Instance Selection Framework for Alzheimer's Disease Classification Using Multiple Regions of Interest and Atlas Integration

Juan A. Castro-Silva[1,2], Maria N. Moreno-Garcia[1], Lorena Guachi-Guachi[3] and Diego H. Peluffo-Ordoñez[4,5]

[1]*Universidad de Salamanca, Salamanca, Spain*

[2]*Universidad Surcolombiana, Neiva, Colombia*

[3]*Department of Mechatronics, International University of Ecuador, Simon Bolivar Avenue 170411, Quito, Ecuador*

[4]*College of Computing, Mohammed VI Polytechnic University, Lot 660, Hay Moulay Rachid Ben Guerir, 43150, Morocco*

[5]*SDAS Research Group*

Keywords: Alzheimer's Disease, Swin Transformer, Weighted Ensemble, Instance Selection, Multiple Region of Interest.

Abstract: Optimal selection of informative instances from a dataset is critical for constructing accurate predictive models. As databases expand, leveraging instance selection techniques becomes imperative to condense data into a more manageable size. This research unveils a novel framework designed to strategically identify and choose the most informative 2D brain image slices for Alzheimer's disease classification. Such a framework integrates annotations from multiple regions of interest across multiple atlases. The proposed framework consists of six core components: 1) Atlas merging for ROI annotation and hemisphere separation. 2) Image preprocessing to extract informative slices. 3) Dataset construction to prevent data leakage, select subjects, and split data. 4) Data generation for memory-efficient batches. 5) Model construction for diverse classification training and testing. 6) Weighted ensemble for combining predictions from multiple models with a single learning algorithm. Our instance selection framework was applied to construct Transformer-based classification models, demonstrating an overall accuracy of approximately 98.33% in distinguishing between Cognitively Normal and Alzheimer's cases at the subject level. It exhibited enhancements of 3.68%, 3.01%, 3.62% for sagittal, coronal, and axial planes respectively in comparison with the percentile technique.

1 INTRODUCTION

Alzheimer's disease (AD) is the leading cause of dementia in older adults. It is a progressive brain disorder that causes nerve cells to die, leading to significant brain volume reduction and affecting almost all brain functions (WHO, 2023). AD affects crucial brain regions, such as the Entorhinal Cortex, Fornix, Hippocampus, Frontal lobe, Temporal lobe, and Parietal lobe, impacting spatiotemporal orientation, cognition, memory, intelligence, judgment, behavior, and language (Yu and Lee, 2020).

Medical imaging, including modalities like MRI, PET, and DTI, aids in diagnosis and treatment by visualizing brain structures. The performance of AD classification models is influenced by data quality and quantity. Instance selection methods, varying in strategy and selected slice numbers, may involve removing slices based on position or informative content (Castro-Silva. et al., 2022).

The mentioned instance selection methods have limitations. Including less informative content can increase computational time and introduce noise, harming model performance. Selecting a low or fixed number of slices per volume may not guarantee MRI representativeness and may exclude AD-related or informative instances.

Data leakage occurs when test data is used in the training process, leading to bias. An incorrect split dataset, lack of independent testing data, or biased transfer learning can cause it (Wen et al., 2020).

Deep learning in computer vision excels at detecting brain structural changes via MRI, using 2D or 3D models for image or ROI analysis. Originally designed for sequence-to-sequence tasks like machine translation, the widely adopted Transformer (Vaswani et al., 2017) is now applied in diverse domains, including NLP, CV, and speech processing. The Swin Transformer (Liu et al., 2021), with its novel hierarchical architecture and Shifted windows, is known for

Castro-Silva, J., Moreno-Garcia, M., Guachi-Guachi, L. and Peluffo-Ordoñez, D.
Instance Selection Framework for Alzheimer's Disease Classification Using Multiple Regions of Interest and Atlas Integration.
DOI: 10.5220/0012469600003654

In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 453-460
ISBN: 978-989-758-684-2; ISSN: 2184-4313

efficient computation, making it suitable for tasks like image classification.

Ensemble methods can outperform single classifiers for AD diagnosis (Young et al., 2018) Homogeneous ensembles use one base classifier with diverse training data, and weighted ensembles assign weights based on member performance.

To overcome the limitations mentioned above, in this paper, we propose a novel framework for selecting the most informative 2D image slices based on the annotations fusion of multiple Regions of Interest (ROIs) such as the entorhinal cortex, fornix, hippocampus, frontal, parietal, and temporal lobes. The ROI analysis involves using multiple atlases or maps to divide the brain into distinct regions, leading to a method called Multiple ROI and Multiple Atlas-based Instance Selection, which classifies Cognitively Normal (CN) and AD cases.

The proposal consists of several components:

- A volume dataset builder to prevent data leakage, conducting an early subject dataset split for creating independent training, validation, and test sets.
- A data generator in charge of producing multiple-view instances from a volume belonging to a preprocessed dataset with skull-stripping and registration.
- A weighted ensemble builder that combines homogeneous methods. This ensemble employs a single base classifier 2D Swin Transformer model, trained on various datasets encompassing ROIs, planes, and hemispheres.

The major contributions of this novel proposal can be summarized as follows.

1. To capture the most informative images, we propose an ROI content extraction method. Using the mode, it identifies the centroid position (x, y) and crops the 2D slice image accordingly.
2. To enhance diagnostic accuracy in AD tasks, we propose Multiple ROI and Multiple Atlas-based Instance Selection. It combines ROI annotations from multiple atlases to select informative ROI slice images and remove useless instances.
3. To tackle insufficient sample utilization and limitations of a single classifier, we propose a homogeneous weighted ensemble using a 2D Swin Transformer model. This method uses multiple-view samples in a single classifier with different weights according to their accuracy performance.
4. The experimental results demonstrate that the proposed method's accuracy outperforms the state-of-the-art related works.

The remainder of this paper is structured as follows: Section 2 presents some related works. The materials and methods used for preprocessing and instance selection are included in Section 3. Section 4 provides a detailed description of the experiments conducted in this work and the parameter settings used. The results of the experiments are discussed in Section 5. Finally, Section 6 summarizes the concluding remarks of this work.

2 RELATED WORKS

Most proposed approaches for image instance selection for AD classification differ in the number of slices selected, and the technique used to obtain the most representative or discard the least informative slices.

In previous studies (Choi and Lee, 2020), a common method involves computing entropy values for slices, sorting by descending entropy, and selecting a fixed number (typically 8 to 32) of top slices. In (Qin et al., 2022), a pre-trained U-Net is used for skull stripping, resulting in 3D MRI images ($64 \times 64 \times 64$). In (Hu et al., 2023; Altay et al., 2021), 96 axial slices are selected after skull stripping and volume registration. In (Lyu et al., 2022), images are resized and down-sampled (70×75×50) after preprocessing. Finally, in (Castro-Silva. et al., 2022), instance selection uses percentile positions for 32 slices.

ROI extraction, a crucial image processing task, varies in approaches. In (Zaabi et al., 2020), the MRI scan is divided into 32*32-pixel blocks, extracting only those containing the hippocampus as ROIs. In (Bae et al., 2020), the complete hippocampus, analyzed with 30 MRI coronal slices. Other studies, such as (Pan et al., 2022; Li et al., 2022), create an ensemble classifier by extracting ROI-based patches from different brain regions, including the hippocampus, amygdala, and insulae.

The previously mentioned instance selection methods have some limitations. The model's performance greatly depends on the number of slices per volume (Castro-Silva. et al., 2022). Adding more image slices with less informative content can result in redundant or less representative information, increases the computational cost (time) of training, introduce noise, and deteriorates model performance. On the other hand, selecting a fixed number of slice images could exclude AD-related or more informative instances. A low number of slices per volume, for example (1, 8), does not ensure the representativeness of the 170-256 slice instances that comprise an MRI volume.

Classification algorithms must be evaluated without bias to ensure clinical relevance. Biased evaluations caused by data leakage issues, such as incorrect dataset splitting, absence of an independent test set, delayed splitting, and biased transfer can lead to misleading results by inflating model performance (Wen et al., 2020).

Vision and Swin Transformers are applied to AD classification using MRI data. Examples include the hybrid 2D model Conv-Swinformer in (Hu et al., 2023), merging a CNN module (VGGNet-16) with a Swin Vision Transformer. In (Lyu et al., 2022), a ViT model pre-trained on ImageNet-21K is used for AD/CN classification. Additionally, (Zhang and Khalvati, 2022) proposes a Convolutional Voxel Vision Transformer (CVVT) for 3D MRI scans, and (Altay et al., 2021) presents a 3D Recurrent Visual Attention Model and an Attention Transformer. Finally, (Xin et al., 2023) introduces ECSnet, a two-stream model combining CNN and Swin Transformer using the 2.5D-subject approach.

This work proposes an innovative framework for selecting informative 2D image slices. It merges annotations from various ROIs (entorhinal cortex, fornix, hippocampus, frontal, parietal, and temporal lobes) from several atlases. To prevent data leakage, the dataset is separated early on at both the subject (volume) and slice levels. A data generator produces multiple perspectives from a volume, using preprocessed datasets with skull-stripping and registration. The network's input diversity is accomplished by combining 2D perspectives, utilizing the cropped ROI. A weighted classification ensemble employs a single Swin Transformer base classifier trained on various datasets to optimize performance and model reliability.

3 MATERIALS AND METHODS

This section presents the dataset used and the framework proposed for building classification models. It also describes the image preprocessing techniques, instance selection, data generator, swin model of transformer classifiers, ensembles, and model performance evaluation used in this work.

3.1 Datasets

This study employs T1-weighted structural MRI images, merging diverse datasets—Alzheimer's Disease Neuroimaging Initiative (ADNI, 2023), Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL, 2023), and Open Access Series of Imaging Studies (OASIS, 2023)—for enhanced model robustness and generalization. The multicenter dataset characterizes subjects using the Clinical Dementia Rating (CDR) scale (ranging from 0 to 3) to determine dementia severity. Cases with CDR zero are Cognitively Normal (CN), while those with CDR one or greater are identified as Alzheimer's Disease (AD) cases. Demographic information is summarized in Table 1.

Table 1: Summary of participant demographics and global clinical dementia rating (CDR) scores of all the study datasets.

Dataset	Class	Subjects	Age	Gender F / M	Total Subjects
ADNI	CN	70	78.63 ± 5.82	34/36	140
	AD	70	78.63 ± 6.50	31/39	
AIBL	CN	70	74.56 ± 5.81	37/33	140
	AD	70	74.87 ± 7.57	43/27	
OASIS	CN	70	69.89 ± 9.38	39/31	140
	AD	70	76.36 ± 9.15	34/36	
ALL-420	CN	210	74.36 ± 8.01	110/100	420
	AD	210	76.62 ± 7.93	108/102	

3.2 Proposed Framework

This framework involves six components: 1) An atlas merging in charge of fusing ROI annotations from multiple atlases; 2) An image preprocessor responsible for obtaining the most informative content of each slice; 3) A dataset builder in charge of avoiding data leakage by early selecting subjects, volumes, and image slice; splitting the data into train, validation, and test sets; 4) A data generator that provides batch-by-batch data to fit in memory; 5) A model builder to train and test different classification models; and 6) A weighted ensemble builder that combines the predictions from two or more models trained on multiple datasets. The proposed framework is shown in Figure 1.

3.2.1 Atlas Merging

In this study, various regions, including the Entorhinal Cortex, Fornix, Hippocampus, and the Frontal, Parietal, and Temporal Lobes, which are associated with cognitive decline in Alzheimer's Disease, are employed. However, due to variations in voxel content within a given ROI across different atlases, our proposal involves the fusion of multiple atlas annotations to address this issue.

The instance selection proposal, which is based on the fusion of ROI annotations from multiple atlases, involves selecting appropriate atlases, merging them, and separating left and right hemisphere structures,

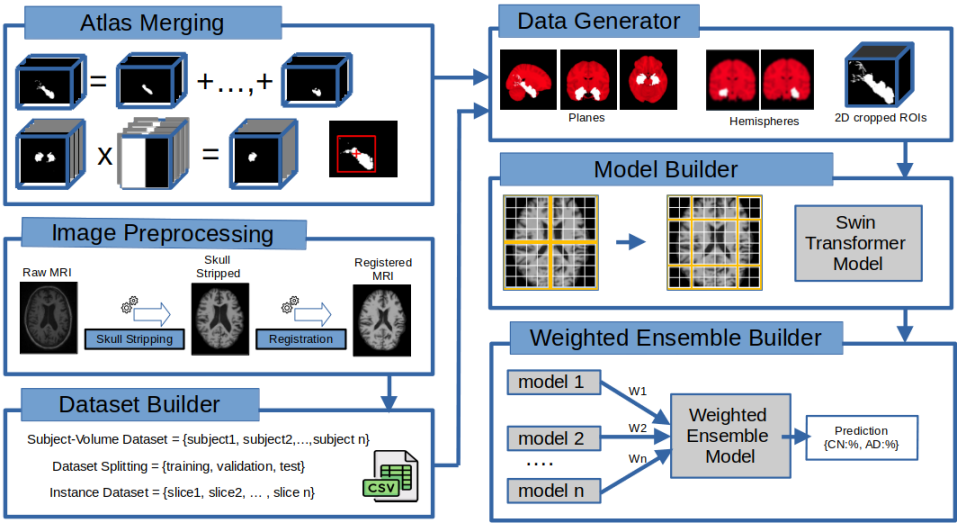

Figure 1: Proposed framework.

outlined as follows:

[Step-1] Atlas Fusion. In this research, a set of atlases ($\boldsymbol{A}$) has been integrated, including JHU DTI-based white-matter, Jülich histological, Talairach, and Harvard-Oxford cortical and subcortical structural atlases (FSL, 2023). These atlases, featuring ROI annotations for both hemispheres, are registered into MNI152 space. The (n) selected atlases ($\boldsymbol{A}$) containing a specific region of interest are merged into a single mean values map ($\boldsymbol{M}^* \in \mathbb{R}^{w \times h \times d}$).

The merged atlas ($\boldsymbol{M}^*$) is binarized, converting to a "1" any voxel having a numerical value greater than zero, as follows: Let ($\boldsymbol{M} \in \mathbb{R}^{w \times h \times d}$) the binarized merged atlas given by:

$$\forall \{i, j, k \in \mathbb{N}, 0 \leq i < h, 0 \leq j < w, 0 \leq k < d\},$$
$$M_{ijk} = \begin{cases} 1, & \text{if} M^*_{ijk} > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where (h) is the slice image height, (w) is the slice image width, and (d) is the number of slices of the $(\boldsymbol{M})$ volume.

[Step-2] Atlas Volume Bounding. The volume ($\boldsymbol{M}^{whd}$) is traversed by each plane in ascending order to obtain the initial slice numbers $(x_1, y_1, \text{and } z_1)$ and in descending order to obtain the final slice numbers $(x_2, y_2, \text{and } z_2)$ for sagittal, coronal, and axial planes respectively.

The slice number (i) of the image ($\boldsymbol{M}_i$) with a mean greater than zero is included in the boundaries list, ($\boldsymbol{A}$) for initial slices and ($\boldsymbol{\Omega}$) for finals, using Equation 2.

$$\begin{aligned} \boldsymbol{A}_{n+1} &= i, \forall \{i \in \mathbb{N}, 0 \leq i < d\}, \text{ and} \\ \boldsymbol{\Omega}_{m+1} &= i, \forall \{i \in \mathbb{N}, d > i \geq 0\}, \\ &\text{if} \left(\frac{\sum_{j=1}^{w} \sum_{k=1}^{h} \boldsymbol{M}_{ijk}}{w \times h} \right) > 0, \\ &\text{with } x_1 = \boldsymbol{A}_1 \text{ and } x_2 = \boldsymbol{\Omega}_1, \end{aligned} \tag{2}$$

where (d) is the number of slice images for a particular plane. The (n) and (m) variables are the number of elements in the $\boldsymbol{A}$ and $\boldsymbol{\Omega}$ lists. (x_1) is the initial slice number and (x_2) the final.

3.2.2 Image Preprocessing

The raw volumes undergo preprocessing, involving skull stripping and registration. The skull-stripped dataset volumes are then registered to the MNI152 T1 template MRI scan, ensuring uniformity in shape, position, and alignment. The resulting scans have normalized intensity, dimensions of $182 \times 218 \times 182$, and a resolution of 1 mm.

3.2.3 Dataset Builder

[Step-1] Subject - Volume Dataset Building. The volumes for each subject are arranged in chronological order based on their visit dates, and the filename of the T1-weighted MRI volume from the last visit is incorporated into the subjects' dataset. This approach guarantees the inclusion of only one volume per subject in the dataset. The subject dataset is balanced using a simple random sampling, including the same (k) number of subjects per class, where (k) is less than or equal to the number of samples from the minority class, thus avoiding class imbalance problems.

[Step-2] Dataset Splitting. The dataset-splitting process ensures reproducible testing and prevents data leakage by first splitting the subject dataset to create independent training, validation, and test sets. The datasets are split randomly, ensuring that an MRI volume per subject is included in only one distribution (training, validation, or test).

[Step-3] Dataset Metadata. Metadata information is generated for each slice containing the ROI, categorized by plane and hemisphere. This metadata includes details such as the group (ADNI, AIBL, or OASIS), volume filename, label, and the position of the ROI's center. The ROI position is represented as a voxel (x, y, z), which includes the slice number (z) and the centroid position (x, y). This centroid position is crucial in cropping the 2D slice image to extract the most informative content from the ROI.

The process involves several steps, including the creation of a list of slice numbers containing the ROI ($\boldsymbol{S}$), the generation of lists for pixel positions along the x and y axes, and the determination of the centroid position (x, y) based on the mode. The ROI center position $(\boldsymbol{\Phi}_i, \boldsymbol{\Lambda}_i)$, is calculated for each slice (i) included in the ROI slice list $\boldsymbol{S}$, using a mode-based approach. The function $(f(\chi))$ obtains the mode from the $\boldsymbol{X}$ and $\boldsymbol{\Upsilon}$ list of pixel positions, as follows:

$$\begin{aligned} &\forall \{i \in \mathbb{N}, i \in \boldsymbol{S}\}, \\ &\boldsymbol{\Phi}_i = f(\boldsymbol{X}_i), \\ &\boldsymbol{\Lambda}_i = f(\boldsymbol{\Upsilon}_i) \end{aligned} \tag{3}$$

where the $f(\chi)$ function gets the mode of the χ list. $(\boldsymbol{\Phi}_i, \boldsymbol{\Lambda}_i)$ represent the x and y mode values for each slice in the ROI slice list $(\boldsymbol{S})$.

3.2.4 Data Generator

This component prepares training data for batch loading by extracting metadata from the instance dataset. It uses the image preprocessor to create an instance image batch in memory based on specified output preferences. The dataset batch size is limited by computational resources like GPU and memory.

3.2.5 Model Builder

This component builds diverse Transformer classification models using datasets that include ROIs, planes, and hemispheres. This task also includes evaluating model performance, as follows:

Swin Transformers. The 2D Swin Transformer (Liu et al., 2021) serves as a weak learner, boosting training speed and diagnostic accuracy. It augments the number of instances and facilitates the use of pretrained models through transfer learning.

Model Performance Evaluation. The Transformer model's performance in Alzheimer's case classification is evaluated through average accuracy and its standard deviation. Assessment is conducted at the subject-patient level by combining classifications from a subject's slice level using majority voting.

3.2.6 Weighted Ensemble Builder

This component selects the highest-accuracy single base classifiers (2D Swin Transformer), trained on diverse datasets, to form a homogeneous weighted ensemble model. The goal is to improve performance, robustness, and reliability in AD classification.

4 EXPERIMENTAL SETUP

The proposed Multiple ROI and Multiple Atlas-Based Instance Selection method is rigorously examined through four experiments.

4.1 Instance Selection

This experiment compares our framework to existing instance selection methods, using a single base classifier (Swin Transformer). The 2D slice image datasets are uniformly derived from the same volumes.
[Method 1] - Percentile Fixed Number. Following (Altay et al., 2021) and (Hu et al., 2023), 96 MRI slices from the middle of all anatomical planes, precisely at the 50th percentile, are chosen to evaluate the instance selection method.
[Method 2] - Multiple ROI and Multiple Atlas-Based Instance Selection (Our Proposal). Atlas annotations were amalgamated to form merged ROIs. The position and the number of slice instances (n) for this experiment vary based on the ROI, anatomical plane, and hemisphere.

4.2 Regions of Interest Datasets

Diverse perspectives are attained by consolidating information from various sources, including ROIs, brain hemispheres, and anatomical planes. This experiment aims to identify the most informative ROI datasets using the proposed framework.

4.3 Weighted Ensemble

Diverse instance datasets, combining data from various sources like ROIs, brain hemispheres, and anatomical planes, train a single 2D Swin Transformer classifier for enhanced diversity. The weighted ensemble model in this experiment selects the most accurate and diverse models, employing a weighted approach for final classification based on each member's performance.

4.4 Performance Comparison

This experiment compares the model performance obtained using the proposed Multiple ROI and Multiple Atlas-based Instance Selection framework with that of state-of-the-art methods. The related works analyzed in this experiment use diverse datasets (ADNI, AIBL, OASIS), input types (2D and 3D), model architectures (Vision Transformer, Swin Transformer, and mixed models combining Convolutional Neural Networks with Transformers), ROIs, and instance selection techniques (Cropping, ROI extraction).

All experiments are reported at the subject level. The volumetric dataset comprises 420 subjects, divided into 70% (300) for training, 15% (60) for validation, and an additional 15% (60) for testing. This subject-volume dataset has been previously randomly selected and partitioned. The instance datasets exclusively contain slices of specific ROIs, planes, and hemispheres. Cropped ROI images consistently maintain a size of 32 x 32 x 3 (width, height, channels). The number of slices per subject varies depending on the specific ROI, plane, and hemisphere.

A single base classifier model, trained on various datasets, is utilized to evaluate their impact on the proposed framework. The 2D Swin Transformer (Lyu et al., 2022) was selected for its effectiveness in image classification.

Hyperparameter optimization was conducted using Hyperband. The values employed to train the proposed framework are as follows: Optimizer Name: Adam, Learning Rate: $1e-04$, Clip Value Rate: 0.5, Dropout: 0.15, Batch Size: 10, and Epochs: 100.

Python libraries NiBabel, TorchIO, PIL, and NumPy preprocess the images. The FreeSurfer tools are used for skull stripping and MRI registration using the MNI152 template. The Keras library is used to build the classification models. All experiments are repeated three times. We carry out the experiments using ten workstations with an Intel Core i9 9900K processor, 32 GB RAM, and 11 GB NVIDIA RTX 2080Ti GPU.

5 RESULTS AND DISCUSSION

Four experiments are conducted to test the proposed framework: I) Compares state-of-the-art instance selection techniques with the proposed Multiple ROI and Multiple Atlas-Based Instance Selection; II) Evaluates the effect of diverse datasets trained with the same base classifier; III) Tests the weighted ensemble model, combining a single base classifier trained on multiple-view datasets; and IV) Compares the models' performance of the proposed method with state-of-the-art related works. The presented experimental results correspond to the model accuracy mean.

5.1 Instance Selection

Since selecting the most informative slices from the original dataset may improve the overall performance of the prediction model (Khan et al., 2019), this experiment compares the proposed instance selection based on a multiple region of interest and multiple atlas with techniques based on percentiles, as shown in Table 2.

Table 2: Accuracy summary from different instance selection techniques, using the same subject-volume dataset (ALL-420). Multi-ROI-Atlas is our proposal.

Technique	Sagittal %	Coronal %	Axial %
Percentile	92.99 ± 0.96	91.43 ± 0.79	91.941 ± 0.79
Multi-ROI-Atlas Hippocampus(Right)	96.67 ± 0.00	94.44 ± 0.79	95.56 ± 0.79

Table 2 shows that all instance selection techniques achieve the highest average accuracy values for the sagittal plane, capturing the most critical information about the regions affected by AD. On the other hand, the average accuracy values for each technique show that our proposed Multiple ROI and Multiple Atlas-Based Instance Selection technique ensures higher accuracy for all three planes (sagittal, coronal, and axial).

5.2 Region of Interest Datasets

In homogeneous ensembles, the main difficulty is generating diversity, despite using the same learning algorithm (Sabzevari et al., 2022). This work experimentally evaluates different MRI datasets trained using a single base classifier.

Table 3 demonstrate that the most informative

ROIs are as follows: a) The Parietal Lobe and the Hippocampus exhibit the highest accuracies, with values of 94.63% and 94.35%, respectively. This is attributed to these ROIs capturing crucial information from regions affected by AD. b) The right hemisphere (93.70%) and the left hemisphere (92.81%) display a marginal difference of 0.89% in their mean accuracy across all planes. This disparity indicates structural asymmetry between the left and right brain hemispheres. Furthermore, c) the sagittal right plane achieves the highest accuracy at 94.26%. This can be attributed to the sagittal plane effectively capturing critical information from regions affected by AD.

These multiple-view dataset models provide variety to build the weighted ensemble model.

Table 3: Summary of accuracy from different ROIs, planes, and hemispheres.

	Left			Right			
ROI	Sagittal	Coronal	Axial	Sagittal	Coronal	Axial	Mean
Entorhinal Cortex	92.78	93.89	95.00	93.89	90.00	95.00	93.43
Fornix	94.44	96.11	93.33	93.33	93.89	92.78	93.98
Frontal Lobe	91.11	93.33	91.67	91.11	90.56	91.67	91.58
Hippocampus	91.67	93.89	93.89	96.67	94.44	95.56	94.35
Parietal Lobe	95.56	97.22	91.67	95.56	95.56	92.22	94.63
Temporal Lobe	88.33	88.33	88.33	95.00	94.44	95.00	91.57
Mean	92.32	93.80	92.32	94.26	93.15	93.71	93.26

5.3 Weighted Ensemble

The optimal ensemble composition is problem-dependent, and determining the number of classifiers for each type remains an open question (Sabzevari et al., 2022). Models from the previous experiments (II) with the highest accuracy and variety are used to create a weighted ensemble, as shown in Table 4.

The contribution of each ensemble member is weighted proportionally to the member's performance to obtain the final classification, creating a weighted ensemble.

Finally, Table 4 shows that combining homogeneous ensemble methods produces accuracy performance results significantly higher (98.33%) than a single learning classification model, providing variety to the weighted ensemble classifier.

5.4 Performance Comparison

Table 5 presents a comparison of the proposed method with the state-of-the-art related works in terms of CN versus AD classification performance.

The experimental results indicate that the proposed Multiple ROI and Multiple Atlas-based Instance Selection method using a weighted ensemble (98.33%) and a single base classifier, such as 2D Swin Transformer, slightly outperforms the state-of-the-art instance selection methods regarding overall results.

This behavior can be attributed to the careful assembly of subject and slice distribution sets, optimal selection of the most significant slice instances, and the most informative content from the ROIs.

Table 4: Summary of accuracy from different model members of the weighted ensemble.

	Ensemble model members accuracy			
Model (ROI)	Plane	Hemisphere	Accuracy %	Weight %
Hippocampus	Sagittal	Right	96.67	8.32
Parietal Lobe	Sagittal	Right	96.67	8.32
Temporal Lobe	Sagittal	Left	96.67	8.32
Temporal Lobe	Sagittal	Right	96.67	8.32
Entorhinal Cortex	Coronal	Left	96.67	8.32
Fornix	Coronal	Left	96.67	8.32
Hippocampus	Coronal	Left	96.67	8.32
Parietal Lobe	Coronal	Left	98.33	8.46
Parietal Lobe	Coronal	Right	96.67	8.32
Entorhinal Cortex	Axial	Right	96.67	8.32
Hippocampus	Axial	Right	96.67	8.32
Temporal Lobe	Axial	Right	96.67	8.32
Ensemble			**98.33**	

Table 5: Performance comparison of the proposed method with other related works for the classification tasks (AD vs. CN).

Author	Model	Dataset	Accuracy (%)
(Lyu et al., 2022)	Vision Transformer	ADNI	96.80
(Hu et al., 2023)	2D CNN+Transformer	ADNI, OASIS	93.56
(Altay et al., 2021)	Vision Transformer	OASIS	91.18
(Xin et al., 2023)	CNN+Swin-Transformer	ADNI, AIBL	93.90
(Huang and Li, 2023)	Swin Transformer	ADNI+AIBL	94.05
(Mora-Rubio et al., 2023)	Vision Transformer	ADNI+OASIS	89.02
Hybrid Ensemble Our Proposal	Swin Transformer	ADNI+AIBL +OASIS	**98.33**

6 CONCLUSIONS AND FUTURE WORK

This work introduces a novel framework for strategically identifying and selecting the most informative 2D image slices based on the fusion of multiple regions of interest annotations from multiple atlases (Multiple ROI and Multiple Atlas-based Instance Selection).

The proposed framework's impact on Transformer-based classification models is experimentally explored. The performance of the 2D Swin Transformer model varies with the dataset (region of interest, plane, and hemisphere), and using 2D slices increases instances, allowing for training with transfer learning or from scratch. The classifications obtained at the slice level are fused to obtain a classification at the subject level. Finally, the weighted ensemble improves the classification model performance and reliability by combining

homogeneous ensemble methods.

For future work, researchers should consider using multiple inputs, mixed data, and 3D transformer model ensembles based on multiple ROIs to enhance the classification model performance and reliability.

REFERENCES

ADNI (2023). Alzheimer's Disease Neuroimaging Initiative. http://adni.loni.usc.edu.

AIBL (2023). Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing. https://aibl.csiro.au.

Altay, F., Sánchez, G. R., James, Y., Faraone, S. V., Velipasalar, S., and Salekin, A. (2021). Preclinical stage alzheimer's disease detection using magnetic resonance image scans. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(17):15088–15097.

Bae, J. B., Lee, S., Jung, W., Park, S., Kim, W., Oh, H., Han, J. W., Kim, G. E., Kim, J. S., Kim, J. H., and Kim, K. W. (2020). Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. *Scientific Reports*, 10(1):1–10.

Castro-Silva., J. A., Moreno-García., M. N., Guachi-Guachi., L., and Peluffo-Ordóñez., D. H. (2022). Instance selection on cnns for alzheimer's disease classification from mri. In *Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods - ICPRAM,*, pages 330–337. INSTICC, SciTePress.

Choi, J. Y. and Lee, B. (2020). Combining of multiple deep networks via ensemble generalization loss, based on mri images, for alzheimer's disease classification. *IEEE Signal Processing Letters*, 27:206–210.

FSL (2023). Templates and Atlases included with FSL), https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases.

Hu, Z., Li, Y., Wang, Z., Zhang, S., and Hou, W. (2023). Conv-swinformer: Integration of cnn and shift window attention for alzheimer's disease classification. *Computers in Biology and Medicine*, 164.

Huang, Y. and Li, W. (2023). Resizer swin transformer-based classification using smri for alzheimer's disease. *Applied Sciences (Switzerland)*, 13.

Khan, N. M., Abraham, N., and Hon, M. (2019). Transfer Learning with Intelligent Training Data Selection for Prediction of Alzheimer's Disease. *IEEE Access*, 7:72726–72735.

Li, C., Cui, Y., Luo, N., Liu, Y., Bourgeat, P., Fripp, J., and Jiang, T. (2022). Trans-resnet: Integrating transformers and cnns for alzheimer's disease classification. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9992–10002.

Lyu, Y., Yu, X., Zhu, D., and Zhang, L. (2022). Classification of alzheimer's disease via vision transformer: Classification of alzheimer's disease via vision transformer. *ACM International Conference Proceeding Series*, pages 463–468.

Mora-Rubio, A., Bravo-Ortíz, M. A., Arredondo, S. Q., Torres, J. M. S., Ruz, G. A., and Tabares-Soto, R. (2023). Classification of alzheimer's disease stages from magnetic resonance images using deep learning. *PeerJ Computer Science*, 9.

OASIS (2023). Open Access Series of Imaging Studies. http://www.oasis-brains.org.

Pan, D., Luo, G., Zeng, A., Zou, C., Liang, H., Wang, J., Zhang, T., Yang, B., and the Alzheimer's Disease Neuroimaging Initiative (2022). Adaptive 3dcnn-based interpretable ensemble model for early diagnosis of alzheimer's disease. *IEEE Transactions on Computational Social Systems*, pages 1–20.

Qin, Z., Liu, Z., Guo, Q., and Zhu, P. (2022). 3d convolutional neural networks with hybrid attention mechanism for early diagnosis of alzheimer's disease. *Biomedical Signal Processing and Control*, 77:103828.

Sabzevari, M., Martínez-Muñoz, G., and Suárez, A. (2022). Building heterogeneous ensembles by pooling homogeneous ensembles. *International Journal of Machine Learning and Cybernetics*, 13(2):551–558.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need.

Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., and Colliot, O. (2020). Overview of classification of Alzheimer's disease. *Medical Image Analysis*, 63.

WHO (2023). World Health Organization. https://www.who.int/news-room/fact-sheets/detail/dementia.

Xin, J., Wang, A., Guo, R., Liu, W., and Tang, X. (2023). Cnn and swin-transformer based efficient model for alzheimer's disease diagnosis with smri. *Biomedical Signal Processing and Control*, 86.

Young, S., Abdou, T., and Bener, A. (2018). Deep super learner: A deep ensemble for classification problems. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10832 LNAI:84–95.

Yu, J. and Lee, T. M. (2020). Verbal memory and hippocampal volume predict subsequent fornix microstructure in those at risk for alzheimer's disease. *Brain Imaging and Behavior*, 14:2311–2322.

Zaabi, M., Smaoui, N., Derbel, H., and Hariri, W. (2020). Alzheimer's disease detection using convolutional neural networks and transfer learning based methods. In *2020 17th International Multi-Conference on Systems, Signals & Devices (SSD)*, pages 939–943.

Zhang, Z. and Khalvati, F. (2022). Introducing vision transformer for alzheimer's disease classification task with 3d input.