Signal, Image and Video Processing (2024) 18:2743–2751
https://doi.org/10.1007/s11760-023-02945-w

ORIGINAL PAPER

CNN-based Alzheimer's disease classification using fusion of multiple 3D angular orientations

Fuat Uyguroğlu[1] · Önsen Toygar[1] · Hasan Demirel[2]

Received: 27 November 2023 / Revised: 27 November 2023 / Accepted: 5 December 2023 / Published online: 2 January 2024

Abstract

Convolutional neural networks (CNN) can extract the features necessary for the recognition and classification of several diseases. Yet, the intricate symptoms encompassing changes in brain anatomy pose challenges for CNN training. While an ideal scenario would leverage a patient's entire magnetic resonance imaging (MRI) data with minimal preprocessing and human involvement, it does not always yield optimal results. To improve the performance of CNNs, researchers utilize much larger and more complex networks, which does not guarantee improvement. In this paper, we propose an innovative way to increase performance, manifested through utilizing multiple distinct 3D orientations of the data, coupled with a multi-classifier framework. The method consists of predictions from networks trained on unique angular orientations of the same data set that combine to offer a unified prediction. The results obtained from the proposed method underscore that these minimalistic, computationally frugal alterations can propel average accuracy rates from 89.84% to a commendable 94.37%, signifying a near 5% performance surge.

Keywords Alzheimer's disease · 3D convolutional neural networks (3D CNN) · Magnetic resonance imaging (MRI) · Multi-classifier systems · Deep learning

1 Introduction

Alzheimer's disease (AD), a progressive neurodegenerative disorder, represents one of the primary causes of cognitive decline in the elderly population worldwide. Early and accurate diagnosis of AD can pave the way for timely intervention, enhancing patients' quality of life and prolonging cognitive function [1]. Magnetic resonance imaging (MRI) has emerged as a pivotal diagnostic tool, capturing intricate details of brain structure, thereby assisting in discerning between AD and healthy controls. However, the vast variability in brain structures and the subtle changes occurring in the early stages of AD present challenges in achieving high diagnostic accuracy [2].

Recent advances in the field of deep learning have demonstrated promise in extracting intricate patterns and features from medical images, such as structural MRIs [3–6]. Deep learning autonomously curates multi-tiered feature representations, streamlining classification processes while minimizing human oversight.

While these networks offer improved diagnostic capabilities, the scarcity of annotated MRI data can hinder their full potential. In navigating this intricate classification terrain, existing research predominantly falls into four categories [7]:

- 2D Slice-level: This method employs slices of an MR image containing frequently affected areas of the brain. While memory-efficient, it demands intricate preprocessing and expert knowledge, potentially constricting its universal application [8, 9].
- Region of Interest: This method isolates localized affected regions as a distinct 3D entity, mirroring the advantages and constraints of the 2D slice-level approach [10].
- 3D Patch-level: In this approach, the MR image undergoes segmentation into multiple, often overlapping,

✉ Fuat Uyguroğlu
fuat.uyguroglu@emu.edu.tr

Önsen Toygar
onsen.toygar@emu.edu.tr

Hasan Demirel
hasan.demirel@emu.edu.tr

1 Department of Computer Engineering, Faculty of Engineering, Eastern Mediterranean University, 99628 Famagusta, North Cyprus, via Mersin 10, Turkey

2 Department of Electrical and Electronic Engineering, Faculty of Engineering, Eastern Mediterranean University, 99628 Famagusta, North Cyprus, via Mersin 10, Turkey

Table 1 Differences in performance based on changing angular orientation and the fusing of predictions to create a multi-classifier

Angular orientation	Angular equivalent	Average accuracy (%)
Axial	(0, 0, 90)	87.36
Sagittal / Default	(0, 0, 0)	89.84
Coronal	(0, 90, 0)	91.62
Multi-classifier with sum rule	(0, 0, 90)+(0, 0, 0)+(0,90,0)	92.45

patches to bolster the sample volume. Yet, a meticulous preprocessing strategy is paramount to sidestep irrelevant brain patches and ensure the network's accurate training [11, 12].

- 3D Subject-level: Encompassing the entire brain as a singular entity, this approach resonates most with deep learning philosophies, albeit at the expense of a more constrained sample size [13, 14].

To navigate the challenges posed by limited sample sizes, data augmentation strategies [15], such as rotation, can be utilized to significantly enrich the data set, thereby enhancing model robustness and generalization capabilities. Another method is using multi-classifier systems, where individual networks bolster one another, ameliorating individual weaknesses.

We tested the ideas above at the beginning of our study by training three different 3D CNNs with the same data set, with three of the most available orientations; sagittal (0,0,0) or the default orientation, coronal (0,90,0), and axial (0,0,90). This experiment resulted in significantly different performances. We, then, fused the predictions of these three networks to create a multi-classifier that performed better than any of the three individual networks. The results are shown in Table 1.

In this paper, we harness the combined strength of unique angular orientations of MRIs and multi-classifier approaches in conjunction with 3D CNNs. Our goal is to achieve a marked improvement in performance, minimizing the need for extensive preprocessing, and offering a more streamlined, efficient approach to Alzheimer's Disease diagnosis using MRIs. The results indicate that training 3D CNNs with the data set fixed to a new orientation can drastically change performance, differing in our case from 84.89% (180, −60, 60) to 91.62% (0, 90, 0). The predictions of the different networks created using this method can further be fused generating a stronger classification framework with a maximum accuracy of 94.37%.

2 Methodology

In this section, the proposed preprocessing methods, classification model, and the novel fusion-based approach using multiple angular orientations will be explained in detail. The method consists of training 3D CNN models with the entire data set oriented to a new angle, and utilizing these networks in a new classification framework where we harness predictions from *n* different iterations of the model.

2.1 Preprocessing

Since our project aimed to improve the performance of CNNs under easily achievable yet somewhat unfavorable conditions, we tried keeping our preprocessing to a minimum. The main aim of the preprocessing procedure was to extract the brain from the MR image of the subject's head. Subsequent to this extraction, we undertook the crucial steps of rotating and resizing the 3D image, ensuring it conformed to a standardized structure.

Data used in the preparation of this article were obtained from the Alzheimer's disease neuroimaging initiative (ADNI) database (adni.loni.usc.edu) and from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL). See www.aibl.csiro.au for further details. AIBL study methodology has been reported previously [16].

The brain extraction part of our preprocessing has been accomplished using the brain extraction tool (BET) [17] with reduced image bias. After this stage, all of the resulting images have been inspected visually, both for overall quality and the inadvertent inclusion of extraneous organ voxels. Any images that repeatedly failed to meet our quality threshold across multiple preprocessing attempts were excluded from the data set. After this step, the FLIRT method has been applied to the samples in order to register them to the MNI152 template [18, 19]. The results of these operations are shown in Fig. 1.

Once the preprocessing stage was completed, we were left with 522 AD images of size $91 \times 101 \times 91$, with 1 mm^3 voxels. We used the same number of HC samples, creating a balanced data set, thereby optimizing the interpretability of ensuing accuracy metrics. These sets have been immediately separated into training (including validation) and test sets in the ratio of 9:1.

2.2 Classification model

A pivotal aspect of our methodology is ensuring minimal standard deviation in results, thus yielding meaningful accuracy differences. In pursuit of this, we tailored 3D adaptations

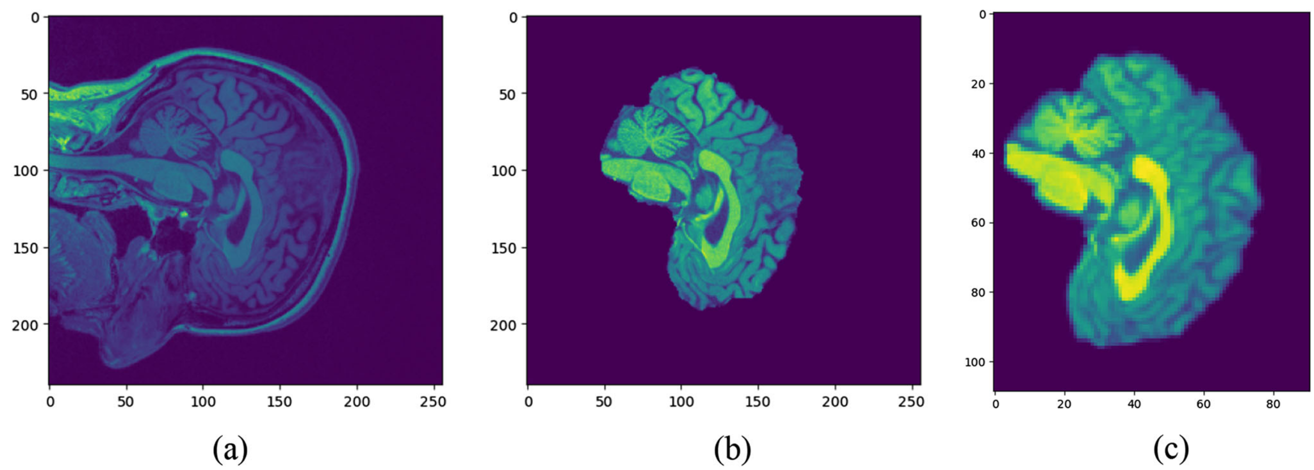

Fig. 1 The results of the BET (**b**), BET and FLIRT (**c**) operations, compared to the original image (**a**)

from renowned networks like VGG16 [20], ResNet [21], and the cutting-edge DenseNet [22]. Among these, the network derived from DenseNet121 stood out, registering the lowest standard deviation at a mere 2.07%. Consequently, DenseNet became our primary choice for the bulk of our testing.

DenseNet uses feature map fusion to combine information of different complexities together. The fusion process is orchestrated through the concatenation of the input and output from convolution groups nestled within its Dense Blocks so that the input to the layer z is:

$$x_z = H_z([x_0, x_1, ..., x_z - 1]) \quad (1)$$

where $[x_0, x_1, ..., x_z - 1]$ is the concatenation of the feature maps produced by the preceding layers, and H_z is the nonlinear transformation applied by the layer z [22].

Each convolution group consisted of the following layers:

- Batch Normalization
- ReLU Activation
- 3D Convolution with kernel size $1 \times 1 \times 1$
- 3D Convolution with kernel size $3 \times 3 \times 3$
- Dropout Layer of value 0.1

The architectural visualization of the overall network is observed in Fig. 3, where Dense Blocks depicted with blue color code consist of several convolution groups which are illustrated in Fig. 2. The detailed network specifications are given in Table 2.

2.3 Angular orientation

Based on the idea of improving performance using proven methods within data augmentation and taking those ideas one step further, we utilize the entire data set, with varied angular orientation. The angular variation is achieved by the rotate function in the scipy.ndimage library, which uses spline interpolation.

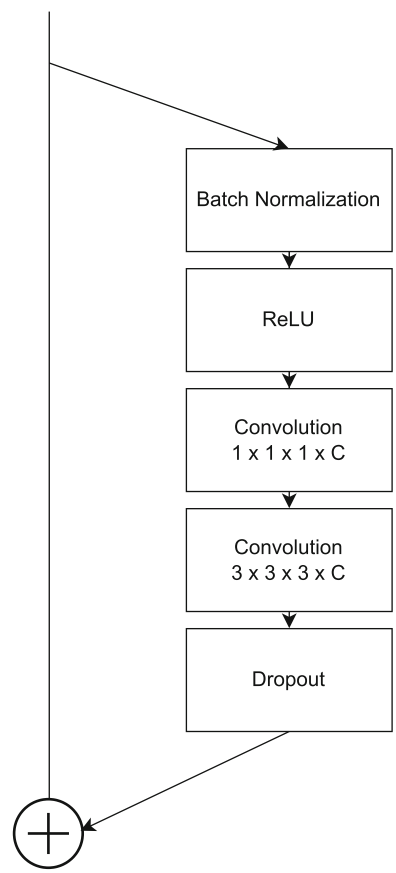

Fig. 2 Example of a convolution group. Each Dense Block within the DenseNet consists of multiple convolution groups creating feature maps to be concatenated with the feature map coming from the pipeline. Such a design facilitates the amalgamation of features from diverse levels. In our experiments, we set the Growth Rate (C) at 32, while maintaining a Dropout value of 0.1

Table 2 3D CNN architecture based on DenseNet121

Layers	Output shape	Properties
Convolution	46 × 51 × 46	Conv 7×7×7×64, Stride = 2
Pooling	23 × 26 × 23	Max pool 3×3×3, Stride = 2
Dense Block	23 × 26 × 23	6 × Conv Group
Transition Layer	23 × 26 × 23	Conv 1×1×1×128
	11 × 13 × 11	Average Pooling 2×2×2
Dense Block	11 × 13 × 11	12 × Conv Group
Transition Layer	11 × 13 × 11	Conv 1×1×1×256
	5 × 6 × 5	Average Pooling 2×2×2
Dense Block	5 × 6 × 5	24 × Conv Group
Transition Layer	5 × 6 × 5	Conv 1×1×1×512
	2 × 3 × 2	Average Pooling 2×2×2
Dense Block	2 × 3 × 2	16 × Conv Group
Pooling	2 × 3 × 2	Batch Norm
	2 × 3 × 2	ReLU
	1 × 1 × 1	Global Average Pool
Classification		1024 Fully Connected+Sigmoid

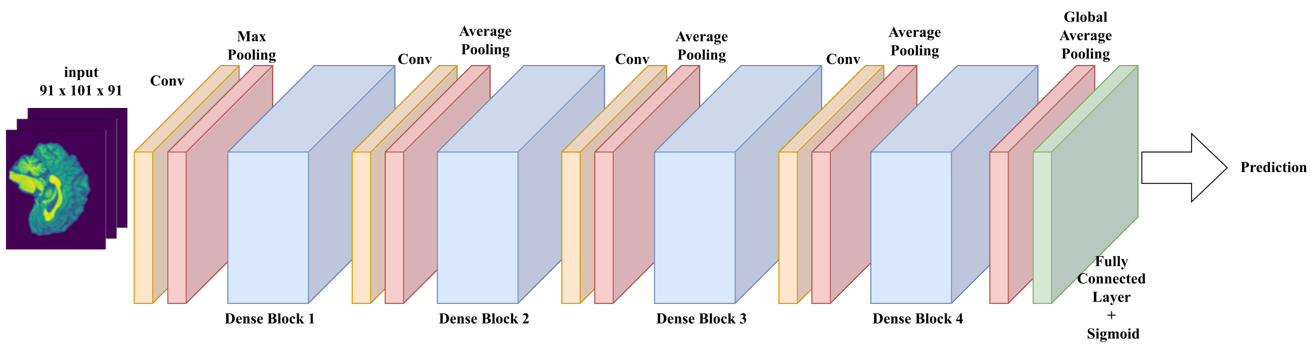

Fig. 3 Graphical representation of DenseNet121 adaptation used in this work

Table 3 Average test performances of networks trained with different orientations

Angle	Average accuracy (%)	Average AUC (%)
(0, 90, 0)	91.62	95.11
(−120, 180, 120)	91.21	93.67
(−120, 120, −120)	90.25	94.04
(0, 60, −60)	89.97	93.29
(0, 0, 0)	89.84	94.12
(60, −120, 0)	89.70	93.36
(−120, 0, 60)	89.70	93.94
(−120, 180, −60)	89.42	93.31
(120, −60, 120)	89.01	93.17
(60, 0, 120)	88.32	93.59
(−120, 0, −120)	87.50	93.76
(0, 0, 90)	87.36	92.97
(0, 120, −60)	86.26	92.19
(180, −60, 60)	84.89	91.57

2.4 Classification

To mitigate performance inconsistencies, we trained seven networks for each distinct test angle. The networks were asked to predict whether a given test sample belonged to the AD or CN groups, which resulted in a value between 0 and 1 because of the sigmoid layer at the end of the network. The predictions as well as the accuracy of each network were kept for further use.

For tighter control over the testing conditions, we consistently used the same training and validation sets for each run. While the contents within these sets were shuffled, their overall composition remained constant. The test set, on the other hand, stayed consistent across all networks but was rotated to match the orientation of the respective training/validation sets.

3 Experimental results

3.1 Results on training with a district orientation

Table 3 shows the average performances of networks trained with the data set with random unique orientations produced with step size 60 degrees in addition to the sagittal (0,0,0), coronal (0,90,0), and axial (0,0,90) planes. The results reported in this table are sorted according to classifier performance based on accuracy. The average accuracy for the worst-performing network (trained with the orientation (180,−60,60)) was 84.89% while the original (sagittal plane) was 89.84%, with 4 networks performing better including the coronal plane with 91.62% average accuracy.

Using the orientation (0,0,0) as our reference baseline, it becomes evident that altering the orientation results in different information being registered into the network. Intriguingly, even a network yielding a comparatively modest accuracy can still hold value when incorporated into the subsequent phase of our methodology.

3.2 Results of combining predictions

Networks trained on data sets with varied angular orientations exhibited an important trait: they correctly classified samples that others misclassified. Capitalizing on this observation, we sought to amalgamate these sets of predictions. The arithmetic mean was employed using the sum rule, a method renowned for its efficacy in integrating multiple classifier systems [23], where the probability of a class C_1 (i.e., AD) is calculated using the formula below and predictions

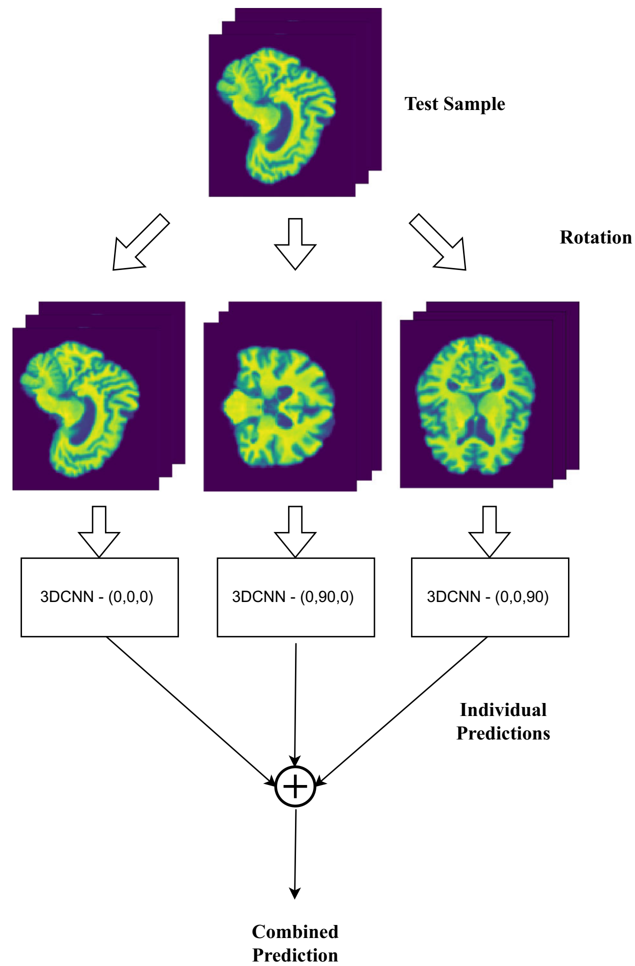

Fig. 4 To amalgamate predictions from various networks, the test sample is initially rotated to the specified target angle of the network that will create the prediction. Subsequent to individual network predictions, the sum rule is employed to compute their arithmetic mean, which is then adopted as the consolidated prediction

made by N networks.

$$p_{combined}(C_1) = \frac{1}{N}\sum_{i=1}^{N} p_i(C_1) \quad (2)$$

The mechanics of this approach are illustrated in Fig. 4.

3.2.1 Combining predictions from two networks

Once a set of combined predictions was created, the accuracy values were obtained and averaged to create a heat map table in Fig. 5. The results are sorted according to the individual performances (accuracy) of the networks. The visualization unequivocally illustrates that fusing predictions from networks trained at different angles consistently amplifies performance. The highest accuracy, 93.96% was achieved by the combination of [(0,90,0) vs (−120,0,60)].

Table 4 Comparison to the state-of-the-art

References	Dataset	Methodology		ACC (%)	SEN (%)	SPE (%)	AUC
Coupé et al. [24]	231 HC,189 AD Subjects	Conventional classifiers (SVM, LDA...)		91.00	87.00	94.00	–
Liu et al. [25]	229 HC, 189 AD Subjects	Conventional classifiers (SVM)		92.00	91.00	93.00	95.20
Suk et al. [26]	101 HC, 93 AD Subjects	Deep boltzman machine		93.52	94.65	95.22	98.70
Liu et al. [27]	61 HC, 50 AD Subjects	3D CNN Subject Level		80.00	–	–	87.00
Lian et al. [28]	429 HC, 358 AD Subjects	Hierarchical FCN		90.00	82.00	97.00	95.00
Nawaz et al. [29]	382 Subjects	AlexNet		92.85	–	–	–
Proposed Method	227 HC, 210 AD Subjects	3D DenseNet	Coronal plane (0, 90, 0)	91.62	91.76	91.48	95.11
			2 network fusion	93.96	94.50	93.40	96.23
			3 network fusion	**94.37**	**96.15**	92.58	96.56

	Coronal (0,90,0)	(-120,180,120)	(-120,120,-120)	(0,60,-60)	Sagittal (0,0,0)	(60,-120,0)	(-120,0,60)	(-120,180,-60)	(120,-60,120)	(60,0,120)	(-120,0,-120)	Axial (0,0,90)	(0,120,-60)	(180,-60,60)	Average
Coronal (0,90,0)		92.86%	93.13%	93.13%	92.03%	93.41%	93.96%	92.72%	92.58%	92.17%	93.27%	92.31%	92.31%	91.35%	92.71%
(-120,180,120)	92.86%		92.99%	92.17%	93.41%	92.86%	92.72%	91.62%	92.17%	91.76%	92.17%	92.03%	92.45%	91.76%	92.38%
(-120,120,-120)	93.13%	92.99%		93.68%	93.41%	92.31%	92.99%	91.62%	92.45%	91.48%	93.41%	90.93%	91.21%	91.07%	92.36%
(0,60,-60)	93.13%	92.17%	93.68%		92.72%	92.58%	92.03%	92.58%	91.76%	92.58%	91.48%	91.62%	91.62%	90.93%	92.22%
Sagittal (0,0,0)	92.03%	93.41%	93.41%	92.72%		92.03%	91.62%	92.17%	91.35%	92.45%	93.13%	91.62%	92.03%	90.11%	92.16%
(60,-120,0)	93.41%	92.86%	92.31%	92.58%	92.03%		92.58%	92.45%	91.48%	91.76%	93.54%	91.48%	91.62%	90.80%	92.22%
(-120,0,60)	93.96%	92.72%	92.99%	92.03%	91.62%	92.58%		91.90%	92.03%	92.31%	92.17%	90.38%	91.48%	90.80%	92.08%
(-120,180,-60)	92.72%	91.62%	91.62%	92.58%	92.17%	92.45%	91.90%		91.76%	90.80%	91.62%	90.38%	90.38%	88.60%	91.43%
(120,-60,120)	92.58%	92.17%	92.45%	91.76%	91.35%	91.48%	92.03%	91.76%		90.11%	91.76%	90.66%	90.93%	89.97%	91.46%
(60,0,120)	92.17%	91.76%	91.48%	92.58%	92.45%	91.76%	92.31%	90.80%	90.11%		91.07%	90.11%	89.97%	88.74%	91.18%
(-120,0,-120)	93.27%	92.17%	93.41%	91.48%	93.13%	93.54%	92.17%	91.62%	91.76%	91.07%		89.56%	89.84%	90.25%	91.79%
Axial (0,0,90)	92.31%	92.03%	90.93%	91.62%	91.62%	91.48%	90.38%	90.38%	90.66%	90.11%	89.56%		89.84%	89.70%	90.82%
(0,120,-60)	92.31%	92.45%	91.21%	91.62%	92.03%	91.62%	91.48%	90.38%	90.93%	89.97%	89.84%	89.84%		89.42%	91.01%
(180,-60,60)	91.35%	91.76%	91.07%	90.93%	90.11%	90.80%	90.80%	88.60%	89.97%	88.74%	90.25%	89.70%	89.42%		90.27%

Fig. 5 For each network orientation, combining its predictions with those of a network from a different orientation yielded an average accuracy value. These are visually represented with color coding in the figure

Remarkably, the combined predictions boasted an average accuracy of 90.27% in the least favorable scenario [(180,−60,60) vs all other angles]-this stands on par with the third highest-performing individual network.

As the color coding in the table accentuates, the angle *(−120, 0, 120)* stands out as a special case. While its standalone accuracy was a modest 87.50%, when synergized with a reasonably accurate network, its combined predictive prowess was notable. This observation underscores a pivotal insight; even networks with sub-optimal performance can, under the right conditions, contribute significantly to combined accuracy. Such a realization attests to the robustness of our method, even without pinpointing the optimal angles.

3.2.2 Combining predictions from three networks

Combining three networks' predictions with the same method results in even better performance. Figure 6a showcases results derived from this approach, using the most successful individual network (0,90,0). The average performance of any combination including this network is 92.99%, almost 1.4% higher than its individual performance.

		(-120,180,120)	(-120,120,-120)	(0,60,-60)	Sagittal (0,0,0)	(60,-120,0)	(-120,0,60)	(-120,180,-60)	(120,-60,120)	(60,0,120)	(-120,0,-120)	Axial (0,0,90)	(0,120,-60)	(180,-60,60)	Average
	(-120,180,120)		93.96%	93.82%	93.68%	93.54%	93.27%	93.13%	92.86%	93.68%	93.82%	92.99%	93.68%	92.72%	93.43%
	(-120,120,-120)	93.96%		93.68%	93.41%	93.68%	93.41%	92.99%	93.13%	93.41%	94.23%	92.31%	93.96%	92.03%	93.35%
	(0,60,-60)	93.82%	93.68%		92.99%	94.23%	93.27%	93.13%	92.58%	92.58%	93.54%	92.58%	93.82%	92.17%	93.20%
	Sagittal (0,0,0)	93.68%	93.41%	92.99%		92.86%	92.86%	92.45%	92.31%	92.99%	93.27%	92.45%	93.41%	92.17%	92.90%
	(60,-120,0)	93.54%	93.68%	94.23%	92.86%		93.41%	92.99%	92.86%	92.72%	93.54%	93.68%	93.13%	92.72%	93.28%
	(-120,0,60)	93.27%	93.41%	93.27%	92.86%	93.41%		93.54%	93.27%	93.68%	93.54%	92.58%	93.27%	92.58%	93.22%
Coronal (0,90,0)	(-120,180,-60)	93.13%	92.99%	93.13%	92.45%	92.99%	93.54%		92.86%	92.58%	93.27%	92.58%	93.41%	91.62%	92.88%
	(120,-60,120)	92.86%	93.13%	92.58%	92.31%	92.86%	93.27%	92.86%		91.90%	92.99%	92.86%	93.13%	91.62%	92.70%
	(60,0,120)	93.68%	93.41%	92.58%	92.99%	92.72%	93.68%	92.58%	91.90%		93.13%	92.31%	92.58%	91.62%	92.77%
	(-120,0,-120)	93.82%	94.23%	93.54%	93.27%	93.54%	93.54%	93.27%	92.99%	93.13%		92.99%	93.13%	92.72%	93.35%
	Axial (0,0,90)	92.99%	92.31%	92.58%	92.45%	93.68%	92.58%	92.58%	92.86%	92.31%	92.99%		92.17%	91.07%	92.55%
	(0,120,-60)	93.68%	93.96%	93.82%	93.41%	93.13%	93.27%	93.41%	93.13%	92.58%	93.13%	92.17%		91.76%	93.12%
	(180,-60,60)	92.72%	92.03%	92.17%	92.17%	92.72%	92.58%	91.62%	91.62%	91.62%	92.72%	91.07%	91.76%		92.07%

(a) (0,90,0) vs all

		Coronal (0,90,0)	(-120,120,-120)	(0,60,-60)	Sagittal (0,0,0)	(60,-120,0)	(-120,0,60)	(-120,180,-60)	(120,-60,120)	(60,0,120)	(-120,0,-120)	Axial (0,0,90)	(0,120,-60)	(180,-60,60)	Average
	Coronal (0,90,0)		93.96%	93.82%	93.68%	93.54%	93.27%	93.13%	92.86%	93.68%	93.82%	92.99%	93.68%	92.72%	93.43%
	(-120,120,-120)	93.96%		93.54%	93.96%	93.54%	93.27%	93.41%	93.13%	92.99%	93.82%	93.54%	92.86%	93.13%	93.38%
	(0,60,-60)	93.82%	93.54%		93.54%	92.99%	92.86%	92.72%	92.86%	93.96%	92.45%	92.58%	92.72%	92.58%	92.98%
	Sagittal (0,0,0)	93.68%	93.96%	93.54%		93.13%	93.82%	92.99%	92.58%	92.86%	93.27%	92.99%	92.86%	92.86%	93.17%
	(60,-120,0)	93.54%	93.54%	92.99%	93.13%		92.99%	93.27%	92.31%	93.27%	93.68%	93.82%	92.45%	92.99%	93.13%
	(-120,0,60)	93.27%	93.27%	92.86%	93.82%	92.99%		92.45%	93.27%	92.58%	92.58%	92.03%	92.99%	92.86%	92.88%
(-120,180,120)	(-120,180,-60)	93.13%	93.41%	92.72%	92.99%	93.27%	92.45%		92.31%	91.76%	92.31%	92.03%	92.31%	92.03%	92.51%
	(120,-60,120)	92.86%	93.13%	92.86%	92.58%	92.31%	93.27%	92.31%		92.86%	93.13%	92.72%	92.86%	91.76%	92.71%
	(60,0,120)	93.68%	92.99%	93.96%	92.86%	93.27%	92.58%	91.76%	92.86%		92.72%	92.45%	92.58%	92.03%	92.73%
	(-120,0,-120)	93.82%	93.82%	92.45%	93.27%	93.68%	92.58%	92.31%	93.13%	92.72%		92.86%	92.86%	93.13%	92.98%
	Axial (0,0,90)	92.99%	93.54%	92.58%	92.99%	93.82%	92.03%	92.03%	92.72%	92.45%	92.86%		92.31%	91.90%	92.66%
	(0,120,-60)	93.68%	92.86%	92.72%	92.86%	92.45%	92.99%	92.31%	92.86%	92.58%	92.86%	92.31%		92.31%	92.64%
	(180,-60,60)	92.72%	93.13%	92.58%	92.86%	92.99%	92.86%	92.03%	91.76%	92.03%	93.13%	91.90%	92.31%		92.51%

(b) (-120,180,120) vs all

		Coronal (0,90,0)	(-120,180,120)	(0,60,-60)	Sagittal (0,0,0)	(60,-120,0)	(-120,0,60)	(-120,180,-60)	(120,-60,120)	(60,0,120)	(-120,0,-120)	Axial (0,0,90)	(0,120,-60)	(180,-60,60)	Average
	Coronal (0,90,0)		93.96%	93.68%	93.41%	93.68%	93.41%	92.99%	93.13%	93.41%	94.23%	92.31%	93.96%	92.03%	93.29%
	(-120,180,120)	93.96%		93.54%	93.96%	93.54%	93.27%	93.41%	93.13%	92.99%	93.82%	93.54%	92.86%	93.13%	93.38%
	(0,60,-60)	93.68%	93.54%		94.09%	93.54%	93.27%	93.41%	92.86%	93.54%	93.41%	92.99%	92.72%	92.17%	93.20%
	Sagittal (0,0,0)	93.41%	93.96%	94.09%		93.41%	93.13%	93.13%	92.99%	93.27%	94.37%	92.17%	92.86%	92.03%	93.15%
	(60,-120,0)	93.68%	93.54%	93.54%	93.41%		93.68%	93.27%	92.45%	92.86%	94.09%	92.31%	92.58%	92.45%	93.06%
	(-120,0,60)	93.41%	93.27%	93.27%	93.13%	93.68%		92.99%	92.86%	93.41%	93.96%	92.03%	92.99%	92.31%	93.06%
(-120,120,-120)	(-120,180,-60)	92.99%	93.41%	93.41%	93.13%	93.27%	92.99%		93.27%	92.72%	93.68%	91.90%	92.45%	91.62%	92.84%
	(120,-60,120)	93.13%	93.13%	92.86%	92.99%	92.45%	92.86%	93.27%		93.13%	93.13%	91.76%	91.90%	90.93%	92.53%
	(60,0,120)	93.41%	92.99%	93.54%	93.27%	92.86%	93.41%	92.72%	93.13%		93.82%	92.31%	91.48%	92.17%	92.87%
	(-120,0,-120)	94.23%	93.82%	93.41%	94.37%	94.09%	93.96%	93.68%	93.13%	93.82%		93.27%	92.86%	93.13%	93.57%
	Axial (0,0,90)	92.31%	93.54%	92.99%	92.17%	92.31%	92.03%	91.90%	91.76%	92.31%	93.27%		91.48%	91.07%	92.13%
	(0,120,-60)	93.96%	92.86%	92.72%	92.86%	92.58%	92.99%	92.45%	91.90%	91.48%	92.86%	91.48%		91.07%	92.24%
	(180,-60,60)	92.03%	93.13%	92.17%	92.03%	92.45%	92.31%	91.62%	90.93%	92.17%	93.13%	91.07%	91.07%		91.90%

(c) (-120,120,-120) vs all

Fig. 6 Combined accuracy averages of three networks trained with three different angles

Figure 6b depicts a similar table, this time based on the angular orientation (−120,190,120). The average performance of combinations including this angle is 92.90%, 1.7% better than the individual performance.

Figure 6c contains the results for combinations including (−120,120,−120), which happen to include the highest performance we have achieved at 94.37%. An intriguing point, this performance was not the result of a fusion with the best-performing networks but rather relatively poor performers (0,0,0) and (−120,0,−120).

Notably, the least favorable accuracy resulting from a tri-network combination, incorporating at least one network from the top five performers, outstrips the best accuracy achieved by any singular network.

A comparison to the state-of-the-art can be done using Table 4. The proposed classification framework based on combining predictions from three networks outperforms the alternative state-of-the-art methods in the literature by means of accuracy and sensitivity (shown in bold) and has comparable performance in specificity and AUC.

4 Discussion

While our empirical investigations reveal that unique orientations of training data sets induce variances in network performance, the root cause of these fluctuations remains under scrutiny. Currently, we are probing into two primary hypotheses:

- **Biological Considerations:** This pertains to the potential topographical alterations in the brain forming a pattern in the MRI scans.
- **Mathematical Aspects:** This delves into factors such as inter-slide rate of change and differential entropy through slides.

Alzheimer's Disease initially targets memory-centric reg ions, notably the hippocampus, and eventually wreaks expansive neuronal damage across the brain [30]. Consequently, rotations enhancing the CNN's ability to discern central brain features-like those from the hippocampus-might yield superior results. Alternatively, this success could emanate from amplifying the contrast between Alzheimer's afflicted regions and healthy brain tissue.

Arguably more intriguing is the notion that the performance might be anchored to the voxel distribution itself, dictated by data orientation. This could pertain to voxel density or information content accessible to the network. A methodology predicated on voxel distribution-devoid of disease-specific biases-would be both cost-effective and universally applicable, transcending the confines of Alzheimer's Disease research.

Our combinatorial results, stemming from aggregating predictions of diverse networks, illuminate a clear trajectory: incorporating an expansive ensemble of network predictions amplifies performance, albeit with escalating computational overheads. Given the industry trend toward deeper, more intricate networks, one might speculate on the practicality of deploying multiple, comparatively simpler networks. This approach could circumvent the spatial complexities and feasibility challenges intrinsic to monolithic network architectures.

5 Conclusion

In this paper, we have illuminated a streamlined yet potent method for enhancing the performance of 3D convolutional neural networks in classifying MR images of Alzheimer's Disease patients and control samples. At the heart of this method lies a strategic blend of large-scale data augmentation and the principles of multiple classifiers. By intentionally minimizing the preprocessing steps, we have laid the groundwork for easy reproducibility of our experiments.

Central to our approach is the innovative use of unique orientations of the entire data set with novel angles, fostering the training of diverse networks. This singular act of rotation amplified the average accuracy notably from 89.84% (utilizing the original orientation) to 91.62%. Venturing deeper into the realm of multiple classifiers, we combined predictions from networks-of pairs and ensembles of three-employing the sum rule. This methodological synergy catapulted our maximum average accuracy figures. Remarkably, a duo of networks culminated in an impressive average accuracy of 93.96%, underscoring consistent performance enhancements. Expanding the ensemble to three networks nudged this to 94.37%. Yet, the true revelation was the bolstered baseline performance, with even the lesser-performing combinations outshining individual network outcomes.

In sum, our research offers a blueprint for an efficient, effective diagnostic tool, underscored by innovative data usage and ensemble predictions, aiming for precision and early diagnosis in the realm of Alzheimer's Disease.

Acknowledgements Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI). Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL).

Author Contributions Not applicable.

Funding This work was supported by the research Grant No. BAPC-02-21-01 provided by the Eastern Mediterranean University Type C (BAP-C) Scientific Research Project Budget.

Data Availability Not applicable.

Code availability Not applicable.

References

1. McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M.: Clinical diagnosis of alzheimer's disease: Report of the nincds-adrda work group* under the auspices of department of health and human services task force on alzheimer's disease. Neurology **34**(7), 939–939 (1984)
2. Dufumier, B., Gori, P., Battaglia, I., Victor, J., Grigis, A., Duchesnay, E.: Benchmarking cnn on 3d anatomical brain mri: architectures, data augmentation and deep ensemble learning. arXiv preprint arXiv:2106.01132 (2021)
3. 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.: Convolutional neural networks for classification of alzheimer's disease: overview and reproducible evaluation. Med. Image Anal. **63**, 101694 (2020)
4. Bernal, J., Kushibar, K., Asfaw, D.S., Valverde, S., Oliver, A., Martí, R., Lladó, X.: Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artif. Intell. Med. **95**, 64–81 (2019)
5. Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C., Wang, J.: Applications of deep learning to MRI images: a survey. Big Data Min. Analyt. **1**(1), 1–18 (2018)
6. Lundervold, A.S., Lundervold, A.: An overview of deep learning in medical imaging focusing on MRI. Z. Med. Phys. **29**(2), 102–127 (2019)
7. 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.: Convolutional neural networks for classification of alzheimer's disease: overview and reproducible evaluation. Med. Image Anal. **63**, 101694 (2020)
8. Islam, J., Zhang, Y.: Brain MRI analysis for alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. Brain Inform. **5**, 1–14 (2018)
9. Qiu, S., Chang, G.H., Panagia, M., Gopal, D.M., Au, R., Kolachalama, V.B.: Fusion of deep learning models of MRI scans, mini-mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment. Alzheimer's Dementia: Diagn., Assess. Dis. Monitor. **10**, 737–749 (2018)
10. Aderghal, K., Boissenin, M., Benois-Pineau, J., Catheline, G., Afdel, K.: Classification of smri for ad diagnosis with convolutional neuronal networks: A pilot 2-d+ study on adni. In: International Conference on Multimedia Modeling, 690–701 (2016). Springer
11. Li, F., Liu, M., Initiative, A.D.N., et al.: Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. Comput. Med. Imaging Graph. **70**, 101–110 (2018)
12. Cheng, D., Liu, M., Fu, J., Wang, Y.: Classification of mr brain images by combination of multi-cnns for ad diagnosis. In: Ninth International Conference on Digital Image Processing (ICDIP 2017), vol. 10420, 875–879 (2017). SPIE
13. Shmulev, Y., Belyaev, M., Initiative, A.D.N.: Predicting conversion of mild cognitive impairments to alzheimer's disease and exploring impact of neuroimaging. In: Graphs in biomedical image analysis and integrating medical imaging and non-imaging modalities: second international workshop, GRAIL 2018 and First International Workshop, Beyond MIC 2018, Held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 2, 83–91 (2018). Springer
14. Senanayake, U., Sowmya, A., Dawes, L.: Deep fusion pipeline for mild cognitive impairment diagnosis. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (isbi 2018), IEEE, pp. 1394–1997 (2018)
15. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big Data **6**(1), 1–48 (2019)
16. Ellis, K.A., Bush, A.I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P., et al.: The Australian imaging, biomarkers and lifestyle (aibl) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease. Int. Psychogeriatr. **21**(4), 672–687 (2009)
17. Smith, S.M.: Fast robust automated brain extraction. Hum. Brain Mapp. **17**(3), 143–155 (2002)
18. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. Med. Image Anal. **5**(2), 143–156 (2001)
19. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage **17**(2), 825–841 (2002)
20. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
21. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
22. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp. 4700–4708 (2017)
23. Kittler, J., Alkoot, F.M.: Sum versus vote fusion in multiple classifier systems. IEEE Trans. Pattern Anal. Mach. Intell. **25**(1), 110–115 (2003). https://doi.org/10.1109/TPAMI.2003.1159950
24. Coupé, P., Eskildsen, S.F., Manjón, J.V., Fonov, V.S., Pruessner, J.C., Allard, M., Collins, D.L., Initiative, A.D.N., et al.: Scoring by nonlocal image patch estimator for early detection of alzheimer's disease. NeuroImage Clin. **1**(1), 141–152 (2012)
25. Liu, M., Zhang, D., Shen, D., Initiative, A.D.N.: Hierarchical fusion of features and classifier decisions for alzheimer's disease diagnosis. Hum. Brain Mapp. **35**(4), 1305–1319 (2014)
26. Suk, H.-I., Lee, S.-W., Shen, D., Initiative, A.D.N., et al.: Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. Neuroimage **101**, 569–582 (2014)
27. Korolev, S., Safiullin, A., Belyaev, M., Dodonova, Y.: Residual and plain convolutional neural networks for 3d brain MRI classification. In: 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017), IEEE, 835–838 (2017)
28. Lian, C., Liu, M., Zhang, J., Shen, D.: Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural mri. IEEE Trans. Pattern Anal. Mach. Intell. **42**(4), 880–893 (2018)
29. Nawaz, H., Maqsood, M., Afzal, S., Aadil, F., Mehmood, I., Rho, S.: A deep feature-based real-time system for alzheimer disease stage detection. Multim. Tools Appl. **80**, 35789–35807 (2021)
30. Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C.E., Cummings, J., Flier, W.M.: Alzheimer's disease. The Lancet **397**(10284), 1577–1590 (2021)