CoMRIAD: A Novel Deep Learning-based Neuroimage Analysis Pipeline for Improved Alzheimer's Disease Detection by Combining Magnetic Resonance Image Planes

Anonymous Author(s)
Affiliation
Address
email

Abstract

Three-dimensional magnetic resonance images (MRI) have emerged as a valuable tool to diagnose and characterise Alzheimer's Disease (AD). Most current MRI analysis pipelines for AD detection focus on a single plane, limiting their ability to capture subtle changes associated with different stages of the disease. This paper proposes a novel deep learning-based pipeline called CoMRIAD that combines the three MRI planes (coronal, axial and sagittal, and referred to as combiplane) for enhanced AD detection and classification. Transfer learning architectures like InceptionV3, InceptionResNetV2, Xception, DenseNet121, and CNN were separately trained and tested on individual planes as well as the combiplane. Experimental results demonstrate that CoMRIAD outperforms single-plane MRI analysis, achieving a 6-8% increase in overall accuracy for two-way and four-way classification tasks. The heatmaps generated using GradCAM and Pearson's correlation coefficient computed between the original MRI and heatmap show high affinity to the predicted class. The CoMRIAD enhances AD detection from 3D MRI, facilitating the monitoring of the disease and relevant interventions. The source code CoMRIAD implementation can be found at: `https://github.com/brai-acslab/comriad`.

1 Introduction

Alzheimer's Disease (AD) is a progressive neurological disorder that affects memory, thinking, and behaviour. It has significant impacts on patients' personal lives, social interactions, and the economy [23]. AD is ranked as the 7th leading cause of death worldwide [6]. Early detection of mild cognitive impairment (MCI), a precursor to AD, is crucial for implementing interventions to slow down or prevent its progression [31]. Structural Magnetic Resonance Imaging (sMRI) is commonly used for clinical diagnosis of AD, serving as a marker for disease progression [28]. However, the current identification process relies on manual assessment by specialists, which is time-consuming and expensive [24].

Figure 1 shows the utilisation of Deep Learning (DL) methodologies and algorithms in the automated classification of diseases, using Positron Emission Tomography (PET), Functional Magnetic Resonance Imaging (fMRI), and MRI as imaging modalities. These algorithms can reveal subtle patterns that can aid in identifying individuals at risk of diseases before clinical symptoms manifest [15]. Convolutional Neural Networks (CNN) applied to single-plane MRI scans can extract features like cortical thickness, hippocampal volume, and ventricle size. Recent advancements include utilising Siamese architecture and VGG16 model as a feature extractor using the triplet-loss function for

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

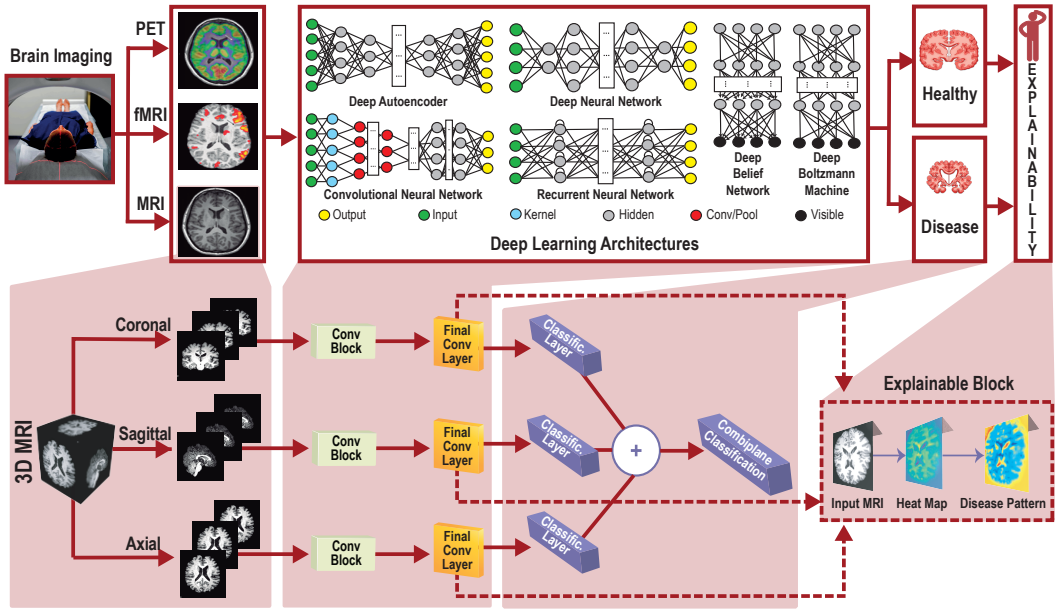

Figure 1: Alzheimer's disease classification pipeline using different deep learning architectures. The shadowed blocks map the proposed pipeline of combiplane MRI analysis for AD classification.

a 4-way classification of AD [8]. Hazarika et al. introduced a modified DenseNet architecture that outperformed other DL models in terms of speed and accuracy for AD classification using sagittal plane MRI datasets [9]. Additionally, a DL framework integrating multimodal data and an explainable model was employed for a 4-way classification of AD, establishing a mapping between computational predictions and pathological indicators of neurodegeneration [17]. These studies collectively demonstrate the potential of DL in the early detection and classification of AD using MRI scans.

In automated AD classification, *combiplane* (term refers to the utilisation of multiplane MRI scans) analysis plays a crucial role by incorporating coronal, axial, and sagittal MRI planes. It enables a comprehensive evaluation of structural and pathological changes, enhancing accuracy and sensitivity in identifying disease-related areas and biomarkers.

Many works have been attempted towards this end. Ryosuke et al. [19] compared the prediction accuracy of individual and combiplanes using pre-trained CNN architectures (AlexNet, VGG16, VGG19, ResNet50, and DenseNet121) on PET images. ResNet50 demonstrated the best performance for a 3-way classification (CN vs MCI vs AD).

Cucun et al. [1] developed a multi-plane CNN for analysing MRI images. They compared the prediction accuracy of single-plane CNN models with multi-plane CNN models and results showed that multi-plane CNN models outperformed single-plane models. Authors further extended the use of multi-plane MRI images using machine learning models which demonstrated combiplane approach significantly enhances classification accuracy. Fei et al. [13] proposed a multi-plane and multi-scale feature fusion network model for AD prediction. The model comprises a feature encoder, attention layers to evaluate feature impact scores, and a feature similarity discriminator for enhancing discrimination of atrophy features by identifying minimally similar features. The study demonstrated improved interpretability, enhanced accuracy, and performance when tested with MRI data from the ADNI dataset.

Jinseong et al. [11] proposed an approach for AD classification using a vision transformer(ViT). The authors demonstrate that the ViT can effectively capture attention relationships among multi-plane and multi-slice images, alongside CNN. The study compares the performance of the proposed model with traditional 3D CNNs. The hybrid model, which integrates all models including a ViT, achieves the best results. While deep learning techniques are known for their robustness in capturing subtle feature differences, Bansal et al.[2] argue that their effectiveness can be hindered by the scarcity of available

data and the domain-agnostic nature inherent to these methods. Their study introduces, Deep3DScan, an ensemble framework for lung cancer analysis. It uses 3D segmentation, deep features, handcrafted descriptors, and achieved segmentation accuracy of 0.927 (outperforming template matching) and detection accuracy of 0.883 (beating the previous state-of-the-art at 0.866) on the LUNA16 dataset.

Despite promising results in the literature, the combiplane approach for AD classification is not widely explored. Furthermore, many current DL studies lack transparency, which hampers interpretability and explainability [29]. This lack of transparency makes it difficult to trust and validate the decisions made by these models, limiting their adoption and understanding of their reasoning. Consequently, it can result in biased or unfair outcomes.

To address the aforementioned limitations, we propose a novel DL model that utilises combiplane sMRI images from coronal, axial, and sagittal planes. Our model aggregates classification results from each plane, effectively leveraging the complementary information across multiple planes to enhance accuracy and assist in early detection and diagnosis. We also employ the explainable Artificial Intelligence tool GradCAM [21] to validate results and provide visual explanations by highlighting the regions of input images contributing most to the model's predictions.

2 Proposed Pipeline

This study improves the accuracy and efficiency of AD diagnosis using combiplane MRI images by utilising the CNN design and integrating convolutional and pooling layers. This section highlights three key components of the proposed pipeline: 1) training single-plane MRI using CNN architecture from scratch, 2) testing and prediction with combiplane images, and 3) the interpretation method employed. These components play crucial roles in achieving improved results.

In this study, the utilisation of MRI is demonstrated as an illustration, with PET and fMRI being alternative options for consideration. The shadowed blocks of Figure 1 shows the implementation of the proposed approach, where a non-pre-trained CNN model is used to perform individual testing and prediction on MRI images from the axial, coronal, and sagittal planes. We trained the CNN model from scratch with the architecture $16C2 - 16C2 - MP2 - 32C2 - 32C2 - 32C2 - MP2 - 64C2 - 16C1 - Flatten - 4N$.

In the next step, the last convolutional layer output and predictions for each plane are obtained and interpreted. The prediction results are then fused to create an ensemble of combiplane MRI images. In the proposed approach, the ensemble is formed by a soft voting process using individual predictions from the three planes. The final prediction is based on the highest probability value among the summed predicted probabilities. For a test sample j, the soft voting approach yields the following probabilities:

$$\Omega_c^j = (\beta_1^{cj}, \beta_2^{cj}, \cdots, \beta_k^{cj}) \quad (1)$$

$$\Omega_a^j = (\beta_1^{aj}, \beta_2^{aj}, \cdots, \beta_k^{aj}) \quad (2)$$

$$\Omega_s^j = (\beta_1^{sj}, \beta_2^{sj}, \cdots, \beta_k^{sj}) \quad (3)$$

where, Ω_c^j, Ω_a^j and Ω_s^j indicate the probability obtained by sample j respectively in planes(p) coronal, axial and sagittal. The β_k^{pj} (where $k = 2$ for binary classification and $k = 4$ for 4-way classification) indicates the probability assigned for each of the k classes in corresponding plane p. The final prediction label, denoted as l^j, is achieved as follows:

$$l^j = argmax(\sum \beta_1^{pj}, \sum \beta_2^{pj}, \cdots, \sum \beta_k^{pj}) \quad \forall p \in [c, a, s] \quad (4)$$

In the final phase of the proposed work, the axial, coronal, and sagittal plane results from the final convolutional layer are analysed using the GradCAM technique. Using the gradient information from the last CNN layer, a heatmap is generated to highlight critical image regions and provide explanations for predictions. The heatmap is then overlaid onto the input image, clearly explaining the influential regions. In the equation,

$L_{GradCAM}^c = ReLU\left(\sum_k \alpha_k^c \cdot A^k\right)$, $L_{GradCAM}^c$ represents the GradCAM heatmap, α_k^c denotes the weights calculated for each feature map A^k, and the summation aggregates the weighted feature

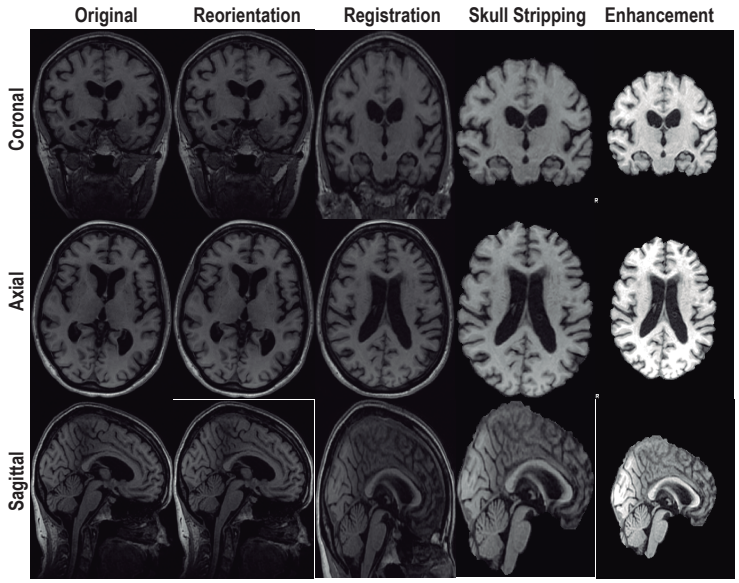

Figure 2: The preprocessing pipeline. Original MRI → Reorientation → Registration → Skull-stripping → Image Enhancement. The reorientation process aligns the image with a standardised anatomical reference frame which might not be perceivable visually.

maps for a target class c. The ReLU function is applied to ensure the heatmap contains only positive values.

This equation summarises the key steps of the GradCAM technique, which identifies the significant regions in the input image that contribute to the predicted class.

To correlate the heatmap with the original image, we explored using Pearson's correlation coefficient (PCC) [27]. The images were resized and converted into 1D vectors represented by X and Y. Normalisation of X and Y was performed using Equations 5 and 6:

$$\hat{X} = \frac{X - \mu_X}{\sigma_X} \tag{5}$$

$$\hat{Y} = \frac{Y - \mu_Y}{\sigma_Y} \tag{6}$$

where μ_X and μ_Y is the mean and σ_X and σ_Y is the standard deviation. The covariance of the normalised images $\hat{X}$ and $\hat{Y}$ are calculated as follows:

$$cov = \frac{\sum_{i=1}^{n} \left((\hat{X} - \bar{X})(\hat{Y} - \bar{Y}) \right)}{n - 1} \tag{7}$$

where n is the number of elements in the vector and $\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively. The standard deviations σ_X and σ_Y of the normalised vectors $\hat{X}$ and $\hat{Y}$ are computed as in Equations 8 and 9.

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{X}_i - \bar{X} \right)^2}{n - 1}} \tag{8}$$

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y} \right)^2}{n - 1}} \tag{9}$$

The Pearson's Correlation Coefficient (PCC), defined by Equation-10, ranges between -1 and +1. A negative value indicates a negative correlation, zero represents no correlation, and a positive value

signifies a positive correlation.

$$P_r = \frac{cov}{\sigma_X \cdot \sigma_Y} \tag{10}$$

The results with Pearson's correlation results are discussed in the experimental section.

3 Results and Discussion

The study utilised T1-weighted MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) in axial, sagittal, and coronal planes [10]. The dataset included 1056 MRI images with 223 AD, 475 EMCI, 262 LMCI, and 96 CN cases. The Synthetic Minority Over Sampling Technique (SMOTE) algorithm was applied, resulting in balanced datasets with approximately 13,822, 13,780, and 13,856 samples for coronal, axial, and sagittal planes, respectively. The data was split into 70% training and 30% testing sets. Furthermore, a low learning rate (0.0001) and Adam activation were used, with a batch size of 128. The number of epochs was determined using EarlyStopping callback of Keras platform. The model implementation and fine-tuning of pretrained models and subsequent evaluation were performed on a Windows system equipped with an NVIDIA RTX 3060 GPU and a 3.2 GHz CPU, ensuring good computational support for the experiments. Model development and testing were carried out using the Keras API within TensorFlow, allowing for efficient and flexible experimentation.

The MRI images underwent preprocessing using the popular FMRIB Software Library (FSL) toolkit, which offers various analytical tools for MRI data. Preprocessing involved three main steps: reorientation, registration, and skull-stripping [20]. The skull-stripping was performed using the brain extraction tool (BET). This widely applied preprocessing pipeline ensures enhancing analytical accuracy and image interpretation. Figure 2 illustrates the intermediate outputs of these preprocessing steps. The code used in this preprocessing pipeline is sourced from the study by Vimbi et al. [30].

Table 1: Performance metrics of Models on Individual planes

N	Plane	Acc	Sp	Se	FNR	FPR
1000	C	0.889	0.9633	0.8884	0.1115	0.0366
	A	0.888	0.9628	0.8870	0.1129	0.0371
	S	0.904	0.9679	0.9027	0.0972	0.0320
	Co	0.979	0.9930	0.9786	0.0213	0.0069
2000	C	0.878	0.9593	0.8784	0.1215	0.0406
	A	0.8885	0.9628	0.8890	0.1109	0.0371
	S	0.9085	0.9695	0.9086	0.0913	0.0304
	Co	0.978	0.9926	0.9782	0.0217	0.0073
4000	C	0.878	0.9592	0.8780	0.1219	0.0407
	A	0.8942	0.9646	0.8940	0.1054	0.0353
	S	0.9147	0.9715	0.9148	0.0851	0.0284
	Co	0.9832	0.9944	0.9831	0.0168	0.0055
8000	C	0.8756	0.9583	0.8760	0.1239	0.0416
	A	0.8942	0.9645	0.8940	0.1059	0.0354
	S	0.9098	0.9698	0.9103	0.0896	0.0301
	Co	0.9807	0.9935	0.9807	0.0192	0.0064

Legend– N: Number of training samples; C: Coronal; A: Axial; S: Sagittal; Co: combiplane; Acc: Accuracy; Sp: Specificity; Se: Sensitivity; FNR: False negative rate; FPR: False positive rate.

The evaluation of the models was done using well-known performance metrics: Accuracy (A_c), Sensitivity (S_e) or True Positive Rate (TPR), and Specificity (S_p) or True Negative Rate (TNR) [22]. The false positive rate (FPR) can be derived as $1 - S_p$. The false negative rate (FNR) can be calculated as $1 - S_e$. All reported values were averaged using the one-vs-all strategy. In addition, paired t-tests were computed to compare the accuracy of the proposed combiplane method over the individual plane method.

In the initial experiment, models were trained using the 70% training pool and testing was carried out by varying the number of test samples (N) from the coronal, axial and sagittal planes with identical labels. We established test pools of varying numbers of N samples each from three planes with

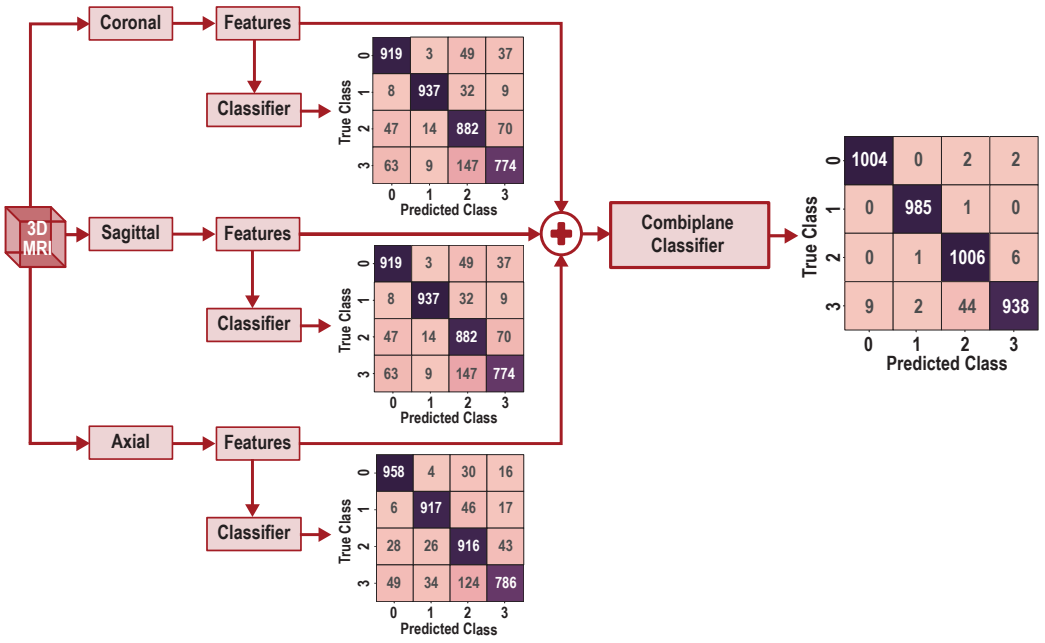

Figure 3: Confusion matrices as classification results of individual planes (i.e., coronal, sagittal and axial) and combiplane.

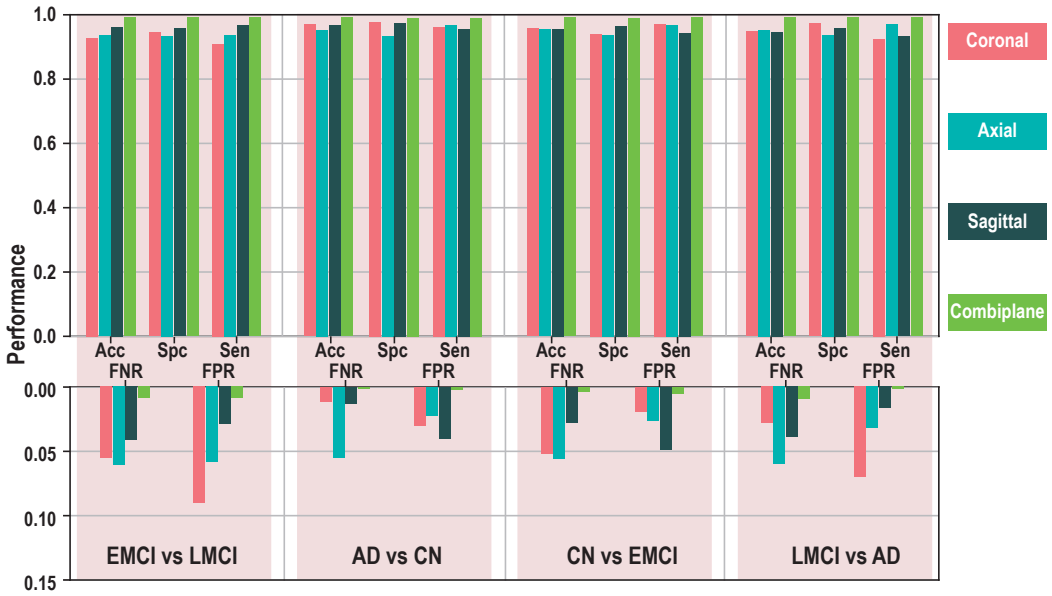

Figure 4: Comparative Analysis of Diagnostic Performance Metrics: Accuracy (Acc), Specificity (Spc), Sensitivity (Sen), False Negative Rate (FNR), and False Positive Rate (FPR) across Coronal, Axial, Sagittal, and combiplanes in Binary Classification of AD.

the same patient ID and, therefore, sharing the same label. In addition to disease characterisation in different planes individually, this experiment also assessed the combiplane model's capacity to analyse multiple MRI images concurrently. Table 1 reveals that the combiplane approach consistently outperformed the individual planes across different numbers of test samples. For instance, with 1000 training samples, the combiplane achieved an accuracy of 0.979, a specificity of 0.993, a sensitivity of 0.9786, an FNR of 0.0213, and an FPR of 0.0069. This is an impressive improvement in accuracy, approximately 9.65%, compared to the average accuracy of the individual planes. These findings

demonstrate the effectiveness of combining information from multiple planes in enhancing disease characterisation accuracy, providing valuable insights for medical imaging analysis and diagnosis. It is important to note that the combiplane's high accuracy was also accompanied by excellent performance in other metrics. The combiplane achieved high specificity (0.9944) with a low FPR (0.0055) and high sensitivity (0.9831) with a low FNR (0.0168), highlighting its effectiveness in correctly identifying positive cases. In summary, the experiment showcases the superiority of the proposed combiplane's ability to identify and classify AD accurately.

Figure 3 shows confusion matrices ($N = 4000$) obtained for individual and combiplane models. The combiplane model was accurate across all classes (CN, EMCI, LMCI, and AD). It identified CN and EMCI instances more accurately than models trained individually. The combiplane model also distinguished LMCI and AD better than individual plane models. This improved discrimination allows early dementia prediction and management, postponing AD development. Thus, the combiplane method provides superior diagnostic capabilities for MCI diagnosis.

Table 2: Comparison of transfer learning models across individual and combiplane. C: Coronal, A: Axial, S: Sagittal, Co: combiplane.

Architectures	C	A	S	Co	p-value
InceptionV3	0.80	0.84	0.94	0.99	0.000138
IRV2	0.39	0.41	0.47	0.57	5.90E-07
Xception	0.73	0.78	0.85	0.93	8.64E-09
DenseNet121	0.73	0.84	0.90	0.95	0.000531
CNN	0.88	0.89	0.91	0.98	4.47E-07

We compared the combiplane technique to prominent transfer learning (TL) architectures to test its consistency. For this experiment, each architecture was trained separately on individual planes and considered 4000 random samples for testing and computed the accuracy on 5 separate runs. This resulted in 5 different accuracy values for planes C, A, S and Co planes. Table 2 shows that the combiplane outperforms the individual planes in all TL networks. The combiplane accuracy yielded an impressive 0.98 accuracy, showing that combining disease patterns from distinct planes improves performance. Among the TL architectures evaluated, InceptionV3 performs well in all image orientations. With accuracies of 0.80 for coronal, 0.84 for axial, and an impressive 0.94 for sagittal orientations, InceptionV3 consistently demonstrates strong predictive capabilities. InceptionResNetV2 obtained worse accuracy than other architectures in this study. Yet the combiplane approach improves accuracy through the complimentary information from several planes. Xception and DenseNet121 achieve 0.73 to 0.90 accuracies across multiplanes. Their integration into the combiplane leads to notable accuracy boosts, further highlighting the effectiveness of the combiplane. To compare the significance of combiplane accuracy over individual planes, we computed paired t-tests to test the significance of the results. In the context of comparing the combiplane model to individual planes, the obtained p-values (see Table 2) imply high statistical significance confirming increased reproducibility of the method and reflecting a consistent pattern in the results rather than a random chance.

For binary classification of different stages of AD (AD vs CN, EMCI vs LMCI, CN vs EMCI, and LMCI vs AD), we evaluated the efficacy of models considering both individual planes and the combiplane. The results, presented in Figure 4, demonstrated that the combiplane achieved superior accuracy (0.99325) compared to the individual planes. Additionally, the combiplane exhibited high specificity (0.99292572) and sensitivity (0.993567541), resulting in a low FNR of 0.00707428 and FPR of 0.006432459. While the individual planes also demonstrated strong performance with accuracies of 0.95 (coronal), 0.955 (axial), and 0.947 (sagittal), the combiplane consistently outperformed them, highlighting its enhanced capability in distinguishing between the targeted stages of AD.

The confusion matrices for all planes involved in the classifications are shown in Figure 5. The combiplane model demonstrates remarkable accuracy in correctly detecting various stages of AD, which is crucial for early disease detection. This underscores the invaluable role of the combiplane in enhancing diagnostic accuracy and enabling timely interventions.

The non-pretrained CNN architecture was fine-tuned and used for prediction with preserved weights. The heatmap was generated using the predictions from the last convolutional layer to explain the

Figure 5: Comprehensive Comparison of confusion matrices for coronal, axial, sagittal, and combiplane in the binary classification of AD. The abscissas denote the predicted classes, while the ordinates denote the true classes.

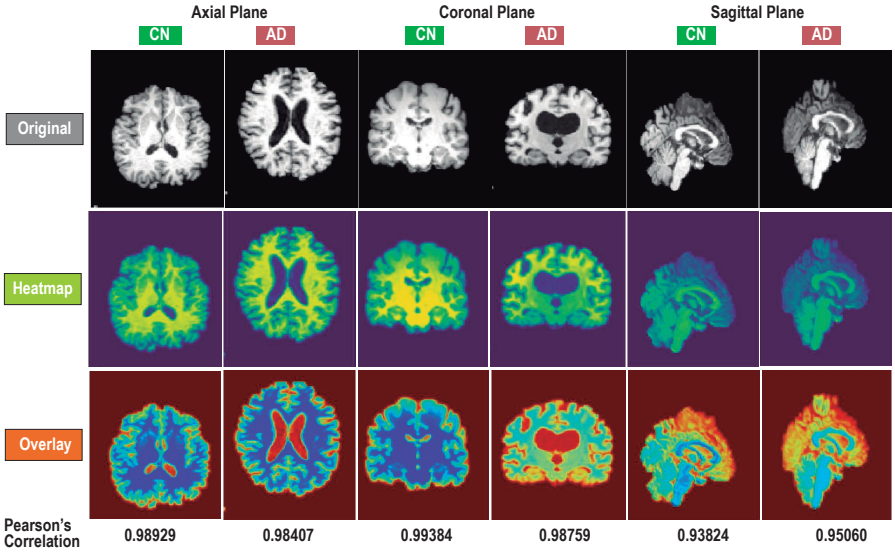

Figure 6: Comparison between CN vs AD for original MRI image, heatmap, overlays and PCC for axial, coronal, and sagittal planes.

results. Figure 6 depicts the original MRI image, GradCAM heatmap, and class overlays (CN vs. AD) for axial, coronal, and sagittal planes. The 'red' regions in the overlays signify a close affinity to the predicted class. The axial plane highlights regions associated with neurodegeneration, atrophy, or abnormal metabolic activity in AD. The coronal heatmap focuses on cortical areas like the temporal and parietal lobes, commonly affected in AD. The sagittal heatmap highlights vulnerable regions like the prefrontal cortex and posterior cingulate cortex. By examining all the heatmaps, clinicians can gain insights into the effects of AD-related changes in the brain.

In this experiment, we employed PCC to measure the affinity between the original MRI image and the heatmap. The results demonstrate a high positive correlation in both CN and AD cases. As shown in Figure 6, the PCC values between the original image and its heatmap are high for both CN (0.98929, 0.99384, and 0.93824) and AD (0.98407, 0.98759, and 0.95060) for the three planes. The PCC measures how well the heatmap captures the features of MRI; hence, high values effectively highlight the relevant regions of interest for the predicted class.

4 Limitations

One limitation of this approach is the increase in computational cost associated with processing multiple planes to achieve enhanced accuracy. Each plane requires additional processing power, introduces more parameters to the model, leading to a cumulative increase in resource demand. In our experiments, which included 8,000 images, parallel processing proved manageable; however, scaling this approach to significantly larger datasets may strain resources, especially in real-time or large-scale applications. Consequently, the feasibility of deploying this method may be restricted by hardware constraints, particularly for real-time applications or large-scale datasets. Future work may explore optimization techniques to further reduce computational costs, making the approach more accessible and scalable across varied computational infrastructures.

5 Conclusion

Recent studies have shown that deep learning techniques with sMRI can effectively classify AD. Our research further improved the accuracy of AD detection by incorporating CNNs and multi-plane MRI images. We discovered that combining classification results from different imaging perspectives (combi-plane models) produced even better results. This novel approach of combining planes, followed by the application of a classification model, is compatible with any machine learning algorithm.

To ensure explainability, we used GradCAM, which provided heatmaps highlighting the specific brain regions most relevant to the classification process. We also found a strong positive correlation between the original MRI and the heatmap using the PCC across the axial, coronal, and sagittal planes. Our future work could further leverage the potential of combi-plane approaches in AD classification addressing challenges related to interpretability and trust. Enhancing trust is paramount to ensuring a reliable integration of AI-driven diagnostic systems into clinical practice.

In summary, utilising deep learning methods for AD classification with sMRI has shown to be highly effective and full of potential. Through ongoing research, we can anticipate substantial enhancements in AD identification and treatment, ushering in a new age of targeted medicine and improved patient care.

References

[1] Cucun Very Angkoso, Hapsari Peni Agustin Tjahyaningtijas, MH Purnomo, and IKE Purnama. Multiplane convolutional neural network (mp-cnn) for alzheimer's disease classification. *Int. J. Intell. Eng. Syst.*, 15:329–340, 2022.

[2] Gaurang Bansal, Vinay Chamola, Pratik Narang, Subham Kumar, and Sundaresan Raman. Deep3dscan: Deep residual network and morphological descriptor based framework forlung cancer classification and 3d segmentation. *IET Image Process.*, 14(7):1240–1247, 2020.

[3] Leon Y Cai, Qi Yang, Colin B Hansen, Vishwesh Nath, Karthik Ramadass, Graham W Johnson, Benjamin N Conrad, Brian D Boyd, John P Begnoche, Lori L Beason-Held, et al. Prequal: An

automated pipeline for integrated preprocessing and quality assurance of diffusion weighted mri images. *Magnetic resonance in medicine*, 86(1):456–470, 2021.

[4] Matthew Cieslak, Philip A Cook, Xiaosong He, Fang-Cheng Yeh, Thijs Dhollander, Azeez Adebimpe, Geoffrey K Aguirre, Danielle S Bassett, Richard F Betzel, Josiane Bourque, et al. Qsiprep: an integrative platform for preprocessing and reconstructing diffusion mri data. *Nature methods*, 18(7):775–778, 2021.

[5] Robert W Cox. Equitable thresholding and clustering: a novel method for functional magnetic resonance imaging clustering in afni. *Brain connectivity*, 9(7):529–538, 2019.

[6] S Gauthier, C Webster, S Sarvaes, JA Morais, and P Rosa-Neto. *World Alzheimer Report 2022: Life After Diagnosis - Navigating Treatment, Care and Support*. Alzheimer's Disease International, London, England, 2022.

[7] Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

[8] Faizal Hajamohideen, Noushath Shaffi, Mufti Mahmud, Karthikeyan Subramanian, Arwa Al Sariri, Viswan Vimbi, et al. Four-way classification of alzheimer's disease using deep siamese convolutional neural network with triplet-loss function. *Brain Inform.*, 10(1):1–13, 2023.

[9] Ruhul Amin Hazarika, Debdatta Kandar, and Arnab Kumar Maji. An experimental analysis of different deep learning based models for alzheimer's disease classification using brain magnetic resonance images. *J. King Saud Univ. Comput. Inf. Sci.*, 34(10):8576–8598, 2022.

[10] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, and et al Thompson, Paul. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *J. Magn. Reson. Imaging*, 27(4):685–691, 2008.

[11] Jinseong Jang and Dosik Hwang. M3t: three-dimensional medical image classifier using multi-plane and multi-slice transformer. In *Proc. CVPR*, pages 20718–20729, 2022.

[12] Stefan J Kiebel, John Ashburner, Jean-Baptiste Poline, and Karl J Friston. Mri and pet coregistration—a cross validation of statistical parametric mapping and automated image registration. *Neuroimage*, 5(4):271–279, 1997.

[13] Fei Liu, Huabin Wang, Shiuan-Ni Liang, Zhe Jin, Shicheng Wei, Xuejun Li, et al. Mps-ffa: A multiplane and multiscale feature fusion attention network for alzheimer's disease prediction with structural mri. *Comput. Biol. Med.*, 157:106790, 2023.

[14] Chandler RL Mongerson, Russell W Jennings, David Borsook, Lino Becerra, and Dusica Bajic. Resting-state functional connectivity in the infant brain: methods, pitfalls, and potentiality. *Frontiers in Pediatrics*, 5:159, 2017.

[15] Manan Binth Taj Noor, Nusrat Zerin Zenia, M Shamim Kaiser, Shamim Al Mamun, and Mufti Mahmud. Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of alzheimer's disease, parkinson's disease and schizophrenia. *Brain Inform.*, 7:1–21, 2020.

[16] Steve Pieper, Michael Halle, and Ron Kikinis. 3d slicer. In *2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No. 04EX821)*, pages 632–635. IEEE, 2004.

[17] Shangran Qiu, Matthew Miller, Prajakta Joshi, Joyce Lee, Yunruo Ni, Yuwei Wang, et al. Multi-modal deep learning for alzheimer's disease dementia assessment. *Nat. Commun.*, 13(1):3404, 2022.

[18] Alessia Sarica, Giuseppe Di Fatta, and Mario Cannataro. K-surfer: a knime extension for the management and analysis of human brain mri freesurfer/fsl data. In *Brain Informatics and Health: International Conference, BIH 2014, Warsaw, Poland, August 11-14, 2014, Proceedings*, pages 481–492. Springer, 2014.

[19] Ryosuke Sato, Yutaro Iwamoto, Kook Cho, Do-Young Kang, and Yen-Wei Chen. Comparison of cnn models with different plane images and their combinations for classification of alzheimer's disease using pet images. In *Proc. KES-InMed and KES-IIMSS*, pages 169–177, 2019.

[20] Florent Ségonne, Anders M Dale, Evelina Busa, Maureen Glessner, David Salat, Horst Karl Hahn, and Bruce Fischl. A hybrid approach to the skull stripping problem in mri. *Neuroimage*, 22(3):1060–1075, 2004.

[21] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proc. IEEE ICCV*, Oct 2017.

[22] Noushath Shaffi, Karthikeyan Subramanian, Viswan Vimbi, Faizal Hajamohideen, Abdelhamid Abdesselam, and Mufti Mahmud. Performance evaluation of deep, shallow and ensemble machine learning methods for the automated classification of alzheimer's disease. *International Journal of Neural Systems*, 2450029, 2024.

[23] Noushath Shaffi, Vimbi Viswan, and Mufti Mahmud. Ensemble of vision transformer architectures for efficient alzheimer's disease classification. *Brain Informatics*, 11(1):25, 2024.

[24] Jayanthi Venkatraman Shanmugam, Baskar Duraisamy, Blessy Chittattukarakkaran Simon, and Preethi Bhaskaran. Alzheimer's disease classification using pre-trained deep networks. *Biomed. Signal Process. Control*, 71:103217, 2022.

[25] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.

[26] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004.

[27] Baris Turkbey et al. Correlation of magnetic resonance imaging tumor volume with histopathology. *J. Urol.*, 188(4):1157–1163, 2012.

[28] Prashanthi Vemuri and Clifford R Jack. Role of structural mri in alzheimer's disease. *Alzheimer's Res. Ther.*, 2(4):1–10, 2010.

[29] Vimbi Viswan, Noushath Shaffi, Mufti Mahmud, Karthikeyan Subramanian, and Faizal Hajamohideen. Explainable artificial intelligence in alzheimer's disease classification: A systematic review. *Cognitive Computation*, 16(1):1–44, 2024.

[30] Vimbi Viswan, Noushath Shaffi, Karthikeyan Subramanian, and Faizal Hajamohideen. Optimizing medical imaging quality: An in-depth examination of preprocessing methods for brain mris. In *International Conference on Applied Intelligence and Informatics*, pages 65–81. Springer, 2023.

[31] Konstantina G Yiannopoulou and Sokratis G Papageorgiou. Current and future treatments in alzheimer disease: an update. *J. Cent. Nerv. Syst. Dis.*, 12:1179573520907397, 2020.

Acknowledgments and Disclosure of Funding

This work didn't receive any funding from external funding agencies.

A Appendix / supplemental material

The supplemental material contains details about code and dataset used in this study.

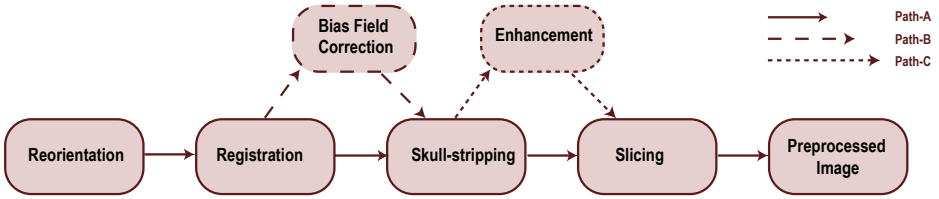

Figure 7: Preprocessing Pipeline for MRI Scans.

A.1 Code

The code used in the study can be accessed from `https://github.com/brai-acslab/comriad/`. The code contains details about required packages, training details including hyperparameter values, train-test split, etc. This code repository has the following Python notebook files:

1. `ModelBuilding.ipynb`: This code uses the dataset (as detailed in Appendix A.2) and builds a model that uses MRI planes for Axial, Coronal and Sagittal planes for the four-way classification of Alzheimer's Disease.
2. `ModelTesting.ipynb`: This code uses the dataset as mentioned in Section-3 to test the efficacy of proposed model for the four-way classification of Alzheimer's Disease.
3. `ModelBuildingBinary.ipynb`: This code uses the dataset (as detailed in Appendix A.2) and builds a model that uses MRI planes for Axial, Coronal and Sagittal planes for the two-way classification of Alzheimer's Disease.
4. `ModelTestingBinary.ipynb`: This code uses the dataset (as detailed in Appendix A.2) and builds a model that uses MRI planes for Axial, Coronal and Sagittal planes for the two-way classification of Alzheimer's Disease.
5. `t-tesst.ipynb`:

A.2 Dataset

A.2.1 Data Extraction

The dataset utilized in this study is sourced from well-established databases such as the Alzheimer's Disease Neuroimaging Initiative dataset (ADNI, `https://adni.loni.usc.edu/`). After the data access request was approved, the ADNI data was downloaded from the LONI Image and Data Archive (IDA). To download the appropriate subset of data from the IDA, the database was searched for T1-weighted structured MRI (sMRI) samples belonging to subjects in ADNI1, ADNI2 and ADNI-GO cohorts aged between 55 and 65 from the CN (Cognitively Normal) AD(Alzheimer's Disease), MCI (Mild Cognitive Impairment), EMCI (Early MCI), and LMCI (Late MCI) categories.

A.2.2 Preprocessing of Extracted MRI

After extraction of the MRI samples, they were preprocessed to enhance image quality. From the quality assessment viewpoint, several preprocessing stages exist, namely, signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), Image Similarity Metrics (ISM), Fractional Anisotropy Analysis (FAA), Chi-squared Analysis (CA), and Mask Quality Analysis (MQA) [3]. Admidst several preprocessing tools, like, Free Surfer[18], Statistical Parametric Mapping (SPM)[12], Advanced Normalisation tools (ANT)[4], and AFNI software[5], the FSL (FMRIB) software library was used [25] which allows the evaluation of each preprocessing stage for data quality.

Several preprocessing steps were used: reorientation, registration, skull stripping, and slicing (see Figure 7, path B).

Reorientation of MRI Reorientation involves applying transformations to align MRIs with a standardized reference space. This is accomplished by aligning an individual's orientation to a standardized anatomical plane by altering their position along certain axes [14]. This practice guarantees consistent orientation of images, facilitating the process of comparing and integrating data

from many participants or research. This process helps align the image with a universally accepted orientation. This alignment simplifies comparison and ensures compatibility across various analyses and software packages. Reorientation, therefore, aids in maintaining uniformity and enhancing the dependability of later studies and data integration.

Image Registration Image registration refers to aligning several images in a shared coordinate space. The image undergoes a linear transformation encompassing translation, rotation, and scaling, which facilitates the alignment of one image with another [7]. The practice of aligning various modalities of imaging data or matching an individual's data to a template or standard space is widely employed. Coherently, the registration algorithms align the spatial attributes of diverse images, thereby compensating for discrepancies in positioning or subject motion that may have occurred during the image acquisition process.

Skull Stripping The MRI image acquired subsequent to the registration procedure comprises extraneous non-cerebral tissues that necessitate removal. Skull stripping refers to separating the brain region from surrounding non-brain tissues, such as the skull, scalp, and non-brain structures [26]. The act of separating brain structures aids in the process of concentrating on the analysis and visualization of those structures. The technique of skull stripping is an essential preprocessing procedure in various neuroimaging investigations, such as brain morphometry, functional connectivity, and diffusion tensor imaging. The skull-stripped images produced by tools such as the Brain Extraction Tool (BET), a technique developed by the Biomedical Engineering and Technology program, exhibit an aesthetically pleasing quality and enhanced interpretability. Consequently, these images are helpful in educational, research, and therapeutic contexts.

Image Slicing The MRI that has been reoriented, registered, and skull-stripped is now available in a three-dimensional format comprising three planes known as axial, coronal, and sagittal. The generation of a two-dimensional representation of a specific plane from an input volume can be achieved through the utilization of the Slicer program [16]. Therefore, the objective is to divide 3D volumetric data into 2D slices for easier visualization and analysis.

A.2.3 MRI Preprocessing with FSL: Setup FSL and Sample Pipeline Walkthrough with Shortcodes

This section presents a guideline for FSL (FMRIB Software Library) setup on macOS-based machines. A comprehensive MRI scan preprocessing methodology is presented and encompasses reorientation, registration, skull stripping, and slicing. The preprocessing procedures were done on an Apple macOS system featuring an Apple M2 CPU, 8 GB of RAM, and a 10-core GPU.

FSL Setup FSL, is a popular software package from the University of Oxford FMRIB Centre (`https://www.win.ox.ac.uk/`). The software offers comprehensive tools and strategies for preprocessing and analyzing neuroimaging data, including functional and structural MRI [26]. This study used FSL version 6.0.6.5, and this release improves the reliability and efficiency of our neuroimaging analysis workflow. FSL was chosen due to its popular use in neuroimaging and consistent research and support by the FMRIB Analysis Group.

The study used FSL for key neuroimaging analyses such as reorientation, registration, skull stripping, and slicing. However, in this study we used the macOS terminal window to preprocess MRI using FSL command-line tools. Shortcodes used in the terminal window can be found in the following section. We also utilised the 'fsleyes' program, a strong neuroimaging tool for examining FSL-processed MRI data. The FSL software setup for macOS is shown in Table 3. Installation guideline for WindowsOS users and further reading can be obtained from (`https://fsl.fmrib.ox.ac.uk/fsl/fslwiki`).

Preprocessing Pipeline with shortcodes In this study, we used FSL to perform tasks like image reorientation, registration, skull stripping, and slicing. To standardize and reorient neuroimaging data, we used the command line tool "fslreorient2std" within FSL. This process guarantees a consistent orientation of the images, facilitating comparing and integrating data from many participants or research. The syntax for reorientation in the command line is provided in the subsequent statement:

`fslreorient2std infname outfname`

Table 3: FSL installation guideline

Installation Steps	**FSL Installation Instructions**
Prerequisites	Install XQuartz to run X Windows system on macOS (`https://www.xquartz.org/`)
Downloader	Register and download the installer (*fslinstaller.py*) (`https://fsl.fmrib.ox.ac.uk/fsldownloads_registration`)
Running the installer on macOS	At terminal window run installer with Python $ *python fslinstaller.py* Open a new terminal window to begin using FSL
Checking installation	At new terminal use code below to display name of installed FSL directory $ *echo $FSLDIR*
Running GUI	To start the main FSL GUI type $ *fsl*

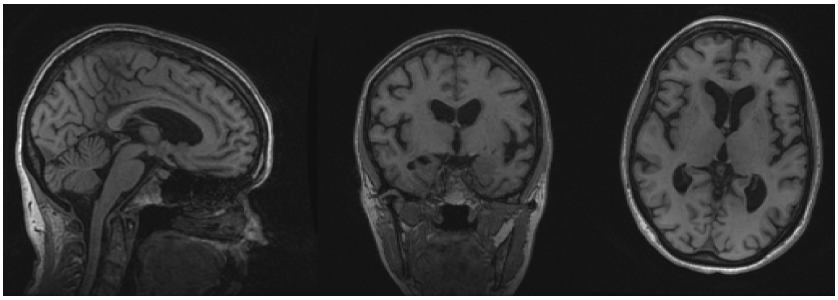

Figure 8: MRI after reorientation.

The input file name, "*infname*", refers to the original MRI image that needs reorientation. The output file name "*outfname*" indicates where the reoriented image will be stored. Figure 8 shows how '*fsleyes*' is used to see the reoriented image. With this command, we aligned the image to a conventional orientation. Keeping data uniform and reliable improves later studies and data integration.

In order to align images to a common coordinate space, also known as image registration, the command line tool within FSL, '*flirt*', was used. This tool is used for a linear transformation of the images, such as translation, rotation, and scaling. Figure 9 shows MRI images after registration, and Table shows the '*flirt*' command with specific parameters for image registration, the breakdown of the commands, and their options. The command line syntax for registration using the FSL tool is shown in the statement :

```
flirt -in rofname -ref referencefname -out outfname -omat fname.mat -bins
   256 -cost corratio -searchrx 0 0 -searchry 0 0 -searchrz 0 0 -dof 12
                              -interp spline
```

The brain region is isolated by removing the skull, scalp, and other non-brain components. Skull-stripped images can be created using the FSL tool BET (Brain Extraction Tool) and are more aesthetically pleasing and simpler to understand. The following statement displays the command line syntax for skull stripping:

```
bet infname outfname -R -f 0.5 -g 0
```

In the command, '*infname*' refers to the input file name, and '*outfname*' is the output file name. The *-f* parameter in the BET command indicates the fractional intensity threshold that determines the algorithm's sensitivity for extraction and can be adjusted to include more or less brain tissue. There are two other options to consider: '*-R*' for robust brain center estimation and '*-g* 0' to refine the skull stripping using local intensity gradients. A higher fractional intensity threshold includes more brain tissue, while a lower value removes more non-brain structures (refer Figure 10).

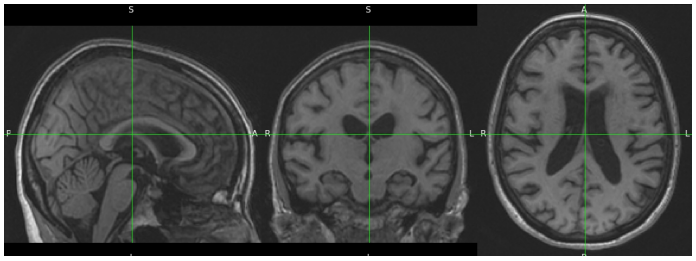

Figure 9: MRI after registration

Table 4: Parameters and its description for the flirt command line tool

Command	**Options**
flirt	This is the command to execute FLIRT
-in rofname	Specifies the input image filename to be registered
-ref referencefname	Specifies the reference image filename to which the input image will be aligned
-out outfname	Specifies the output image filename after registration
-omat fname.mat	Specifies the output matrix filename that stores the transformation matrix
-bins 256	Sets the number of histogram bins used for image intensity matching
-cost corration	Specifies the cost function to be used for registration, and this case, correlation ratio
-searchrx 0 0	Sets the search range for registration in the x direction which is set to 0 indicating no search
-searchry 0 0	Sets the search range for registration in the y direction which is set to 0 indicating no search
-searchrz 0 0	Sets the search range for registration in the z direction which is set to 0 indicating no search
-dof 12	Sets the degrees of freedom for the transformation model and 12 indicates a full affine transformation.
-interp spline	Specifies the interpolation method to be used during resampling.

Reoriented, registered, and skull-stripped MRIs are now available in 3D axial, coronal, and sagittal planes. A Slicer tool can transform an input volume into a 2D view of any plane. Slicer is an open-source medical image processing, research, and visualization application. Example slicing command line syntax:

```
slicer infname -z -90 outfname
```

The command line parameters are: '*infname*' is the name of the input volume file containing the MRI data; '*-z -90*' selects the slice at *-90mm* along the z-axis for an axial or transverse view; and '*outfname*' is the output file name to be saved. The orientations *-y* and *-x* represent the coronal and

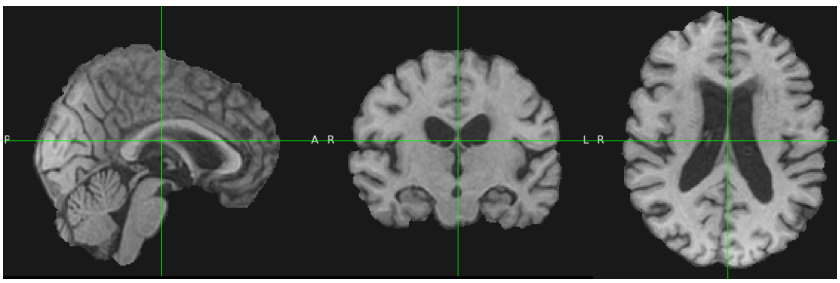

Figure 10: MRI after skull stripping

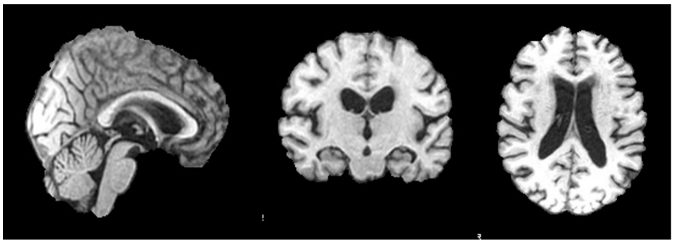

Figure 11: MRI after slicing

sagittal planes, respectively. This command loads the input volume and extracts the desired slice at the specified location. The 2D axial view image (refer Figure 11) can be processed, studied, or visualized for that slice of interest.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: **[TODO]**The claims made in abstract is reflected in Introduction and Results section.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: **[TODO]**There is a separate Limitations section provided in the paper.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: **[TODO]**Our paper focuses exclusively on an approach aimed at enhancing classification accuracy. Since the paper does not present new theoretical results no formal assumptions or proofs are required.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: **[TODO]**Section 2 describes the steps taken to reproduce the model. Abstract provides link to GitHub that has access to reproducible code to make results verifiable

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: **[TODO]**The authors have detailed the steps to access and preprocess the raw data. The preprocessing code used in this study is outlined in the supplementary section. The environment needed to run to reproduce the results is mentioned in the article.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: **[TODO]**Section 3 provides complete details about all training and testing parameters. Supplementary material provides full details.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: **[TODO]**Results are accompanied by error bars and statistical significance test in Section 3

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: **[TODO]**Section 4 indicates the type of compute utilized in the paper with specific model details.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: **[TODO]**The MRI dataset used in the study does not reveal personally identifiable information of subjects/patients

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: **[TODO]**No direct path to negative impact

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: **[TODO]**The paper has no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: **[TODO]**The paper cites the original work that produced the ADNI datset used in the paper (Reference - [10])

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: **[TODO]**The code has been made available through a public GitHub repository and the link of the same is provided in the abstract. The details of the dataset extraction process and preprocessing are provided in the supplementary material.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: **[TODO]**The paper does not involve crowdsourcing nor research with direct human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: **[TODO]**The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.