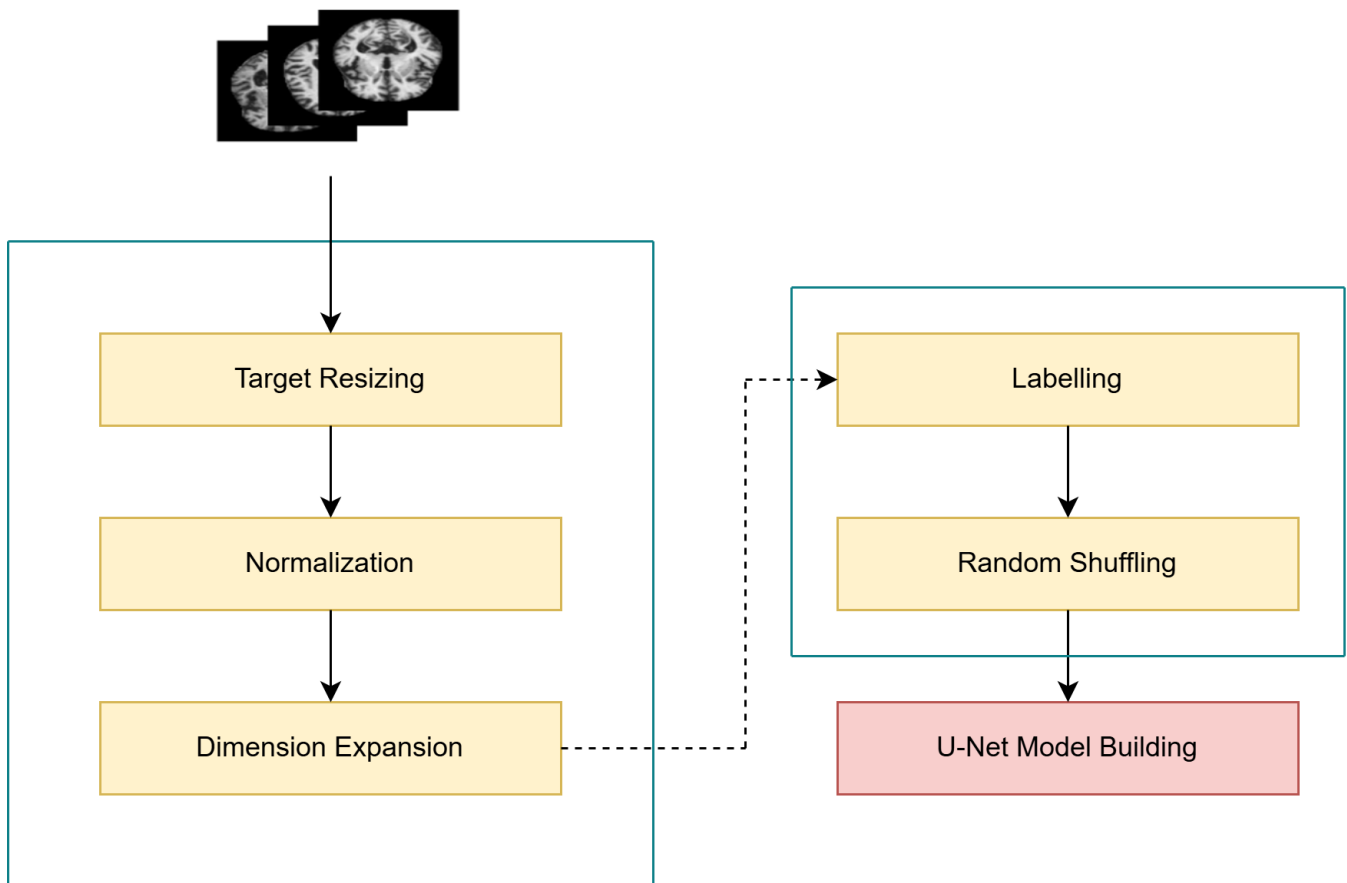


Fig. 3: Workflow of the proposed U-Net-based segmentation model, including preprocessing, labeling, shuffling, and model building steps.

connections, specifically designed for semantic segmentation tasks in biomedical imaging applications. Its configuration facilitates the simultaneous learning of local and global contextual features by combining multi-scale feature maps at corresponding encoder and decoder levels. The complete structural overview of the proposed U-Net model is presented in Fig. 4.

The encoder path of the network progressively reduces the spatial resolution of the input images while extracting hierarchical feature representations through a series of convolutional layers followed by max pooling operations. Each convolutional stage, denoted as conv1 to conv5, utilizes the Rectified Linear Unit (ReLU) activation function to introduce non-linearity and improve feature discrimination. Max pooling layers perform spatial downsampling, effectively reducing the dimensions of the feature maps by a factor of two while preserving essential semantic information necessary for reliable segmentation.

At the center of the U-Net lies the bottleneck layer, which acts as a bridge connecting the encoder and decoder pathways. This section consists of additional convolutional operations designed to capture high-level, abstract features by further compressing the spatial information. The abstracted feature representations generated at this stage are critical for distinguishing subtle structural differences in biomedical images, particularly for identifying pathological brain regions.

Following the bottleneck, the decoder path gradually restores the original spatial resolution of the feature maps through successive upsampling operations. At each stage,

the upsampled feature maps are concatenated with their corresponding encoder outputs via skip connections. This fusion mechanism ensures the preservation of high-resolution spatial details that are typically lost during the downsampling process. The concatenated feature maps then undergo further convolutional operations, labeled conv6 to conv9, which refine the restored spatial information and enable precise reconstruction of anatomical boundaries and abnormal regions.

The final output layer consists of a 1×1 convolutional layer equipped with a sigmoid activation function, generating a binary segmentation mask. Each pixel value in the output mask represents the probability of the corresponding pixel in the input image belonging to the target class. This facilitates accurate, pixel-wise classification and visualization of affected brain regions.

The model is compiled using the Adam optimizer, selected for its adaptive learning rate capabilities, in conjunction with a mean squared error (MSE) loss function. This combination ensures stable convergence and efficient optimization. Model accuracy is used as the primary evaluation metric during training to monitor segmentation performance across epochs. The U-Net model is trained over 10 epochs with a batch size of 16.

All computations, including both training and validation, are conducted on a system equipped with an Intel(R) Core(TM) i5-8265U CPU. The total training duration recorded was 2629 seconds, averaging approximately 21 seconds per training step, while validation computations required a total of 641 seconds, averaging 19 seconds per

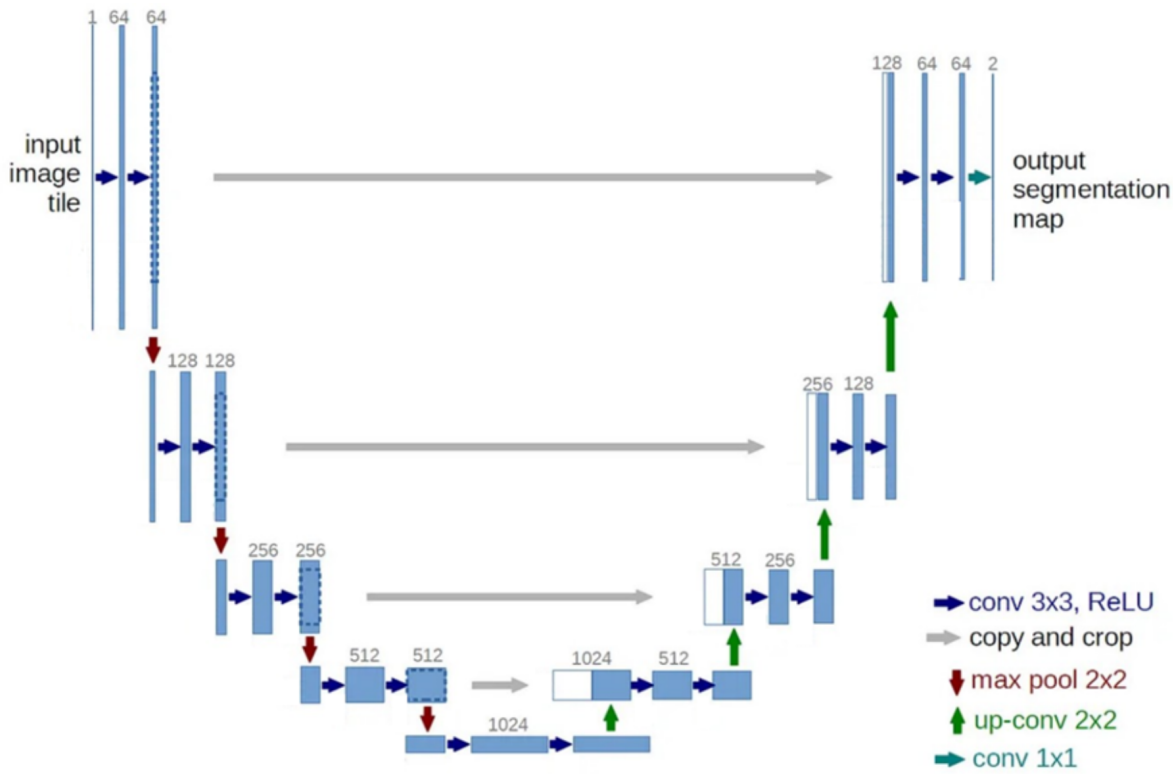


Fig. 4: Proposed U-Net architecture illustrating the symmetric encoder-decoder structure with skip connections, convolutional layers, pooling, upsampling, and final segmentation output

step. A detailed timing analysis of the U-Net training and validation processes is presented in Fig. 5.

F. Classification Models

This study utilizes two classification models: a CNN classifier applied to images segmented through overlay analysis as shown in Fig. 1, and an ensemble classifier applied to images segmented using U-Net segmentation as shown in Fig. 3. The subsequent sections provide a detailed discussion of these approaches.

1) *CNN Classification on Image Dataset Segmented Using Overlaying Analysis*: Following overlay segmentation, a CNN model is trained using two optimizers: Adam and RMSProp. The input images are rescaled and augmented using ImageDataGenerator. The CNN architecture comprises stacked Conv2D and MaxPooling2D layers, followed by dense layers and a softmax output. The model is trained with a batch size of 32. Table II summarizes the architecture parameters.

The performance of the proposed multi-class classification model is evaluated using the Categorical Cross-Entropy (CCE) loss function. CCE is widely adopted in multi-class classification problems where the model outputs a probability distribution over multiple mutually exclusive classes. It measures the dissimilarity between the predicted probability distribution and the actual distribution represented by the one-hot encoded ground truth labels.

The categorical cross-entropy loss for a single observation is defined as:

$$L = - \sum_{i=1}^C y_i \log(\hat{y}_i) \quad (1)$$

TABLE II: CNN layers and parameters

LAYER	PARAMETERS
Conv2D Layer	32 filters, kernel size 3x3, ReLU activation
MaxPooling2D Layer	2x2 pool size
Conv2D Layer	64 filters, kernel size 3x3, ReLU activation
MaxPooling2D Layer	2x2 pool size
Conv2D Layer	128 filters, kernel size 3x3, ReLU activation
MaxPooling2D Layer	2x2 pool size
Flatten Layer	Default
Dense Layer	128 units, ReLU activation
Dense Layer	num_classes units, softmax activation

where C represents the total number of classes, y_i is the true label (1 if the class is correct, otherwise 0), and $\hat{y}_i$ denotes the predicted probability for class i .

For a batch of N training samples, the average categorical cross-entropy loss is computed as:

$$L = - \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^C y_{ij} \log(\hat{y}_{ij}) \quad (2)$$

where y_{ij} indicates the ground truth label for sample j and class i , and $\hat{y}_{ij}$ is the corresponding predicted probability.

This loss function penalizes incorrect predictions by increasing the loss value when the predicted probability for the true class is low. It reaches its minimum when the predicted probability distribution exactly matches the ground truth labels. By minimizing the categorical cross-entropy loss, the model improves its predictive accuracy and classification reliability.

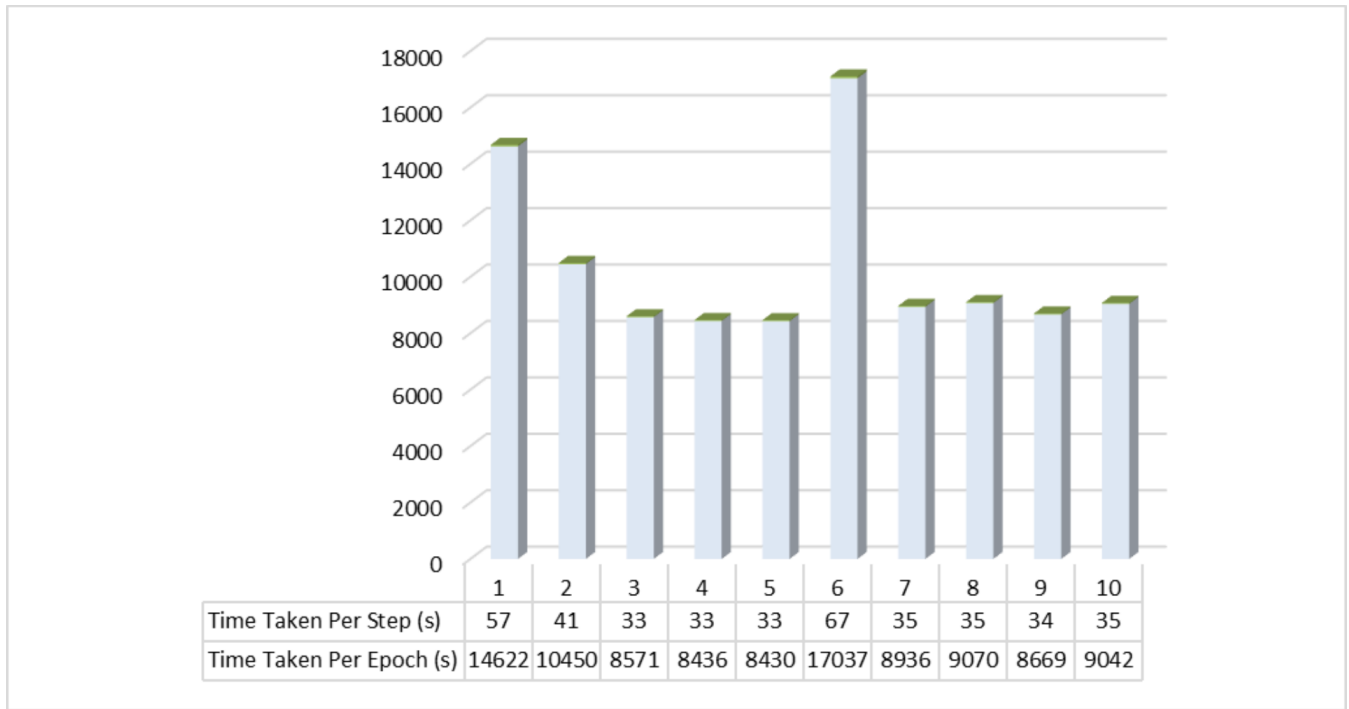


Fig. 5: Timing analysis of U-Net segmentation model: total time taken per epoch and per step during training and validation phases

2) *Ensemble Classification on Images Segmented using U-Net Segmentation*: The U-Net model outputs are processed for ensemble classification. The segmentation outputs are reshaped to fit the input dimensions required by traditional classifiers. Using the `sklearn` library, three classifiers—RF, DT, and KNN are applied.

To improve classification reliability and reduce individual model biases, ensemble classification is performed by combining predictions from these base classifiers through a majority voting mechanism. In this approach, the class label receiving the highest number of votes from the individual classifiers is selected as the final predicted class. The majority voting decision rule is mathematically expressed as:

$$\hat{y} = \arg \max_{c \in C} \sum_{m=1}^M \mathbb{I}(y_m = c) \quad (3)$$

where C denotes the set of possible classes, M represents the total number of base classifiers, y_m is the predicted class label by the m^{th} classifier, and $\mathbb{I}(\cdot)$ is the indicator function that returns 1 if the condition is true and 0 otherwise.

This ensemble classification strategy enhances robustness by leveraging the complementary strengths of RF, DT, and KNN classifiers. By combining their individual predictions, the ensemble model improves overall generalization performance in the context of multi-class AD diagnosis using segmented MRI images.

IV. RESULTS AND DISCUSSION

We utilized an Intel(R) Core(TM) i5-8265U CPU running at 1.80 GHz (boosting to 1.99 GHz) with 16 GB of RAM to conduct all our experiments. The AD classification was implemented using Jupyter Notebook with Python 3 on a 64-bit OS 10 operating system. This section presents a comparative evaluation of two segmentation-classification

TABLE III: Accuracy progression of optimizers across epochs

Epoch	Adam(%)	RMSProp(%)
1	37.81	36.87
4	45.62	40.00
8	65.62	55.62
12	77.50	65.93
16	89.06	80.31
20	92.81	83.12

pipelines: (i) Threshold-based overlay segmentation with CNN classification, and (ii) U-Net segmentation followed by ensemble classification using RF, DT, and KNN. The goal is to assess the performance of each approach in the context of AD classification from MRI scans.

In the first approach, segmentation was achieved using pixel-wise thresholding and overlay analysis, followed by classification through a CNN model. To optimize training and mitigate overfitting, the model was trained for 16 epochs, a value empirically determined to balance convergence and generalization. Two optimization algorithms, RMSProp and Adam, were employed to compare learning efficiency. Adam, which adapts learning rates based on first and second moment estimates, consistently outperformed RMSProp in both convergence rate and final accuracy.

Fig. 6 and Table III illustrate the accuracy progression of both optimizers across epochs. At epoch 16, CNN with the Adam optimiser achieved a classification accuracy of 89.06%, compared to 80.31% with RMSProp. This makes Adam the preferred optimizer for the proposed CNN configuration.

The accepted model over 16 epochs is Adam optimizer in CNN classification of images segmented using overlay analysis with 89.06% accuracy. Furthermore, a detailed evaluation

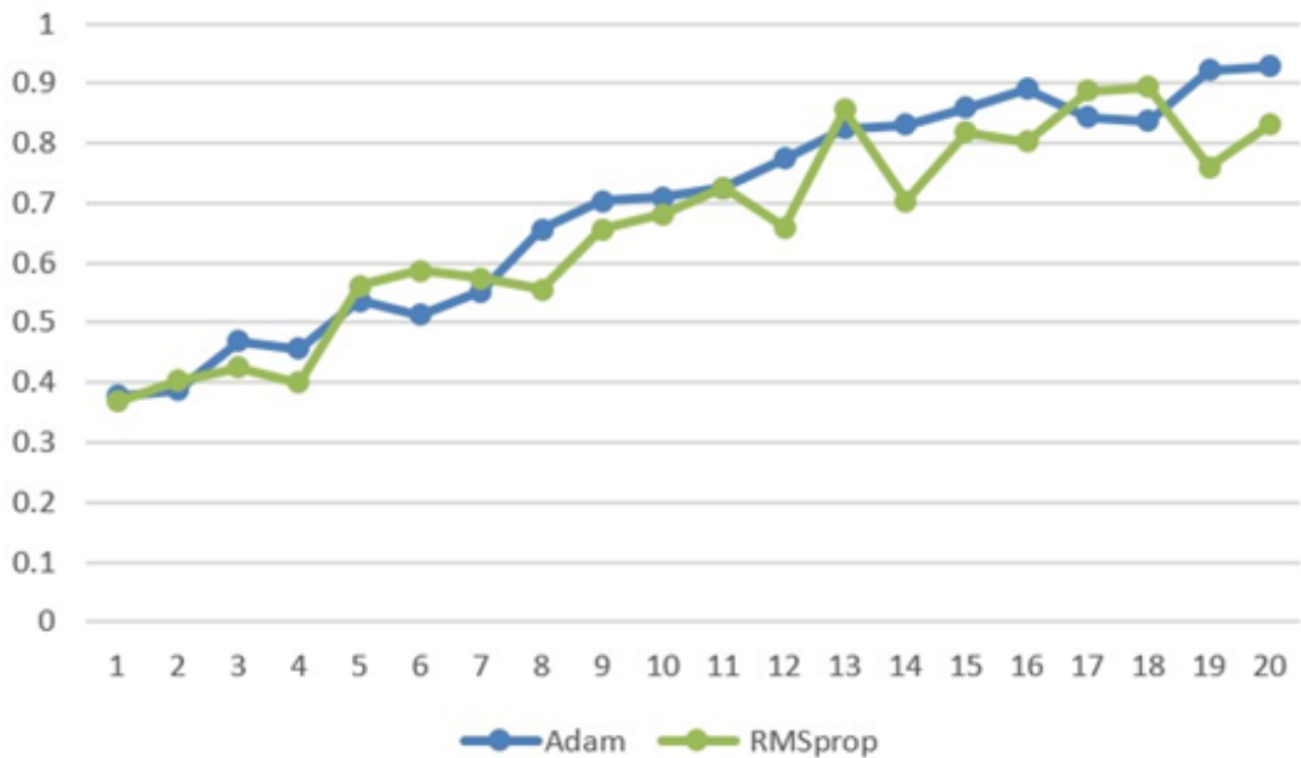


Fig. 6: Graphical depiction of accuracy score of Adam and RMSProp optimizers over 20 Epochs for CNN classification

TABLE IV: Accuracy scores of Random Forest, Decision Tree, and K-Nearest Neighbors along with Ensemble classification

Classifier	Accuracy (%)
Random Forest	51
Decision Tree	67
K-Nearest Neighbours	75
Ensemble Classification	77

of various metrics reinforces this outcome. CNN with overlay segmentation and Adam optimizer achieved an optimal precision of 89.0% with a Recall (Sensitivity) of 89.1%. These results confirm that CNN can effectively extract meaningful spatial and hierarchical features from segmented brain MRI regions, enabling accurate classification across AD severity stages. In contrast, the second approach applied U-Net for image segmentation, followed by classification using an ensemble of traditional machine learning algorithms. The individual classifiers, RF, DT, and KNN achieved accuracy scores of 51%, 67%, and 75% respectively. When combined using majority voting, the ensemble classifier achieved an improved accuracy of 77%. Table IV summarizes these outcomes.

Alongside accuracy, the ensemble classifier achieved a precision of 76.0%, recall of 77.0%, specificity of 85.0%, and an F1-score of 76.5%. The G-Mean for this approach was calculated to be 81.0%, reflecting reasonable performance but clearly lower than that of the CNN-based method. The Precision-Recall curve presented Fig. 7 clearly reflects the performance differences among the evaluated models, and these trends are consistent with the quantitative results sum-

marized in Table V.

The CNN model with Adam optimizer demonstrates the best overall performance, maintaining consistently high precision values across a broad range of recall values in the curve. This is supported by its superior metrics in Table IV, achieving the highest precision (89%), sensitivity (89.1%), specificity (93%), and F1-score (89%). The CNN model with RMSprop optimizer follows closely, with slightly lower precision and recall values in the curve, which correspond well to its reported precision (80%), sensitivity (80.3%), and F1-score (80.2%) in the table.

In contrast, the U-Net based models, especially those combined with Random Forest and Decision Tree classifiers, exhibit comparatively lower precision and recall throughout the curve, indicating weaker performance. This observation is reinforced by the table values, where U-Net + Random Forest and U-Net + Decision Tree report lower precision (50% and 66%, respectively) and F1-scores (50.5% and 66.5%). The U-Net Ensemble Classification model shows intermediate performance, with the curve depicting a gradual decline in precision as recall increases. This trend aligns well with its metrics in Table V, where it achieves a precision of 76%, sensitivity of 77%, and F1-score of 76.5%. Overall, both the graphical and tabular results consistently demonstrate that the CNN model with Adam optimizer outperforms the other models, followed by CNN with RMSprop, while the U-Net based models, despite reasonable performance in ensemble settings, generally lag behind the CNN configurations. The confusion matrices in Fig. 8, 9, 10 illustrate the classification performance of different models on the AD dataset. The relatively stronger performance of the CNN model trained on overlay-segmented images can be attributed to the efficacy of thresholding and pixel-wise difference techniques in isolating

TABLE V: Performance comparison of different models on AD dataset

Model Description	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
CNN (Overlay Segmentation, RMSProp)	80.0	80.3	89.0	80.2
CNN (Overlay Segmentation, Adam)	89.0	89.1	93.0	89.0
U-Net + Random Forest	50.0	51.0	75.0	50.5
U-Net + Decision Tree	66.0	67.0	80.0	66.5
U-Net + KNN	74.0	75.0	84.0	74.5
U-Net + Ensemble Classification	76.0	77.0	85.0	76.5

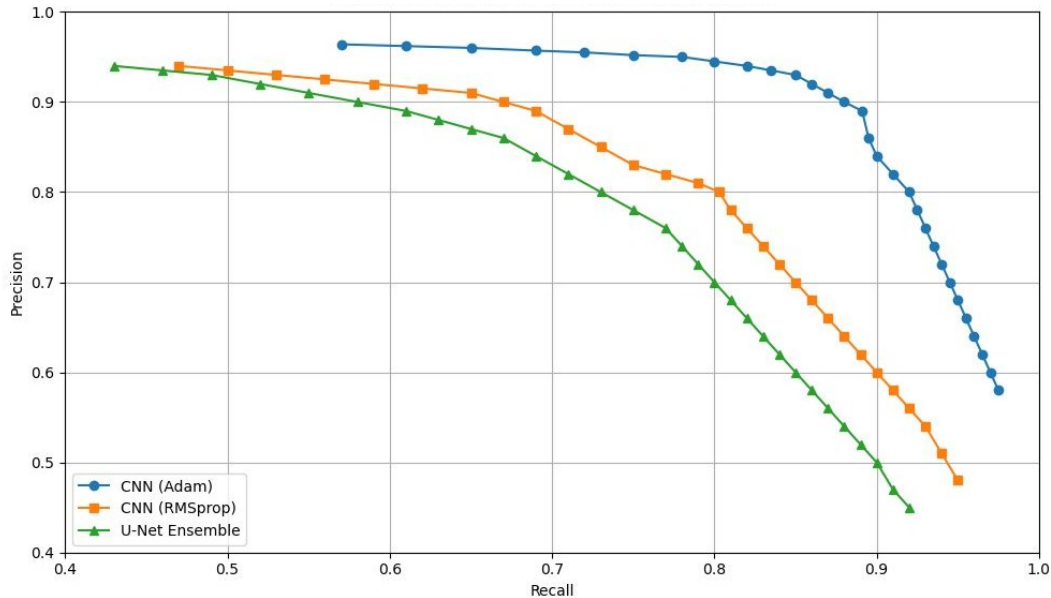


Fig. 7: Precision-Recall curves for CNN (Adam, RMSprop) and U-Net Ensemble on the AD dataset.

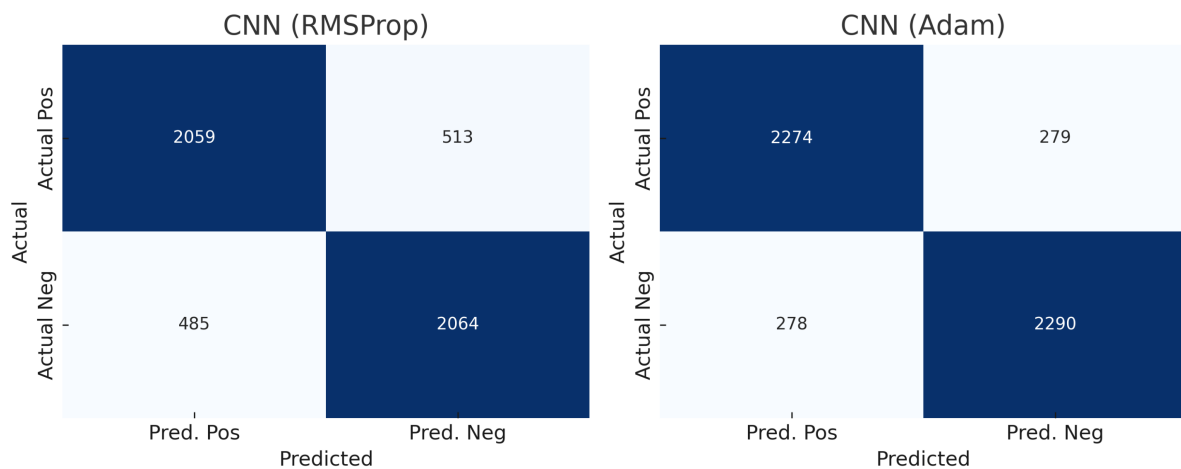


Fig. 8: Confusion Matrices for the proposed CNN models

AD relevant features. By highlighting structural deviations in brain regions, this segmentation method provided the CNN model with high-quality input, allowing the network to effectively extract and interpret both low-level features (such as edges and textures) and higher-level patterns (such as shapes and anatomical structures). The CNN's convolutional layers, pooling mechanisms, and ReLU activations enabled a multi-scale representation that proved particularly useful for distinguishing between the four dementia classes.

Conversely, while U-Net is a well-regarded architecture for biomedical segmentation, its effectiveness in this study may have been limited by segmentation precision as observed in Fig. 11. Any inconsistencies in delineating the regions of

interest could have propagated to the ensemble classifiers, reducing their effectiveness. While ensemble classification benefits from the combination of diverse models, its ultimate accuracy is contingent on the quality of the segmented input data. In conclusion, the CNN-based approach using overlay segmentation visible in Fig. 9 and the Adam optimizer demonstrated the most promising results in terms of classification accuracy and comprehensive performance metrics. Fig. 12 and Fig. 13 demonstrate the performance of all the compared models. The ensemble approach, while beneficial in leveraging model diversity, was hindered by its reliance on the quality of U-Net's segmentation outputs. This highlights the importance of both robust segmentation and effective

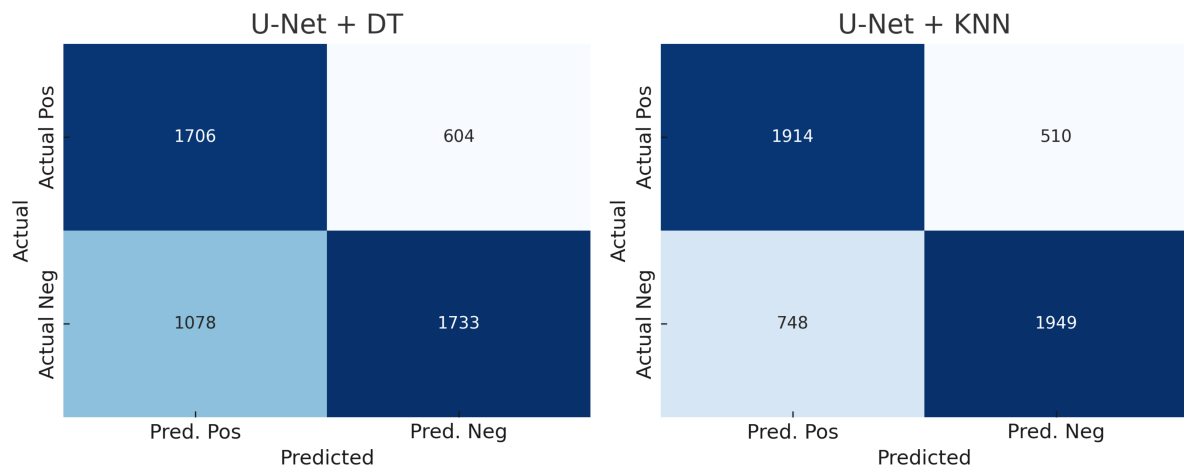


Fig. 9: Confusion Matrices for the proposed U-Net models with DT and KNN

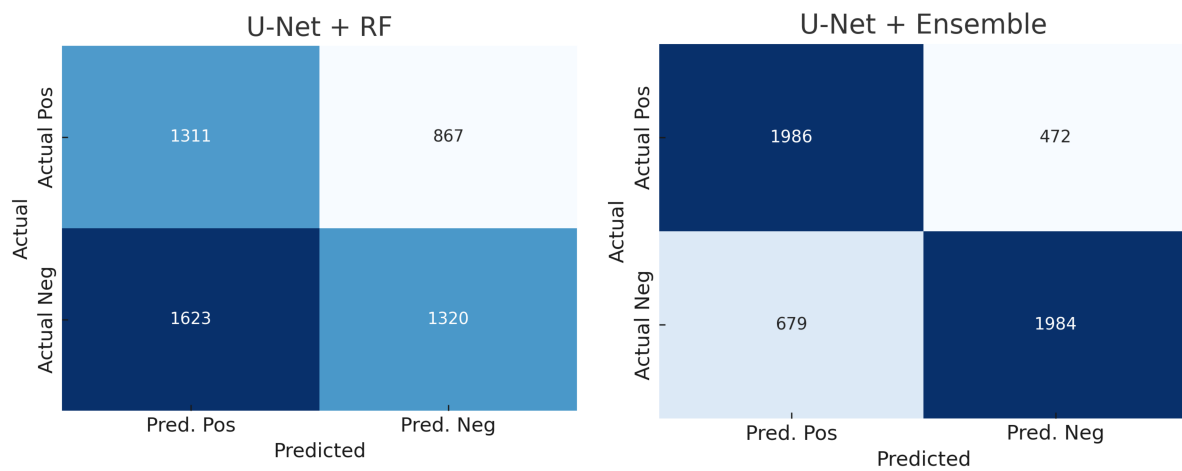


Fig. 10: Confusion Matrices for the proposed U-Net models with RF and Ensemble classification

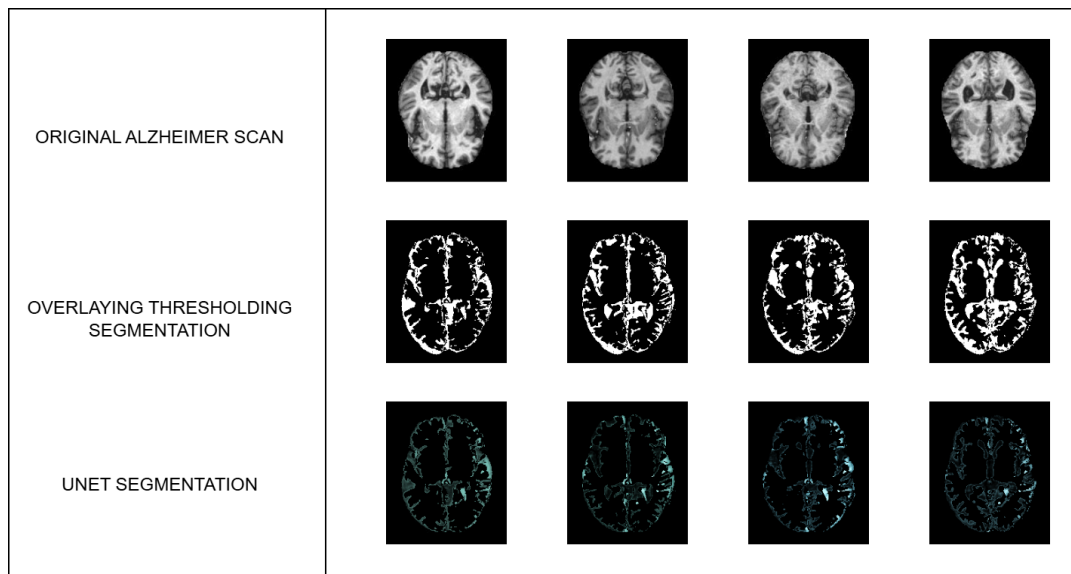


Fig. 11: Image segmentation on AD scan using Overlay Thresholding and UNET Segmentation

classification architectures when developing diagnostic tools for neurodegenerative diseases such as AD.

V. CONCLUSION AND FUTURE WORK

This study presents a comparative framework for AD detection using two segmentation-classification pipelines: overlay-based segmentation integrated with CNN classifica-

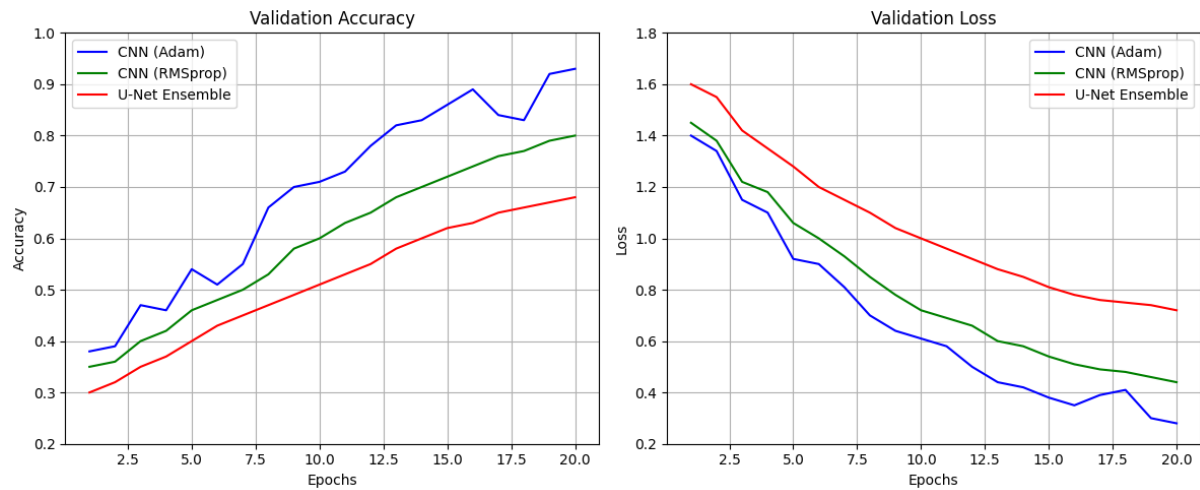


Fig. 12: Validation Training and Validation Loss

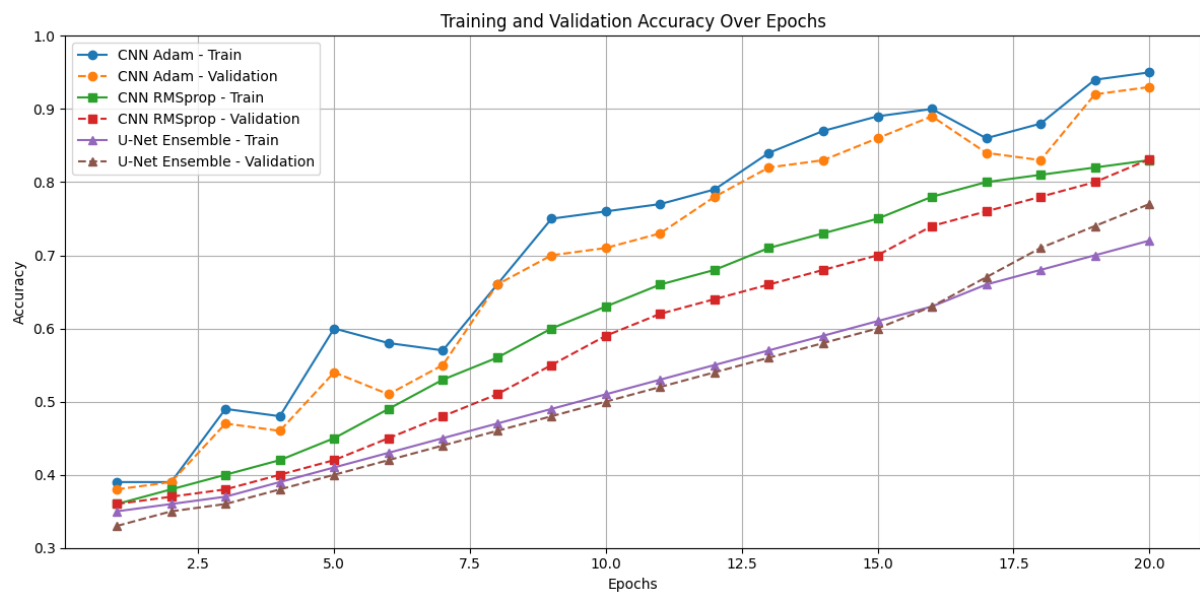


Fig. 13: Validation accuracy vs Training Accuracy

tion, and U-Net segmentation followed by ensemble classification. The results demonstrate that the CNN model applied to overlay-segmented images achieves superior performance, with an accuracy of 89.06% using the Adam optimizer. In contrast, the ensemble classifier comprising RF, DT, and KNN achieved a peak accuracy of 77% when applied to U-Net segmented images.

In addition to accuracy, the evaluation incorporated precision, recall, specificity, sensitivity, F1-score, and G-Mean to offer a holistic view of model performance. The CNN model achieved 89.0% precision, 89.1% recall, 93.0% specificity, 89.0% F1-score, and a G-Mean of 91.0%. These results underscore the model's robustness in correctly identifying both positive and negative cases, which is critical in medical diagnostics where the cost of false negatives can be substantial. The proposed methods offer a scalable and effective approach to brain MRI classification and can serve as a valuable decision-support tool for radiologists. Furthermore, the CNN-based pipeline significantly reduces computational complexity and training time compared to more elaborate segmentation-based classifiers, while maintaining high clas-

sification accuracy across all evaluated metrics. While the current framework focuses on four-class classification based solely on MRI scans, several enhancements can be explored in future work. Incorporating advanced architectures such as vision transformers or attention-based models may improve feature learning and segmentation granularity. The integration of multi-modal imaging data (e.g., PET and MRI) could enrich spatial and functional understanding of disease progression. Expanding the dataset size and applying transfer learning from pre-trained models may improve generalization to diverse patient cohorts. Additionally, the development of a web-based interface for real-time inference and visualization could increase the clinical applicability of the proposed system. These future directions aim to create a more precise, reliable, and deployable tool for early AD detection, contributing meaningfully to computer-aided diagnosis in neurodegenerative disorders.

REFERENCES

- [1] M. W. Oktavian, N. Yudistira, and A. Ridok, "Classification of alzheimer's disease using the convolutional neural network (cnn) with

- transfer learning and weighted loss,” *arXiv preprint arXiv:2207.01584*, 2022.
- [2] W. H. Organization, “Dementia,” <https://www.who.int/news-room/fact-sheets/detail/dementia>, 2021, accessed: 20 May 2024.
- [3] D. Holilah, A. Bustamam, and D. Sarwinda, “Detection of alzheimer’s disease with segmentation approach using k-means clustering and watershed method of mri image,” *Journal of Physics: Conference Series*, vol. 1725, no. 1, p. 012009, 2021.
- [4] C. Taoussi, I. Hafidi, and A. Metrane, “Machine learning and deep learning in health informatics: Advancements, applications, and challenges,” *Engineering Letters*, vol. 33, no. 5, pp. 1448–1461, 2025.
- [5] K. S. Biju, S. S. Alfa, K. Lal, A. Antony, and M. K. Akhil, “Alzheimer’s detection based on segmentation of mri image,” *Procedia Computer Science*, vol. 115, pp. 474–481, 2017.
- [6] H. Mittal, A. C. Pandey, M. Saraswat, S. Kumar, R. Pal, and G. Modwel, “A comprehensive survey of image segmentation: Clustering methods, performance parameters, and benchmark datasets,” *Multimedia Tools and Applications*, vol. 81, no. 24, pp. 35 001–35 026, 2021.
- [7] A. Shah, D. Lalakiya, S. Desai, Shreya, and V. Patel, “Early detection of alzheimer’s disease using various machine learning techniques: A comparative study,” in *Proceedings of the Fourth International Conference on Trends in Electronics and Informatics (ICOEI)*, 2020, pp. 522–526.
- [8] L. Lazli, M. Boukadoum, and O. Ait Mohamed, “Computer-aided diagnosis system for alzheimer’s disease using fuzzy-possibilistic tissue segmentation and svm classification,” in *2018 IEEE Life Sciences Conference (LSC)*. IEEE, 2018, pp. 33–36.
- [9] F. Razavi, M. J. Tarokh, and M. Alborzi, “An intelligent alzheimer’s disease diagnosis method using unsupervised feature learning,” *Journal of Big Data*, vol. 6, no. 1, p. 32, 2019.
- [10] 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, 2021.
- [11] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song, “Early detection of alzheimer’s disease using magnetic resonance imaging: A novel approach combining convolutional neural networks and ensemble learning,” *Frontiers in Neuroscience*, vol. 14, p. 259, 2020.
- [12] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, and G. R. Kwon, “Early diagnosis of alzheimer’s disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of mri t1 brain images,” *PLOS One*, vol. 14, p. e0222446, 2019.
- [13] H. A. Helaly, M. Badawy, and A. Y. Haikal, “Deep learning approach for early detection of alzheimer’s disease,” *Cognitive Computation*, vol. 14, no. 5, pp. 1711–1727, 2021.
- [14] J. F. Beltrán, B. M. Wahba, N. Hose, D. Shasha, and R. P. Kline, “Inexpensive, non-invasive biomarkers predict alzheimer transition using machine learning analysis of the alzheimer’s disease neuroimaging (adni) database,” *PLOS One*, vol. 15, p. e0235663, 2020.
- [15] H. Li, M. Habes, D. A. Wolk, and Y. Fan, “A deep learning model for early prediction of alzheimer’s disease dementia based on hippocampal magnetic resonance imaging data,” *Alzheimer’s & Dementia*, vol. 15, no. 8, pp. 1059–1070, 2019.
- [16] M. Orouskhani, C. Zhu, S. Rostamian, F. S. Zadeh, M. Shafiei, and Y. Orouskhani, “Alzheimer’s disease detection from structural mri using conditional deep triplet network,” *Neuroscience Informatics*, vol. 2, no. 4, p. 100066, 2022.
- [17] J. B. Bae, S. Lee, W. Jung, S. Park, W. Kim, H. Oh, J. W. Han, G. E. Kim, J. S. Kim, J. H. Kim, and K. W. Kim, “Identification of alzheimer’s disease using a convolutional neural network model based on t1-weighted magnetic resonance imaging,” *Scientific Reports*, vol. 10, no. 1, p. 22252, 2020.
- [18] J. Islam and Y. Zhang, “Brain mri analysis for alzheimer’s disease diagnosis using an ensemble system of deep convolutional neural networks,” *Brain Informatics*, vol. 5, no. 2, 2018.
- [19] J.-H. Cai, Y. He, X.-L. Zhong, H. Lei, F. Wang, G.-H. Luo, H. Zhao, and J.-C. Liu, “Magnetic resonance texture analysis in alzheimer’s disease,” *Academic Radiology*, vol. 27, no. 12, pp. 1774–1783, 2020.
- [20] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, “Medical image segmentation based on u-net: A review,” *Journal of Imaging Science & Technology*, vol. 64, no. 2, 2020.
- [21] L.-C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, “Attention to scale: Scale-aware semantic image segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3640–3649.
- [22] C. Kanan and G. W. Cottrell, “Color-to-grayscale: Does the method matter in image recognition?” *PLOS One*, vol. 7, no. 1, p. e29740, 2012.
- [23] K. Lopez-de Ipiña, J. B. Alonso, N. Barroso, M. Faundez-Zanuy, M. Ecay, J. Solé-Casals, C. M. Travieso, A. Estanga, and A. Ezeiza, “New approaches for alzheimer’s disease diagnosis based on automatic spontaneous speech analysis and emotional temperature,” in *International Workshop on Ambient Assisted Living*. Springer, 2012, pp. 407–414.
- [24] R. M. Haralick and L. G. Shapiro, “Image segmentation techniques,” *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 1, pp. 100–132, 1985.
- [25] N. Yamanakkanavar, J. Y. Choi, and B. Lee, “Mri segmentation and classification of human brain using deep learning for diagnosis of alzheimer’s disease: A survey,” *Sensors*, vol. 20, no. 11, p. 3243, 2020.
- [26] P. Benedetti, M. Femminella, and G. Reali, “Mixed-sized biomedical image segmentation based on u-net architectures,” *Applied Science*, vol. 13, no. 1, p. 329, 2022.
- [27] X.-Y. Zhou and G.-Z. Yang, “Normalization in training u-net for 2-d biomedical semantic segmentation,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1792–1799, 2019.