Research article

Novel hippocampus-centered methodology for informative instance selection in Alzheimer's disease data

Juan A. Castro-Silva [a,b,*], María N. Moreno-García [a], Lorena Guachi-Guachi [c], Diego H. Peluffo-Ordóñez [d,e,f]

[a] *Universidad de Salamanca, Salamanca, Spain*
[b] *Universidad Surcolombiana, Neiva, Colombia*
[c] *Universidad Internacional del Ecuador, Quito, Ecuador*
[d] *College of Computing, Mohammed VI Polytechnic University, Lot 660, Hay Moulay Rachid Ben Guerir, 43150, Morocco*
[e] *SDAS Research Group (https://sdas-group.com/), Ben Guerir 43150, Morocco*
[f] *Faculty of Engineering, Corporación Universitaria Autónoma de Nariño, Pasto 520001, Colombia*

ARTICLE INFO

ABSTRACT

The quantity and quality of a dataset play a crucial role in the performance of prediction models. Increasing the amount of data increases the computational requirements and can introduce negligible variations, outliers, and noise. These significantly impact the model performance. Thus, instance selection techniques are crucial for building prediction models with informative data, reducing the dataset size, improving performance, and minimizing computational costs. This study proposed a novel methodology for identifying the most informative two-dimensional slices derived from magnetic resonance imaging (MRI) to study Alzheimer's disease. The efficacy of our methodology was attributable to a hippocampus-centered analysis using data from multiple atlases. The methodology was evaluated by constructing convolutional neural networks to identify Alzheimer's disease, using a consolidated dataset constructed from three standard datasets: Alzheimer's Disease Neuroimaging Initiative, Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing, and Open Access Series of Imaging Studies. The proposed methodology demonstrated a commendable subject-level classification accuracy of approximately (95.00%) when distinguishing between normal cognition and Alzheimer's.

1. Introduction

Alzheimer's disease (AD), the leading cause of dementia in older adults, affects more than 55 million people worldwide. This progressive brain disorder causes nerve cell death, leading to a significant reduction in brain volume and affecting nearly all brain functions [1]. Medical imaging techniques such as magnetic resonance imaging (MRI), positron emission tomography (PET), and diffusion tensor imaging (DTI) are crucial for visualizing brain structures for AD diagnosis and treatment.

A region of interest (ROI) refers to a specific anatomical structure or area within an image selected for further analysis or processing based on its relevance. Key regions affected by AD include the entorhinal cortex, fornix, and hippocampus. This results in cognitive impairment [2–4]. Hippocampal atrophy, which is a critical MRI marker of AD progression, has been studied extensively [5–8].

Deep learning methods, particularly convolutional neural networks (CNNs), have demonstrated potential for detecting brain structural changes via MRI and analyzing entire images or specific ROIs using two-dimensional (2D) or three-dimensional (3D) models. However, the selection of informative instances to train these models remains challenging. The challenge in instance selection lies in the optimization of the selection from large datasets to improve the model performance and efficiency. Balancing the dataset size reduction, model effectiveness, and computational costs remains difficult [9,10].

The existing selection methods based on a fixed number of instances are plagued by significant limitations. The performance of the model depends significantly on the location and number of slices per volume [11]. The entropy technique sorts slices by using entropy values and retains the top slices [12]. This is unsuitable for 3D CNNs because it results in the loss of 3D information and includes dispersed slices from regions not associated with AD. Similarly, the slice percentile position method, which selects slices based on their percentile position within the volume [11], cannot determine the ROI accurately and often selects slices from irrelevant regions. Including excessive number of non-informative slices increases the computational cost and noise, whereas too few slices may exclude vital information necessary for AD diagnosis.

This study proposed a groundbreaking methodology to address these limitations by selecting the most informative 2D image slices. In contrast to existing methods, the proposed approach merged hippocampus-containing regions from multiple atlases and calculated an adjusted centroid (x, y) to capture the most informative content. This innovative technique preserved 3D information and leveraged context by selecting only the adjacent slice instances related to the ROI. The number of instances selected was dependent on the plane, hemisphere, and ROI, thereby providing a more accurate and representative data selection.

Our study employed a hybrid ensemble that merged homogeneous and heterogeneous methods. This ensemble incorporated the most diverse and highest-accuracy CNN models to evaluate the proposed instance selection method. Homogeneous ensemble methods employ a single-base classifier with diverse training data to assess the versatility, whereas heterogeneous ensemble methods use different classifiers with the same data to evaluate the dataset quality.

This proposed method comprised three main phases.

- Data preparation (Phase 1): This phase involved splitting the dataset into training, validation, and test sets. In addition, raw MRI volumes were subjected to skull-stripping and registration preprocessing to create two refined datasets.
- Instance selection (Our proposal) (Phase 2): This phase included slice instance selection to identify slices containing the ROI (hippocampus) and ROI position localization to calculate the centroid position (x, y) based on the mode.
- Instance selection validation (Phase 3): This phase included batch data generation, training and testing diverse classification models, and hybrid ensemble integration. Predictions from accurate models trained on varied datasets and algorithms were combined.

The significant contributions of this study are summarized as follows:

1. Multi-atlas ROI-based instance selection: The hippocampus ROI annotations from multiple atlases were integrated, thereby retaining informative slices and ensuring representativeness.
2. ROI content extraction: This method utilized the statistical mode to determine the most informative content by adjusting the centroid (x, y) position to obtain the content of the ROI for precise 2D slice cropping.

The proposed methodology aimed to enhance the diagnostic accuracy, reduce computational costs, and improve the efficiency of AD tasks, thereby offering significant advancements in the field and potentially improving patient outcomes.

2. Related works

Most proposed approaches for image instance selection for AD classification differ in terms of the number of slices selected and the technique used to obtain the most informative slices or discard the least informative slices. For example, in [13], the interclass variance criterion was used to select a single slice from 3D volumetric data. In [14], skull stripping was performed on raw MRIs using a trained U-Net model, and 3D images were cropped and resized to $64 \times 64 \times 64$ pixels for the proposed model. In [15], the initial and final slices were discarded and the 3D MRI scans were downsized to $96 \times 96 \times 96$. Consequently, randomly selected slices were used to train the 2D CNN model.

In [16,17,12,18], the authors used entropy calculation for each image slice, sorted them in descending order of entropy value, and maintained the top fixed number of slices (8, 16, and 32). In [11], the percentile position of an image slice was used, including a fixed number of instances (32).

ROI extraction is a critical image processing task. For instance, in [19], an MRI scan was divided into square blocks with dimensions of 32×32 pixels. Consequently, only the blocks that included the hippocampus were extracted as ROIs. Other studies, such as [20–22], extracted ROI-based patches from different brain regions, including the hippocampus, amygdala, and insula, to create an ensemble classifier. In [23], a classification framework used ROIs and landmarks to avoid large attribute vectors. Thereafter, they were classified based on their distance to a 3D atlas with gray matter, white matter, and cerebrospinal fluid attributes. In [24], a

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 [45]	CN	70	78.63 ± 5.82	34/36	140
	AD	70	78.63 ± 6.50	31/39	
AIBL [46]	CN	70	74.56 ± 5.81	37/33	140
	AD	70	74.87 ± 7.57	43/27	
OASIS [47]	CN	70	69.89 ± 9.38	39/31	140
	AD	70	76.36 ± 9.15	34/36	
Merged	CN	210	74.36 ± 8.01	110/100	420
	AD	210	76.62 ± 7.93	108/102	

method using Explainable AI with Grad-Cam and a 3D CNN automatically detected patient-specific ROIs rather than fixed ROIs for all patients.

In most previous studies, MR image preprocessing involved (a) skull stripping, segmentation, and nonlinear registration using FreeSurfer [25,26]; (b) brain extraction and tissue segmentation using FSL [27,28]; and (c) realignment, spatial normalization, and smoothing using SPM [29,30]. In [31], noise was removed using a curvelet transform with an adaptive threshold. This was followed by hybrid histogram equalization and segmentation using a multiscale pooling residual autoencoder to facilitate feature extraction.

Skull stripping eliminates non-brain tissues from MR images, which is a crucial step in the automated evaluation of neurodegenerative diseases [32]. Medical image registration, which involves linear and nonlinear registration, calculates the coordinate transforms between images to align them with a predefined coordinate system, such as an anatomical atlas (MNI152 Template) [33].

Deep Learning methods such as CNNs have been used to solve AD classification problems. Recent studies have utilized CNN architectures such as DenseNet [34,28,30,25], InceptionV3 [35,30], InceptionV4 [36], MobileNetV2 [35], ResNet [30,37,29,38], Xception [30], and certain custom CNN [25,17,26,35] for AD diagnosis. Moreover, the studies cited in [39–42] demonstrated that the prediction outcomes of deep neural networks (DNNs) were significantly superior to those of conventional machine learning approaches.

Various approaches have been proposed that utilize CNNs to analyze MRI images for AD diagnosis. These include the use of 2D slices from 3D MRI volumes [29,17,38], ROI-based CNNs that focus on specific informative brain regions [19–22], 3D patches extracted from MR images [26,28], and 3D subject-level CNNs that simultaneously analyze entire MRI images [27,26].

Ensembles of deep CNN models that combined multiple weak learners to form a single strong learner were used by [12,43]. Ensemble deep learning (EDL) improves the accuracy and robustness compared to individual CNNs [44].

Several previous studies on instance selection for Alzheimer's disease diagnosis using MRI have been plagued significant limitations. Many methods cannot determine the ROI accurately, often selecting dispersed, nonsequential slices from irrelevant regions that are not associated with AD. Entropy-based selection cannot leverage the context from adjacent slices and is unsuitable for CNN 3D models owing to the loss of 3D information. Further, ROI-based selection often employs a single atlas and cannot capture the most informative ROI content for precise 2D slice cropping. In addition, these methods often include excessive non-informative slices, which increases the computational cost and noise, whereas including too few slices may exclude vital information for AD diagnosis.

3. Datasets

In this study, we used T1-weighted structural MRI (3D-MRI) from three prominent datasets for AD research: the Alzheimer's Disease Neuroimaging Initiative (ADNI) [45], Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) [46], and Open Access Series of Imaging Studies (OASIS) [47]. The subjects in these datasets were characterized using the Clinical Dementia Rating (CDR) scale, where 0 indicates cognitively normal (CN) and values 1–3 indicate different stages of Alzheimer's disease (AD) [48]. Table 1 lists the demographic information for these datasets.

3.1. Merged dataset

To mitigate the challenge of performance reduction in prediction models across diverse demographic groups, we encompassed various study populations across different age groups (young and older people) by consolidating ADNI, AIBL, and OASIS datasets. This yielded a merged dataset of 420 instances. Each dataset contributed 70 volumes per class (3 datasets × 70 samples × 2 classes).

4. Proposed methodology

This methodology aimed to identify the most informative images by analyzing the pixels belonging to the hippocampus, which represents the ROI. As illustrated in Fig. 1, our methodology began by preprocessing the images to remove non-informative content and align them appropriately. Next, our instance selection method identified slices containing the defined ROI and determined the centroid position (x, y) based on this mode. Finally, the instance selection method was validated using 2D CNN, 3D CNN, and Hybrid ensemble models.
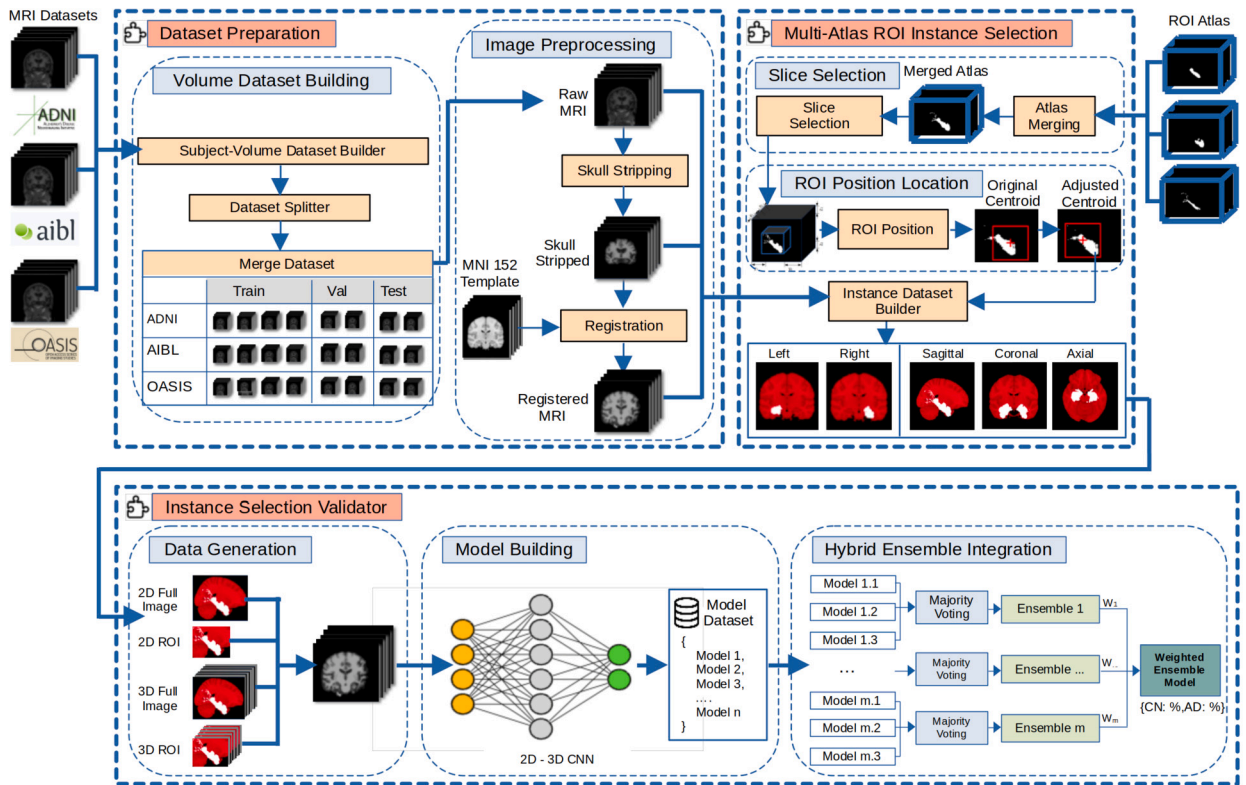
Fig. 1. Proposed methodology.

4.1. Data preparation

This phase selected one volume per subject and split the dataset into training, validation, and testing datasets. The selected raw MRI volumes were subjected to skull stripping and registration preprocessing to create two new datasets.

4.1.1. Volume dataset building

Volumes per subject were sorted by visit date, with only the last visit included in the dataset to ensure one volume per subject. The datasets were then randomly split to ensure reproducible tests and prevent data leakage, with each MRI volume per subject exclusively included in one distribution (training, validation, or testing). The subject distribution was balanced using undersampling. This ensured the same number of subjects per class (k), where k was less than or equal to the number of samples from the minority class, thus addressing class imbalance issues.

4.1.2. Image preprocessing

The Image preprocessing step produced the following three datasets:

- Raw dataset: This included the raw MRI volumes selected in the previous "Volume Dataset Building" step, sourced directly from ADNI, AIBL, or OASIS repositories without preprocessing. The raw dataset varied in volume size.
- Skull-stripped dataset: Raw MRI volumes were subjected to skull stripping as part of the preprocessing step. This yielded a new preprocessed dataset wherein skull-stripped MRI scans were standardized to dimensions of $256 \times 256 \times 256$.
- Registered dataset: Following skull stripping, volumes were registered to the MNI152 T1 template, creating a new preprocessed dataset standardized to $182 \times 218 \times 182$ with a spatial resolution of 1 mm.

4.2. Proposed instance selection method centered on hippocampus content using multi-atlas

This phase orchestrated the creation of diverse datasets by amalgamating preprocessing methods, planes, and hemispheres. It employed multiple atlases to select instances based on ROI boundary positions, retaining only slices featuring the hippocampus. The resulting instance dataset files included metadata such as the MRI filename, slice number, ROI centroid position, and class.

Considering the hippocampus's association with cognitive decline, this study centered its analysis on this region, leveraging atlases aligned with the MNI152 standard template. Images querying these atlases were registered in the MNI152 space to ensure consistency. Multiple annotations were harmonized to reconcile voxel disparities in hippocampal representations across atlases.
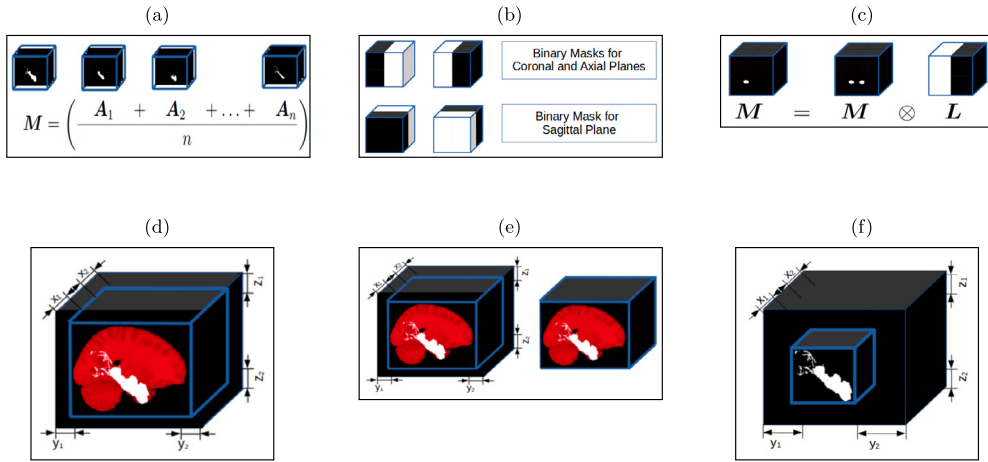
(a)

(b)

(c)

(d)

(e)

(f)

Fig. 2. Steps of the proposed instance selection method centered on hippocampus content using multi-atlas: (a) Atlas merge process, (b) Left and right hemisphere binary masks for each plane, (c) Atlas hemisphere separation, (d) MNI152 template and merged atlas boundaries, (e) Merged atlas trimmed, (f) ROI boundaries.

A slice instance selection method grounded in ROI standardized the image space, expressing the initial and final slice numbers as percentiles. This process ensured a clear and understandable standardization of our data.

The instance selection proposal based on the fusion of ROI (Hippocampus) annotations from multiple atlases, Multi-Atlas ROI, involves the following two steps.

4.2.1. Slice selection

The selection of the slice instances based on the ROI involved selecting appropriate atlases, merging them, and detecting the atlas boundaries to determine the range of slices wherein the ROI (hippocampus) was present.

1. **Atlas selection:** The following atlases, transformed into MNI152 space and containing hippocampus annotations, were used in this study [49]. 1) JHU DTI-based white-matter atlases from Johns Hopkins University. 2) Jülich histological atlas from the Research Center Jülich. Talairach atlas (Daemon Labels) from UTHSCSA, Texas. 4) Harvard-Oxford cortical and subcortical atlases from the Harvard Center for Morphometric Analysis.

2. **Merged atlas:** The (n) selected atlases (A) containing the ROI were merged into a single mean values map ($M \in \mathbb{R}^{w \times h \times d}$), as shown in Fig. 2a.

   The merged atlas (M) was binarized, converting to a "1" any voxel having a numerical value greater than zero, as shown in the Equation (1):

   $$\forall \{i, j, k \in \mathbb{N}, 0 \leq i < h, 0 \leq j < w, 0 \leq k < d\},$$

   $$M_{ijk} = \begin{cases} 1, & \text{if } M_{ijk} > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

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

   The left and right hemisphere structures were merged separately for each orthogonal plane, resulting in 6 merged atlases. Fig. 2b shows the binarized masks (L) and (R) created to ensure that a merged atlas only included the pixels of the left or right hemisphere, using the Equation (2):

   $$\forall \{i, j, k \in \mathbb{N}, 0 \leq i < h, 0 \leq j < w, 0 \leq k < d\},$$

   $$L_{ijk} = \begin{cases} 1, & \text{if } i < \frac{w}{2}, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

   $$R_{ijk} = 1 - L_{ijk},$$

   where (L) and (R) are the binarized masks for the left and right hemispheres, respectively, (h) is the slice image height, (w) is the slice image width, and (d) is the number of slices of the (M) merged atlas volume.

   The merged atlas volume (M) for a specific plane and hemisphere was multiplied by the corresponding binary mask (L or R), as shown in Fig. 2c.

3. **Get volume boundaries:** The volume (M^{whd}) was traversed by each plane in ascending order to obtain the initial slice numbers ($x_1, y_1,$ and z_1). The same was done in descending order to obtain the final slice numbers ($x_2, y_2,$ and z_2) for sagittal, coronal, and axial planes respectively. The results are shown in Fig. 2d.

Table 2
Summary of Hippocampus percentile slice position per plane and hemisphere.

| | Percentiles by hemisphere | | | |
| | Left | | Right | |
Plane	x_1	x_2	x_1	x_2
Sagittal	19.01	48.59	50.70	87.32
Coronal	7.26	69.83	8.94	79.33
Axial	5.00	75.00	15.13	49.34

The slice number (i) of the image (M_i) with a mean greater than zero was included in the boundaries list, (A) for initial slices and (Ω) for finals, using Equation (3).

$$A_{n+1} = i, \forall \{i \in \mathbb{N}, 0 \le i < d\}, \text{ and}$$

$$\Omega_{m+1} = i, \forall \{i \in \mathbb{N}, d > i \ge 0\},$$

$$\text{if} \left(\frac{\sum_{j=1}^{w} \sum_{k=1}^{h} M_{ijk}}{w \times h} \right) > 0, \tag{3}$$

$$\text{with } x_1 = A_1 \text{ and } x_2 = \Omega_1,$$

where (d) is the number of slice images for a particular plane. Further, (n) and (m) are the number of elements in the A and Ω lists, respectively, and (x_1) and (x_2) are the initial and final slice numbers, respectively.

4. **MNI152 template and merged atlas trimming:** The whole-brain MNI152 template was trimmed to correct and standardize the brain area by removing black pixel outliers. As the merged atlas and the MNI152 template volumes shared a common space, with the same shape (182 x 218 x 182) and resolution (1 mm), the merged atlas was trimmed to match the MNI152 template boundaries, as shown in Fig. 2e.
5. **Percentile ROI boundaries:** Equation (3) was used to obtain the merged atlas volume boundaries, as shown in Fig. 2f.
   The merged atlas ROI boundaries are expressed in percentile (p), using the Equation (4):

$$p = \left(\frac{x}{w} \right) \times 100, \tag{4}$$

where (x) is the slice number and (w) is the total number of slices per plane. Table 2 presents the percentile ROI boundaries per plane and hemisphere.

6. **Slice subsets:** An MR image has several slices. The slice subsets were built using the slice positions of the ROI, expressed in percentiles, as presented in Table 2.

The initial α and final ω slice numbers were obtained using the equation given by $\alpha = \varphi(\beta)$ and $\omega = \varphi(\gamma)$; where, $\varphi(z) = n \times (z/100)$. Here, n is the total number of slices, and (β, γ) are the initial and final slice positions, respectively, for a specific plane and hemisphere, expressed as percentiles.

The subset (S) includes the slice instances from H_α to H_ω, using the Equation (5):

$$H_i = i, \forall \{i \in \mathbb{N}, 0 \le i < d\},$$

$$S = \{H_\alpha, H_{\alpha+1}, H_{\alpha+2}, \cdots, H_{\omega-1}, H_\omega\}. \tag{5}$$

where (H) is the list, (d) is the total of slice numbers for a specific plane and hemisphere, and (S) is the slice numbers containing the ROI. The initial and final slice positions where the ROI is present are assigned to (α) and (ω).

4.2.2. ROI position localization

The ROI position includes the slice number and centroid position (x, y). The centroid position was used to crop the 2D slice images and extract the most informative content from the ROI. The process involved generating a list of slice numbers containing the hippocampus region, creating lists of pixel positions for x and y, removing any outlier values from the position lists, obtaining the centroid position (x, y) based on the mode, and converting the centroids into percentiles to consider varying volumes.

1. **Slice selection:** Equation (5) was used to obtain the list of slice numbers (S) where the ROI (Hippocampus) is present.
2. **Pixel position list:** A list of x (X) and y (Y) position values is created for each 2D slice included in (S). Here, only the pixels with a value of "1" were added, as shown in the Equation (6):

$$\forall \{i, j, k \in \mathbb{N}, i \in S, 0 \le j < h, 0 \le k < w\},$$

$$\left\{ (X_{n+1} = j, Y_{m+1} = k) \quad \text{if } M_{ijk} = 1 \right. \tag{6}$$

where (i) is the slice number included in the ROI slice list (S), (h) is the slice image height, (w) is the slice image width, (X) and (Υ) lists store the x and y positions of the pixels included in the ROI, respectively, and (n) and (m) are the number of elements in the (X) and (Υ) lists, respectively.

3. **Outliers removing:** The outlier values from the x and y pixel position lists (X, Υ) were removed using a percentile-based threshold, excluding the pixel positions with values less than the 25th and greater than the 75th percentile, using the Equation (7):

$$\forall \{i, j, k \in \mathbb{N}, i \in S, 0 \leq j < h, 0 \leq k < w\},$$
$$\alpha_{25} = f\left(X_i, 25\right), \alpha_{75} = f\left(X_i, 75\right),$$
$$\omega_{25} = f\left(\Upsilon_i, 25\right), \omega_{75} = f\left(\Upsilon_i, 75\right), \tag{7}$$
$$X_i = \left\{X_i - \{j\}, \forall j \mid j < \alpha_{25} \vee j > \alpha_{75}\right\}$$
$$\Upsilon_i = \left\{\Upsilon_i - \{k\}, \forall k \mid k < \omega_{25} \vee k > \omega_{75}\right\}$$

where i is the slice number included in ROI slice list S, h is the slice image height, w is the slice image width, the function $f(\chi, \eta)$ obtains the η percentile from χ list, α and ω values store the percentile threshold position for x and y lists, respectively, and the X and Υ lists include the non-outlier values of x and y, respectively.

4. **Mode-based centroid:** The ROI center position (Φ_i, Λ_i), was calculated for each slice (i) included in the ROI slice list S, using a mode-based approach. The function ($f(\chi)$) obtained the mode from the X and Υ list of pixel positions, as shown in the Equation (8):

$$\forall \{i \in \mathbb{N}, i \in S\},$$
$$\Phi_i = f\left(X_i\right), \tag{8}$$
$$\Lambda_i = f\left(\Upsilon_i\right)$$

where the $f(\chi)$ function gets the mode of the χ list. (Φ_i, Λ_i) represent the x and y mode values for each slice in the ROI slice list (S), respectively, as shown in the Equation (9).

The (Φ_i, Λ_i) mode values were adjusted using the Algorithm 1.

$$\Phi_i = g\left(\Phi_i, X_i, \delta, \lambda\right), \tag{9}$$
$$\Lambda_i = g\left(\Lambda_i, \Upsilon_i, \delta, \lambda\right)$$

where the function $g\left(\Phi_i, X_i, \delta, \lambda\right)$ adjusts the mode-based centroid values (x, y) by including most of the ROI area in the resulting square image. The desired ROI square size (δ) in this study was 48 pixels for the coronal and axial planes and 96 pixels for the sagittal plane. The λ value corresponded to the slice image width or height in pixels.

Algorithm 1 Adjusted mode.

Input: ϕ // the mode of the x or y values
    X // list of x or y positions
    δ // desired ROI square size in pixels
    λ // the image width or height in pixels
1: $\kappa \leftarrow max(X) - min(X)$ // the ROI width or height
2: $\psi \leftarrow \delta/2$ // the half of the desired ROI square size
3: **if** $\kappa \leq \delta$ **then**
4:     $\phi \leftarrow min(X) + int(\kappa/2)$ // if the ROI size is less than the square size
5: **else**
6:     $\alpha \leftarrow min(X) - (\phi - \psi)$ // the left or top difference
7:     $\beta \leftarrow max(X) - (\phi + \psi)$ // the right or bottom difference
8:     $\gamma \leftarrow int((\alpha + \beta)/2)$ // the x or y difference
9:     $\phi \leftarrow \phi + \gamma$ // adjust the center position
10: **end if**
11: **if** $(\phi - \psi) < 0$ **then**
12:     $\phi \leftarrow \psi$ // if the left or top position is less than 0
13: **end if**
14: **if** $(\phi + \psi) > \lambda$ **then**
15:     $\phi \leftarrow \lambda - \psi$ // if the right or bottom position is greater than the image width or height
16: **end if**
Output: ϕ // the adjusted mode

Fig. 3 shows an example of the adjustment of the mode for an ROI atlas slice. The left image 3a presents the original ROI center position. The right image 3b shows the movement of the x and y positions to the left and top to view most of the ROI.

5. **ROI position:** Owing to the different sizes of the dataset volumes, ROI center positions were expressed in percentiles relative to each other. The ROI center position (X_i, Υ_i), was calculated for each slice (i) included in the ROI slice list S, using the Equation (10):

Fig. 3. ROI centroid positions: (a) Original centroid and (b) Adjusted centroid.

$$\forall \{i \in \mathbb{N}, i \in \boldsymbol{S}\},$$

$$\boldsymbol{X}_i = \left(\frac{\boldsymbol{\Phi}_i}{w}\right) \times 100,$$

$$\boldsymbol{\Upsilon}_i = \left(\frac{\boldsymbol{\Lambda}_i}{h}\right) \times 100,$$

(10)

where (i) is the slice number included in the ROI slice list ($\boldsymbol{S}$), (h) is the slice image height, (w) is the slice image width. ($\boldsymbol{X}_i$) and ($\boldsymbol{\Upsilon}_i$) are the lists of ROI center positions expressed in percentiles.

4.3. Instance selection validation

This phase included a) data generation, b) model building, and c) hybrid ensemble integration.

4.3.1. Data generation:
Data generation began by reading the instance dataset metadata, including the volume filename, slice number, ROI centroid, and class. Data were then loaded batch-by-batch to manage the memory constraints. Depending on the image size and desired CNN input type, the images were either resized to obtain the entire 2D image or cropped to extract the hippocampus, which was the ROI used in this study.

4.3.2. Model building:
This step involved the construction of multiple CNN classification models using diverse CNN input types and architectures. The building process integrated various datasets with different pre-processing types, planes, and hemispheres. In addition, it encompassed hyperparameter optimization and the evaluation of model performance, as detailed below.

1. Convolutional Neural Network (CNN): 2D convolutional filters analyzed each slice independently, thereby missing contextual information from adjacent slices. This independent analysis may result in 2D CNNs losing essential 3D spatial data. In contrast, 3D CNNs extract features from consecutive 2D feature maps at three scales. Consequently, they capture feature dimensions and spatial information [50].
   The 2D CNNs DenseNet (121) [51], EfficientNetV2 (M) [52], Inception ResNet V2 [53], and ResNet [54] were employed to increase the number of instances and facilitate the use of pre-trained models with transfer learning from ImageNet. This study proposed a custom 3D CNN model called Custom3D adapted from [55].
2. Hyperparameter optimization: The effectiveness of the machine learning algorithms is reliant on the identification of the optimal hyperparameters. Hyperband, an advanced technique, optimizes random search by allocating predefined resources to randomly sampled configurations and implementing early stopping [56]. This study employed it to select the optimal hyperparameters for training the AD classification models.
3. Model performance evaluation: The receiver operating characteristic curve (ROC Curve) illustrates the trade-off between the true positive rate (TPR) and false positive rate (FPR) at different classification thresholds. TPR (sensitivity) indicates the proportion of actual positives correctly identified by the model. Whereas, FPR shows the proportion of actual negatives incorrectly classified as positives. The overall performance of the binary classification model was quantified using the area under the curve (AUC) metric, which represents the area under its ROC curve and the accuracy metric's average and standard deviation.
   We proposed the evaluation of the performance of the model at the subject-patient level. This is because, in 2D CNN types, multiple images of a subject's 3D volume were used in the training process. The classifications obtained from a subject's slice level were combined using majority voting.

4.3.3. Hybrid ensemble integration:
The versatility of the proposed instance selection method and the quality of the generated datasets were tested using a hybrid ensemble model. Top-performing and diverse base classifiers trained on various datasets and models were selected to assess their performance and improve AD classification. A weighted ensemble was then formed by proportionally weighing each member's contribution to the predictions based on their confidence or performance, following the methodology outlined in [57].

As shown in Fig. 4a, our weighted ensemble assigned weights to models based on their accuracy. The process is outlined as follows. 1) A set containing the accuracy of each model was created. $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_{m-1}, \mathbf{A}_m\}$, where ($m$) denotes the number
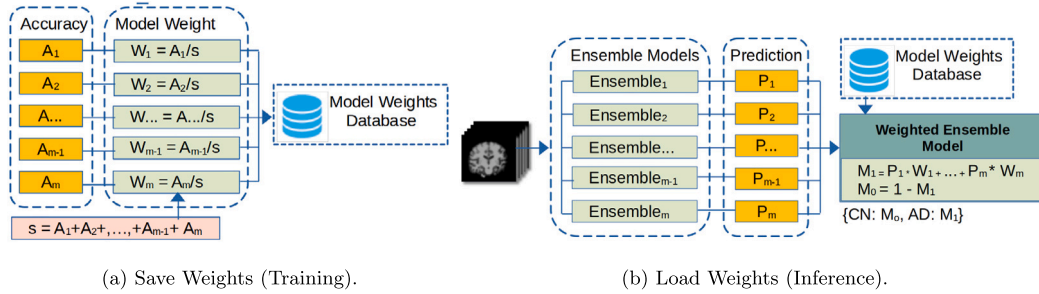
(a) Save Weights (Training).   (b) Load Weights (Inference).

Fig. 4. Save and Load Weights.

of model members. 2) The sum of the accuracies was calculated $s = \sum_{i=1}^{m} A_i$. 3) The weights of each model were calculated $\mathbf{W} = \{\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{m-1}, \mathbf{W}_m\}$, where $\mathbf{W}_i = \frac{A_i}{s}$, 4) Finally, the model weight database ($\mathbf{W}$) was saved onto the disk.

During the inference, as depicted in Fig. 4b, the hybrid ensemble model predictions were weighted by combining the predictions of the model members multiplied by their respective weights. The steps are outlined as follows. 1) A set of predictions ($\mathbf{P}$) were created from (m) model members (CN = 0, AD = 1). $\mathbf{P} = \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_{m-1}, \mathbf{P}_m\}$, 2) The weighted sum of predictions for each class was calculated $\mathbf{M}_1 = \sum_{i=1}^{m} (\mathbf{P}_i \times \mathbf{W}_i)$ $\mathbf{M}_0 = 100 - \mathbf{M}_1$ and 3) The predicted label (ψ) was determined. $\psi = \begin{cases} 1 & \text{if } \mathbf{M}_1 > \mathbf{M}_0, \\ 0, & \text{otherwise.} \end{cases}$

This approach ensured that the final prediction leveraged the strengths of the most accurate models. This enhanced the robustness and accuracy of AD classification.

4.4. Implementation details

All the evaluated models were trained sequentially on 10 workstations equipped with an Intel Core i9 9900K processor, 32 GB RAM, and an 11 GB NVIDIA RTX 2080Ti GPU. Our instance selection methodology was implemented using Python software.

5. Experimental setup

5.1. Instance selection

This experiment compared the performance of the proposed methodology with that of state-of-the-art instance-selection methods. A single-base classifier (DenseNet121) was trained to evaluate the impact of the instance selection methods on the model performance. The 2D slice image datasets were created using the same volume.

1. Entropy fixed number: The entropy-based instance selection method [16,17,12,18] involved calculating the image entropy of each slice, sorting them in descending order based on their entropy values, and retaining the top 32 image slices.
2. Percentile fixed number: The instance selection method based on slice percentile position [11] selected 32 instances from various positions in a percentile distribution of the entire volume. The percentile positions per anatomical plane used were sagittal = 65, coronal = 35, and axial = 35.
3. Multiple atlas ROI-based instance selection (our proposal): The ROI (hippocampus) annotations from multiple atlases were fused. The 2D image slices belonging to the merged ROI were included in the dataset. The number of slice instances was variable and dependent on the anatomical plane.

5.2. Datasets

This experiment assessed the influence of the proposed methodology on various datasets trained using a single-base classifier (DenseNet121), thereby introducing diversity to the homogeneous ensemble. Three distinct MRI datasets were generated from identical volumes using skull-stripped, skull-stripped registered image pre-processing techniques, and raw data. Multiple perspectives from these datasets were acquired by amalgamating the brain hemispheres (left, right, and left+right) and anatomical planes (sagittal, coronal, and axial).

5.3. CNN input type

The proposed instance selection method was tested using 2D and 3D network inputs by employing the entire image or a cropped ROI. Using DenseNet121 [51], 2D CNNs processed the image slices extracted from the 3D MRI volume, and their predictions were fused to yield a subject-level classification. Considering the 3D nature of MRI, Custom3D CNNs were used to analyze all subject slices collectively. The 2D and 3D CNNs were trained on the same dataset, thereby enriching the heterogeneous ensemble with diversity.

Table 3
Summary of parameters of the Transfer Learning 2D CNN Model.

2D CNN Transfer learning Model	
Layer Type	Parameters
pre-trained_model	pre-trained_model = {'DenseNet121', 'ResNet50V2', 'InceptionResNetV2', ... }, weights = 'imagenet', input_shape = (160,160,3), include_top = False, trainable = True
GlobalAveragePooling2D Dropout	0.75
Dense	n_classes, activation = 'softmax'

5.4. CNN model

This experiment evaluated the impact of the proposed instance selection method on state-of-the-art model architectures, such as DenseNet (121) [51], EfficientNetV2 (M) [52] Inception ResNet V2 [53] ResNet [54], These models used the same training data.

5.5. Hybrid ensemble

The overall effect of the proposed instance selection methodology was assessed using a hybrid ensemble model. This classifier selected the most diverse and highest-accuracy models from the homogeneous and heterogeneous ensemble methods (Experiments II-IV). Each member of the hybrid set comprised ensembles, each comprising three individual models. Classification within each hybrid ensemble was determined by majority voting, aggregating the predictions from the three individual models. A weighted ensemble approach was employed for the final classification, with each set member contributing to a proportionally weighted prediction based on its performance.

5.6. Performance comparison

This experiment compared the performance of the proposed Multi-Atlas ROI-based instance selection methodology with that of the state-of-the-art methods. The related studies analyzed in this experiment used diverse datasets (ADNI, AIBL, and OASIS), CNN input types (2D and 3D), model architectures (ResNet, VGG, customs, and ensembles), preprocessing methods (skull stripping, registration, segmentation), and instance selection techniques (cropping, entropy, ROI extraction).

5.7. Implementation details

The subject dataset was divided into training (70%), validation (15%), and testing (15%) sets. All experiments were repeated three times. Random seeds were set for the NumPy, TensorFlow, Random, and OS libraries to obtain reproducible results. The Python libraries NiBabel, TorchIO, PIL, and NumPy preprocessed the images without saving the results on the disk. The FreeSurfer tools were used for skull stripping and MRI registration using the MNI152 template. The Keras library was used to construct classification models.

5.8. Transfer learning

Transfer learning was used to customize a model pretrained on ImageNet, and all layers were used as feature extraction components of a new model. The pretrained CNN was loaded without the classifier part of the model by specifying the top argument included in False and the preferred shape of the images in our new dataset as $160 \times 160 \times 3$. The CNN model was defined via the addition of a new average pooling layer following the last pooling layer. A new classifier model with a dense, fully connected layer and an output layer that predicts the probability of the two classes was added, as presented in Table 3. The weights of the pre-trained and new models were all trained together on the new dataset.

5.9. Hyperparameters tuning

Hyperband hyperparameter optimization was used to determine the optimal hyperparameters for training the models. The DenseNet121 model and the merged dataset (ADNI+AIBL+OASIS) were utilized to tune the hyperparameters. Table 4 defines the hyperparameters search spaces.

Table 5 presents the five best results of Hyperband hyperparameters tuning.

DenseNet121 [51] was selected to explore the influence of the proposed methodology on the overall performance of CNN-based classification models. This CNN model was selected because of its successful contribution to the computer vision field of image classification, object detection and localization, scene understanding, and other related tasks [58,59].

Table 6 outlines the relevant information regarding the hyperparameter values chosen for training the proposed methodology and all the selected CNNs to classify AD cases from MRI, except for the 3D custom CNN that used a dropout of 0.15.

Table 4
Hyperparameters search space.

Hyperparameters	Values
Optimizer name	{Adam, SGD, RMSprop, Adadelta, Adagrad, Adamax, Ftrl}
Learning rate	{0.01, 0.001, 0.0001, 1e-05, 1e-06}
Dropout	{0.20, 0.25, 0.30, ... 0.70, 0.75, 0.80}
Batch size	{8, 16, 32}
Epochs	{10, 15, 20}

Table 5
Hyperparameter values for Trial Summary Hyperband.

Optimizer	Learning Rate	Dropout	Batch Size	Epochs	Score
Adam	$1e-05$	0.75	8	15	95.03
SGD	0.001	0.40	16	15	94.91
Adam	$1e-05$	0.70	32	15	94.25
Adam	$1e-05$	0.20	32	15	94.02
RMSprop	0.0001	0.55	8	15	94.02

Table 6
Hyperparameter values for DenseNet121 [51] on merged dataset.

Hyperparameter	Value	Description
Dataset	Merged	70 Subjects each dataset (ADNI+AIBL+OASIS).
Image Size	160×160	Image size in pixels.
Number of channels	3	Number of channels (3 = RGB, 1 = Gray scale).
Transfer learning	ImageNet	Dataset name.
Learning rate	LRS	LearningRateSchedule exponential decay.
Decay steps	100000	Drops the learning rate by a factor.
Decay rate	0.9	The rate at which the learning rate is decayed.
Optimizer Name	Adam	The optimizer name.
Initial Learning Rate	$1e-05$	The initial learning rate.
Dropout	0.75	The rate in which the learning rate is decayed.
Batch Size	8	The number of instances per batch.
Epochs	20	The number of epochs.

5.10. Data analysis

All experiments were reported at the subject level by fusing all the classifications obtained from a subject at the slice level by majority voting.

One-way ANOVA test used to assess significant differences between the means of three or more groups. The t-test was used to check if two samples had identical averages. The significance level was set at 0.05% or 5%.

In this study, the following null and alternative hypotheses were tested.

- H_0 (null hypothesis): $\mu_1 = \mu_2 = \mu_3 \cdots = \mu_k$ (It implies that the mean accuracy of all models is equal).
- H_1 (alternative hypothesis): It states that at least one model accuracy mean will differ from the rest.

6. Results and discussion

Our proposed methodology was tested in six experiments to achieve the following objectives (1) Compare our Multi-Atlas ROI-Based Instance Selection technique with state-of-the-art methods. (2) Evaluate the versatility of our method to generate diverse datasets using the same base classifier. (3) Examine the ability of the proposed technique to generate 2D and 3D datasets from full-size or cropping images. (4) Evaluate the quality of the datasets generated by our method using state-of-the-art pre-trained CNN models. (5) Examine the multi-view datasets produced by the proposed method using a hybrid ensemble model. (6) Compare the performance of the proposed method with that of related state-of-the-art methods.

The experimental results corresponded to the mean and standard deviation of the model accuracy.

6.1. Experiment I (instance selection)

The results presented in Fig. 5 demonstrate that the proposed multi-atlas ROI-based instance selection technique outperformed both the entropy [16,17,12,18] and percentile [11] methods in terms of accuracy. These significant differences allowed us to reject the null hypothesis of equality, with a p-value of 0.005.
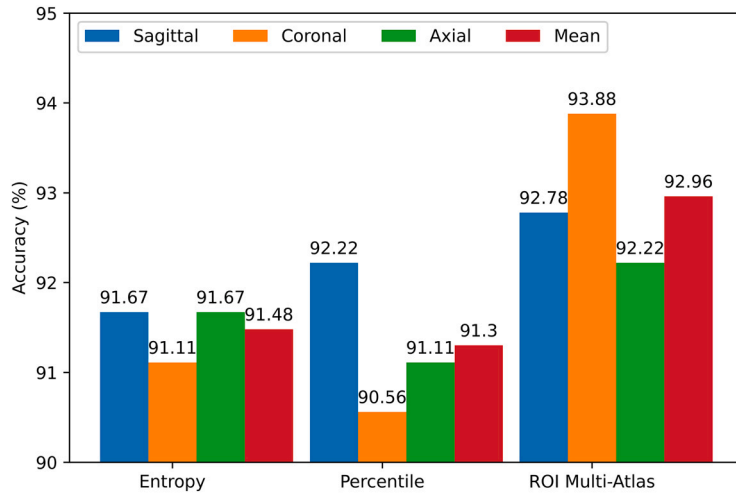
Fig. 5. Accuracy summary from different instance selection techniques, using the same subject-volume dataset (Merged).

(a) ROC Curve.   (b) Accuracy and Loss.
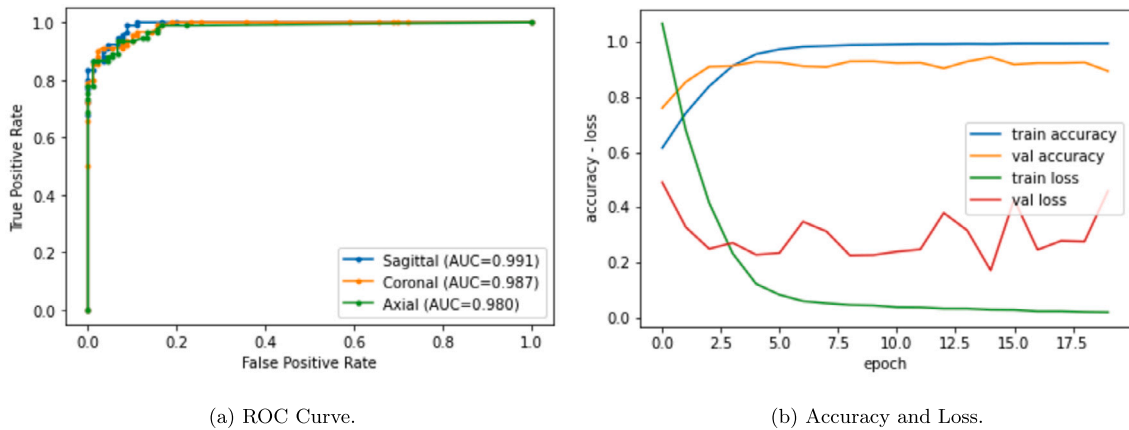
Fig. 6. Performance evaluation.

Substantial differences were observed in the following aspects.

a) Instance selection method: In terms of mean accuracy across all planes, our proposed method exhibited significant differences of 1.47% and 1.05% compared to the entropy and percentile methods, respectively, with p-values of 0.003 and 0.006, respectively.

b) Anatomical planes: Significant differences were observed in the coronal plane between the groups. Our proposed method revealed a difference of 2.78% and 3.32% compared with the entropy and percentile methods, respectively, with p-values of 0.024 and 0.013, respectively.

The ROC curves and AUC values of the model, shown in Fig. 6a, evaluated the performance of the CNN models using datasets from our proposed method. High AUC values in the sagittal (0.991), coronal (0.987), and axial (0.980) planes indicated excellent model performance and a strong capability to differentiate between CN and AD classes. In addition, the Accuracy and Loss curves for the CNN model (using the sagittal plane as an example), shown in Fig. 6b, identified a convergence point between the 15th and 20th epochs. This indicated that the model effectively learned the underlying patterns in the data and that further training was unlikely to yield significant improvements.

The performance of the prediction models was heavily influenced by the quality and quantity of the dataset, and the selection of informative slices improved the results [17]. In this experiment, the proposed instance selection method, based on a multi-atlas region of interest, was compared with state-of-the-art methods using the same subject-volume datasets.

The entropy technique [16,17,12,18] was outperformed by our proposed regarding accuracy rate. In our method, 3D information was preserved, and the context was leveraged by selecting only adjacent slice instances related to the ROI. The number of instances selected was variable and dependent on the plane, hemisphere, and ROI. By contrast, entropy techniques selected a fixed number of slices, calculated the image entropy of each slice, sorted them in descending order based on their entropy values, and retained the top (n) image slices. However, this approach may be unsuitable for 3D CNNs because of the loss of 3D information and the inclusion of dispersed slices from extreme regions not associated with AD. In [17], datasets were created by retaining the top 8, 16, and 32 image slices per subject using an entropy-based sorting algorithm. In [12,18], the authors selected the 32 most informative slices from each projection of 3D MRI data. In this study, 32 slice images were used.

Table 7

Summary of model accuracy from different experiments. As the table highlights, the most diverse and highest-accuracy performance results are included in the hybrid ensemble.

CNN Input Type	CNN Model	Preprocess	Hemisphere	Model Accuracy by Plane		
				Sagittal %	Coronal %	Axial %
Experiment II - Datasets						
2D Full Image	DenseNet121	Raw	Left+Right	**93.89 ± 0.79**	91.67 ± 1.36	90.00 ± 1.36
		Skull Stripped	Left+Right	91.11 ± 0.79	**93.33 ± 1.36**	91.67 ± 1.36
		Registered	Left	89.44 ± 0.79	90.56 ± 0.79	91.67 ± 0.00
			Right	93.33 ± 2.72	92.22 ± 0.79	**93.33 ± 1.36**
			Left+Right	92.78 ± 0.79	93.89 ± 0.79	92.22 ± 0.79
Experiment III - CNN input types						
2D Full Image	DenseNet121	Registered	Left+Right	92.78 ± 0.96	93.88 ± 0.96	92.22 ± 0.96
2D ROI	DenseNet121	Registered	Left	91.11 ± 2.08	**91.67 ± 2.36**	91.67 ± 0.00
			Right	92.22 ± 0.79	90.56 ± 0.79	87.78 ± 0.79
3D Full Image	Custom3D	Registered	Left+Right	89.44 ± 0.78	91.67 ± 2.36	**92.22 ± 0.79**
3D ROI	Custom3D	Registered	Left	88.89 ± 2.08	78.89 ± 1.92	86.11 ± 0.79
			Right	**93.33 ± 1.36**	76.67 ± 3.48	89.44 ± 2.83
Experiment IV - CNN models						
2D Full Image	DenseNet121	Registered	Left+Right	92.78 ± 0.96	93.88 ± 0.96	92.22 ± 0.96
	ResNet50V2			93.89 ± 1.57	92.22 ± 0.79	**91.11 ± 0.79**
	InceptionResNetV2			**94.44 ± 0.79**	90.0 ± 1.36	91.67 ± 0.00
	EfficientNetV2M			93.33 ± 1.36	**91.11 ± 0.79**	89.44 ± 0.79

The instance selection method based on the slice percentile position [11] was also outperformed by our proposed method. This method selected a fixed number (n) of instances from various positions in the percentile distribution of the entire volume. The experimental percentile positions per anatomical plane were as follows: sagittal = 65, coronal = 35, and axial = 35. In contrast, only slices belonging to the ROI were selected using the proposed method. In the aforementioned studies [60], images measuring 182 × 218 × 182 pixels were produced using skull stripping and volume registration. Here, 96 representative MRI slices from the 43rd to 139th were selected from the axial plane. In [61], all raw images were resized and normalized and 96 slices from the middle of the axial plane were selected.

The aforementioned selection methods, which are based on a fixed number of instances, have certain limitations. The performance of the model was significantly dependent on the number of slices per volume [11]. The addition of more image slices with less informative content can result in redundant or less representative information, increase the computational cost (time) of training, introduce noise, and deteriorate the model performance. However, the selection of a fixed number of slice images can 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.

This approach enhanced the model performance by reducing computational costs (time) and preventing excessive noise. This method guaranteed the representativeness of MRI and ensured the inclusion of AD-related information. These findings suggest that the proposed instance selection method is crucial in diagnosing AD and normal cognition.

6.2. Experiment II (datasets)

The results in Table 7 (Experiment II - Datasets) indicate significant differences between the models. The null hypothesis that the models were equal was rejected with a p-value of 0.012.

Substantial differences were observed in the following aspects.

a) Preprocessing: The mean accuracy of the sagittal plane varied significantly across datasets with different preprocessing methods. Specifically, the raw dataset's combined left and right hemispheres differed by 2.78% compared with the skull-stripped dataset and by 4.45% compared to the registered dataset's left hemisphere, with p-values of 0.024 and 0.005, respectively.

b) Hemispheres: In the registered dataset, a significant difference of 2.40% was observed between the mean accuracies of all planes in the left (90.56%) and right (92.96%) hemispheres (p = 0.007).

c) Anatomical planes: Significant differences were observed between the sagittal and axial planes. A notable difference of 3.89% was obtained between the raw dataset's combined left and right hemispheres in the sagittal plane. Conversely, a 2.23% difference was explicitly observed in the left hemisphere of the registered datasets. These disparities corresponded to p-values of 0.025 and 0.016, respectively. In addition, a 3.34% difference was evident between the left and the combined left and right hemispheres in the sagittal plane. In comparison, a 3.33% difference was observed in the coronal plane within the registered datasets with a p-value of 0.013.

J.A. Castro-Silva, M.N. Moreno-García, L. Guachi-Guachi et al.

Heliyon 10 (2024) e37552

Homogeneous ensembles encounter the challenge of generating diversity despite using the same learning algorithm [62]. This study experimentally evaluated the ability of the proposed approach to produce various MRI datasets when trained with a single-base classifier.

The mean accuracy varied significantly among the datasets subjected to different preprocessing methods (skull-stripped and registered) and those that were not preprocessed (raw). Similar results were obtained by [63].

Our study revealed significant differences between the left and right hemispheres. Similarly, in [20], notable disparities were observed in the accuracies of the left and right hemispheres across three ROIs: the amygdala, hippocampus, and insula. Specifically, the Hippocampus showed a difference of 9.93

The studies in [64,65] reported differences in the model accuracy across anatomical planes. Similarly, our proposed method demonstrated significant differences in model accuracy when the same base classifier was trained using datasets from the sagittal, coronal, and axial planes.

Our proposed method generated diverse datasets using different preprocessing methods, anatomical planes, and hemispheres. These datasets exhibited significant variations in model accuracy and offered multiple approaches for constructing a hybrid ensemble model.

6.3. Experiment III (CNN input types)

Table 7 (Experiment III - CNN input types) reveals significant differences between 2D and 3D CNNs. Differences were observed when considering the entire image and ROI and when combining hemispheres and planes. The null hypothesis was rejected with a p-value of 0.000. A significant difference of 17.22% was observed between the highest model accuracy in the right hemisphere of the 3D ROI and the lowest accuracy in the combined left and right hemispheres of the 2D full image, with a p-value of 0.001.

Substantial differences were observed in the following aspects.

a) CNN input type: The mean accuracy of all planes was significantly different by 1.85% between the 2D (92.96%) and 3D (91.11%) full-image models, with a p-value of 0.029. In the combined left and right hemispheres, there was a significant difference of 4.13% between the 2D (91.54%) and 3D (87.41%) full-image models and 5.27% between the 2D (90.83%) and 3D (85.56%) ROI models, with p-values of 0.001 and 0.002, respectively.

b) Anatomical planes: Significant differences were observed between the planes for the various model configurations. In the right hemisphere of 2D ROI models, the differences between sagittal and axial planes were 4.44%, and between the coronal and axial planes were 2.78%, with p-values of 0.005 and 0.024, respectively.

c) Hemispheres: In 3D ROI models of the left hemisphere, the differences between sagittal and coronal planes were 10.0%, and between the coronal and axial planes were 7.22%, with p-values of 0.024 and 0.006, respectively. In the right hemisphere of 3D ROI models, the differences between sagittal and coronal planes were 16.66%, and between the coronal and axial planes were 12.77%, with p-values of 0.002 and 0.011, respectively. Finally, in the combined left and right hemispheres of 3D full-image models, a 2.78% difference existed between sagittal and axial planes, with a p-value of 0.024.

In this study, we experimentally evaluated the impact of 2D and 3D CNNs on the construction of heterogeneous ensembles of various classifiers. Subject-level 3D CNNs analyzed all slices of a subject simultaneously, leveraging the 3D nature of MRI scans. Conversely, slice-level 2D CNNs processed individual 2D slices extracted from a 3D MRI volume and fuse the resulting classifications to achieve a subject-level classification [66].

In [67,68], the 2D CNN models outperformed 3D approaches in classifying AD while reducing training times. Similarly, our study revealed significant differences in the performance between 2D and 3D models for both complete images and ROIs. In addition, significant differences were observed between the planes across the various model configurations.

The size of the dataset significantly influenced the performance of the 2D and 3D CNNs in AD classification. A small dataset may hinder the performance of 3D models because more data may be required to tune the model end-to-end [67] effectively. The improvement in 2D CNNs' performance and the reduction in training time were primarily owing to a reduction in model complexity, fewer trainable parameters, and the application of transfer learning. The notable differences observed in the anatomical planes were likely because the sagittal plane captures essential information.

In practice, combining the outputs of individual classifiers results in more precise predictions. However, these improvements required diverse ensemble members [62]. Our proposed datasets, including full-size and cropped images focusing on the hippocampus ROI, demonstrated significant variations in model accuracy between 2D and 3D CNNs. These differences offer diverse options for constructing the hybrid ensemble models.

6.4. Experiment IV (CNN models)

From Table 7 (Experiment IV - CNN models), a significant difference was observed between the CNN model architectures, thereby rejecting the null hypothesis with a p-value of 0.000. A significant difference of 5.0% was observed between the highest model accuracy in the combined left and right hemispheres of the skull-stripped registered dataset with a p-value of 0.003. The Inception-ResNetV2 model achieved the highest accuracy (94.4%) in the sagittal plane. Conversely, the EfficientNetV2M model yielded the lowest accuracy (89.44%) in the axial plane.

Substantial differences were observed in the following aspects.

a) CNN models: Regarding the CNN model differences, the mean accuracy of all planes for the EfficientNetV2m model (89.44%) showed a significant difference of 2.78% with DenseNet121 (92.22%), and 2.23% with InceptionResNetV2 (91.67%), with p-values

Table 8
Summary of accuracy from different model members of the hybrid ensemble.

| Ensemble Model | Ensemble model members accuracy | | | | | | |
	Preprocess	Plane	Hemisphere	CNN Type	Ensemble CNN Model	Ensemble Accuracy %	Weight %
Homogeneous							
Ensemble 1	Raw	Sagittal	Left+Right	2D Full Image	DenseNet121	95.00	11.33
Ensemble 2	Skull Stripped	Coronal	Left+Right			93.33	11.13
Ensemble 3	Registered	Axial	Right			93.33	11.13
Heterogeneous							
Ensemble 4	Registered	Coronal	Left	2D ROI	DenseNet121	91.67	10.93
Ensemble 5		Axial	Left+Right	3D Full Image	Custom3D	91.67	10.93
Ensemble 6		Axial	Right	3D ROI	Custom3D	93.33	11.13
Ensemble 7		Axial	Left+Right	2D Full Image	ResNet50V2	91.67	10.93
Ensemble 8		Sagittal	Left+Right		InceptionResNetV2	95.00	11.33
Ensemble 9		Coronal	Left+Right		EfficientNetV2M	93.33	11.13
Hybrid							
Weighted Ensemble						**95.00**	

of 0.024 and 0.016, respectively. Concerning the coronal plane, significant differences of 3.89% and 2.78% were observed when comparing the DenseNet121 model with InceptionResNetV2 and EfficientNetV2m, with p-values of 0.025 and 0.024, respectively. Similarly, the axial plane showed significant differences when comparing EfficientNetV2m with DenseNet121 of 2.78% and 2.23% with InceptionResNetV2, with p-values of 0.024 and 0.016, respectively.

b) Anatomical planes: Significant differences were found in the anatomical planes when comparing the means of all CNN models. The sagittal plane (93.61%) revealed a difference of 1.80% in the coronal plane (91.81%) and 2.5% in the axial plane (91.11%) with p-values of 0.012 and 0.000, respectively. The InceptionResNetV2 model accuracy in the sagittal plane showed significant differences of 4.44% in the coronal plane and 2.76% in the axial plane, with p-values of 0.016 and 0.007, respectively. In addition, a significant difference of 3.89% was observed between the EfficientNetV2m sagittal and axial planes, with a p-value of 0.025.

In this experiment, we evaluated the quality of the datasets generated by the proposed method using state-of-the-art 2D CNN classification models and architectures, all of which were built using the same dataset (merged). This approach enables a thorough comparison of deep learning algorithms and enriches the diversity of hybrid ensembles.

In [69], the authors observed significant differences in the performance of various 2D CNN architectures such as InceptionV3, ResNet with 50 and 101 layers, and DenseNet with 169 layers for AD classification. DenseNet169 exhibited the best results for AD classification. Similarly, in [70], DenseNet169 outperformed the ResNet50 architecture. Our proposed method also identified significant variations in the mean accuracy across different 2D CNN architectures, both overall and within specific anatomical planes.

The accuracy of 2D CNN architectures for AD classification is influenced by model complexity, architectural design, number of trainable parameters, and network depth. These factors affect the learning efficiency and generalization ability of the models. The effectiveness of a model depends on its structure and implementation.

Heterogeneous ensemble methods utilize a variety of base classifiers, each trained on the same dataset, to capitalize on their distinct strengths and counterbalance their weaknesses. The significant differences between the 2D CNN architectures in this study contributed to the diversity of the ensemble model and enhanced its robustness, thereby improving its overall performance and reliability.

6.5. Experiment V (hybrid ensemble)

Table 8 presents that combining homogeneous and heterogeneous ensemble methods produces significantly higher accuracy (95.00%) than a single-learning classification model, providing variety to the hybrid ensemble classifier.

The optimal composition of the ensemble is problem-dependent, and determining the number of classifiers of each type that must be used has yet to be addressed [62]. Because weak classifiers must be combined adequately to fully exploit the ensemble, this experiment evaluated the impact of multiple-view datasets trained on different base classifiers.

The models from the previous experiments (II-IV) with the highest accuracy performance and variety were used to create a hybrid ensemble. This included different base classifiers trained on diverse datasets, as shown in Table 8.

Hybrid ensemble members are sets of three classifiers that combine the predictions from majority voting. Each member's contribution was weighted proportionally to its performance to obtain the final classification and create a weighted ensemble.

6.6. Experiment VI (performance comparison)

Table 9 compares the proposed method with state-of-the-art related works regarding CN and AD classification performance.

Table 9
Performance comparison of the proposed method with other related works for the classification tasks (AD vs. NC).

Study - Dataset	MRIs	Preprocess	Instance Selection Technique	CNN Input Type	Model	Accuracy
[14] - ADNI	212	Skull Stripping	Cropping	3D	ResNet50	82.93
					HA-ResUNet	92.68
[71] - ADNI	405	(Skull Stripping, Registration)		3D CNN 3D DenseNet	Ensemble	89.46
[12] - ADNI	315	(Skull Stripping, Registration)	Entropy(32)	2D	Ensemble VGG 16	85.27
[21] - ADNI	200	(Skull Stripping, Registration)	ROI	3D	Custom	89.00
- OASIS	196				3DCNN+EL+GA	76.00
[18] - ADNI 1	510	Normalization	Entropy (32)	2D	Ensemble VGG-16 GoogLeNet AlexNet	93.15
[22] - ADNI	563	(Skull Stripping, Bias Correction, Segmentation, Registration)	ROI ROI ROI ROI	3D	ResNet Trans-ResNet ResNet Trans-ResNet	92.61 93.85 92.60 93.17
- AIBL	567					
Proposed - Merged (ADNI + AIBL + OASIS)	420	(Skull Stripping, Registration)	ROI Multi-Atlas	2D Full Image	ResNet50V2 InceptionResNetV2 InceptionResNetV2 DenseNet121	93.89 94.44 93.33 93.88
				2D ROI		92.22
				3D Full Image	Custom3D	92.22
				3D ROI		93.33
				2D-3D	Hybrid Ensemble	**95.00**

The experimental results indicated that the proposed Multi-Atlas ROI-based instance selection method using a hybrid ensemble (95.00%) and single-base classifiers such as DenseNet121 (93.88%), ResNet50V2 (93.89%), and InceptionResNetV2 (94.44%) slightly outperformed the entropy-based and ROI-extraction methods (93.15% and 93.85%, respectively).

Compared with the best related work reported in [22], our proposed method offered several advantages.

(a) Reduced preprocessing tasks: We used fewer preprocessing techniques. Specifically, our dataset was preprocessed with skull stripping and registration, while the comparison method uses segmentation and bias correction.
(b) More versatile: Our method could produce multiple view datasets with 2D or 3D full images or ROIs compared with a single view.
(c) Enhanced robustness and generalization: Our method demonstrated greater robustness and generalization capacity. We tested our proposal using a merged dataset from multiple centers (ADNI, AIBL, and OASIS) in contrast to the single-center dataset (ADNI) used in the comparison.
(d) Improved performance: Our proposal slightly outperformed the best-reported results. The accuracy of our hybrid ensemble (95.00%) and single base classifiers, such as DenseNet121 (93.88%), ResNet50V2 (93.89%), and InceptionResNetV2 (94.44%), was slightly higher than the 93.85% achieved using a mixed model with Transformers and CNNs.

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

7. Limitations

The proposed method relied on a template with ROI annotations to identify the (x, y) centroids of an ROI and its slices. This dependence on preexisting annotated templates limits the accuracy and applicability of the method, which is contingent on the availability and quality of these annotations.

AD affects critical brain regions such as the entorhinal cortex, fornix, and hippocampus, thereby affecting cognitive functions [2–4]. Although our methodology is adaptable to multiple ROIs, it is currently limited to extracting informative instances from only the hippocampus.

In addition, the relatively small dataset size poses a limitation, particularly affecting the performance of 3D models. Thus, 3D CNNs require significant amounts of data to prevent overfitting and to effectively learn comprehensive 3D features.

J.A. Castro-Silva, M.N. Moreno-García, L. Guachi-Guachi et al.

Heliyon 10 (2024) e37552

8. Conclusions and future work

This study proposed a novel methodology for identifying and selecting the most informative 2D slices by merging hippocampus ROI annotations from multiple atlases (an instance selection method centered on hippocampus content using Multi-Atlas). Further, an innovative approach for extracting ROI content by calculating the adjusted centroid (x, y) and capturing the most informative image content has been developed.

A multisource domain adaptation method was proposed to address domain shift and improve model robustness. It merged three diverse population datasets (ADNI, AIBL, and OASIS) into a single domain, creating a multicenter dataset with registered MRI skull-stripped images. The participants were characterized using the clinical dementia rating (CDR) scale for integrated data comparison.

The impact of the proposed methodology on the overall performance of CNN-based classification models was explored experimentally. Our instance selection method could select instances for 2D and 3D datasets from an ROI or an entire image. The dataset generated using our proposed method achieved state-of-the-art performance when trained using a combination of homogeneous and heterogeneous ensemble methods.

The model performance varied according to the dataset (preprocessing technique, hemisphere, and plane) and model (CNN input type and model architecture). The use of 2D slices increased the number of instances and enabled existing 2D CNNs to train models through transfer learning or from scratch.

In future work, it would be interesting to explore the use of custom CNNs and model ensembles that utilize multiple ROIs and segmentation techniques to improve the performance and reliability of classification models. Moreover, increasing the size of the dataset is crucial for enhancing performance and accuracy, generalizing new data more effectively, and accessing a more diverse and representative sample. This approach aids in the identification of a broader range of patterns and features associated with Alzheimer's disease.

CRediT authorship contribution statement

Juan A. Castro-Silva: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **María N. Moreno-García:** Writing – review & editing, Supervision, Investigation, Formal analysis, Conceptualization. **Lorena Guachi-Guachi:** Writing – review & editing, Supervision, Formal analysis. **Diego H. Peluffo-Ordóñez:** Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data. However, the datasets used in this manuscript are publicly available and can be accessed at https://www.oasis-brains.org/ and https://adni.loni.usc.edu/.

References

[1] W.H. Organization, Dementia, https://www.who.int/news-room/fact-sheets/detail/dementia, 2023.

[2] National Institute on Aging, https://www.nia.nih.gov/health/what-happens-brain-alzheimers-disease. (Accessed 15 November 2023), 2023.

[3] Alzheimer's Society, https://www.alzheimers.org.uk/about-dementia/symptoms-and-diagnosis/how-dementia-progresses/symptoms-brain#content-start. (Accessed 15 November 2023), 2023.

[4] J. Yu, T.M. Lee, Verbal memory and hippocampal volume predict subsequent fornix microstructure in those at risk for Alzheimer's disease, Brain Imaging Behav. 14 (2020) 2311–2322, https://doi.org/10.1007/s11682-019-00183-8.

[5] Y.L. Rao, B. Ganaraja, B.V. Murlimanju, T. Joy, A. Krishnamurthy, A. Agrawal, Hippocampus and its involvement in Alzheimer's disease, a review, https://doi.org/10.1007/s13205-022-03123-4, 2 2022.

[6] C. Salvatore, A. Cerasa, P. Battista, M.C. Gilardi, A. Quattrone, I. Castiglioni, Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach, Front. Neurosci. 9 (SEP) (2015) 1–13, https://doi.org/10.3389/fnins.2015.00307.

[7] A. Khvostikov, K. Aderghal, A. Krylov, G. Catheline, J. Benois-Pineau, 3D inception-based CNN with sMRI and MD-DTI data fusion for Alzheimer's disease diagnostics, arXiv:1809.03972.

[8] W. Lin, T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, X. Qu, Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment, Front. Neurosci. 12 (NOV) (2018) 1–13, https://doi.org/10.3389/fnins.2018.00777.

[9] M. Blachnik, M. Kordos, Comparison of instance selection and construction methods with various classifiers, Appl. Sci. (Switzerland) 10, https://doi.org/10.3390/app10113933.

[10] V.R. Christo, H.K. Nehemiah, J. Brighty, A. Kannan, Feature selection and instance selection from clinical datasets using co-operative co-evolution and classification using random forest, IETE J. Res. 68 (2022) 2508–2521, https://doi.org/10.1080/03772063.2020.1713917.

[11] J.A. Castro-Silva, M.N. Moreno-García, L. Guachi-Guachi, D.H. Peluffo-Ordóñez, 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, INSTICC, SciTePress, 2022, pp. 330–337.

[12] M. Tanveer, A.H. Rashid, M.A. Ganaie, M. Reza, I. Razzak, K.-L. Hua, Classification of Alzheimer's disease using ensemble of deep neural networks trained through transfer learning, IEEE J. Biomed. Health Inform. 26 (4) (2022) 1453–1463, https://doi.org/10.1109/JBHI.2021.3083274.

[13] S.-H. Wang, Y. Zhang, Y.-J. Li, W.-J. Jia, F.-Y. Liu, M.-M. Yang, Y.-D. Zhang, Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization, Multimed. Tools Appl. 77 (9) (2018) 10393–10417, https://doi.org/10.1007/s11042-016-4222-4.

[14] Z. Qin, Z. Liu, Q. Guo, P. Zhu, 3d convolutional neural networks with hybrid attention mechanism for early diagnosis of Alzheimer's disease, Biomed. Signal Process. Control 77 (2022) 103828, https://doi.org/10.1016/j.bspc.2022.103828.

[15] F.U.R. Faisal, G.R. Kwon, Automated detection of Alzheimer-s disease and mild cognitive impairment using whole brain MRI, IEEE Access 10 (2022) 65055–65066, https://doi.org/10.1109/ACCESS.2022.3180073.

[16] E. Jabason, M.O. Ahmad, M.N. Swamy, Classification of Alzheimer's disease from MRI data using an ensemble of hybrid deep convolutional neural networks, in: Midwest Symposium on Circuits and Systems 2019-Augus (Mci), 2019, pp. 481–484.

[17] N.M. Khan, N. Abraham, M. Hon, Transfer learning with intelligent training data selection for prediction of Alzheimer's disease, IEEE Access 7 (2019) 72726–72735, https://doi.org/10.1109/ACCESS.2019.2920448, arXiv:1906.01160.

[18] J.Y. Choi, B. Lee, Combining of multiple deep networks via ensemble generalization loss, based on mri images, for Alzheimer's disease classification, IEEE Signal Process. Lett. 27 (2020) 206–210, https://doi.org/10.1109/LSP.2020.2964161.

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

[20] S. Ahmed, B.C. Kim, K.H. Lee, H.Y. Jung, Ensemble of ROI-based convolutional neural network classifiers for staging the Alzheimer disease spectrum from magnetic resonance imaging, PLoS ONE 15 (12 December), https://doi.org/10.1371/journal.pone.0242712.

[21] D. Pan, G. Luo, A. Zeng, C. Zou, H. Liang, J. Wang, T. Zhang, B. Yang, The Alzheimer's disease neuroimaging initiative, adaptive 3dcnn-based interpretable ensemble model for early diagnosis of Alzheimer's disease, IEEE Trans. Comput. Soc. Syst. (2022) 1–20, https://doi.org/10.1109/TCSS.2022.3223999.

[22] C. Li, Y. Cui, N. Luo, Y. Liu, P. Bourgeat, J. Fripp, T. Jiang, Trans-resnet: integrating transformers and cnns for Alzheimer's disease classification, in: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, pp. 1–5.

[23] K.M. Poloni, R.J. Ferrari, Automated detection, selection and classification of hippocampal landmark points for the diagnosis of Alzheimer's disease, Comput. Methods Programs Biomed. 214 (2022) 106581, https://doi.org/10.1016/j.cmpb.2021.106581.

[24] A. Aghaei, M.E. Moghaddam, Smart ROI Detection for Alzheimer's Disease prediction using explainable AI, Tech. Rep.

[25] F. Ren, C. Yang, Q. Qiu, N. Zeng, C. Cai, C. Hou, Q. Zou, Exploiting discriminative regions of brain slices based on 2D CNNs for Alzheimer's disease classification, IEEE Access 7 (2019) 181423–181433, https://doi.org/10.1109/ACCESS.2019.2920241.

[26] K. Backstrom, M. Nazari, I.Y.H. Gu, A.S. Jakola, An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images, in: Proceedings - International Symposium on Biomedical Imaging 2018-April (Isbi), 2018, pp. 149–153.

[27] N.T. Duc, S. Ryu, M.N.I. Qureshi, M. Choi, K.H. Lee, B. Lee, 3D-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI, Neuroinformatics 18 (1) (2020) 71–86, https://doi.org/10.1007/s12021-019-09419-w.

[28] Y. Zhao, B. Ma, P. Jiang, D. Zeng, X. Wang, S. Li, Prediction of Alzheimer's disease progression with multi-information generative adversarial network, IEEE J. Biomed. Health Inform. 25 (3) (2021) 711–719, https://doi.org/10.1109/JBHI.2020.3006925.

[29] A. Farooq, S. Anwar, M. Awais, M. Alnowami, Artificial intelligence based smart diagnosis of Alzheimer's disease and mild cognitive impairment, in: 2017 International Smart Cities Conference, ISC2 2017, 2017, pp. 0–3.

[30] Z. Guan, R. Kumar, Y.R. Fung, Y.J. Wu, M. Fiterau, A comprehensive study of Alzheimer's disease classification using convolutional neural networks, arXiv:1904.07950 [abs].

[31] M. Menagadevi, S. Mangai, N. Madian, D. Thiyagarajan, Automated prediction system for Alzheimer detection based on deep residual autoencoder and support vector machine, Optik 272 (November 2022) (2023) 170212, https://doi.org/10.1016/j.ijleo.2022.170212.

[32] P. Kalavathi, V.B. Prasath, Methods on skull stripping of MRI head scan images—a review, J. Digit. Imag. 29 (3) (2016) 365–379, https://doi.org/10.1007/s10278-015-9847-8.

[33] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (December 2012) (2017) 60–88, https://doi.org/10.1016/j.media.2017.07.005, arXiv:1702.05747.

[34] E. Jabason, M. Omair Ahmad, M.N. Swamy, Hybrid feature fusion using RNN and pre-trained CNN for classification of Alzheimer's disease (poster), in: FUSION 2019 - 22nd International Conference on Information Fusion, 2019, pp. 2019–2022.

[35] E. Hussain, M. Hasan, S.Z. Hassan, T. Hassan Azmi, M.A. Rahman, M. Zavid Parvez, Deep learning based binary classification for Alzheimer's disease detection using brain MRI images, in: Proceedings of the 15th IEEE Conference on Industrial Electronics and Applications, ICIEA 2020, 2020, pp. 1115–1120.

[36] J.B. Bae, S. Lee, W. Jung, S. Park, W. Kim, H. Oh, J.W. Han, G.E. Kim, J.S. Kim, J.H. Kim, K.W. Kim, Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging, Sci. Rep. 10 (1) (2020) 1–10, https://doi.org/10.1038/s41598-020-79243-9.

[37] C. Yang, A. Rangarajan, S. Ranka, Visual explanations from deep 3D convolutional neural networks for Alzheimer's disease classification, arXiv:1803.02544v3.

[38] B. Khagi, G.R. Kwon, CNN model performance analysis on MRI images of an OASIS dataset for distinction between healthy and Alzheimer's patients, IEIE Trans. Smart Process. Comput. 8 (4) (2019) 272–278, https://doi.org/10.5573/IEIESPC.2019.8.4.272.

[39] A. Raza, J. Uddin, S. Almuhaimeed, S. Akbar, Q. Zou, A. Ahmad, AIPs-SnTCN: predicting anti-inflammatory peptides using fastText and transformer encoder-based hybrid word embedding with self-normalized temporal convolutional networks, J. Chem. Inf. Model. 63 (21) (2023) 6537–6554, https://doi.org/10.1021/acs.jcim.3c01563.

[40] S. Akbar, A. Raza, T.A. Shloul, A. Ahmad, A. Saeed, Y.Y. Ghadi, O. Mamyrbayev, E. Tag-Eldin, pAtbP-EnC: identifying anti-tubercular peptides using multi-feature representation and genetic algorithm-based deep ensemble model, IEEE Access 11 (December) (2023) 137099–137114, https://doi.org/10.1109/ACCESS.2023.3321100.

[41] S. Akbar, M. Hayat, M. Tahir, S. Khan, F.K. Alarfaj, cACP-DeepGram: classification of anticancer peptides via deep neural network and skip-gram-based word embedding model, Artif. Intell. Med. 131 (July) (2022) 102349, https://doi.org/10.1016/j.artmed.2022.102349.

[42] S. Akbar, S. Khan, F. Ali, M. Hayat, M. Qasim, S. Gul, iHBP-DeepPSSM: identifying hormone binding proteins using PsePSSM based evolutionary features and deep learning approach, Chemom. Intell. Lab. Syst. 204 (June) (2020) 104103, https://doi.org/10.1016/j.chemolab.2020.104103.

J.A. Castro-Silva, M.N. Moreno-García, L. Guachi-Guachi et al.

Heliyon 10 (2024) e37552

[43] R. Sharma, T. Goel, M. Tanveer, P.N. Suganthan, I. Razzak, R. Murugan, Conv-ervfl: convolutional neural network based ensemble rvfl classifier for Alzheimer's disease diagnosis, IEEE J. Biomed. Health Inform. (2022) 1–9, https://doi.org/10.1109/JBHI.2022.3215533.

[44] M. Ganaie, M. Hu, A. Malik, M. Tanveer, P. Suganthan, Ensemble deep learning: a review, Eng. Appl. Artif. Intell. 115 (2022) 105151, https://doi.org/10.1016/j.engappai.2022.105151.

[45] Alzheimer's Disease Neuroimaging Initiative (ADNI), http://adni.loni.usc.edu.

[46] Australian imaging, biomarker & lifestyle flagship study of ageing (AIBL), https://aibl.csiro.au.

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

[48] C.P. Hughes, L. Berg, W. Danziger, L.A. Coben, R.L. Martin, A new clinical scale for the staging of dementia, Br. J. Psychiatry 140 (6) (1982) 566–572, https://doi.org/10.1192/bjp.140.6.566.

[49] Templates and Atlases included with FSL, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases.

[50] L. Zhang, Z. Chen, B. Zou, Y. Gao, Polarimetric sar terrain classification using 3d convolutional neural network, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 4551–4554.

[51] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, https://doi.org/10.48550/ARXIV.1608.06993, 2016, https://arxiv.org/abs/1608.06993.

[52] M. Tan, Q.V. Le, Efficientnetv2: smaller models and faster training, https://doi.org/10.48550/ARXIV.2104.00298, 2021, https://arxiv.org/abs/2104.00298.

[53] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, https://doi.org/10.48550/ARXIV.1602.07261, 2016, https://arxiv.org/abs/1602.07261.

[54] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, https://doi.org/10.48550/ARXIV.1512.03385, 2015, https://arxiv.org/abs/1512.03385.

[55] H. Zunair, A. Rahman, N. Mohammed, J.P. Cohen, Uniformizing techniques to process ct scans with 3d cnns for tuberculosis prediction, https://doi.org/10.48550/ARXIV.2007.13224, 2020, https://arxiv.org/abs/2007.13224.

[56] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, A. Talwalkar, Hyperband: a novel bandit-based approach to hyperparameter optimization, J. Mach. Learn. Res. 18 (185) (2018) 1–52, http://jmlr.org/papers/v18/16-558.html.

[57] J. Brownlee, Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions, Machine Learning Mastery With Python, vol. 1 (2), 2018, p. 539.

[58] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.

[59] A. Rosebrock, Deep Learning for Computer Vision, PyImageSearch, 2017.

[60] Z. Hu, Y. Li, Z. Wang, S. Zhang, W. Hou, Conv-swinformer: integration of cnn and shift window attention for Alzheimer's disease classification, Comput. Biol. Med. 164, https://doi.org/10.1016/j.compbiomed.2023.107304.

[61] F. Altay, G.R. Sánchez, Y. James, S.V. Faraone, S. Velipasalar, A. Salekin, Preclinical stage Alzheimer's disease detection using magnetic resonance image scans, Proc. AAAI Conf. Artif. Intell. 35 (17) (2021) 15088–15097, https://doi.org/10.1609/aaai.v35i17.17772, https://ojs.aaai.org/index.php/AAAI/article/view/17772.

[62] M. Sabzevari, G. Martínez-Muñoz, A. Suárez, Building heterogeneous ensembles by pooling homogeneous ensembles, Int. J. Mach. Learn. Cybern. 13 (2) (2022) 551–558, https://doi.org/10.1007/s13042-021-01442-1.

[63] B.M. Cobbinah, C. Sorg, Q. Yang, A. Ternblom, C. Zheng, W. Han, L. Che, J. Shao, Reducing variations in multi-center Alzheimer's disease classification with convolutional adversarial autoencoder, Med. Image Anal. 82 (2022) 102585, https://doi.org/10.1016/j.media.2022.102585, https://www.sciencedirect.com/science/article/pii/S1361841522002237.

[64] A. Kabir, F. Kabir, M.A. Hasib Mahmud, S.A. Sinthia, S.M. Rakibul Azam, E. Hussain, M.Z. Parvez, Multi-classification based Alzheimer's disease detection with comparative analysis from brain MRI scans using deep learning, in: IEEE Region 10 Annual International Conference, Proceedings/TENCON 2021-December, 2021, pp. 905–910.

[65] R. Mendoza-Léon, J. Puentes, L.F. Uriza, M. Hernández Hoyos, Single-slice Alzheimer's disease classification and disease regional analysis with supervised switching autoencoders, Comput. Biol. Med. 116 (October 2019), https://doi.org/10.1016/j.compbiomed.2019.103527.

[66] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, O. Colliot, Overview of classification of Alzheimer's disease, Med. Image Anal. 63, arXiv:1904.07773.

[67] G. Liang, X. Xing, L. Liu, Y. Zhang, Q. Ying, A.L. Lin, N. Jacobs, Alzheimer's disease classification using 2D convolutional neural networks, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2021, pp. 3008–3012.

[68] X. Xing, G. Liang, H. Blanton, M.U. Rafique, C. Wang, A.L. Lin, N. Jacobs, Dynamic Image for 3D MRI Image Alzheimer's Disease Classification, Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS, vol. 12535, 2020, pp. 355–364, arXiv:2012.00119.

[69] S. Pallawi, D.K. Singh, Review and analysis of deep neural network models for Alzheimer's disease classification using brain medical resonance imaging, Cogn. Comput. Syst. 5 (1) (2023) 1–13, https://doi.org/10.1049/ccs2.12072.

[70] W. Al Shehri, Alzheimer's disease diagnosis and classification using deep learning techniques, PeerJ Comput. Sci. 8, https://doi.org/10.7717/PEERJ-CS.1177.

[71] A. Gamal, M. Elattar, S. Selim, Automatic early diagnosis of Alzheimer's disease using 3d deep ensemble approach, IEEE Access 10 (2022) 115974–115987, https://doi.org/10.1109/ACCESS.2022.3218621.