Reducing variations in multi-center Alzheimer's disease classification with convolutional adversarial autoencoder

Bernard M. Cobbinah [a], Christian Sorg [c], Qinli Yang [a], Arvid Ternblom [a], Changgang Zheng [a], Wei Han [a], Liwei Che [a], Junming Shao [a,b,d,*]

[a] School of Computer Science and Engineering, University of Electronic Science and Technology of China, 611731 Chengdu, China
[b] Center for Information in BioMedicine, University of Electronic Science and Technology of China, 611731 Chengdu, China
[c] Department of Neuroradiology, TUM-NIC Neuroimaging Center of Technical University Munich, Germany
[d] Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou 313001, China

ARTICLE INFO

ABSTRACT

Based on brain magnetic resonance imaging (MRI), multiple variations ranging from MRI scanners to center-specific parameter settings, imaging protocols, and brain region-of-interest (ROI) definitions pose a big challenge for multi-center Alzheimer's disease characterization and classification. Existing approaches to reduce such variations require intricate multi-step, often manual preprocessing pipelines, including skull stripping, segmentation, registration, cortical reconstruction, and ROI outlining. Such procedures are time-consuming, and more importantly, tend to be user biased. Contrasting costly and biased preprocessing pipelines, the question arises whether we can design a deep learning model to automatically reduce these variations from multiple centers for Alzheimer's disease classification? In this study, we used T1 and T2-weighted structural MRI from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset based on three groups with 375 subjects, respectively: patients with Alzheimer's disease (AD) dementia, with mild cognitive impairment (MCI), and healthy controls (HC); to test our approach, we defined AD classification as classifying an individual's structural image to one of the three group labels. We first introduced a convolutional adversarial autoencoder (CAAE) to reduce the variations existing in multi-center raw MRI scans by automatically registering them into a common aligned space. Afterward, a convolutional residual soft attention network (CRAT) was further proposed for AD classification. Canonical classification procedures demonstrated that our model achieved classification accuracies of 91.8%, 90.05%, and 88.10% for the 2-way classification tasks using the RAW aligned MRI scans, including AD vs. HC, AD vs. MCI, and MCI vs. HC, respectively. Thus, our automated approach achieves comparable or even better classification performance by comparing it with many baselines with dedicated conventional preprocessing pipelines. Furthermore, the uncovered brain hotpots, i.e., hippocampus, amygdala, and temporal pole, are consistent with previous studies.

1. Introduction

Alzheimer's disease (AD) is characterized by increasing cognitive and behavioral impairments, eventually resulting in dementia syndrome; it is a neurodegenerative disease, which is the most common cause of age-related dementia (Jack et al., 2018; Association, 2019). For example, about 5.7 million people lived with AD in the United States in 2018 (Association et al., 2018), with predicted numbers for 2040 more than doubled. Early AD intervention needs foregoing diagnosis, and several biomarkers have been proposed to identify the disease. Amongst these markers, magnetic resonance imaging (MRI) provides a promising cost-efficient tool to improve AD diagnosis (Wen et al., 2020). However,

multiple variations in the MRI-approach on AD classification, ranging from MRI scanners to center-specific parameter settings, imaging protocols, and often manual brain region-of-interest (ROI) definition, pose a big challenge for reliable diagnostic procedures (Mayerhoefer et al., 2009; Dong et al., 2017; Sedeno et al., 2017). Modeling an automatic AD prediction pipeline to handle such MRI-approach-based variability might improve AD diagnosis. Thereby, data-driven-based automated AD characterization and classification have gained more attention in recent years (Li et al., 2020; Hett et al., 2021). The current study contributes to this topic with a special focus on multi-center AD classification problems based on deep learning approaches.

Recent research has demonstrated the effectiveness, timeliness, and reliability of deep learning techniques in diagnosing AD (Jo et al., 2019b; Myszczynska et al., 2020; Iizuka et al., 2019; Zhang et al., 2021b). Deep learning techniques, such as convolutional neural networks (CNNs), can automatically extract in-depth representative features from the raw images without needing prior manual feature selection processes (Goceri, 2021; Vieira et al., 2017; Wang et al., 2021a). In this way, deep learning techniques reduce the over-reliance on domain expertise in the user-biased handcrafted design and extraction of vital discriminative features in AD diagnosis.

Among the most current deep learning techniques in Alzheimer's disease prediction, attention mechanism has proven to be a potent tool, and an important component in predicting the disease using MRI scans (Zhang et al., 2021a, 2020; Liu et al., 2022). The attention mechanism technique has shown high accuracy for the prediction of AD, as reported in these works (Liu et al., 2022; Wang et al., 2021b; Jin et al., 2019). Attention mechanism has proved efficient in capturing global dependencies because it learns the dependency relationship between each AD feature and assigns corresponding weights to map out the most discriminative features.

Acquisition of multi-center MRI scans in AD diagnosis is progressively considered a reliable, worthwhile, and promising approach to facilitate a robust diagnosis (Hao et al., 2020; Gradin et al., 2010; Basaia et al., 2019). Multi-center studies allow better generalization of diagnosis performance, not only limited to different subjects scanned on the same scanner but also different subjects scanned on different scanners (Mwangi et al., 2012). However, the massive challenge of multi-center AD classification is variations in multi-center MRI raw scans (Dong et al., 2017; Sedeno et al., 2017; Teipel et al., 2010; Potvin et al., 2019) and the user-biased manual intervention of experts to focus on selected characteristics. These variations are based on MRI scanners' differences, center-specific parameter settings, imaging protocols, different individual brain morphology, MRI scans acquisition imperfections, and patient positioning across various clinics and hospitals (Schnack et al., 2010; Gilmore et al., 2019). Because of these variations, typical multi-step conventional preprocessing techniques, including skull stripping, segmentation, registration, cortical reconstruction, ROI outlining, and statistical correction, are applied to the acquired MRI scans before AD classification (Zhu et al., 2021; Qiu et al., 2019). Critically, such preprocessing tends to be user-biased and thus become difficult for AD prediction models to generalize well on unseen patient data with variations. Thereby, a fully automated model for AD prediction without any user intervention to reduce the inter-and intra-variability in multi-center MRI scans is highly recommended.

In this study, our principal interest is harnessing the capabilities of deep learning techniques in handling multi-center MRI variations and making AD classification directly from raw MRI data. Concerning data, we used T1-weighted and T2-weighted structural MRI from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset based on three groups, respectively: patients with Alzheimer's disease (AD) dementia, with mild cognitive impairment (MCI), and healthy controls (HC). To test our approach, we defined AD classification as classifying an individual's structural image to one of three group labels. Concerning our analytical approach, an unsupervised convolutional adversarial autoencoder (CAAE) is proposed for the automatic feature extraction and alignment of raw multi-center MRI scans into a common space to reduce the variations existing in multi-center raw MRI scans. We further performed AD classification using a CNN with a soft residual attention mechanism (i.e., convolutional residual soft attention network, CRAT). Furthermore, we also validated our CAAE alignment model by using the output of its latent representations/embedding (reduced dimension) on selected statistical machine learning models as an incentive method. Building upon our model, it offers an intuitive way to handle multi-center MRI variations, which does not involve any user-biased preprocessing procedure in AD classification. Beyond the soft residual attention mechanism, we explore the brain regions relevant for AD classification and compare them with well-known regions most affected by AD. The contributions of this paper can be summarized as follows:

Table 1
Demographic and clinical characteristics.

	AD	MCI	HC	p
Sex (M/F)	205/170	219/156	190/185	0.34
Age (y)	75.2 ± 7.2	75.0 ± 7.3	74.4 ± 6.0	0.31
CDR	1.2 ± 0.6	0.5 ± 0.2	0 ± 0	<0.01
MMSE	24.09 ± 2.15	26.09 ± 1.15	28.40 ± 1.18	<0.01

Values are presented as mean ± SD. AD = Alzheimer's disease patient; MCI = Mild cognitive impairment; HC = Cognitively normal patient; MMSE = Mini-Mental State Examination; CDR = Clinical Dementia Rating; p = *p*-value calculated from Chi-squared test.

1. We propose an adversarial autoencoder method to reduce multi-center MRI variations in AD classification.
2. The proposed framework is robust and provides a time-efficient fully automatic end-to-end approach for AD classification.
3. An attention-based mechanism is leveraged to uncover the brain hotspots discriminative to AD, which is consistent with the literature.
4. The performance of the proposed approach is empirically validated, and reported results are encouraging for raw-MRI scans without typical MRI preprocessing pipeline.

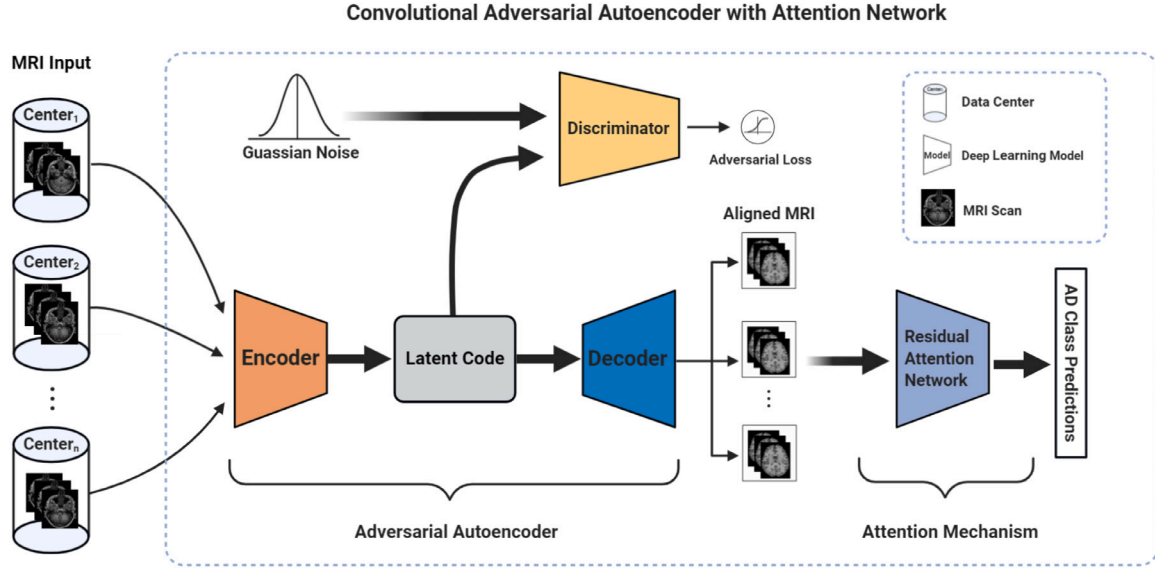## 2. Multi-center neuroimaging data

2.1. Subjects

The dataset used in this research was obtained from the ADNI database (http://adni.loni.usc.edu/methods/documents). The ADNI is a public project that makes reliable multi-center clinical and imaging data available to Alzheimer's disease researchers. An extensive review of ADNI datasets can be found in Weiner et al. (2015). In the current study, we included 1125 participants (Table 1) from the ADNI database. The dataset includes all subjects from ADNI-1, ADNI-2, and ADNI-GO, who had baseline T1-weighted and T2-weighted scans. The datasets were acquired from different sites with different MRI scanners having different acquisition protocols. We included three groups of participants, each group with 375 subjects: cognitively normal persons (Mini-Mental State Examination (MMSE) > 24, Clinical Dementia Rating (CDR) = 0, non-depressed), patients with Alzheimer's disease (MMSE < 26, CDR > 0.5), and patients with mild cognitive impairment (MMSE > 24, CDR = 0.5, with objective memory loss).

2.2. Data acquisition

Original structural MRI (sMRI) scans, both T1 and T2 weighted, were obtained from different 1.5T or 3T scanners (General Electric Medical System, Philips Medical Systems, and Siemens MRI scanners) in all ADNI centers. The MRI data were acquired from different scanners to develop a reliable model to classify Alzheimer's disease irrespective of the MRI scans' variability. Specifically, structural T1-weighted and T2-weighted MRI scans were acquired using 1.5T and 3T scanners. We obtained the following sequences: (i) 1.5T T1-weighted; acquisition parameters were inversion time (TI) = 1000 ms, slice thickness = 2 mm, repetition time (TR) = 2400 ms, field-of-view (FOV) = 240×240 mm^2, minimum full echo time (TE), and flip angle = $\angle 8$ or $\angle 9$; (ii) 1.5T T2-weighted; the acquisition parameters were FOV= 260×260 mm^2, slice thickness = 3 mm or 4 mm, flip angle = $\angle 8$ or $\angle 9$, minimum full TE, and TI = 900 ms, TR = 2300 ms. (iii) 3T T2-weighted; the acquisition parameters were echo time (TE) = 85 ms, FOV = 230×208 mm^2, flip angle = $\angle 90$, repetition time (TR) = 3000 ms, and slice thickness = 3 mm. All scans were acquired with varying acquisition matrix in the x, y, and z dimensions. In Table 2, we present typical sequence differences between MRIs captured with 1.5T and 3T scanners. The scans comprised full brain coverage axial, coronal, and sagittal acquisition planes. Since several patients have MRI scans taken

Table 2
Typical differences between MRIs captured with 1.5T and 3T scanners.

	1.5T	3T
1	Poor image quality.	Better image quality.
2	Robust to artifacts.	Susceptible to artifacts.
3	Less specific absorption rate.	Increased specific absorption rate.
4	Less spatial resolution.	Increased spatial resolution.
5	Less signal-to-noise ratio (SNR).	Increased signal-to-noise ratio.
6	Less acoustic noise.	Increased acoustic noise.
7	The capture of anatomical structure is not very precise.	Produces precise anatomical capture.

Fig. 1. A Flow diagram illustrating the combined architecture for multi-center MRI variability handling and attention-based convolutional network for Alzheimer's disease classification. MRI scans from multiple centers are aligned to a common space, and discriminative features are extracted using a convolutional adversarial autoencoder. The aligned MRI scans are then passed through a convolutional soft attention network to further handle variability by attending significant AD prediction regions. Finally, three binary classification tasks (Alzheimer's disease vs. cognitively normal, Alzheimer's disease vs. mild cognitive impairment, mild cognitively normal vs. cognitively normal) are performed.

at different visits or centers in the ADNI database over the years, we ensured that the same patient's images could be only part of either the training or test data set. We focus on the original raw MRI scans without any typical MRI preprocessing like skull stripping, head motion correction, registration, segmentation, and the like.

3. Methodology

We developed a CAAE architecture to automatically reduce the variations from multiple centers and a convolutional attention network for Alzheimer's disease classification. Specifically, the proposed methodology consists of two parts: (1) Multicenter MRI variations handling via convolutional adversarial autoencoder, and (2) Classification with convolutional residual attention network. For an overview, see Fig. 1.

3.1. Data augmentation

We applied data augmentation techniques (Wang et al., 2021a) to the raw multi-center MRI scans belonging to the training set. This was to help increase the dataset and also achieve model robustness. In this work, size augmentation and location augmentation were applied. In size augmentation, after applying the cropping augmentation technique, the dimension of the MRI scans were resized to fit the desired dimension for the model. For location augmentation, we applied gaussian blur with sigma value ranging between 0 and 1.3, both vertical and horizontal flip with a probability of 0.5, and image rotation between the angle of -45 and 45. We produced an increasing amount of dataset for each class with the equation:

$$T_c = n_c \times K \qquad (1)$$

where T_c is the new amount of dataset for each class c, n_c is the initial number of each class before data augmentation, K is the increment factor and $c \epsilon \{AD, MCI, NC\}$. In this work, the increment factor K is 2. It means that each input data would be augmented once and increase the initial training dataset with a factor of 2.

3.2. Multi-center MRI variations handling via convolutional adversarial autoencoder

To circumvent the multi-center MRI variance-induced problem, we developed a CAAE architecture to automatically project all raw three-dimensional (3D) MRI scans into a common aligned space. The adversarial autoencoder (Makhzani et al., 2015) was used to extract discriminative features that align each raw MRI scan from different domains or multiple centers into a common (or standard) space. We utilized the reconstruction and regularization phase of adversarial autoencoders to perform the extraction of unique features that capture the variability in raw MRI scans from multiple centers. We formulated the convolutional autoencoder into the following parts: autoencoder and adversarial network. See Fig. 2 for the CAAE architecture.

3.2.1. Autoencoder

In this paper, we term the handling of the variability among MRI scans from multiple centers as 'MRI alignment'. The MRI alignment is handled with an adversarial autoencoder technique. The multi-center MRI scans are passed as inputs to the architecture for a common transformed output (aligned MRI scans) for further classification tasks. In adversarial autoencoders, we simultaneously trained an autoencoder (Masci et al., 2011; Pulgar et al., 2020) to minimize the reconstruction loss of an input MRI scan and a discriminative (adversarial)
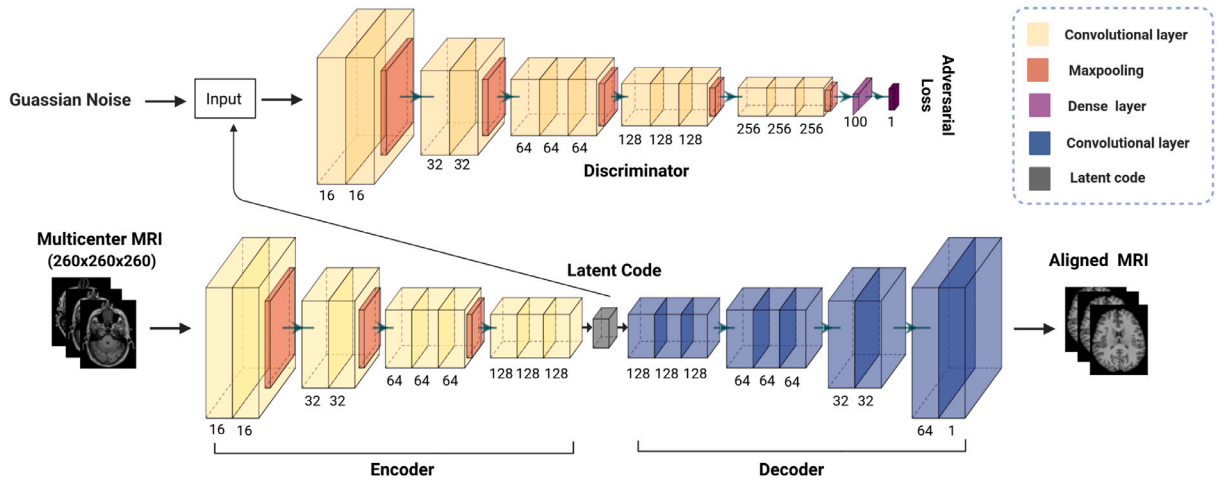
Fig. 2. The CAAE architecture for extracting generic features and aligning MRI scans into a common space for AD classification.

network to distinguish between true samples and generated samples. The autoencoder is made up of an encoder and a decoder. The encoder part aims at encoding raw MRI data into reduced latent features of deep representations. The encoder learns to convert the raw MRI scans distribution to a prior distribution. To this end, the encoded feature comprises hidden code information of the raw MRI data. The latent code generated by the encoder is represented as

$$z = \sigma(W * x + b) \tag{2}$$

where z is the encoder's output, x is the MRI input, $*$ is the convolutional operation, and W, b, and σ are weights, bias, and Rectified Linear Unit (ReLU) activation function respectively of the convolutional operation. The encoder is also the generator of this network.

The decoder network projects the latent code of the encoder network into a new MRI space. In brief, the decoder maps the intended prior distribution to the raw MRI scans distribution. Furthermore, the decoder network guides all the raw MRI scans into a new common space. The decoder objective function is represented as

$$\hat{x} = \sigma(W * z + b) \tag{3}$$

where $\hat{x}$ is the reconstructed MRI, and z is the latent code from the encoder. The reconstruction loss function of the autoencoder is calculated as

$$L(x, \hat{x}) = \frac{1}{n} \sum_{i=1}^{n} \|x_i - \hat{x}_i\|^2 \tag{4}$$

where $\hat{x}$ represents reconstructed MRI scans, and x is the raw MRI scans as input to the encoder. The autoencoder network of the convolutional autoencoder is a variant of VGG16 convolutional network architecture (Simonyan and Zisserman, 2014) with no 5th convolutional block. The network's basic building blocks are 3D convolutional layers, ReLU (He et al., 2015), batch normalization (Ioffe and Szegedy, 2015), and 3D Max-pooling. The encoder has ten 3D convolutional layers with a batch normalization succeeding each convolutional layer. The network's decoder is designed with ten 3D transpose convolution with each followed by a batch normalization. See Table A.4 and Table A.5 at Appendix.C, for more information on the architectural design.

3.2.2. Adversarial network

The adversarial network consists of a discriminator model that guides the generated output (latent code) of the encoder network to match a random Gaussian prior distribution to ensure that the autoencoder network generates an output (transformed MRI scans) similar to the intended distribution. Let x be the raw MRI scans, z the latent code of the encoder, $p(z)$ be an arbitrary Gaussian prior to being

imposed on the latent code as the real input to the discriminator, $q(z)$ be an aggregated posterior generated by the encoder distribution $q(z|x)$, and $p(x|z)$ the decoding distribution. Here $q(z|x)$ follows a Gaussian distribution whose mean and variance is predicted by the encoder network, $z \sim \mathcal{N}(\mu(x), \sigma(x))$.

The adversarial autoencoder's training phase is regularized by matching the aggregated posterior to the arbitrary Gaussian prior. The adversarial network's generator is the encoder of the autoencoder. The encoder ensures the aggregated posterior distribution will trick the discriminative adversarial network into believing that the latent code comes from the real Gaussian prior distribution. This encoder operation allows robust convergence of the network and ensures that the network generates the best-desired MRI outputs. The cost function used to train the discriminator, or the adversarial loss is

$$L_{Adver} = E_{x \sim p_{data}(x)}(\log D(x)) + E_{z \sim p_z(z)}(\log(1 - D(G(z)))) \tag{5}$$

where $D(x)$ denotes the probability that a point is from the prior distribution, and $G(z)$ maps samples z from the prior $p(z)$ to the new MRI space. The discriminator network is composed of thirteen 3D convolutional layers. We train the autoencoder and the adversarial network simultaneously with the Adam Optimizer (Da, 2014). Although, various optimizations like Sobolev gradient based methods (Goceri, 2019c,b) have been applied in some recent works, we applied Adam to provide efficiency in terms of computational cost.

3.3. Classification with convolutional residual attention network

Inspired by the recent application of attention mechanisms (Bahdanau et al., 2014; Yang et al., 2016; Yao et al., 2020) in deep neural networks and AD diagnosis, we propose CRAT for Alzheimer's disease classification. This convolutional network adopts a mixed attention mechanism with residual networks (He et al., 2016), which significantly affect the performance of a classifier (Goceri, 2019a), in classifying Alzheimer's disease using the new transformed raw MRI scans. Incorporating the attention mechanism (Fig. 3(a)) allows the model to focus specifically on only important AD features to discriminate and attend to the prediction task's relevant information. Soft attention (Bahdanau et al., 2014), a type of attention, calculates a contextual weight over all the model's input features. The calculated weights represent the relative importance of each feature for the given classification tasks. An overall representation of the input is then computed with the weights as a weighted combination of all the input features. During prediction, attention weights with greater values are given higher priorities in determining the corresponding significant
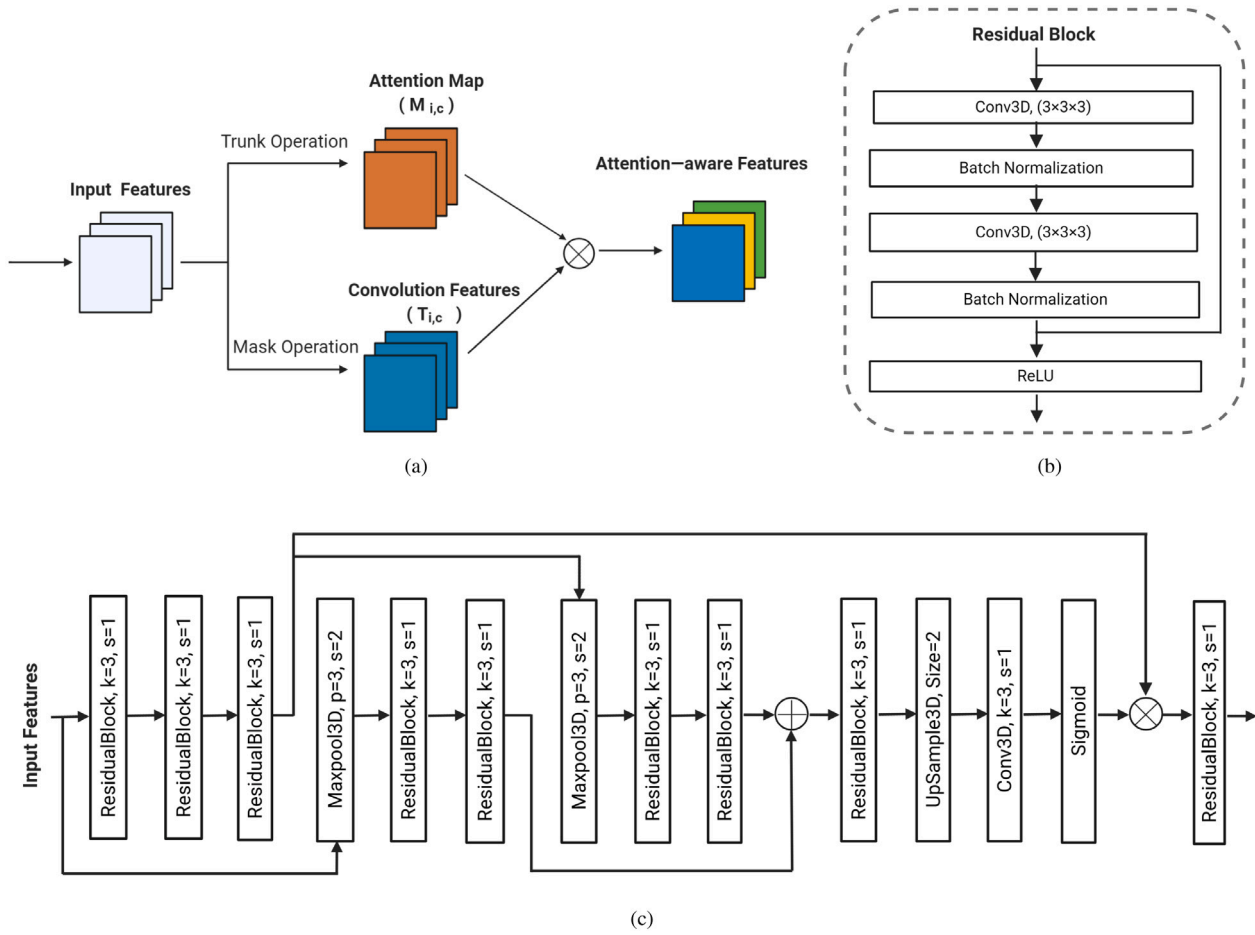
Fig. 3. Residual attention network block: (a) The attention-based mechanism in our network. (b) Residual block in the network. (c) The simplified single attention module in the network. The k and s notation represents the kernel size and Stride Size of layer respectively. The Residual block is made of two 3D convolutional layers, two batch normalization to handle overfitting in the network, and a rectifier activation function (ReLU) as the non-linear function.

input feature. The proposed model, CRAT, comprises multiple attention modules (Fig. 3(c)) that generate attention-aware discriminative features for AD diagnosis' binary classification tasks.

The 3D supervised convolutional residual attention network is modeled by heaping multiple attention modules. It is worth noting that the attention architecture is composed of several residual blocks (Fig. 3(b)), which captures inter-layer information. Each attention module consists of two branches: mask branch, $M(x)$, and trunk branch, $T(x)$. The mask branch's function is to down-sample the MRI-input volume using the operations of average pooling or max pooling and up-sample the output of the down-sampled. The output of the mask operation is passed through a sigmoid activation. This operation indicates the importance of each MRI voxel. The trunk branch handles feature processing and extraction. Mathematically, the operation of the convolutional attention mechanism is given as

$$F_{i,c} = M_{i,c}(x) \times T_{i,c}(x) \tag{6}$$

where $F_{i,c}(x)$ is the output of the attention mechanism, $T_{i,c}(x)$ the output of the trunk branch, $M_{i,c}(x)$ the output of the mask branch, and i ranges overall spatial positions (x, y, z) and $c \epsilon \{1, \ldots, C\}$ is the channel's index. The soft mask branch has the potential of breaking the good property of the trunk branch. A typical example is the identity mapping of the residual block. To overcome this, we modify the attention module operation in Eqs. (6) to (7) as proposed in (Wang et al., 2017).

$$F_{i,c} = (1 + M_{i,c}(x)) \times T_{i,c}(x) \tag{7}$$

$M(x)$ has a range of 0 to 1, and $F(x)$ will approximate original features $T(x)$. During training, the parameters present in the two

branches of the attention module is updated at the back-propagation step. The proposed convolutional residual attention architecture for AD classification is constructed with twenty-one residual blocks. Out of the twenty-one residual blocks, eighteen were used in building the attention modules; each single attention module having nine residual blocks. From a broader perspective, the network comprises two attention networks and three residual blocks. Three parameters: f, m, and l were used for the design of the attention module. The parameter f represents the number of residual preprocessing units before branching into the mask branch and trunk branch, m represents the number of residual units in the trunk branch, and l represents the number of residual units between adjacent pooling layer in the mask branch. The soft mask residual unit and its corresponding trunk branches have the same number of channels. The attention module is designed to reduce noise in the MRI scans, which is not preprocessed, while keeping discriminative information by applying the dot product between input features and the soft mask. Further information on the convolutional residual soft attention network is shown in Fig. 4 and Table A.6 at Appendix.C.

3.4. Exploration of critical brain regions for classification with Class Activation Mapping (CAM)

We utilized the architectural properties of our 3D CRAT in the AD classification tasks to visualize the activations of the last convolutional layer to indicate the brain regions relevant for AD classification. In visualizing the activations for the network's decision, we explored class activation mapping (CAM) (Zhou et al., 2016; Budding et al.,
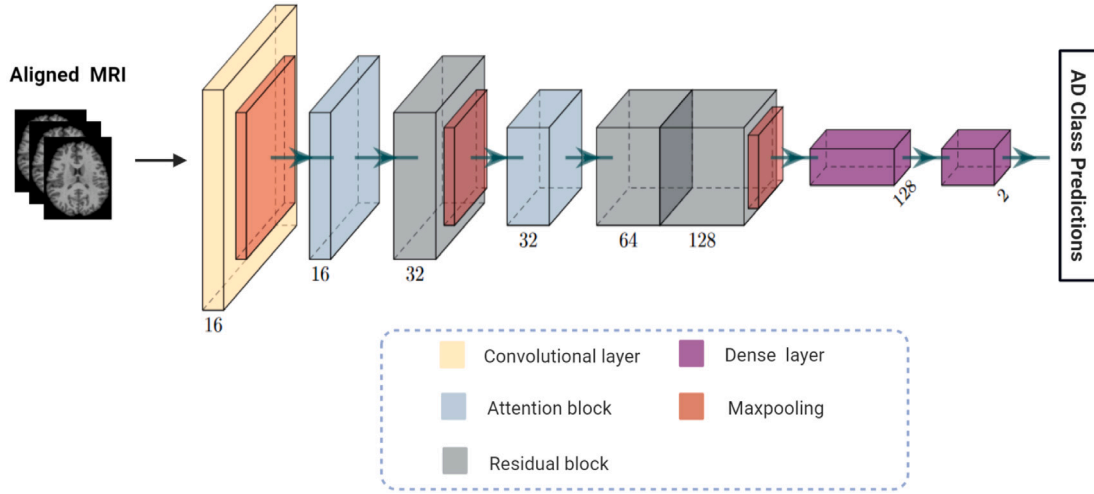
Fig. 4. The convolutional attention network for the AD classification tasks. The aligned raw MRI scans from the convolutional adversarial network is passed to the network.

2020), a technique that incorporates a global average pooling (GAP) layer succeeding the last convolutional layer in any image classification task. We visualize the activations to interpret our proposed model's robustness and performance without treating it as a black-box. This technique provides remarkable localization performance on discriminative features in the images by generating heat maps to buttress the network's performance. Specifically, we applied the CAM to highlight the parts of the brain that are discriminative for AD classification and medically validate our proposed method's robustness. The CAM technique is extended to a 3D architecture to produce the activations in the AD prediction task. The class activation map harnesses the activations produced by the last convolutional layers in visualizing the discriminative features. The method projects the class weights of the output layer onto the activation maps in the last convolutional layer. Furthermore, a weighted sum of the features in the last convolutional layer generates the activations. In implementing the 3D CAM for this work, we modified the last block of the CRAT architecture by replacing the max-pooling layer with the global average layer to have the desired architecture in generating a class activation map. Mathematically, for a given MRI image, let $f_n(x, y, z)$ denotes the activation of the unit n at the last convolutional layer at a spatial location (x, y, z). At the unit n, the global average pooling F_n is $\sum_{(x,y,z)} f_n(x, y, z)$. For a given AD class c, we calculate the softmax S_c as $\sum_n w_n^c F_n$ where w_n^c is the corresponding weight for AD class c, at unit n. Specifically, w_n^c shows the importance of F_n, given an AD class. Eventually, we calculate the output of the softmax for a specified AD class c, O_c as $\frac{exp(S_c)}{\sum_c exp(S_c)}$. Substituting F_n into AD class score S_c, we derive

$$S_c = \sum_n w_n^c \sum_{x,y,z} f_n(x, y, z) = \sum_{x,y,z} \sum_n w_n^c f_n(x, y, z) \quad (8)$$

We finally define the class activation map A_c for a given AD class as

$$A_c(x, y, z) = \sum_n w_n^c f_n(x, y, z) \quad (9)$$

Intuitively, $\sum_{x,y,z} A_c(x, y, z)$ and $A_c(x, y, z)$ highlight the significant activations at the spatial location (x, y, z) of the MRI image to interpret the importance of regions in predicting AD. After incorporating the 3D CAM in our AD prediction model, we overlayed the 3D CAM on the MNI152 brain template (Evans et al., 2012; Brett et al., 2001).

3.5. MRI alignment with Hausdorff distance

We also measured the average accuracy of the alignment between the outputs of 3D CAAE and raw MRI scans with different acquisition protocols by extracting the edges of the raw scans and the common aligned space, respectively, using the Canny edge detector (Ding and Goshtasby, 2001) for the assessment of the MRI alignment (Tang et al., 2000; Fedorov et al., 2008; Agaian and Almuntashri, 2009; Othman et al., 2009; Archip et al., 2007). The alignment assessment invariably measures the degree of variations between MRI scans. The Hausdorff distance (HD) metric (Huttenlocher et al., 1993; Morain-Nicolier et al., 2007; Fedorov et al., 2008) is then applied to the recovered edges to evaluate the accuracy of MRI alignment as an estimate for registration accuracy. The Hausdorff distance defines the maximum distance of a set to the nearest points in the other set. The Hausdorff distance is considered as an image similarity metric. We calculated the average HD between the two-point sets or images across all MRI scans in this work. Mathematically, we define Hausdorff distance from set A to set B, $H(A, B)$ as a directed HD between the two sets of points, $h(A, B)$ representing the maximum distance from any of the points in set A to set B.

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (10)$$

where

$$h(A, B) = \max_{a \in A} \{ \min_{b \in B} \{ d(a, b) \} \} \quad (11)$$

where a and b are the points for A and B, respectively, and $d(A, B)$ is the Euclidean distance between the two points. In this work, A and B represent two different MRI images used to calculate the alignment accuracy. We then calculated the HD between the two sets of points representing the edges, which represent the accuracy of the alignment. The ideal case for no variation or perfect alignment is when the HD value is zero.

It is worth noting that the same algorithm for the 2D canny edge detection was extended for the 3D scenario. Thus, the edge detection is performed in 2D for each MRI slice and the final results for each slice put together to form a 3D volume. The edges are superimposed on the original 3D MRI data volume. Additional information could be referenced from the work (Fedorov et al., 2008) in performing the detection of edges in 3D MRI images.

4. Experiments

4.1. Experimental settings

The performance of our automatic multi-center variability handling approach for predicting Alzheimer's disease was evaluated by a 5-fold stratified cross-validation. We partitioned 1125 MRI raw input scans into a training dataset (N = 338) and a test dataset (N = 37) for each of the MRI data labels (AD, MCI, HC). We trained and evaluated
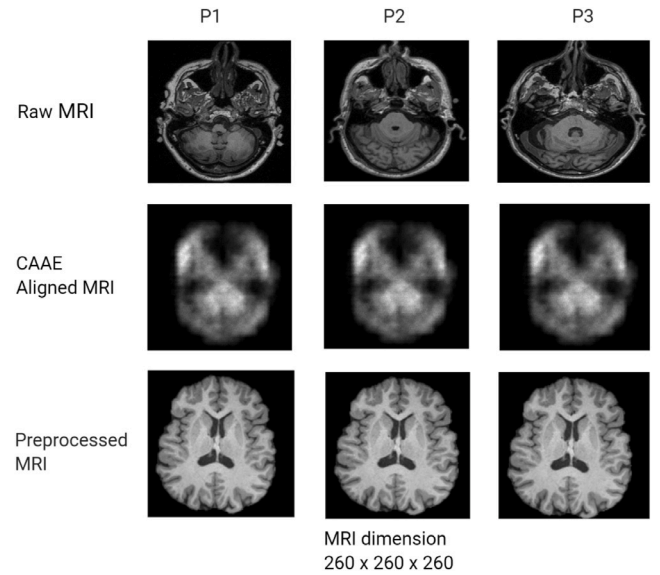
our method on three binary classifiers: AD vs. HC, AD vs. MCI, and MCI vs. HC. Specifically, the dataset was randomly partitioned into a training and a hold-out test set, where 10% or 111 MRI scans of all subjects were assigned to the test set. In brief, the experimental setting comprises validation and testing. In the validation stage, we performed the 5-fold stratified cross-validation to help tune the hyperparameter whiles training the CRAT model and evaluated our model on the test dataset. We followed the data augmentation procedure for the training set. The model was then trained and validated using the training dataset and finally tested using the test dataset. To prevent data leakage, we maintain the same data splitting strategy both for the CAAE and the CRAT at the very beginning of the experiments. For a complete cross-validation experiment, the model was reset, and the best checkpoints were saved for testing. Sensitivity and specificity were also calculated on the test dataset. Both the 3D CAAE and 3D CRAT models are trained for 150 epochs to minimize the softmax cross-entropy loss using the Adam Optimizer. We also performed an ablation study to ascertain the effect on the proposed model without the CAAE component (See Appendix.D). Furthermore, we performed additional experiments to verify the effectiveness of the proposed method by testing the trained model on two external datasets. The training was done with a batch size of five because of the high dimensionality of the MRI scans ($260 \times 260 \times 260$). Learning rates of 0.0001 and 0.00001 were used for the CAAE and CRAT training, respectively. We applied a dropout of 0.8 as the probability of retaining the output of each node in the hidden layers of the network to prevent over-fitting during training. In this current research, the soft attention hyper-parameters used for this experiment are $\{f = 1, m = 2, l = 1\}$.

We developed the deep learning framework using Tensorflow 1.15 python version. The experiments were conducted on a machine running Linux with three Xeon E5-2678 V3 12 core CPU, processor speed of 2.50 GHz, 125 GB main memory and Nvidia Geforce RTX 2080Ti GPU of 11 GB memory. In ensuring computational memory efficiency, we did not load all the MRI data into memory, but only loaded the MRI data batch into memory during the mini-batch training.

4.2. Comparison methods

To evaluate the discriminative ability of our proposed approach, we thoroughly compared it with five baseline classification models. Specifically, sparse autoencoders and 3D Convolutional Neural Networks (Payan and Montana, 2015), 3D Convolution Neural Network (Luo et al., 2017), Random Forest (RF) (Maggipinto et al., 2017), Support Vector Machine (SVM) (Ortiz et al., 2015), Naïve Bayes (NB) (Gupta et al., 2019), Transfer Component Analysis (TCA) (Pan et al., 2010), Subspace Alignment (SA) (Fernando et al., 2013), and Correlation Alignment (CORAL) (Kumar et al., 2017) were selected. TCA, SA, and CORAL are typical hand-crafted feature-based domain adaptation methods, which learn based on the MRI image features with different domain protocols. We used both the raw MRI scans and the MRI scans' preprocessed form for all baseline methods. The preprocessing pipeline (including skull stripping, registration, normalization, segmentation, etc.) applied on the MRI scans for these baseline methods are provided in the Appendix.A. We further evaluated the average variation between the MRI scans by computing the alignment between the aligned outputs of the CAAE and raw MRI scans with different scanner acquisition protocols using the Canny edge detector and Hausdorff distance.

As incentive methods, we also reported the classification performance of selected statistical machine learning classifiers (SVM, RF, and NB) on the discriminative latent representation/embeddings of the CAAE instead of the full 3D aligned MRI scans, and the classification performance on the manually preprocessed MRI scans with our proposed method. Although, in this current research, our aim is to automatically aligned the raw multi-center MRI scans in its 3D space without any intermediate representations, but 3D MRI data has the problem of curse of dimensionality, which hold a lot of indiscriminative

Fig. 5. Multi-center MRI alignment mapped onto a common space with the CAAE. MRI scans with different acquisition protocols (P1, P2, P3), specifically the repetition time parameter, slice thickness, and the dimension of the scans from different centers, are the raw MRIs at the top. The common aligned MRI scans with the dimension of $260 \times 260 \times 260$ are generated for each varying raw MRI scan from different centers. The common aligned MRI scans by CAAE in the middle also show discriminative features extracted by the convolutional adversarial autoencoder. Lastly, the Preprocessed MRI scans show the MNI152 registered scans with different acquisition protocols.

features which misleads most classification models. For an ideal case, the performance of the incentive methods should surpass or equal the performance of the CRAT on the full CAAE aligned 3D MRI scan outputs.

To investigate how different brain imaging data regions contributed to the classification, class relevance heatmaps with respect to the AD cohorts were obtained. We evaluated the interpretability of our model visually by using the 3D CAM. After incorporating the 3D CAM in our AD prediction model, we overlayed the 3D CAM on the MNI152 brain template.

4.3. MRI alignment evaluation

The transformed aligned raw MRI scans shown in Fig. 5 illustrate our proposed 3D CAAE deep learning model's generalizability in aligning the multi-center MRI scans into a common template. The selected slices of the transformed MRI subgroups showed the common learned generic features amidst the multi-center variability, making it challenging to automatically predict the disease. Our proposed CAAE architecture could capture features related to AD biomarkers and discard almost all other features not related to AD diagnosis. Fig. 5 shows that no typical preprocessing was applied to the multi-center MRI scans before the diagnostics process. The CAAE showed a novel capability of generating features that put the different MRI scans into a common space amidst the multi-center variability. The subsequent task-specific AD prediction model, CRAT, further enhances the three disease subgroups' discriminability.

In evaluating the shared space of the output produced by our CRAT model, we denote the MRI scans captured with different acquisition protocols as P1 (TR = 2400 ms, MRI dimension = $208 \times 240 \times 256$, slice thickness = 2 mm), P2 (TR = 3000 ms, MRI dimension = $166 \times 256 \times 256$, slice thickness = 3 mm), and P3 (TR = 2300 ms, MRI dimension = $176 \times 240 \times 256$, slice thickness = 4 mm) for the different AD classes (see Table A.1 Appendix.A). Table 3 represents the calculated average Hausdorff distance between MRI scans with different acquisition protocols for (1) raw multi-center scans, (2) CAAE

Table 3
Alignment and variation assessment of the CAAE aligned common space with different MRI acquisition protocol (P) for different AD class.

	HD measurement (mm)		
	P1 vs. P2	P1 vs. P3	P2 vs. P3
	AD		
Raw MRI	16.55	16.97	17.26
CAAE aligned common space	4.10	4.32	4.97
MNI152 registered	3.05	3.16	3.60
	MCI		
Raw MRI	16.24	17.66	17.74
CAAE aligned common space	5.19	6.12	6.56
MNI152 registered	4.12	6.00	5.74
	HC		
Raw MRI	16.12	17.14	16.64
CAAE aligned common space	4.20	4.09	4.10
MNI152 registered	3.00	3.16	3.61

AD = Alzheimer's disease patient; MCI = Mild cognitive impairment; HC = Cognitively normal patient; HD = Hausdorff distance; P1 (TR = 2400 ms, MRI dimension = 208 × 240 × 256, slice thickness = 2 mm), P2 (TR = 3000 ms, MRI dimension = 166 × 256 × 256, slice thickness = 3 mm), and P3 (TR = 2300 ms, MRI dimension = 176 × 240 × 256, slice thickness = 4 mm).

aligned common space, and (3) original scans registered to MNI152 template for the different AD class.

The average computed HD values between the various acquisition protocols show that our proposed model's alignments achieved an edge performance over the raw MRI scans. It recorded lower HD values for the three-protocol comparison (4.10 mm, 5.19 mm and 4.09 mm for AD, MCI, and HC, respectively), making it a better choice than the raw MRI scans (16.55 mm, 16.24 mm and 16.12 mm for AD, MCI, and HC respectively). This affirms its role in reducing the variability that exists between the MRI scans from multiple centers. Though our model had a better alignment score over the raw MRI scans, the preprocessed MRI scanned recorded the best alignment score with small margins, as showed in Table 3. Also, we show evidence of the varying feature distribution among the MRI acquisition protocols with the kernel density estimation (KDE) (Kim and Scott, 2012) plot. Specifically, we used the KDE method to visualize the distribution observations of the raw MRI scans before and after the CAAE for the various acquisition protocols. Regarding Fig. 6, it could be seen that the relative difference between the density feature distribution for each protocol in the CAAE space is smaller and more uniform as compared to the raw MRI. Even though the KDE of the preprocessed MRI is better than our proposed method, it possesses the capability of learning a near to perfect common transformation space for the raw MRI scans. Most importantly, these new transformation contains AD discriminative features for the robust prediction of the disease.

4.4. Multicenter automatic Alzheimer's disease classification

We analyzed the dependency on using multi-center raw MRI scans without typical preprocessing and the efficiency of our proposed 3D CNNs on Alzheimer's disease classification. The results in Table 4 show that our deep learning model accurately classifies AD on multi-center raw MRI scans. From the 2-way classification tasks results (accuracy: 91.5%, 90.0%, and 88.1% for AD vs. HC, AD vs. MCI, and MCI vs. HC, respectively for the full 3D aligned RAW MRI scan, and best accuracy: 97.23%, 96.30%, and 96.10% for AD vs. HC, AD vs. MCI, and MCI vs. HC, respectively with the latent representations of the CAAE with the statistical machine learning classifiers), it is evident that the extracted features by the attention convolutional layer possess strong discriminative power.

We evaluated the three binary classification tasks' performance using three main metrics: accuracy, sensitivity, and specificity. We found out that our model achieved the best performance in all comparison with the five-baseline machine learning algorithms with both

(a) Raw MRI

(b) CAAE

(c) Preprocessed MRI

Fig. 6. Kernel density estimation feature plot for the different acquisition protocols for the raw MRI scans, output of CAAE and preprocessed scans.

the common aligned feature space extracted by the 3D CAAE and the preprocessed MRI scans. Due to the nature of traditional machine learning algorithms (SVM, Naïve Bayes, and Random Forest) needing feature-engineered input because of the MRI scans' complexity, the traditional classifiers did not exhibit excellent performance both on the full raw and preprocessed 3D MRI scans. Our model, specifically the convolutional residual attention network, made complete use of the AD-related information and utilized the new common aligned space

Table 4

Performance (%) of the proposed method and selected baseline methods. The values in the parentheses are the standard deviations of each respective metric. Acc = Accuracy; Sen = Sensitivity; Spe = Specificity; Avg = Average; RF = Random Forest; NB = Naïve Bayes; SAE = Sparse Autoencoder, CNN = Convolutional neural network, TCA = Transfer Component Analysis, SA = Subspace Alignment, CORAL = Correlation Alignment.

	Method	AD vs. HC				AD vs. MCI				MCI vs. HC			
		Acc.	Sen.	Spe.	AUC	Acc.	Sen.	Spe.	AUC	Acc.	Sen.	Spe.	AUC
Preprocessed													
Payan and Montana (2015)	SAE	92.0	91.89	94.59	92.46	90.0	89.20	86.48	88.16	87.90	85.90	82.63	84.35
		(0.10)	(0.07)	(0.08)	(0.05)	(0.15)	(0.11)	(0.08)	(0.06)	(0.05)	(0.06)	(0.05)	(0.10)
Luo et al. (2017)	3D CNN	91.0	90.51	91.78	88.74	88.43	87.0	85.65	86.05	85.12	84.90	81.69	84.10
		(0.08)	(0.05)	(0.07)	(0.06)	(0.09)	(0.10)	(0.10)	(0.04)	(0.05)	(0.07)	(0.06)	(0.08)
Ortiz et al. (2015)	SVM	86.49	88.20	87.46	86.96	83.0	84.0	82.41	85.16	78.16	78.90	77.10	79.92
		(0.06)	(0.03)	(0.04)	(0.04)	(0.08)	(0.09)	(0.01)	(0.12)	(0.07)	(0.09)	(0.06)	(0.07)
Gupta et al. (2019)	NB	85.13	83.20	84.0	85.15	82.34	81.45	80.0	84.64	80.19	81.32	80.20	79.43
		(0.06)	(0.05)	(0.05)	(0.03)	(0.04)	(0.07)	(0.05)	(0.10)	(0.06)	(0.03)	(0.04)	(0.08)
Maggipinto et al. (2017)	RF	83.78	82.20	83.10	85.21	81.56	80.49	79.78	83.49	81.49	80.90	79.69	80.14
		(0.09)	(0.06)	(0.10)	(0.06)	(0.06)	(0.07)	(0.05)	(0.06)	(0.03)	(0.04)	(0.07)	(0.03)
Pan et al. (2010)	TCA	88.39	86.72	88.10	86.42	85.52	82.04	80.92	84.95	83.10	80.35	82.27	82.51
		(0.05)	(0.08)	(0.06)	(0.01)	(0.05)	(0.03)	(0.04)	(0.06)	(0.04)	(0.05)	(0.06)	(0.05)
Fernando et al. (2013)	SA	89.64	88.92	89.35	88.24	86.03	85.13	86.43	85.01	83.51	81.0	83.01	81.82
		(0.06)	(0.05)	(0.01)	(0.05)	(0.06)	(0.07)	(0.04)	(0.05)	(0.04)	(0.03)	(0.06)	(0.02)
Kumar et al. (2017)	CORAL	88.11	85.04	86.41	86.04	84.67	83.21	84.06	84.63	82.93	79.12	80.14	80.45
		(0.04)	(0.04)	(0.03)	(0.04)	(0.05)	(0.03)	(0.05)	(0.03)	(0.04)	(0.06)	(0.06)	(0.05)
Ours	CAAE + CRAT	**93.10**	**95.40**	**96.11**	**94.86**	**91.41**	**92.32**	**93.12**	**92.20**	**89.26**	**89.45**	**91.13**	**90.23**
		(0.07)	(0.06)	(0.06)	(0.04)	(0.07)	(0.09)	(0.05)	(0.04)	(0.05)	(0.06)	(0.09)	(0.09)
Raw MRI													
Payan and Montana (2015)	SAE	82.43	80.14	82.17	79.06	81.56	83.20	80.21	79.83	76.67	75.34	73.17	74.37
		(0.03)	(0.06)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.03)	(0.06)	(0.04)	(0.05)	(0.07)
Luo et al. (2017)	3D CNN	81.08	80.01	80.90	78.27	80.73	79.24	78.48	79.13	73.05	72.27	72.18	73.47
		(0.06)	(0.07)	(0.05)	(0.06)	(0.08)	(0.09)	(0.06)	(0.07)	(0.04)	(0.04)	(0.06)	(0.04)
Ortiz et al. (2015)	SVM	74.0	73.23	74.0	75.06	72.0	72.34	73.12	73.38	42.0	51.11	49.70	54.07
		(0.05)	(0.08)	(0.08)	(0.07)	(0.04)	(0.06)	(0.06)	(0.05)	(0.06)	(0.04)	(0.06)	(0.05)
Gupta et al. (2019)	NB	75.0	74.12	75.34	75.18	68.0	69.0	67.23	72.15	46.0	56.78	52.60	55.62
		(0.06)	(0.04)	(0.07)	(0.06)	(0.04)	(0.08)	(0.05)	(0.04)	(0.05)	(0.07)	(0.06)	(0.06)
Maggipinto et al. (2017)	RF	74.0	72.78	73.50	74.27	66.0	67.15	65.20	69.45	55.0	54.13	53.56	56.69
		(0.07)	(0.10)	(0.09)	(0.08)	(0.05)	(0.06)	(0.05)	(0.05)	(0.07)	(0.08)	(0.08)	(0.06)
Pan et al. (2010)	TCA	75.61	76.01	77.52	75.82	70.25	69.72	70.04	71.88	62.57	61.46	63.91	60.14
		(0.03)	(0.05)	(0.04)	(0.05)	(0.07)	(0.04)	(0.05)	(0.05)	(0.02)	(0.04)	(0.05)	(0.01)
Fernando et al. (2013)	SA	77.04	76.01	77.52	76.02	73.12	72.85	73.27	73.41	64.70	64.03	66.42	62.91
		(0.02)	(0.05)	(0.07)	(0.05)	(0.03)	(0.07)	(0.05)	(0.05)	(0.06)	(0.02)	(0.06)	(0.03)
Kumar et al. (2017)	CORAL	75.26	73.62	75.20	75.79	69.80	69.24	71.50	70.84	61.08	62.20	63.28	59.86
		(0.09)	(0.06)	(0.10)	(0.06)	(0.06)	(0.07)	(0.05)	(0.06)	(0.03)	(0.04)	(0.07)	(0.03)
Ours	CAAE + CRAT	**91.90**	**94.596**	**89.18**	**91.51**	**90.05**	**91.89**	**89.18**	**88.94**	**88.10**	**86.48**	**83.78**	**86.19**
		(0.04)	(0.0)	(0.05)	(0.04)	(0.06)	(0.05)	(0.07)	(0.04)	(0.04)	(0.03)	(0.05)	(0.05)
Raw CAAE Latent													
Ours (Latent)	CAAE + SVM	**97.23**	**96.02**	**98.15**	**96.79**	**96.30**	**96.18**	**97.52**	**95.84**	94.37	93.15	95.10	95.16
		(0.06)	(0.07)	(0.06)	(0.05)	(0.05)	(0.07)	(0.04)	(0.05)	(0.03)	(0.04)	(0.05)	(0.04)
Ours (Latent)	CAAE + NB	96.12	95.12	97.31	96.10	95.0	93.74	94.30	94.21	95.0	95.10	95.89	95.01
		(0.03)	(0.05)	(0.03)	(0.03)	(0.02)	(0.04)	(0.06)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)
Ours (Latent)	CAAE + RF	96.12	95.23	97.03	96.26	95.04	95.10	96.45	95.13	**96.10**	**96.40**	**97.43**	**95.42**
		(0.05)	(0.06)	(0.04)	(0.07)	(0.08)	(0.03)	(0.06)	(0.07)	(0.04)	(0.06)	(0.05)	(0.06)

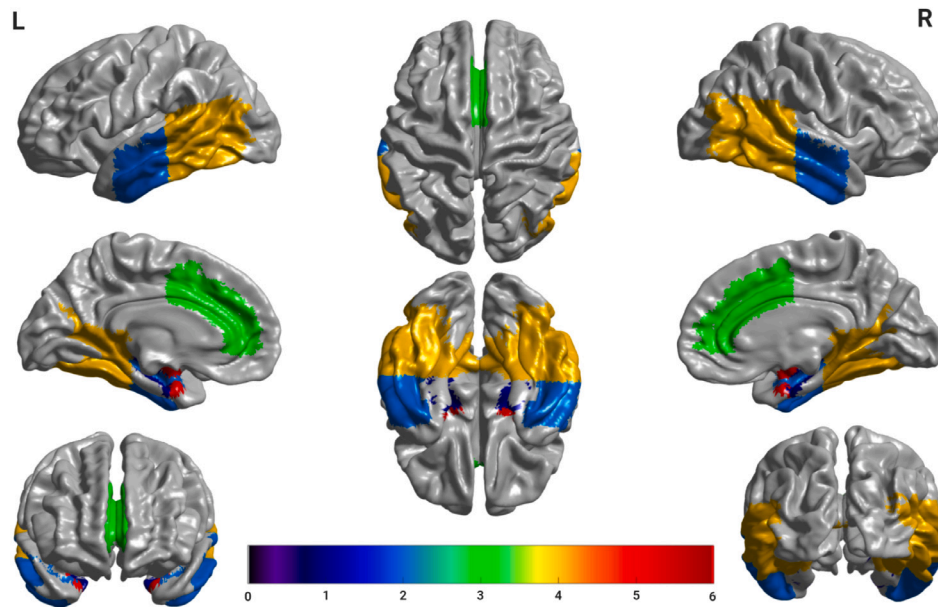(in the 3D space) giving it superior performance over the traditional classifiers.

4.5. Model interpretation and visualization

Fig. 7 shows the captured overlapping salient discriminative regions for AD, MCI, and HC by applying the 3D CAM to our 3D CRAT for the three binary classification tasks (AD vs. HC, AD vs. MCI, and MCI vs. HC). Aforementioned in section 3.4, we modified the last block of the CRAT architecture by replacing the max-pooling layer with the global average layer to have the desired architecture in generating the class activation map for each binary classification task. The heatmaps generated by the 3D CAM differentiate the importance of the different brain regions contributing to the prediction of each AD group. The 3D CAM generated by each AD binary classification task is then overlayed on the MNI152 template as discussed in Section 3.4. The distinct brain regions generated by each classification task's 3D CAM are compared with the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). Ten representative regions (Squire et al., 2007; Lerch et al., 2008; Eskildsen et al., 2015) of interest in the

AAL atlas are compared with the generated heatmaps. The following regions of interest were used: left and right hippocampus, amygdala, left and right middle temporal gyrus, left and right superior temporal pole, parahippocampus, insula, and precuneus. See Appendix.B for the ten representative regions. Specifically, six overlapping regions were identified by our model relating to AD, when compared with all the three binary classification task overlays. The captured regions include the left and right hippocampus, parahippocampus, amygdala, left and right middle temporal gyrus in predicting AD, MCI, and HC. See Table 5 for significant brain regions contributing to Alzheimer's disease's diagnosis in order of importance. We observed that the areas of the hippocampus stand out as the most discriminative parts. It should be noted that the activated heatmap is upsampled from a $22 \times 22 \times 22$ coarse heatmap.

4.6. Ablation study on external datasets

To further verify the proposed method's effectiveness, we tested the trained models on two popular external AD datasets. The datasets are the Australian Imaging, Biomarker and Lifestyle Flagship Study of

Fig. 7. Salient 3D attention-map overlapping regions relating to Alzheimer's disease classification. An Increasing colormap scale indicates that the region is more significant for classifying the disease. Heatmaps from CAM were generated and overlaid on the Montreal Neurological Institute's standard brain for the three binary classification tasks (AD vs. HC, AD vs. MCI, MCI vs. HC). Six salient overlapping regions of interest (left and right hippocampus, parahippocampus amygdala, and left and right superior temporal pole) were observed for AD prediction. The red color indicates the brain morphometric region playing the most pivotal role in the prediction of AD. L = Left; R = Right.

Table 5
Ordered significant regions from the Class Activation Map related to the prediction of AD patients, mild cognitive impairment patients, and cognitively normal individuals. L = Left; R = Right.

No.	AD vs. HC	AD vs. MCI	MCI vs. HC
1	L Hippocampus	L Hippocampus	Parahippocampus
2	R Hippocampus	R Hippocampus	L Hippocampus
3	Parahippocampus	Parahippocampus	R Hippocampus
4	L. Temporal pole	L. Temporal pole	L. Temporal pole
5	R. Temporal pole	R. Temporal pole	R. Temporal pole
6	Amygdala	Amygdala	Amygdala
7	Precuneus	Precuneus	–
8	Insula	Insula	–
9	Middle temporal gyrus	Middle temporal gyrus	–
10	Cingulate gyrus	Cingulate gyrus	–

Ageing (AIBL) (Ellis et al., 2010) and the National Alzheimer's Coordinating Center (NACC) (Beekly et al., 2004). For each external dataset, we used 195 subjects where 65 belong to each of the classifying groups. The performance is illustrated in Table 6. It can be seen that although the prediction on the preprocessed MRI scans achieved better results than the raw scans aligned to a common space, the performance is still comparable. The experimental results prove that the convolutional adversarial model and the attention mechanism can predict AD without a robust preprocessing pipeline. Also, the proposed model is likely to have improved performance when trained from scratch on the external dataset and a larger dataset.

5. Discussion

This current study proposed a new model for handling variations in multi-center AD classification with raw MRI scans without any conventional preprocessing. Compared with other baselines, results have demonstrated the out-performance of capturing the significant brain regions for AD on raw MRI scans without typical preprocessing.

5.1. Towards an end-to-end approach for multicenter analysis

Alzheimer's disease classification models using MRI scans are mostly constructed using multi-center databases. Such datasets pose additional

challenges for statistical analysis but offer opportunities for model generalizability to different centers. AD classification using deep learning methods tailored to multi-center raw neuroimaging data are understudied. There is a lack of transparency in developing a generalized model using preprocessed MRI scans. Hence, multi-center AD classification approaches using raw neuroimaging data to handle center variability have untapped potential. It is not surprising that many AD researchers use multi-center neuroimaging datasets as a substrate in developing deep learning prediction models for the disease. The caveat to these previous researches (Zhang et al., 2021b; Hosseini-Asl et al., 2016; Li et al., 2019b,a; Jo et al., 2019a) is the incorporation of typical neuroimaging preprocessing pipeline to the disease characterization process and making the generalizability of the model depend on many factors such as the wholesome quality of the preprocessing step used, the AD domain expert involved, and the manually extracted region of interest (ROI) related to the disease. For an AD classification model to be more useful, it must be reliable in new centers (different clinics or hospitals), time-saving, and involving fewer or no domain experts. The fact that variability exists in multi-center MRI scans shed light on the need for a more robust deep learning model to generalize AD prediction irrespective of the neuroimaging data source. Our multi-center variability handling approach enables efficient and reliable AD prediction using multi-center raw MRI scans while eliminating typical neuroimaging preprocessing pipeline. Specifically, our 3D convolutional adversarial autoencoder network can eliminate processes such as skull stripping, head motion correction, registration, and parcellation of brain regions in a typical AD classification pipeline, introduced as a result of multi-center MRI variations. It is worth noting that our approach's automatic variability handling is significant in classifying Alzheimer's disease (Crutch et al., 2012; Ossenkoppele et al., 2015). Hence, automatic multi-center AD prediction is of significant interest if characterized by the skipping of MRI preprocessing while handling the variations that exist. Our experiments also demonstrated the reduction of variation between raw MRI scans and the common aligned space generated by our CAAE by the calculated HD distance metric. The ablation study results (see Table A.8 Appendix.D) show that CAAE plays a critical role in the multi-center variability handling and is also the Achilles' heel of the overall performance of the proposed method. Specifically, when CAAE is removed from the proposed method, the

Table 6

Ablation performance (%) of the trained model on AIBL and NACC datasets. Acc = Accuracy; Sen = Sensitivity; Spe = Specificity.

	AD vs. HC				AD vs. MCI				MCI vs. HC			
	Acc.	Sen.	Spe.	AUC	Acc.	Sen.	Spe.	AUC	Acc.	Sen.	Spe.	AUC
AIBL												
Preprocessed MRI + CAAE + CRAT	**83.42**	**72.09**	**88.75**	**89.02**	**80.90**	**70.34**	**87.10**	**86.02**	**73.59**	**69.36**	**84.08**	**83.71**
Raw MRI + CAAE + CRAT	79.13	69.82	85.73	87.43	76.52	67.61	85.37	85.91	70.67	67.35	83.04	81.60
NACC												
Preprocessed MRI + CAAE + CRAT	**81.47**	**73.51**	**89.34**	**88.52**	**77.88**	**70.13**	**86.94**	**84.81**	**72.17**	**67.91**	**83.34**	**81.98**
Raw MRI + CAAE + CRAT	78.56	71.20	86.70	86.03	75.45	67.87	84.02	82.05	69.73	66.0	80.97	78.05

classification results drops drastically. Though our model had a better alignment score over the raw MRI scans, the preprocessed MRI scanned recorded the best alignment score, as showed in Table 3 and needs more future research works in this domain for a complete multi-center MRI variation handling with deep learning techniques.

5.2. AD relevant regions with attention mechanism

In evaluating our results, we report four established metrics: accuracy, specificity, sensitivity, and AUC. We classified AD vs. HC, AD vs. MCI, and MCI vs. HC with excellent accuracy of 91.5%, 90.0%, and 88.1%, respectively for the raw fully aligned MRI scans (3D space), based on the train-and-test setting. From the baseline approaches implemented, the classification accuracy for AD vs. HC, the work of Payan and Montana (2015) performed slightly better than our approach when using the fully aligned raw MRI scans in their 3D space, but it is probably from the prior extraction of ROI from the MRI scans before passing it through their deep learning architecture. When our proposed two-stage method was applied on the preprocessed MRI scans, the results (accuracy of 93.10%, 91.41%, and 89.26% for AD vs. HC, AD vs. MCI, and MCI vs. HC, respectively) surpassed all the 3D space MRI scans results. Our aligned raw MRI scans results were comparable with the preprocessed form more than the other baseline methods on the 3D space MRI scans.

From Table 4, the classification performance of our proposed method consistently outperforms the conventional handcrafted feature-based harmonization methods (TCA, SA, and CORAL).

We also explored the performance of the three statistical machine learning classifiers (SVM, NB, and RF) on the CAAE latent representations/embeddings of the aligned MRI scans in their one-dimensional space (1D). Surprisingly, the classifiers performed greatly on the latent representations produced by our CAAE model. In Table 4, the classification performance on the latent representations produced the best accuracies of 97.23%, 96.30%, and 96.10% for AD vs. HC, AD vs. MCI, and MCI vs. HC, respectively. The classification performance on the latent representations of the raw aligned MRI scans performed better than the full raw aligned MRI scans in the 3D space, but we are more interested in the alignment capability of our CAAE producing more discriminative features for these classification models. This experiment confirms that MRI scans in the 3D space is hugely affected by the curse of dimensionality problem. Although the aligned MRI scans in the 3D space contains more features, they are very sparse making it difficult for learning models to extract relevant information as compared to its latent representations. The incorporation of the automatic CAAE alignment model provides a promising research path and further works in handling automatic MRI variations in AD diagnosis.

We further established the AD-specific discriminative pattern of atrophy with our deep learning model's classifiers to predict these classes. The residual modules and the attention components of the 3D-CRAT architecture played a pivotal role in discriminating the pivotal regions of the brain for AD prediction. The AD-related brain regions that best-discriminated AD from other groups (MCI, HC) are identified as the hippocampus, parahippocampus, amygdala, anterior cingulate gyrus, medial temporal lobe, and the areas of the subcortical thalamic-basal

ganglia, which is consistent with previous findings of characterizing structural brain changes in AD (Budding et al., 2020; Hao et al., 2020; Oh et al., 2019; Hu et al., 2021; Su et al., 2021).

In general, computational cost is not an issue for our proposed MRI alignment method. The average runtime (in hours) of the typical MRI preprocessing pipeline is 172 h. The runtime for the proposed CAAE method is approximately 108 h making it the best option in respect to runtime and performance. The training of the deep learning CRAT classification model took roughly 17 h.

With the ubiquity of data preprocessing in neuroimaging, almost all previous research works handled the variability between multi-center MRI data in AD prediction by applying the typical preprocessing pipeline. This makes it difficult to make a holistic comparison with our current work. Notwithstanding this difficulty, our recorded results performed well, if not exceptional, from studies in this domain. It is worth noting that the present study is based on a limited number of subjects, and further works using a larger multi-center dataset is necessary for future works. Admittedly, the training of our deep learning model was computationally intensive and time-consuming.

5.3. Nonparametric statistical analysis

To ascertain whether there exist significant differences among the baseline methods, the Friedman ranking test and the Nemenyi post-hoc test (Hodges and Lehmann, 2012; Demšar, 2006), are used to statistically analyze the results for the test prediction accuracy and the HD variation assessment. We first perform the Friedman test on the averaging rankings of each results to calculate their p-values at the 5% significance level. When the calculated p-value is smaller than 0.05, it rejects the null hypothesis that there is no significant difference between the comparison methods. If the null hypothesis is rejected, we then perform the Nemenyi post-hoc test to ascertain the significant differences between these methods. Table 7 shows the average rankings of the Friedman test for the AD prediction tasks for all methods.

We observed that our proposed method has the highest average rankings (best four ranks). The reported p value is $2.41e^{-09}$, which indicates there are significant differences among all the baseline methods at the 5% significance level. This p-value rejects the null hypothesis that there is no statistical difference between the baseline methods since the p-value is smaller than the 5% significance level. Furthermore, we use the Nemenyi post-hoc test to further confirm the significant differences by constructing the critical difference (CD) diagram (Fig. 8) using the average ranks for all methods.

We also performed the statistical analysis of our method with only the preprocessed baseline methods. The Friedman ranking test recorded a p-value of $6.77e^{-06}$, which shows there is a statistical difference between this comparison. The associated Nemenyi post-hoc test is shown in Fig. 9. Figs. 8 and 9 show the post-hoc Nemenyi test produced a critical difference of 17.12 and 10.53 on the average ranks of all baseline methods and preprocessed methods respectively, revealing which algorithms were statistically significantly different from each other. Specifically, it denotes that our proposed method ranked higher than other competing methods.

Finally, we statistically analyzed the HD variation assessment among the MRI scans with different acquisition protocol, as showed in Table 8.
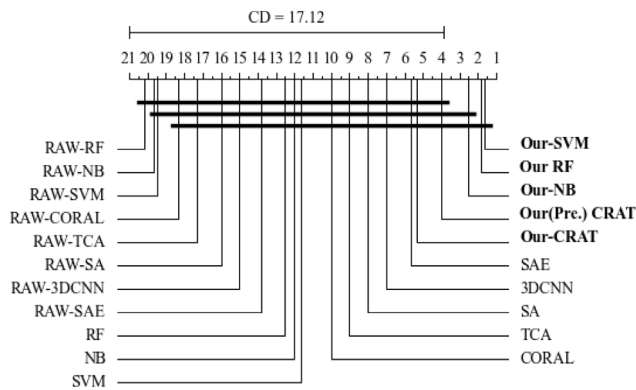
Table 7

Average rankings of the Friedman test of the proposed method and selected baseline methods for AD prediction with p-value $= 2.41e^{-09}$ and Avg. R = Average Rank.

	AD vs. HC.	AD vs. MCI	MCI vs. HC.	Avg. R
Our CRAT	91.90	90.05	88.10	5.33
Our(Pre.) CRAT	93.10	91.41	89.26	4.0
Our SVM	97.23	96.30	94.37	**1.67**
Our NB	96.12	95.0	95.0	2.5
Our RF	96.12	95.04	96.10	1.83
Payan and Montana (2015) SAE	92.00	90.00	87.90	5.67
Luo et al. (2017) 3DCNN	91.00	88.43	85.12	7.0
Ortiz et al. (2015) SVM	86.49	83.0	78.16	11.67
Gupta et al. (2019) NB	85.13	82.34	80.19	12.0
Maggipinto et al. (2017) RF	83.78	81.56	81.49	12.50
Pan et al. (2010) TCA	88.39	85.52	83.10	9.0
Fernando et al. (2013) SA	89.64	86.03	83.51	8.0
Kumar et al. (2017) CORAL	88.11	84.67	82.93	10.0
Raw SAE	82.43	81.56	76.67	13.83
Raw 3DCNN	81.08	80.73	73.05	15.0
Raw SVM	74.0	72.0	42.0	19.50
Raw NB	75.0	68.0	46.0	19.66
Raw RF	74.0	66.0	55.0	20.16
Raw TCA	75.61	70.25	62.50	17.33
Raw SA	77.04	73.12	64.70	16.0
Raw CORAL	75.26	69.80	61.08	18.33

Table 8

Average rankings of the Friedman test variation assessment on HD (mm) for the different acquisition protocols for AD classes with p-value = 0.096.

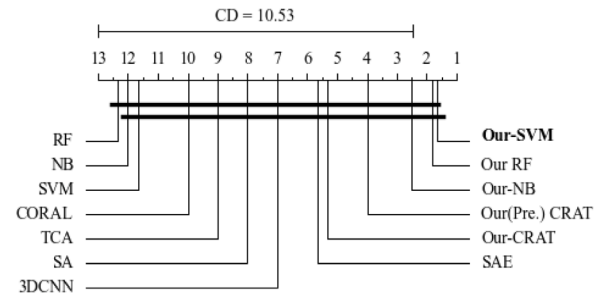|	Raw MRI	CAAE aligned space	MNI152 registered
P1 vs. P2.	48.91	13.49	10.17
P1 vs. P3.	51.77	14.53	12.32
P2 vs. P3.	51.64	15.63	12.95
Avg	50.77	14.38	11.81
Avg. Rank	3.0	2.0	**1.0**

Fig. 8. Nemenyi post-hoc test on all prediction tasks for all baseline methods. Our(Pre)-CRAT is the CRAT method used on the preprocessed MRI scans.

Fig. 9. Nemenyi post-hoc test on all prediction tasks on preprocessed baseline methods.

The *p*-value for the Friedman ranking test is 0.096 on the average rankings. Although the Friedman ranking test does not reject the null hypothesis, the smaller *p*-value indicates the competitiveness of the proposed CAAE model in respect to the manual registered MRI scans.

5.4. Limitation and future works

This study is not without limitations. The proposed approach's limitations are three-fold: (i) The dataset used in this study is size-limited and future works need validation on a vast multi-center database where MRI heterogeneity is likely to increase; (ii) The proposed architecture cannot handle multi-modal Alzheimer's dataset (MRI, PET, clinical records, etc.); and (iii) Although the computational time for the convolutional adversarial autoencoder is better than the typical manual MRI preprocessing pipeline, we admit that it is computationally expensive (time complexity) and needs an improved optimization procedure.

Therefore, in our future works, we would consider the following aspects: (i) Expand the dataset's size and validate our proposed model on a larger multi-center heterogeneous dataset; (ii) Modified our proposed methodology to work on an advanced fusion of multi-modal data set for robust heterogeneity handling; and (iii) Consider how to optimize the proposed method more efficiently.

6. Conclusion

In this paper, we proposed a robust soft-attention-based 3D residual network, which automatically predicts AD on multi-center raw structural brain MRI scans and spares conventional MRI preprocessing effort compared to several recent proposed approaches. The 3D convolutional adversarial autoencoder network enhances the AD features' generalization, captures its biomarkers, and aligns the raw multi-center MRI scans into a common space. Finally, the aligned MRI scans from the unsupervised network are used for AD prediction using the supervised 3D soft attention network on a subset of the ADNI dataset. Also, we showed the robustness of our CAAE aligned method by performing prediction on its latent embedding by using traditional statistical classifiers. The results demonstrated that representational feature extraction, such as anatomical variations, was enhanced in hidden layers of the 3D CAAE by better discriminating between AD, MCI, and HC raw MRI scans without typical user-based preprocessing with our trained classifiers. This research provides novel insights in handling multi-center neuroimaging data variabilities and warrants thorough independent studies in the future. The results have demonstrated the out-performance of capturing the significant brain regions for AD on raw MRI scans without typical preprocessing.

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

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2022.102585.

References

Agaian, S., Almuntashri, A., 2009. Noise-resilient edge detection algorithm for brain MRI images. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 3689–3692.

Archip, N., Clatz, O., Whalen, S., Kacher, D., Fedorov, A., Kot, A., Chrisochoides, N., Jolesz, F., Golby, A., Black, P.M., et al., 2007. Non-rigid alignment of pre-operative MRI, fMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery. Neuroimage 35 (2), 609–624.

Association, A., 2019. 2019 Alzheimer's disease facts and figures. Alzheimer's Dement. 15 (3), 321–387.

Association, A., et al., 2018. 2018 Alzheimer's disease facts and figures. Alzheimer's Dement. 14 (3), 367–429.

Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., Filippi, M., Initiative, A.D.N., et al., 2019. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. NeuroImage: Clin. 21, 101645.

Beekly, D.L., Ramos, E.M., van Belle, G., Deitrich, W., Clark, A.D., Jacka, M.E., Kukull, W.A., et al., 2004. The national Alzheimer's coordinating center (NACC) database: an alzheimer disease database. Alzheimer Dis. Assoc. Disord. 18 (4), 270–277.

Brett, M., Christoff, K., Cusack, R., Lancaster, J., et al., 2001. Using the Talairach atlas with the MNI template. Neuroimage 13 (6), 85.

Budding, C., EitelAlbrecht, F.-P., Ritter, K., 2020. Visualization of deep learning relevance maps for AD detection: Doctor AI: Making computers explain their decisions. Alzheimer's Dement. 16, e037352.

Crutch, S.J., Lehmann, M., Schott, J.M., Rabinovici, G.D., Rossor, M.N., Fox, N.C., 2012. Posterior cortical atrophy. Lancet Neurol. 11 (2), 170–178.

Da, K., 2014. A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7, 1–30.

Ding, L., Goshtasby, A., 2001. On the Canny edge detector. Pattern Recognit. 34 (3), 721–725.

Dong, A., Toledo, J.B., Honnorat, N., Doshi, J., Varol, E., Sotiras, A., Wolk, D., Trojanowski, J.Q., Davatzikos, C., Initiative, A.D.N., 2017. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. Brain 140 (3), 735–747.

Ellis, K.A., Rowe, C.C., Villemagne, V.L., Martins, R.N., Masters, C.L., Salvado, O., Szoeke, C., Ames, D., Group, A.R., 2010. Addressing population aging and Alzheimer's disease through the Australian imaging biomarkers and lifestyle study: Collaboration with the Alzheimer's disease neuroimaging initiative. Alzheimer's Dement. 6 (3), 291–296.

Eskildsen, S.F., Coupé, P., Fonov, V.S., Pruessner, J.C., Collins, D.L., Initiative, A.D.N., et al., 2015. Structural imaging biomarkers of alzheimer's disease: predicting disease progression. Neurobiol. Aging 36, S23–S31.

Evans, A.C., Janke, A.L., Collins, D.L., Baillet, S., 2012. Brain templates and atlases. Neuroimage 62 (2), 911–922.

Fedorov, A., Billet, E., Prastawa, M., Gerig, G., Radmanesh, A., Warfield, S.K., Kikinis, R., Chrisochoides, N., 2008. Evaluation of brain MRI alignment with the robust Hausdorff distance measures. In: International Symposium on Visual Computing. Springer, pp. 594–603.

Fernando, B., Habrard, A., Sebban, M., Tuytelaars, T., 2013. Unsupervised visual domain adaptation using subspace alignment. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2960–2967.

Gilmore, A., Buser, N., Hanson, J.L., 2019. Variations in structural mri quality impact measures of brain anatomy: Relations with age and other sociodemographic variables. Biorxiv 581876.

Goceri, E., 2019a. Analysis of deep networks with residual blocks and different activation functions: classification of skin diseases. In: 2019 Ninth International Conference on Image Processing Theory, Tools and Applications. IPTA, IEEE, pp. 1–6.

Goceri, E., 2019b. CapsNet topology to classify tumours from brain images and comparative evaluation. IET Image Process. 14 (5), 882–889.

Goceri, E., 2019c. Diagnosis of Alzheimer's disease with Sobolev gradient-based optimization and 3D convolutional neural network. Int. J. Numer. Methods Biomed. Eng. 35 (7), e3225.

Goceri, E., 2021. Deep learning based classification of facial dermatological disorders. Comput. Biol. Med. 128, 104118.

Gradin, V., Gountouna, V.-E., Waiter, G., Ahearn, T.S., Brennan, D., Condon, B., Marshall, I., McGonigle, D.J., Murray, A.D., Whalley, H., et al., 2010. Between-and within-scanner variability in the CaliBrain study n-back cognitive task. Psychiatry Res.: Neuroimaging 184 (2), 86–95.

Gupta, Y., Lee, K.H., Choi, K.Y., Lee, J.J., Kim, B.C., Kwon, G.-R., 2019. Alzheimer's disease diagnosis based on cortical and subcortical features. J. Healthc. Eng. 2019.

Hao, X., Bao, Y., Guo, Y., Yu, M., Zhang, D., Risacher, S.L., Saykin, A.J., Yao, X., Shen, L., Initiative, A.D.N., et al., 2020. Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease. Med. Image Anal. 60, 101625.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1026–1034.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hett, K., Ta, V.-T., Oguz, I., Manjón, J.V., Coupé, P., Initiative, A.D.N., et al., 2021. Multi-scale graph-based grading for Alzheimer's disease prediction. Med. Image Anal. 67, 101850.

Hodges, J., Lehmann, E.L., 2012. Rank methods for combination of independent experiments in analysis of variance. In: Selected Works of EL Lehmann. Springer, pp. 403–418.

Hosseini-Asl, E., Keynton, R., El-Baz, A., 2016. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In: 2016 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 126–130.

Hu, J., Qing, Z., Liu, R., Zhang, X., Lv, P., Wang, M., Wang, Y., He, K., Gao, Y., Zhang, B., 2021. Deep learning-based classification and voxel-based visualization of frontotemporal dementia and Alzheimer's disease. Front. Neurosci. 14, 1468.

Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J., 1993. Comparing images using the Hausdorff distance. IEEE Trans. Pattern Anal. Mach. Intell. 15 (9), 850–863.

Iizuka, T., Fukasawa, M., Kameyama, M., 2019. Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies. Sci. Rep. 9 (1), 1–9.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR, pp. 448–456.

Jack, Jr., C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., et al., 2018. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's Dement. 14 (4), 535–562.

Jin, D., Xu, J., Zhao, K., Hu, F., Yang, Z., Liu, B., Jiang, T., Liu, Y., 2019. Attention-based 3D convolutional network for Alzheimer's disease diagnosis and biomarkers exploration. In: 2019 IEEE 16Th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, pp. 1047–1051.

Jo, T., Nho, K., Risacher, S.L., Saykin, A.J., 2019a. P1-398: Multimodal-3DCNN: Diagnostic classification of Alzheimer's disease using deep learning on neuroimaging, genetic, and demographic data. Alzheimer's Dement. 15, P409–P410.

Jo, T., Nho, K., Saykin, A.J., 2019b. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Front. Aging Neurosci. 11, 220.

Kim, J., Scott, C.D., 2012. Robust kernel density estimation. J. Mach. Learn. Res. 13 (1), 2529–2565.

Kumar, D., Kumar, C., Shao, M., 2017. Cross-database mammographic image analysis through unsupervised domain adaptation. In: 2017 IEEE International Conference on Big Data (Big Data). IEEE, pp. 4035–4042.

Lerch, J.P., Pruessner, J., Zijdenbos, A.P., Collins, D.L., Teipel, S.J., Hampel, H., Evans, A.C., 2008. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. Neurobiol. Aging 29 (1), 23–30.

Li, X., Gu, Y., Dvornek, N., Staib, L.H., Ventola, P., Duncan, J.S., 2020. Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results. Med. Image Anal. 65, 101765.

Li, H., Habes, M., Wolk, D.A., Fan, Y., Initiative, A.D.N., et al., 2019a. A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. Alzheimer's Dement. 15 (8), 1059–1070.

Li, H., Wolk, D.A., Fan, Y., 2019b. P4-593: Early prediction of cognitive decline based on brain MRI images using a deep learning survival analysis model. Alzheimer's Dement. 15, P1551–P1552.

Liu, Z., Lu, H., Pan, X., Xu, M., Lan, R., Luo, X., 2022. Diagnosis of Alzheimer's disease via an attention-based multi-scale convolutional neural network. Knowl.-Based Syst. 238, 107942.

Luo, S., Li, X., Li, J., 2017. Automatic Alzheimer's disease recognition from MRI data using deep learning method. J. Appl. Math. Phys. 5 (9), 1892–1898.

Maggipinto, T., Bellotti, R., Amoroso, N., Diacono, D., Donvito, G., Lella, E., Monaco, A., Scelsi, M.A., Tangaro, S., Initiative, A.D.N., et al., 2017. DTI measurements for Alzheimer's classification. Phys. Med. Biol. 62 (6), 2361.

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B., 2015. Adversarial autoencoders. arXiv preprint arXiv:1511.05644.

Masci, J., Meier, U., Cireşan, D., Schmidhuber, J., 2011. Stacked convolutional auto-encoders for hierarchical feature extraction. In: International Conference on Artificial Neural Networks. Springer, pp. 52–59.

Mayerhoefer, M.E., Szomolanyi, P., Jirak, D., Materka, A., Trattnig, S., 2009. Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study. Med. Phys. 36 (4), 1236–1243.

Morain-Nicolier, F., Lebonvallet, S., Baudrier, E., Ruan, S., 2007. Hausdorff distance based 3D quantification of brain tumor evolution from MRI images. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 5597–5600.

Mwangi, B., Ebmeier, K.P., Matthews, K., Douglas Steele, J., 2012. Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. Brain 135 (5), 1508–1521.

Myszczynska, M.A., Ojamies, P.N., Lacoste, A.M., Neil, D., Saffari, A., Mead, R., Hautbergue, G.M., Holbrook, J.D., Ferraiuolo, L., 2020. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nat. Rev. Neurol. 16 (8), 440–456.

Oh, K., Chung, Y.-C., Kim, K.W., Kim, W.-S., Oh, I.-S., 2019. Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning. Sci. Rep. 9 (1), 1–16.

Ortiz, A., Munilla, J., Álvarez-Illán, I., Górriz, J.M., Ramírez, J., Initiative, A.D.N., et al., 2015. Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's disease diagnosis. Front. Comput. Neurosci. 9, 132.

Ossenkoppele, R., Pijnenburg, Y.A., Perry, D.C., Cohn-Sheehy, B.I., Scheltens, N.M., Vogel, J.W., Kramer, J.H., van der Vlies, A.E., Joie, R.L., Rosen, H.J., et al., 2015. The behavioural/dysexecutive variant of Alzheimer's disease: clinical, neuroimaging and pathological features. Brain 138 (9), 2732–2749.

Othman, Z., Haron, H., Kadir, M.R.A., Rafiq, M., 2009. Comparison of Canny and Sobel edge detection in MRI images. Comput. Sci. Biomech. Tissue Eng. Group Inf. Syst. 133–136.

Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q., 2010. Domain adaptation via transfer component analysis. IEEE Trans. Neural Netw. 22 (2), 199–210.

Payan, A., Montana, G., 2015. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. arXiv preprint arXiv:1502.02506.

Potvin, O., Khademi, A., Chouinard, I., Farokhian, F., Dieumegarde, L., Leppert, I., Hoge, R., Rajah, M.N., Bellec, P., Duchesne, S., et al., 2019. Measurement variability following MRI system upgrade. Front. Neurol. 10, 726.

Pulgar, F.J., Charte, F., Rivera, A.J., del Jesus, M.J., 2020. Choosing the proper autoencoder for feature fusion based on data complexity and classifiers: Analysis, tips and guidelines. Inf. Fusion 54, 44–60.

Qiu, S., Heydari, M.S., Miller, M.I., Joshi, P.S., Wong, B.C., Au, R., Kolachalama, V.B., 2019. P1-119: ENHANCING deep learning model performance for ad diagnosis using roi-based selection. Alzheimer's Dement. 15, P280–P281.

Schnack, H.G., van Haren, N.E., Brouwer, R.M., van Baal, G.C.M., Picchioni, M., Weisbrod, M., Sauer, H., Cannon, T.D., Huttunen, M., Lepage, C., et al., 2010. Mapping reliability in multicenter MRI: Voxel-based morphometry and cortical thickness. Human Brain Mapp. 31 (12), 1967–1982.

Sedeno, L., Piguet, O., Abrevaya, S., Desmaras, H., García-Cordero, I., Baez, S., Alethia de la Fuente, L., Reyes, P., Tu, S., Moguilner, S., et al., 2017. Tackling variability: A multicenter study to provide a gold-standard network approach for frontotemporal dementia. Human Brain Mapp. 38 (8), 3804–3822.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Squire, L.R., Wixted, J.T., Clark, R.E., 2007. Recognition memory and the medial temporal lobe: a new perspective. Nat. Rev. Neurosci. 8 (11), 872–883.

Su, L., Surendranathan, A., Huang, Y., Bevan-Jones, W.R., Passamonti, L., Hong, Y.T., Arnold, R., Rodríguez, P.V., Wang, Y., Mak, E., et al., 2021. Relationship between tau, neuroinflammation and atrophy in Alzheimer's disease: The NIMROD study. Inf. Fusion 67, 116–124.

Tang, H., Wu, E., Ma, Q., Gallagher, D., Perera, G., Zhuang, T., 2000. MRI brain image segmentation by multi-resolution edge detection and region selection. Comput. Med. Imaging Graph. 24 (6), 349–357.

Teipel, S.J., Ewers, M., Wolf, S., Jessen, F., Kölsch, H., Arlt, S., Luckhaus, C., Schönknecht, P., Schmidtke, K., Heuser, I., et al., 2010. Multicentre variability of MRI-based medial temporal lobe volumetry in Alzheimer's disease. Psychiatry Res.: Neuroimaging 182 (3), 244–250.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15 (1), 273–289.

Vieira, S., Pinaya, W.H., Mechelli, A., 2017. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. Neurosci. Biobehav. Rev. 74, 58–75.

Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X., 2017. Residual attention network for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3156–3164.

Wang, S.-H., Nayak, D.R., Guttery, D.S., Zhang, X., Zhang, Y.-D., 2021a. COVID-19 classification by CCSHNet with deep fusion using transfer learning and discriminant correlation analysis. Inf. Fusion 68, 131–148.

Wang, S.-H., Zhou, Q., Yang, M., Zhang, Y.-D., 2021b. ADVIAN: Alzheimer's disease VGG-inspired attention network based on convolutional block attention module and multiple way data augmentation. Front. Aging Neurosci. 13, 313.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Cedarbaum, J., Donohue, M.C., Green, R.C., Harvey, D., Jack Jr., C.R., et al., 2015. Impact of the Alzheimer's disease neuroimaging initiative, 2004 to 2014. Alzheimer's Dement. 11 (7), 865–884.

Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O., et al., 2020. Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. Med. Image Anal. 63, 101694.

Yang, Z., He, X., Gao, J., Deng, L., Smola, A., 2016. Stacked attention networks for image question answering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 21–29.

Yao, Q., Wang, R., Fan, X., Liu, J., Li, Y., 2020. Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network. Inf. Fusion 53, 174–182.

Zhang, Y., Wang, S., Xia, K., Jiang, Y., Qian, P., Initiative, A.D.N., et al., 2021b. Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. Inf. Fusion 66, 170–183.

Zhang, L., Wang, L., Zhu, D., 2020. Jointly analyzing alzheimer's disease related structure-function using deep cross-model attention network. In: 2020 IEEE 17th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 563–567.

Zhang, J., Zheng, B., Gao, A., Feng, X., Liang, D., Long, X., 2021a. A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification. Magn. Reson. Imaging 78, 119–126.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2921–2929.

Zhu, Y., Kim, M., Zhu, X., Kaufer, D., Wu, G., Initiative, A.D.N., et al., 2021. Long range early diagnosis of Alzheimer's disease using longitudinal MR imaging data. Med. Image Anal. 67, 101825.