Automatic Early Diagnosis of Alzheimers Disease Using 3D Deep Ensemble Approach 2022

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 October 2022, accepted 24 October 2022, date of publication 1 November 2022, date of current version 8 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3218621

RESEARCH ARTICLE

Automatic Early Diagnosis of Alzheimer's Disease Using 3D Deep Ensemble Approach

AYA GAMAL[1], MUSTAFA ELATTAR[1,2], AND SAHAR SELIM[1,2]
[1]Medical Imaging and Image Processing Research Group, Center for Informatics Science, Nile University, Giza 12677, Egypt
[2]School of Information Technology and Computer Science, Nile University, Giza 12677, Egypt

Corresponding author: Aya Gamal (ay.gamal@nu.edu.eg)

ABSTRACT Alzheimer's disease (AD) is considered the 6th leading cause of death worldwide. Early diagnosis of AD is not an easy task, and no preventive cures have been discovered yet. Having an accurate computer-aided system for the early detection of AD is important to help patients with AD. This study proposes a new approach for classifying disease stages. First, we worked on the MRI images and split them into an appropriate format to avoid data leakage. Subsequently, a simple and fast registration-free preprocessing pipeline was applied to the dataset. Numerous experiments were conducted to analyze the performances of different 3D classification architectures. Finally, an ensemble learning approach is applied to the top-performing models. The outstanding performance of the proposed method was demonstrated using augmentation of the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Our proposed ensemble approach outperforms studies in literature for distinguishing between people with AD and mild cognitive impairment (MCI), and MCI and cognitive normal (CN) with an AUC score of 91.28% and 88.42%, respectively. We also targeted the multiclass task, which was marginalized in previous work, by differentiating between the three stages of the disease.

INDEX TERMS MRI, Alzheimer's disease classification, convolutional neural network, ensemble learning.

I. INTRODUCTION

Alzheimer's disease (AD) is a chronic neurodegenerative disease. It is the major cause of dementia, which is a more general term that defines a group of symptoms that affect cognitive tasks, such as memory, thinking, and behavior. A subject with Alzheimer's disease survives for four to eight years on average after being diagnosed [1], [2]. In the United States, AD is considered the 6th leading cause of death [3] and is expected to affect one out of 85 people in the world by 2050. As a result, the cost of caring for AD patients is expected to increase dramatically. Commonly, Alzheimer's is associated with normal aging. An increase in age is the main risk factor for AD, and most people with Alzheimer's are 65 years and older. However, Alzheimer's is not a part of normal aging [4]. These changes are associated with changes in mental cognition, functional connectivity, and brain volume. As AD is a progressive disease, all of these changes are expected to proceed faster in AD than in normal aging, where symptoms gradually worsen over time. Therefore, having an accurate computer-aided system for the early detection of AD as soon as possible is important and urgent to help patients with AD and their families.

Early detection of AD is not an easy task, and no prevention methods or cures have yet been discovered. It has been proven that the early stages of AD may begin before the clinical onset of symptoms or visible behavioral changes arise for up to 20 years [5]. So, there is a great need for biological biomarkers to help predict the onset of Alzheimer's disease instead of the traditional diagnosis methods. These methods are limited to clinical cognitive tests, which mainly focus on clinical function only. Cognitive tests are primarily neuropsychological examinations that are performed by clinical experts. These tests determine whether a person is aware of their symptoms and their surroundings. The most extensively used cognitive tests are the Mini-Mental State Exam (MMSE) [6], Clinical Dementia Rating sum of boxes (CDR) [7], and Rey's Auditory Verbal Learning Test (RAVLT) [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

Several biomarkers have been used as indicators for AD prediction [9] such as β-amyloid, neurodegeneration indicated by tau protein measured in cerebrospinal fluid, or synaptic dysfunction, brain atrophy, memory loss, and clinical function. Due to this, there are different modalities used for identifying these biomarkers as imaging, genetics, clinical and demographic modalities. A lot of research on this topic was done to fuse several modalities [10], [11], [12], [13], [14], [15]. In our study, we focused on imaging modalities, especially magnetic resonance imaging (MRI). Magnetic resonance imaging (MRI) is widely used to image the anatomy, structure, and physiological processes of the brain and other body parts. It is an important biomarker for quantifying atrophy, as it is widely available and noninvasive [16]. In addition, it is a good indicator of disease progression because a structural MRI scan shows the contrast between gray matter (GM) and white matter (WM) tissues. Therefore, it can be successfully used for volume measurement.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) (adni.loni.usc.edu), Open Access Series of Imaging Studies (OASIS), and Australian Imaging Biomarker & Lifestyle Study of Ageing (AIBL) (https://aibl.csiro.au) are the most used databases in this research topic. There have been approximately 2567 publications in recent years, distributed as follows:1709 for ADNI, 747 for OASIS, and 141 for AIBL. [17], [18], [19] are represented as samples of studies that utilized the ADNI dataset in their papers. Referenes [20], [21], [22] used the OASIS dataset in their work. While [23], [24] presented their work on the AIBL with ADNI dataset. References [25], [26] utilized all these datasets by using one of them for training and the rest for external testing. In our work, we focused only on the ADNI dataset, since it is one of the most used benchmarks in this field.

Alzheimer's stage determination involves classifying subjects as cognitively normal (CN), mild cognitive impairment (MCI), or Alzheimer's disease (AD) patients. MCI is a condition that can be an early sign of Alzheimer's, but not everyone with MCI will develop the disease. MCI can be further divided into progressive mild cognitive impairment (pMCI) and stable mild cognitive impairment (sMCI). Patients with pMCI are subjects with MCI who will progress to AD, whereas patients with sMCI will remain with MCI. Different classification experiments can be performed, depending on the disease stage. We present the main tasks considered in the literature according to these stages as binary classification and multiclass tasks. The binary classification tasks included AD vs. CN, AD vs. MCI, MCI vs. CN, and sMCI vs. pMCI tasks. However, the sMCI and pMCI classes were not included in our study. In a multiclass task, they differentiated between the three stages of the disease as AD vs. MCI vs. CN task, which is considered the most difficult one. Several studies focused on different binary classification tasks only, while others targeted their studies on both binary and multiclass tasks [27].

To deal with MRI images, there are different approaches for the input type of the network, such as 2D slice-level, 3D patch-level, ROI-based, and 3D subject-level [25]. Each of these approaches has advantages and disadvantages. In the 2D slice-level approach, the 3D MRI volumes are converted into 2D images, which helps in feeding the network with more images. However, many slices have irrelevant information for the disease [15], [18], [20], [28], [29], [30], [31]. At the 3D patch level, each volume was divided into small patches with fixed dimensions. In the 3D subject-level approach, we consider the 3D volume as a whole image for each subject. Based on the studies that compared the two approaches, there is no significant difference between them because they generate almost the same information but in different ways [19], [27], [32], [33], [34], [35]. However, the 3D subject-level approach outperforms the other approaches because 3D patches require significant computational power. Finally, another approach is called the ROI-based. In this approach, prior studies [32], [36], [37], [38] only segmented a region of interest from the entire 3D volume to be used as input to the network. Therefore, this approach helps the network extract features that are relevant to the disease faster than using the entire volume.

Over the last few decades, machine learning and deep learning have gained immense potential in computer-aided diagnosis systems [39]. Many researchers have used neuroimaging data to predict AD using typical machine learning (ML) classifiers [40], [41], [42]. The most widely used algorithms are the support vector machine (SVM) and random forest after extracting the features from the image preprocessing pipeline. Recently, there are a lot of studies that use deep learning (DL) as it has made a great contribution to the medical imaging domain [43], [44], [45]. We report several DL network architectures used in the literature. Convolutional Neural Network (CNN) is the most widely used architecture, either for dealing with 2D slices or for 3D volumes [25], [27], [34]. It proved its superior performance in the diagnosis of AD [33], [34], [46], [47]. Transfer learning has also been applied and has proven to outperform in this problem, either with 2D images using pre-trained architectures on ImageNet or with 3D volumes using pre-trained autoencoders [25], [28], [31]. Other studies have applied different architectures such as DenseNet and Resnet [33], [48], [49], [50], [51], [52], [53]. Others have focused on using different attention modules in their work, as in [17], [19], [37], [54], [55], [56], and [57].

Several variations have been discovered in previous studies. For example, in dataset type and size, pre-processing pipelines, various targeted classification tasks, various architectures with various hyperparameters for the models, various approaches for dealing with MRI images, evaluation metrics, and strategies.

[25] examined the numerous studies that used CNNs for AD classification. They then extended their framework to reduce ambiguity between studies by implementing a modular set of image-preprocessing procedures, classification architectures, and evaluation strategies. Their efforts in their study were a good starting point to categorize the papers based on their approach to dealing with MRI images and

which papers had a clear data leakage in their studies. However, their training was not deterministic, and overfitting occurred in their experiments owing to the limited data size. [58] presented two different 3D CNN architectures for brain MRI classification. They avoided the issue of data leaks by selecting only the first images taken for each subject. However, their preprocessing pipeline fails to be simple as it involves two complex steps: skull stripping and image alignment. Reference [19] proposed a 3D Residual U-Net model having a hybrid attention mechanism (3D HA-ResUNet). Their framework includes a U-Net network that performs feature extraction at different scales of the receiving input images and a hybrid attention module that is incorporated with the skip connection in the U-Net model, which combines channel domain attention and spatial domain attention. Reference [17] proposed a framework based on the pyramid squeeze attention mechanism for Alzheimer's disease classification. They only classified the AD and CN classes, without considering the MCI disease stage. In [27], different approaches and computational techniques (3D-Subject level, 3D-patch based, slice level) were used to deal with MRI images by building a 3D ConvNet framework for AD, MCI, and CN classification. Reference [10] proposed a fusion framework that combines the features from multiple modalities, including volumetric MRI and neuropsychological measures, using a deep learning network inspired by state-of-the-art architectures such as Google Net, ResNet, and DenseNet. Reference [11] is another study that examined the fusion between different modalities. They proposed an end-to-end classification system based on a CNN and convolutional autoencoder using MRI and FDG-PET images. However, this study witnessed a clear data leakage issue. Reference [12] is also another multi-modal data fusion study. For MRI scans, they used a 3D CNN only for feature representation learning instead of end-to-end training. Despite the improved performance, their study has limitations, such as small dataset sizes.

All these earlier studies suffer from clarifying the type of the used dataset from ADNI except [17] which in turn makes the number of scans used variable and makes the comparison or reproduction of their experiments difficult. Regarding preprocessing pipelines, most studies have attempted to find the best-performing pipeline based on its speed, simplicity, and enhancement of the received input MRI images. Reference [17] proposed a pipeline in which the impact of different image filtering approaches was quantified on model performance. Reference [11], [12], [17], [27] presented their preprocessing pipeline with the registration step which is a complex step as it is considered an optimization problem [59] and takes much time to be implemented. Reference [25] presented two preprocessing pipelines including several types of registration. Reference [10], [19] did not mention any alignment step in their pipeline. Regarding the skull stripping step, many tools are presented in [60] that are used in brain extraction, such as the machine learning-based brain extraction tool ROBEX algorithm. Furthermore, [61] demonstrated that ROBEX outperformed other commonly used MRI processing tools (FSL-BET, BSE, FreeSurfer, AFNI, BridgeBurner, and GCUT). It does not require hyperparameter tuning and can accurately achieve skull stripping while retaining the entire brain structure of subjects to the greatest extent possible.

Regarding the massive number of studies on this topic, it was necessary to define criteria for selecting papers that are closer to our work. Therefore, we are interested in the studies that followed the 3D subject-level approach for only the MRI modality and used the baseline scans or the first scans from baseline visits on the same classes from the ADNI dataset.

In our study, we try to remove the ambiguity that showed up clearly in the previous work and overcome the limitations stated previously. Accordingly, we focus on designing an end-to-end system that can differentiate between CN, MCI, and AD individuals through four classification tasks. Our research contributions are listed as follows:

1) We utilized two different types of field magnetization strength, 1.5T and 3T images from the ADNI dataset, and tested the performance of our framework using each of them.
2) In image preprocessing, we introduced a simple and fast pipeline by quantifying the impact of excluding the registration from the steps by recording a time analysis. We also measured the effectiveness of applying augmentation in improving the results.
3) We evaluated the most effective 3D deep learning architectures and compared them in several experiments.
4) We propose an ensemble mechanism that fuses the three top-performing architectures.

The remainder of this paper is organized as follows. In Section 2, the proposed methodology is introduced along with the dataset description, pre-processing pipeline, augmentation, and details of our experiments on different architectures. We report the analysis and results of these experiments in Section 3. Finally, the discussion and conclusions are presented in Sections 4 and 5, respectively.

II. MATERIALS AND METHODS

In this section, we present our proposed methodology for Alzheimer's detection by discussing each component required to structure the entire system. First, we examined baseline MRI images. Subsequently, a simple registration-free preprocessing pipeline was applied to the dataset. Extensive experiments were conducted to analyze the performance of different 3D classification architectures. Finally, an ensemble learning approach was applied to the top-performing models. The workflow of the components was designed as a block diagram, as shown in fig. 1.

A. DATASET

In this study, the ADNI database was obtained for analysis, which is publicly available on their website. The ADNI is a global longitudinal multicenter research study that actively works on the early detection and tracking of Alzheimer's disease. ADNI is composed of two different field magnetization

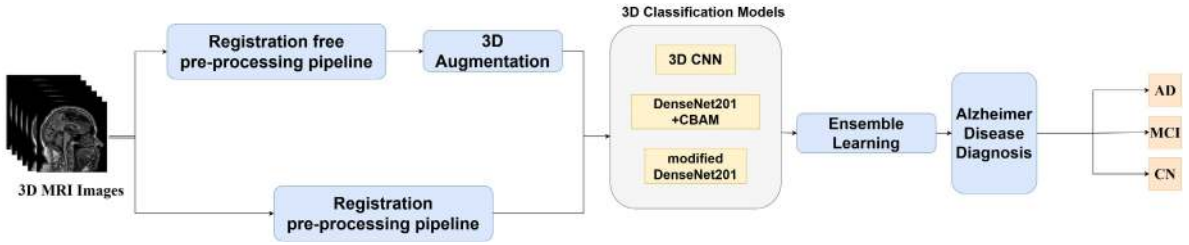

FIGURE 1. Workflow of our proposed system.

TABLE 1. Summary of participant demographics of the small dataset (3T images) in ADNI.

	Subjects	Age	Gender
AD	31	74.23 ± 8.41 [65.82, 82.64]	M: 11 / F: 20
MCI	73	75.22 ± 7.95 [67.27, 83.17]	M: 45 / F: 28
CN	46	75.32 ± 3.99 [71.33, 79.31]	M: 18 / F: 28
Total Number of scans = 150			

Values are presented as mean ± SD [range]. M: male, F: female.

TABLE 2. Summary of participant demographics of the large dataset (1.5T images) in ADNI.

	Subjects	Age	Gender
AD	133	75.63 ± 7.59 [68.04, 83.22]	M: 69 / F: 64
MCI	311	75.93 ± 7.13 [68.8, 83.06]	M: 201 / F: 110
CN	195	77.08 ± 5.22 [71.86, 82.3]	M: 102 / F: 93
Total Number of scans = 639			

Values are presented as mean ± SD [range]. M: male, F: female.

strengths, 1.5T and 3T, which result in two sub-datasets. The 3T dataset was small, with 150 structural T1-weighted MRI images. The 1.5T dataset is the largest one, which includes a total of 639 structural T1-weighted images. This study included both types of images in the analysis pipeline. These images were collected from subjects and included three diagnosis groups of patients: MCI, AD, and CN. Most of the subjects have multiple scans based on the follow-up visits over 36 months. In the literature on using this data, there were different methods of dealing with the scans, such as using only one scan for each subject, taking the first scan and the last one from the follow-up visits of each subject, or taking all the visits' scans and treating them independently. We chose baseline MRI images corresponding to all visits of each subject by considering only the scan of the first visit of the patient. Because of this, we were able to avoid data leakage, which occurs when each subject appears in both the training and testing sets, by taking multiple scans. The distribution of the participants in each group and a summary of their demographic information are shown in Tables 1 and 2.

B. DATA PREPROCESSING

The structure of biomedical images is quite different from that of natural images because they measure the physical properties of the human body. Owing to the variation in the acquisition type, biomedical images require special preprocessing steps. As mentioned, we are particularly interested in MRI modalities from the ADNI dataset. MRI scans are typically volumetric (3D) images. They have already undergone pre-processing steps, including multiplanar reconstruction (MPR), Gradwarp, B1 non-uniformity correction, and N3 intensity normalization during the data acquisition process [62].

In the literature, many different procedures have been proposed for preprocessing MRI images. However, one of the objectives of this study is to define a clear and fast preprocessing pipeline. First, we started with whole-brain extraction from the MRI images, also known as skull stripping. This is the first and most required component of most neuroimaging pipelines. This is a pivotal step towards the robustness of the full preprocessing pipeline and the performance of the entire system. This step aims to simplify the input of the training model by eliminating the noise from the images, as the skull is irrelevant information for AD prediction. Many algorithms and techniques were used in this step. In this study, we used the Robust Brain Extraction (ROBEX) algorithm [61]. This method includes discriminative and generative models to achieve the required function. The discriminative model was a random forest classifier trained to detect the brain boundary by finding the contour with the highest likelihood. The generative model is a point distribution model that ensures that the result is reasonable through contour refinement using graph cuts. Aligning several 3D brain volumes for optimized deep learning training is controversial [59], [63], which has motivated us to study its effect as a preprocessing step on the whole pipeline. As it is commonly employed as a tool to preprocess data for subsequent tasks like object detection, segmentation, or classification. We utilized nonlinear registration by adapting all images into the same space. We used a nonlinear image registration algorithm based on diffusion imaging in Python (Dipy) [64]. The MNI 305 atlas [65] was used as the reference template for the registration step. There are also many problems specific to MRI, such as inhomogeneous image intensities in MR images due to the scanner, and inconsistent tissue intensities across different MR scanners. Therefore, we performed voxel-intensity normalization. Additionally, we accounted for voxel spacing, which may vary between the images. To mark only the important regions from the volume, 3D cropping was applied to extract only the brain from the black background. Finally, all the images must be resized to the same output size.

TABLE 3. Time analysis for each step in the preprocessing pipelines.

Steps	Time recorded (seconds)
Skull Stripping	125.076
Intensity Normalization	0.1
Resampling	0.11
Cropping	0.15
Resizing	0.09
Non-linear Registration	662.665
Augmentation	1.156

We present two preprocessing pipelines. The first excludes the registration step from the pipeline and the second includes the registration step after skull stripping. Figure 2 shows the preprocessing pipelines. We also demonstrated a time analysis for each step included in the two pipelines, as shown in table 3.

C. AUGMENTATION

Augmentation is an essential tool for artificially increasing the dataset size. This alleviates the lack of labeled data, which is a major issue in the medical field. Transformations, including rotations, reflections, and elastic deformations, are widely used in augmentation methods to produce training images that closely mimic specific training examples. In this study, we used a 3D rotation strategy to augment the training dataset. We provide a rotation function with an angle range from zero to 90 °. The augmenting factors differ from the 1.5T to the 3T images and between the three classes. As there is a clear imbalance between classes, using augmentation helped overcome this issue. For the 1.5T images, the dataset sizes after applying the augmentation were 904 AD, 1060 MCI, and 996 CN. For the 3T images, the size of the images after applying the augmentation is as follows: 74 AD, 59 MCI, and 74 CN.

D. 3D CLASSIFICATION MODELS

In this section, we present the different 3D classification models used in this study. 3D convolutional neural networks (3D CNN), modified DenseNet201-based transfer learning, vgg-19, vision transformers, and attention modules are among the models included.

1) 3D CONVOLUTION NEURAL NETWORKS

CNNs are multilayered structures that can successfully capture spatial and temporal dependencies in an image by applying suitable filters. They include the convolution layer, pooling layer, number of consecutive fully connected layers, and output layer. The modern CNN-based model was manually designed by researchers using several different layers. In this paper, we presented, reproduced, and modified different 3D CNN basic architectures from previous studies [25], [34], [58]. The best model architecture was selected based on results reported by Zunair et al. [66]. The architectural details of the 3D CNN model are shown in fig. 3.

2) MODIFIED 3D DenseNet201 BASED TRANSFER LEARNING

Transfer learning occurs when existing models are repurposed to address new challenges or problems. In medical imaging, such as cancer classification, heart function quantification, and lung disease classification, transfer learning has yielded promising results [67], [68]. All the outcomes were demonstrated using transfer learning, delivered excellent classification accuracy in medical domains, and reached maximum results on AD classification with a small number of datasets.

In this study, we used classification models pre-trained on ImageNet data [69] with 1000 classes. We selected the DenseNet201 architecture for this purpose. The rationale behind the selection of the DenseNet architecture is that it produces high-accuracy results and more effective performance for computer-aided diagnosis problems. In most of the previous work, it proved a superior performance on the same task [33], [49], [50], [51]. It has been proven that DenseNet architectures utilize parameters more efficiently than alternative architectures [70].

The primary focus of the DenseNet architecture is the modification of the standard CNN. In this architecture, each layer is connected to another layer. For the L layers, there were L(L+1)/2 direct connections. The feature maps of all the previous layers were utilized as inputs for each layer, and their feature maps were used as inputs for the subsequent layers. Each architecture in the DenseNets consists of four dense blocks with varying numbers of layers. Transition layers were employed between the four dense blocks. The fourth dense block is followed by a classification layer that accepts the feature maps of all layers of the network to perform classification. To utilize transfer learning, we extracted all convolutional layers of the model as transfer layers and added the classifier part to these layers by modifying the fully connected layers and the output classification layer. Therefore, they learn the class-specific features of the Alzheimer's dataset by applying fully connected layer compression and dropout addition. The proposed network architecture is shown in fig. 4. The first layer was trained using ImageNet, whereas the ADNI dataset was used to train the remaining adaptation layers. The size of the output layer was set equal to the class labels.

3) VISION TRANSFORMERS

The vision transformer (ViT) is a transformer used in computer vision that operates similarly to transformers used in natural language processing. The transformer learns internally by measuring the relationship between input token pairs. In computer vision, we can use image patches as tokens. Inspired by [71], they applied a pure transformer-based model for video classification by suggesting four variants of the vision transformer architectures. In our work, we chose the spatio-temporal attention model for the dataset tasks for Alzheimer's diagnosis. The model captures the spatial and temporal dependencies across slices from each volume.

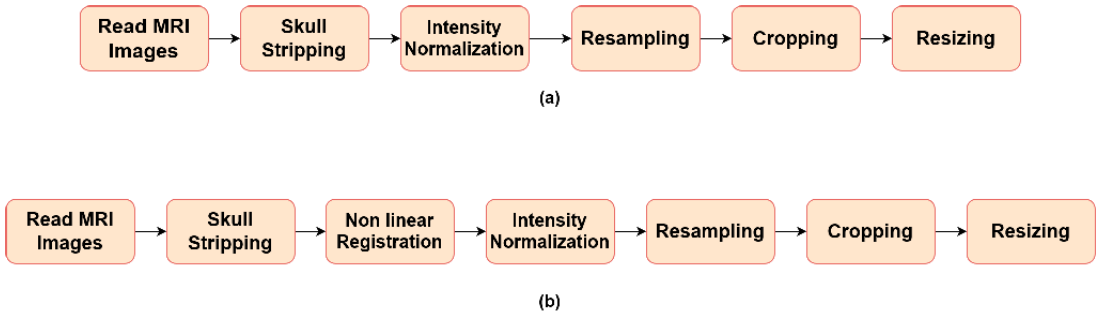

FIGURE 2. The preprocessing pipeline: (a) Registration free preprocessing pipeline, (b) Registration Preprocessing pipeline.

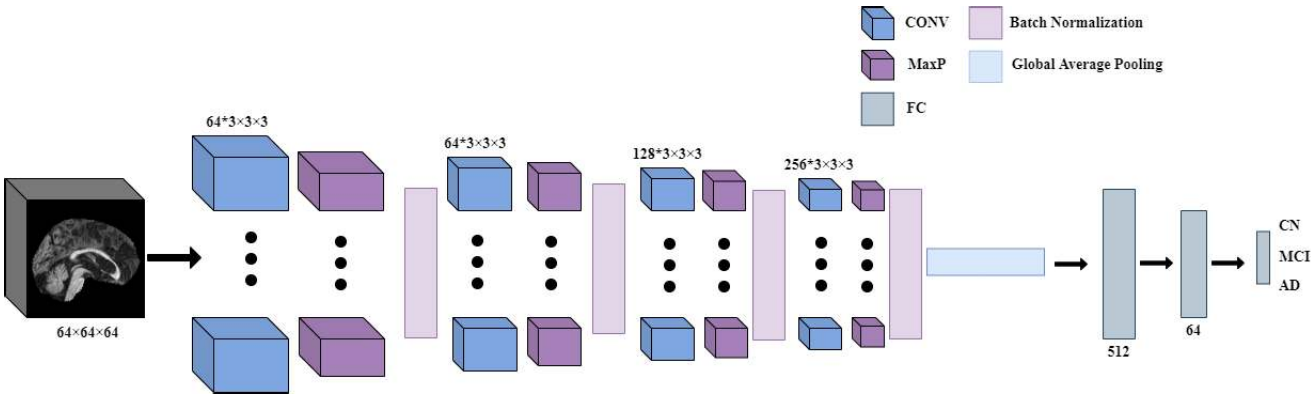

FIGURE 3. 3D CNN model architecture.

Hence, it models the long-range interactions across the volume from the first layer of the transformer. To obtain the input to the Spatio-temporal transformer, each volume is mapped to a sequence of tokens using a tublet embedding, and then a positional embedding is added. This type of transformer is explored by [72] in their "Joint Space-Time" model. Another reason for selecting this transformer is its speed in training on our dataset, in contrast to different complex 3D CNN architectures.

4) CHANNEL AND SPATIAL ATTENTION-BASED DenseNet

The convolutional block attention module (CBAM) is an effective attention module for convolutional neural networks [73]. It consists of two sequential submodules: channel and spatial modules, as illustrated in fig. 5. The focus of channel attention is on 'what' is meaningful given an input image. In contrast to channel attention, spatial attention focuses on 'where' as an informative component that complements channel attention. Every convolutional block of a convolutional neural network model includes CBAM as a layer. It accepts a tensor containing the feature maps from the previous convolutional layers and refines them using a channel attention module. Subsequently, the refined tensor was passed to the spatial attention module, which applied spatial attention, resulting in the output of the refined feature maps. Therefore, at each convolutional block of the deep network, the intermediate feature map is adaptively refined by the CBAM module. In this study, CBAM was added as a module after each dense block in the modified 3D DenseNet201 architecture.

5) EVALUATION METRICS

In this study, we evaluated the performance of the models using the following metrics: accuracy (ACC), area under the receiver operating characteristic (ROC) curve (AUC), Balanced Accuracy (BA), and f1-score. The performance can be obtained from the confusion matrices. The confusion matrix of the multiclass classification is shown in Table 4.

1) Accuracy shows the percentage of the correctly classified subjects among the entire subgroups

$$Acc(binary) = \frac{(T_P + T_N)}{(T_P + F_P + T_N + F_N)} \tag{1}$$

where T_P, F_P, T_N, and F_N are the true positive, false positive, true negative, and false negative, respectively.

$$Acc(multiclass) = \frac{(T_A + T_B + T_C)}{(T + F)} \tag{2}$$

2) AUC is the summary of the ROC curve that measures the ability of a classifier to distinguish between classes. As ROC plots the true positive rate against the false positive rate at different threshold values.

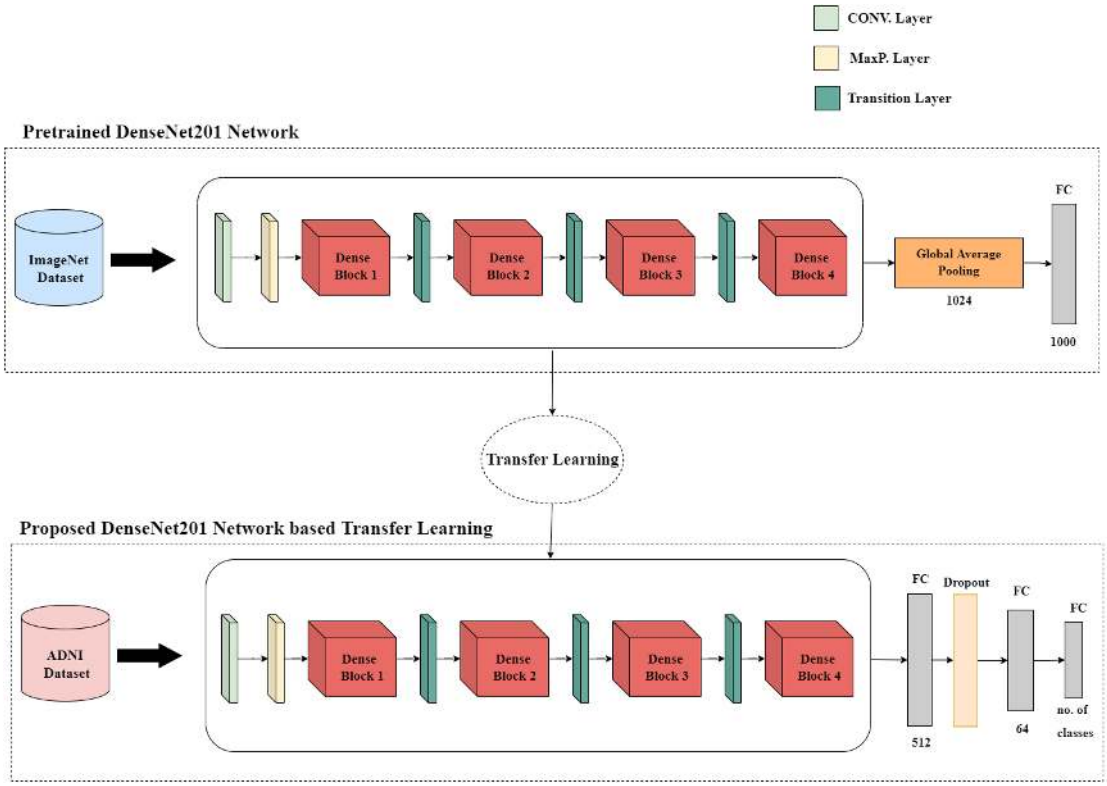

FIGURE 4. Proposed 3D DenseNet201 architecture.

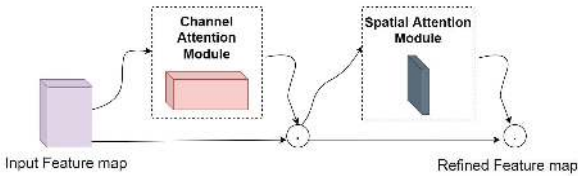

FIGURE 5. Overview of CBAM architecture.

3) F1-score considers both precision and recall and performs a synthetic evaluation of the model performance whereas precision calculates the percentage of samples properly classified and the recall is the percentage of relevant instances recovered compared to the total number of relevant occurrences.

$$Precision(binary) = \frac{T_p}{(T_P + F_P)} \tag{3}$$

$$Precision(classA) = \frac{T_A}{(T_A + F_{BA} + F_{CA})} \tag{4}$$

$$Recall(binary) = \frac{T_P}{(T_P + F_N)} \tag{5}$$

$$Recall(classA) = \frac{T_A}{(T_A + F_{AB} + F_{AC})} \tag{6}$$

$$F1 - score = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \tag{7}$$

TABLE 4. Confusion matrix of multiclass classification.

	Predicted Positive(class A)	Predicted Negative(class B)	Predicted Negative(class C)
Actual positive (class A)	True A (TA)	False AB (FAB)	False AC (FN)
Actual negative (class B)	False BA (FP)	True B (TB)	True BC (TN)
Actual negative (class C)	False CA (FP)	True CB (TN)	True C (TC)

4) BA is a metric that is used when the classes are imbalanced. It evaluates how well the classifier is whether the task is binary or multiclass.

$$BA = \frac{(Precision + Recall)}{2} \tag{8}$$

where precision and recall differ from binary and multiclass cases as illustrated in equations 3,4,5, and 6.

III. EXPERIMENTS DETAILS AND RESULTS

In this section, the dataset and preprocessing setup, an ablation study to illustrate the effectiveness of each module in our method, the experimental setting, and the results of the summarization are introduced.

A. DATASET AND PREPROCESSING SETUP

A total of 789 3D MRI images from 1.5T and 3T images were used. We randomly selected training and testing subjects. Only one scan was obtained for each participant during the first visit. A 5-fold cross-validation strategy was performed, which resulted in a fold (20%) of the data for validation and

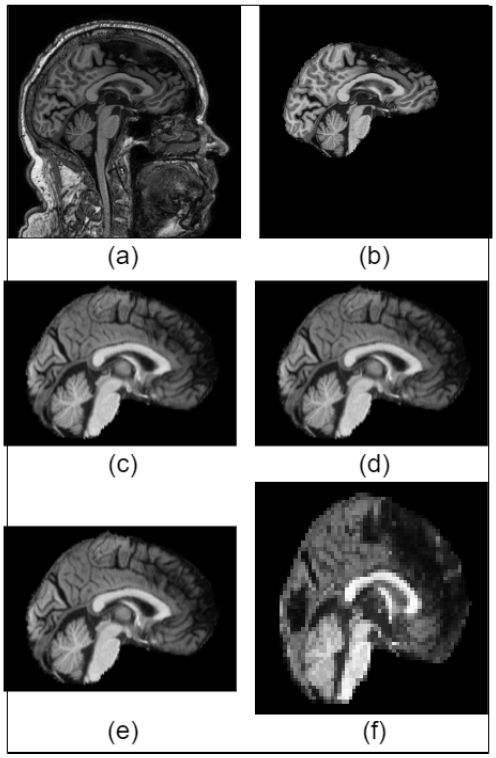

FIGURE 6. Output of the preprocessing pipeline with registration step: (a) the original image, (b) Skull Stripped image, (c) Registered image, (d) Normalized image, (e) Resampled image, (f) Resized image.

the rest for training. These images were then fed into the preprocessing pipeline.

To apply the skull stripping algorithm, we used the Pyrobex library in Python [74]. This software package was used for robust brain extraction. We used Dipy for the registration. It is a Python package that is used for diffusion imaging. It includes general image registration algorithms, including affine and nonlinear registration. Subsequently, a zero-mean, unit-variance normalization method was applied to the images for intensity normalization. For resampling, we resampled the images to an isotropic resolution of 1 mm^2. Finally, we cropped and resized the images to $64\times64\times64$ as the output dimension, according to the required input size for the used models. The output image of each step of the pipeline is shown in fig.6.

B. ABLATION STUDY

This study is focused to investigate the independent effect of the following parameters on our main method's performance such as the registration and augmentation preprocessing steps, different types of field magnetization's strength from the ADNI dataset, different 3D encoding architectures, and the different classification tasks.

1) THE REGISTRATION AND AUGMENTATION STEPS

First, we studied the effect of excluding the registration step and applied the augmentation strategy through a registration-free pipeline. We applied augmentation using the 3D image rotation method, and we created augmented data for each fold only on the training part and concatenated them to the original images. The second pipeline included a registration step to check its impact on the performance of the entire system. We did not apply both registration and augmentation at the same time because they will cancel the effect of each other. To verify the effectiveness of these two steps, we applied this experiment by utilizing the large dataset (1.5T images).

TABLE 5. The models' complexity.

Model	# Parameters
3D CNN	1,388,802
Modified Densenet201	25,603,776
Modified Densenet201 with CBAM	35,315,690
Vgg-19	60,062,144
Vision transformer	3,247,778

2) DIFFERENT TYPES OF FIELD MAGNETIZATION STRENGTH FROM THE ADNI DATASET

We aimed to investigate the effect of using images that were obtained from different MRI scanners with different magnetization field strengths. Therefore, we applied our experiments to 1.5T and 3T and merged these two image types. Using 3T images helped us illustrate the influence of using images of higher quality than the 1.5T images. As the cost is a critical factor in the market, we also focused on applying our experiments on 1.5T images because their scanners are more affordable than 3T scanners [75], [76], and their size is larger than that of 3T images. We also tried to generalize our method by applying our experiments to the merged dataset by concatenating the 1.5T and 3T images.

3) DIFFERENT 3D ARCHITECTURES

We studied the effect of using different 3D effective architectures in our experiments. We evaluated the 3D CNN, 3D modified DenseNet, and 3D denseNet201 with the CBAM attention module, vgg-19, and 3D vision transformer. We selected these 3D models based on their superior performance reported in the literature. For example, we focused on applying the 3D denseNet 201 architecture in two ways, as it has the advantages of feature propagation improvement and efficient parameter utilization. The complexity of the models is a critical parameter for deep learning models. Accordingly, we reported the number of parameters of each of these five models in table 6 as a metric for the model's complexity.

4) DIFFERENT CLASSIFICATION TASKS

Four classification tasks are investigated in this study, listed as follows: (1) AD vs CN, (2)AD vs MCI, (3) MCI vs CN, and (4) AD vs MCI vs CN. Understanding each stage is crucial for the accurate diagnosis of the disease at each level. Most previous studies focused only on binary classification tasks,

and there was a deficiency in multiclass classification tasks [77]. Hence, we targeted both binary and multiclass tasks.

C. EXPERIMENTAL SETTING

In this study, there are three groups of experiments, as shown in the experimental structure in fig.7: experiments using a small dataset (3T images), using a large dataset (1.5T Images), and finally using the merged dataset (3T and 1.5T images). In our ablation study, we fixed some parameters at each level of the modules in our proposed system based on the results obtained from each experiment. We began with an investigation to check the impact of using registration and augmentation techniques. We applied this experiment only on the 1.5T images, as it has more advantages than using 3T or merged data. In Table 6, we obtained the AD vs. CN task results by showing the effect of using augmentation and without it versus using registration. The table shows that the best results were obtained using the proposed Densenet201 model on the augmented dataset, with an AUC score of 95.82%. The second-best performance results were obtained for the Densenet201 with CBAM architecture, with an AUC score of 95.37%. We can see that the registration results were lower than the augmentation results. Based on the analysis of the remaining results, we conclude that our proposed DenseNet201, DenseNet201 with CBAM, and 3D CNN are the best-performing models compared to the rest of the models. Therefore, we applied ensemble learning to these three models based on an averaging technique to our proposed algorithm and recorded the results. Each model may perform well on certain data while performing poorly on others. Therefore, ensemble learning was efficient. When all of them are combined, their flaws are balanced out.

Subsequently, we fixed our next experiments to be applied using augmentation and only the three top-performing models, as concluded from the previous experiment. We tested the performance of these three models on the merged dataset for all four classification tasks, as listed in Tables 7–10.

To check the impact of using 1.5T, 3T, and merged images, we assessed the three top-performing models on the three datasets for all classification tasks and reported the ensemble results in Table 11.

All experiments were performed using 50 epochs with a batch size of 8, and the Adam optimizer was adopted with a LearningRateSchecule that uses a piecewise constant decay schedule. The learning rate was 1e−3 for the first 6000 steps, 1e−4 for the next 6000 steps, 5e−4 for the next 6000 steps, and 1e−5 for any additional step. The general network parameter configuration for each experiment was the same. These parameters were selected based on many experiments and our computational resources' limitations. The implementation of our experiments was performed using the Keras library in Python 3.7, on K80, P100, and T4 GPU machine.

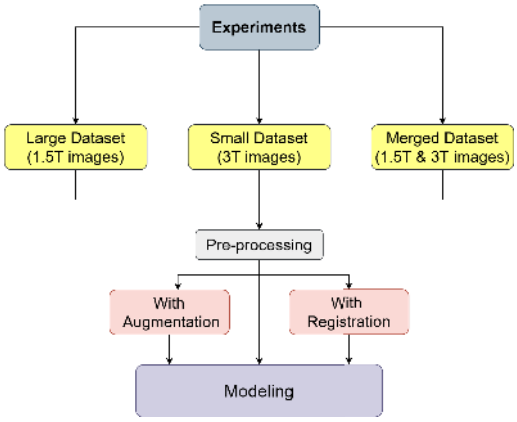

FIGURE 7. The experiments' structure.

IV. DISCUSSION

In this study, we introduced an end-to-end system to recognize the three AD stages using four classification tasks. We conducted extensive experiments to highlight the contributions of this study. Based on the experimental results presented in Section 3, we can infer that

We assessed the performance using the ADNI dataset with different image quality types. As shown in Table 11, the results of the AD vs. CN and AD vs. CN vs. MCI tasks improved when using the merged data. However, 3T images achieved higher results on the AD vs. MCI and MCI vs. CN tasks. The classification of the MCI class is much more complicated. Therefore, AD vs. MCI and MCI vs. CN tasks were considered challenging to the model. Therefore, a dataset with good image quality is an immersive factor for model learning. We assume that the quality factor in the 3T images helps the model learn these demanding patterns from the classes and achieve better results.

Another aim of this study was to set a clear, simple, and fast preprocessing pipeline for images. In our proposed pipeline, we focused on how to prepare the images to be fed into the model in a few steps. As demonstrated in table 6, the results of using only augmentation outperform the experiment of applying the registration. We concluded that the skull stripping and the registration are the most time-hungry steps, as reported in table 3. Hence, we did not perform any alignment for the images because we noticed that the time taken for each volume to be registered with 170 slices was approximately 11 min. This time is huge compared to the time taken to apply the augmentation function on each volume, which is less than 1 minute. Based on the time analysis that is reported in table 3, the total required time for the registration-free pipeline is 126.547 seconds, while the total required time for the registration pipeline is 788.191 seconds. By studying the influence of augmentation in the experiments, all experiments performed with augmentation outperformed the experiments that were performed without applying augmentation. Hence, we conclude that the more data samples generated, the better

TABLE 6. All experiments Results for AD vs CN task on 1.5T dataset.

Model	With Augmentation				Without Augmentation				With Registration			
	ACC	BA	AUC	F1-score	ACC	BA	AUC	F1-score	ACC	BA	AUC	F1-score
3D CNN	88.3	88.178	94.62	AD: 87.44, CN: 89.06	73.82	70.539	77.38	AD: 60.54,CN: 80.04	83.5	82.01	85.42	AD:77.18, CN:86.92
Modified Densenet201	**90.74**	**90.734**	**95.82**	**AD: 90.16, CN:91.24**	76.34	73.46	81.78	AD: 65.28, CN: 81.58	80.26	79.278	83.65	AD: 75.2, CN: 83.58
Densenet201 with CBAM	88.66	88.576	95.378	AD:87.68, CN: 89.48	74.57	71.407	80.74	AD: 65.06, CN:78.12	81.32	80.06	84.799	AD: 76, CN: 84.4
Vgg-19	52.42	50	50.14	AD: 0, CN: 68.75	58.025	50	52.945	AD: 0,CN:73.35	59.48	50	36.33	AD:0, CN:74.57
Vision transformer	65.74	64.81	69.98	AD: 60.22,CN: 67.7	65.22	59.689	58.96	AD: 37.26, CN: 75.38	68.06	72.676	76.794	AD: 65.52, CN: 78.52
Proposed ensemble method	89.23	89.16	95.27	AD:88.42,CN: 89.92	74.91	71.80	79.96	AD:63.62, CN:79.913	81.69	80.44	84.62	AD:76.12, CN:84.96

TABLE 7. All experiments Results for AD vs CN task on merged dataset.

Model	With Augmentation			
	ACC	BA	AUC	F1-score
3D CNN	88.04	87.97	94.25	AD:87.52, CN:88.44
Modified DenseNet201	**90.62**	**90.59**	**95.65**	**AD:90.22, CN:90.96**
DenseNet201 with CBAM block	89.74	89.67	95.39	AD:89.09, CN:90.29
Proposed ensemble method	89.46	89.41	95.09	AD:88.94, CN:89.89

TABLE 8. All experiments Results for AD vs MCI task on merged dataset.

Model	With Augmentation			
	ACC	BA	AUC	F1-score
3D CNN	77.32	76.95	84.13	AD:74.60 MCI:79.13
Modified DenseNet201	**80.30**	**80.25**	**87.36**	**AD:79, MCI:81.36**
DenseNet201 with CBAM block	78.20	77.60	85.96	AD:74.55, MCI:80.88
Proposed ensemble method	78.60	78.26	85.81	AD:76.05, MCI:80.45

TABLE 9. All experiments Results for MCI vs CN task on merged dataset.

Model	With Augmentation			
	ACC	BA	AUC	F1-score
3D CNN	78.39	78.52	85.19	MCI:78.36, CN:78.34
Modified DenseNet201	**79.62**	**79.35**	**86.92**	**MCI:80.04, CN:78.80**
DenseNet201 with CBAM block	78.57	78.78	84.78	MCI:79.13, CN:77.96
Proposed ensemble method	78.86	78.88	85.63	MCI:79.17, CN:78.36

TABLE 10. All experiments Results for AD vs MCI vs CN task on merged dataset.

Model	With Augmentation		
	ACC	BA	F1-score
3D CNN	68.72	68.87	AD:69.8, MCI:63.26, CN:72.92
Modified DenseNet201	**71.32**	**71.43**	**AD:74.04, MCI:65.76, CN:74.56**
DenseNet201 with CBAM block	70.96	71.11	AD:73.7, MCI:66.24, CN:74.46
Proposed ensemble method	70.33	70.47	AD:72.51, MCI:65.08, CN:73.98

TABLE 11. Ensemble learning results of different ADNI's data type for the four classification tasks.

Dataset	AD vs CN			
	ACC	BA	AUC	F1-score
1.5T	89.23	89.16	95.27	AD:88.42, CN: 89.92
3T	76.33	76.90	76.7	AD:73.51, CN:76.71
Merged	**89.46**	**89.41**	**95.09**	**AD:88.94, CN:89.89**
	AD vs MCI			
Dataset	ACC	BA	AUC	F1-score
1.5T	77.76	77.49	84.73	AD:75.7, MCI:79.24
3T	**88.01**	**88.06**	**91.28**	**AD:88.48, MCI:86.02**
Merged	78.60	78.26	85.81	AD:76.05, MCI:80.45
	MCI vs CN			
Dataset	ACC	BA	AUC	F1-score
1.5T	74.7	74.65	80.90	MCI:75.58, CN:73.43
3T	**85.93**	**83.71**	**88.42**	**MCI:81.5, CN:86.72**
Merged	78.86	78.88	85.63	MCI:79.17, CN:78.36
	AD vs MCI vs CN			
Dataset	ACC	BA	AUC	F1-score
1.5T	69.41	69.69	-	AD:72.04, MCI:64.10, CN:72.16
3T	64.57	64.78	-	AD:64.36, MCI:68.18, CN:59.21
Merged	**70.33**	**70.47**	**-**	**AD:72.51, MCI:65.08, CN:73.98**

the performance of the models. Augmentation helps not only in increasing the data size but also in resolving the problem of imbalance between classes. We can see the effect of augmentation on this problem by showing the BA and F1-score metric values at the class level. In all experiments, there was a noticeable change in the values with and without the augmentation experiments. We were able to achieve superior results with a few steps of analysis of the complex MRI images instead of the heavy multistep pipelines that currently prevail in studies, as shown in Table 6. It was difficult to do a time analysis of other preprocessing tools that are frequently used in the literature, as each study works on different machine resources with a different setup and all these variables are hard to be fixed.

Five different models were evaluated in our study using different dataset types. Our proposed DenseNet201, DenseNet201 with channel-and spatial attention-based, and 3D CNN models master the rest of the models. The best results were obtained from the experiment using the merged dataset by applying augmentation with an AUC score of 95.27% for the AD versus CN task. However, vgg-19, which is the most complex model compared to the others, as illustrated in Table 5, achieved worse results, as shown in Table 6. There are different approaches to applying ensemble learning. The ensemble members, or the models that contribute

TABLE 12. Algorithm comparisons for the four classification tasks.

Classification Task	Study	Dataset Type	No. of scans	Registration	Augmentation	ACC (%)	BA (%)	AUC (%)	F1-score (%)
AD vs CN	Korolev et al. [58]	unclear	111	Yes	No	79	N/A	88	N/A
	Senanayake et al. [10]	unclear	322	No	No	76	N/A	N/A	N/A
	Vu et al. [11]	unclear	400	Yes	No	86.25	N/A	N/A	N/A
	Wen et al. [25]	unclear	666	Yes	No	82	N/A	N/A	N/A
	Venugopalan et al. [12]	unclear	398	Yes	No	86	N/A	N/A	86
	Qin et al. [19]	unclear	212	No	No	**92.68**	N/A	N/A	**91.89**
	Proposed ensemble approach	merged	405	No	Yes	89.46 ± 1.03	89.41 ± 0.88	95.09 ± 0.77	AD:88.94 ± 0.78, CN:89.89 ± 1.14
AD vs MCI	Senanayake et al. [10]	unclear	354	No	No	76	N/A	N/A	N/A
	Vu et al. [11]	unclear	408	Yes	No	76.52	N/A	N/A	N/A
	Proposed ensemble approach	3T	104	No	Yes	**88.01 ± 0.59**	88.06 ± 1.23	91.28 ± 0.37	AD:88.48 ± 0.07, MCI:86.02 ± 1.67
MCI vs CN	Senanayake et al. [10]	unclear	354	No	No	75	N/A	N/A	N/A
	Vu et al. [11]	unclear	422	Yes	No	85.66	N/A	N/A	N/A
	Proposed ensemble approach	3T	119	No	Yes	**85.93 ± 0.71**	83.71 ± 1.81	88.42 ± 1.54	MCI:81.5 ± 2.73, CN:86.72 ± 0.61
AD vs MCI vs CN	**Proposed ensemble approach**	merged	509	No	Yes	70.33 ± 1.15	70.47 ± 1.13	-	AD:72.51 ± 1.92, MCI:65.08 ± 1.25, CN:73.98 ± 0.75

to the ensemble, might be of the same or distinct type. In our study, we used several different models. We focused on applying both approaches to the ensemble. We applied the ensemble to every single model using 5-fold cross-validation. We have also combined the top three most performed models. The proposed ensemble results for the different data types for the four classification tasks are summarized in Table 11.

As mentioned before, there is a clear ambiguity in the previous work. We compared our proposed approach to the studies that followed our selection criteria, which were mentioned in the introduction section. We consider the dataset used and its size, the existence of registration or augmentation in their work, and the evaluation metrics used. Table 12 shows a comparison of our proposed system with previous methods for the four classification tasks. All studies mentioned in Table 12 follow the 3D subject-level approach with baseline scans. In most studies, it was difficult to detect the type of dataset used in ADNI. In Table 12, unclear means that we cannot detect the exact type of the dataset, and N/A means that the paper did not provide this evaluation metric. The uncertainty of the predictive models is important in the scope of our problem which is dependent on two aspects: the data uncertainty and modeling process uncertainty. Regarding the data uncertainty, labeling was confirmed by a clinical based diagnosis not image based. Image data acquisition quality is also an important aspect where in ADNI, the structural MRI scans with image data have been released to the public after passing quality control (QC) tests at the Mayo Clinic's Aging and Dementia Imaging Research Laboratory [78]. For the modeling uncertainty, we designed our experiments to incorporate k-fold analysis and we have added the mean ± std to represent the central tendency of the model performance as well as its variance as shown in table 12.

In [25], the authors studied the performance of different approaches for MRI images using minimal and extensive preprocessing pipelines. Their minimal pipeline included a linear (affine) registration step. We compared our proposed method, which does not include the registration step in the pipeline, with their experiment, which applies the minimal preprocessing pipeline on baseline images of the AD vs. CN task.

According to the comparison of our method with other studies, our method outperforms the previous studies for distinguishing between people with AD and MCI, and MCI and CN, with an AUC score of 91.28 % and 88.42%, respectively. Based on our selection criteria, we did not find any study that worked on the multiclass task. Accordingly, we were able to determine the severity of the disease and aid in its early detection. We achieved an ACC of 70.33% for this task. Our work can be extended to further research using datasets other than ADNI.

V. CONCLUSION

In this paper, we present an early diagnosis system for Alzheimer's disease that recognizes AD, MCI, and CN using MRI images. To address the problem of data leakage, we demonstrated a system that uses only the first visit of each patient and ignores the others. Extensive comparative experiments were performed to study the effects of using various data types from the ADNI and the size of the dataset. We investigated different pre-processing pipelines and evaluated them using the most effective 3D classification models. Furthermore, the proposed method alleviates the ambiguity of previous studies. Regarding the AD vs. MCI and MCI vs. CN tasks, we achieved the best AUC scores using the proposed ensemble approach. It combines the three top-performed models with a simple and fast preprocessing pipeline and augmentation. Our proposed approach also gives a good classification accuracy of 70.33%, for all categories, indicating that it is a promising approach hence, contributing to the early prediction of the disease. In future work, we will use the clinical data from ADNI as another modality besides MRI to increase the performance of our results and to design a more reliable assistant tool for physicians.

REFERENCES

[1] S. C. Waring, R. S. Doody, V. N. Pavlik, P. J. Massman, and W. Chan, "Survival among patients with dementia from a large multi-ethnic population," *Alzheimer Disease Assoc. Disorders*, vol. 19, no. 4, pp. 178–183, 2005.
[2] R. Brookmeyer, M. M. Corrada, F. C. Curriero, and C. Kawas, "Survival following a diagnosis of Alzheimer disease," *Arch. Neurol.*, vol. 59, no. 11, pp. 1764–1767, 2002.

[3] *FastStats—Deaths and Mortality*. Accessed: Jul. 26, 2022. [Online]. Available: https://www.cdc.gov/nchs/fastats/deaths

[4] M. Toepper, "Dissociating normal aging from Alzheimer's disease: A view from cognitive neuroscience," *J. Alzheimer's Disease*, vol. 57, no. 2, pp. 331–352, Mar. 2017.

[5] W. Jagust, "Imaging the evolution and pathophysiology of Alzheimer disease," *Nature Rev. Neurosci.*, vol. 19, no. 11, pp. 687–700, Nov. 2018.

[6] *Mini-Mental State Exam (MMSE) Test for Alzheimer's Dementia*. Accessed: Jul. 26, 2022. [Online]. Available: https://www.dementiacarecentral.com/mini-mental-state-exam

[7] S. E. O'Bryant, S. C. Waring, C. M. Cullum, J. Hall, L. Lacritz, P. J. Massman, P. J. Lupo, J. S. Reisch, and R. Doody, "Staging dementia using clinical dementia rating scale sum of boxes scores: A Texas Alzheimer's research consortium study," *Arch. Neurol.*, vol. 65, no. 8, pp. 1091–1095, 2008.

[8] E. Moradi, I. Hallikainen, T. Hänninen, and J. Tohka, "Rey's auditory verbal learning test scores can be predicted from whole brain MRI in Alzheimer's disease," *NeuroImage, Clin.*, vol. 13, pp. 415–427, Jan. 2017.

[9] R. Craig-Schapiro, A. M. Fagan, and D. M. Holtzman, "Biomarkers of Alzheimer's disease," *Neurobiol. Disease*, vol. 35, no. 2, pp. 128–140, 2009.

[10] U. Senanayake, A. Sowmya, and L. Dawes, "Deep fusion pipeline for mild cognitive impairment diagnosis," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1394–1997.

[11] T.-D. Vu, N.-H. Ho, H.-J. Yang, J. Kim, and H.-C. Song, "Non-white matter tissue extraction and deep convolutional neural network for Alzheimer's disease detection," *Soft Comput.*, vol. 22, no. 20, pp. 6825–6833, Oct. 2018.

[12] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, "Multimodal deep learning models for early detection of Alzheimer's disease stage," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, Dec. 2021.

[13] J. Song, J. Zheng, P. Li, X. Lu, G. Zhu, and P. Shen, "An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis," *Frontiers Digit. Health*, vol. 3, Feb. 2021, Art. no. 637386.

[14] N. Seshadri, D. S. McCalla, and R. Shah, "Early prediction of Alzheimer's disease with a multimodal multitask deep learning model," *J. Student Res.*, vol. 10, no. 1, pp. 1–8, Mar. 2021.

[15] S. Dwivedi, T. Goel, M. Tanveer, R. Murugan, and R. Sharma, "Multimodal fusion-based deep learning network for effective diagnosis of Alzheimer's disease," *IEEE MultimediaMag.*, vol. 29, no. 2, pp. 45–55, Apr. 2022.

[16] K. Mizuno and K. P. Giese, "Towards a molecular understanding of sex differences in memory formation," *Trends Neurosci.*, vol. 33, no. 6, pp. 285–291, Jun. 2010.

[17] B. Yan, Y. Li, L. Li, X. Yang, T.-Q. Li, G. Yang, and M. Jiang, "Quantifying the impact of pyramid squeeze attention mechanism and filtering approaches on Alzheimer's disease classification," *Comput. Biol. Med.*, vol. 148, Sep. 2022, Art. no. 105944.

[18] R. Sharma, T. Goel, M. Tanveer, and R. Murugan, "FDN-ADNet: Fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans," *Appl. Soft Comput.*, vol. 115, Jan. 2022, Art. no. 108099.

[19] Z. Qin, Z. Liu, Q. Guo, and P. Zhu, "3D convolutional neural networks with hybrid attention mechanism for early diagnosis of Alzheimer's disease," *Biomed. Signal Process. Control*, vol. 77, Aug. 2022, Art. no. 103828.

[20] M. El-Geneedy, H. E.-D. Moustafa, F. Khalifa, H. Khater, and E. AbdElhalim, "An MRI-based deep learning approach for accurate detection of Alzheimer's disease," *Alexandria Eng. J.*, vol. 63, pp. 211–221, Feb. 2023.

[21] S. Shaji, J. F. A. Ronickom, A. K. Ramaniharan, and R. Swaminathan, "Study on the effect of extreme learning machine and its variants in differentiating Alzheimer conditions from selective regions of brain MR images," *Expert Syst. Appl.*, vol. 209, Dec. 2022, Art. no. 118250.

[22] M. Orouskhani, C. Zhu, S. Rostamian, F. S. Zadeh, M. Shafiei, and Y. Orouskhani, "Alzheimer's disease detection from structural MRI using conditional deep triplet network," *Neurosci. Informat.*, vol. 2, no. 4, Dec. 2022, Art. no. 100066.

[23] S. A. Mofrad, A. Lundervold, and A. S. Lundervold, "A predictive framework based on brain volume trajectories enabling early detection of Alzheimer's disease," *Computerized Med. Imag. Graph.*, vol. 90, Jun. 2021, Art. no. 101910.

[24] X. Zhou, S. Qiu, P. S. Joshi, C. Xue, R. J. Killiany, A. Z. Mian, S. P. Chin, R. Au, and V. B. Kolachalama, "Enhancing magnetic resonance imaging-driven Alzheimer's disease classification performance using generative adversarial learning," *Alzheimer's Res. Therapy*, vol. 13, no. 1, pp. 1–11, Dec. 2021.

[25] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, and O. Colliot, "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101694.

[26] L. Bloch and C. M. Friedrich, "Machine learning workflow to explain black-box models for early Alzheimer's disease classification evaluated for multiple datasets," 2022, *arXiv:2205.05907*.

[27] N. Goenka and S. Tiwari, "AlzVNet: A volumetric convolutional neural network for multiclass classification of Alzheimer's disease through multiple neuroimaging computational approaches," *Biomed. Signal Process. Control*, vol. 74, Apr. 2022, Art. no. 103500.

[28] A. Mehmood, S. Yang, Z. Feng, M. Wang, A. S. Ahmad, R. Khan, M. Maqsood, and M. Yaqub, "A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images," *Neuroscience*, vol. 460, pp. 43–52, Apr. 2021.

[29] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song, "Early detection of Alzheimer's disease using magnetic resonance imaging: A novel approach combining convolutional neural networks and ensemble learning," *Frontiers Neurosci.*, vol. 14, p. 259, May 2020.

[30] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informat.*, vol. 5, no. 2, pp. 1–14, 2018.

[31] A. Loddo, S. Buttau, and C. Di Ruberto, "Deep learning based pipelines for Alzheimer's disease diagnosis: A comparative study and a novel deep-ensemble method," *Comput. Biol. Med.*, vol. 141, Feb. 2022, Art. no. 105032.

[32] W. Lin, T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, and X. Qu, "Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment," *Frontiers Neurosci.*, vol. 12, p. 777, Nov. 2018.

[33] H. Wang, Y. Shen, S. Wang, T. Xiao, L. Deng, X. Wang, and X. Zhao, "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease," *Neurocomputing*, vol. 333, pp. 145–156, Jan. 2019.

[34] S. Liu, C. Yadav, C. Fernandez-Granda, and N. Razavian, "On the design of convolutional neural networks for automatic detection of Alzheimer's disease," in *Proc. Mach. Learn. Health Workshop*, 2020, pp. 184–201.

[35] F. U. R. Faisal and G.-R. Kwon, "Automated detection of Alzheimer's disease and mild cognitive impairment using whole brain MRI," *IEEE Access*, vol. 10, pp. 65055–65066, 2022.

[36] K. Aderghal, M. Boissenin, J. Benois-Pineau, G. Catheline, and K. Afdel, "Classification of sMRI for AD diagnosis with convolutional neuronal networks: A pilot 2-D+ϵ study on ADNI," in *Proc. Int. Conf. Multimedia Modeling*. Cham, Switzerland: Springer, 2017, pp. 690–701.

[37] Z. Liu, H. Lu, X. Pan, M. Xu, R. Lan, and X. Luo, "Diagnosis of Alzheimer's disease via an attention-based multi-scale convolutional neural network," *Knowl.-Based Syst.*, vol. 238, Feb. 2022, Art. no. 107942.

[38] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, and M. Xu, "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *NeuroImage*, vol. 208, Mar. 2020, Art. no. 116459.

[39] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-learning-based disease diagnosis: A comprehensive review," *Healthcare*, vol. 10, no. 3, p. 541, Mar. 2022.

[40] F. Falahati, E. Westman, and A. Simmons, "Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging," *J. Alzheimer's Disease*, vol. 41, no. 3, pp. 685–708, Jul. 2014.

[41] S. Haller, K. O. Lovblad, and P. Giannakopoulos, "Principles of classification analyses in mild cognitive impairment (MCI) and Alzheimer disease," *J. Alzheimer's Disease*, vol. 26, no. s3, pp. 389–394, Oct. 2011.

[42] A. Raut and V. Dalal, "A machine learning based approach for early detection of Alzheimer's disease by extracting texture and shape features of the hippocampus region from MRI scans," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 6, no. 6, pp. 320–325, Jun. 2017.

[43] J. Bernal, K. Kushibar, D. S. Asfaw, S. Valverde, A. Oliver, R. Martí, and X. Lladó, "Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: A review," *Artif. Intell. Med.*, vol. 95, pp. 64–81, Apr. 2019.
[44] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.
[45] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification in BioApps*, 2018, pp. 323–350.
[46] M. Sethi, S. Ahuja, S. Rani, D. Koundal, A. Zaguia, and W. Enbeyle, "An exploration: Alzheimer's disease classification based on convolutional neural network," *BioMed Res. Int.*, vol. 2022, Jan. 2022, Art. no. 8739960.
[47] M. Khojaste-Sarakhsi, S. S. Haghighi, S. F. Ghomi, and E. Marchiori, "Deep learning for Alzheimer's disease diagnosis: A survey," *Artif. Intell. Med.*, vol. 130, Aug. 2022, Art. no. 102332.
[48] M. Odusami, R. Maskeliūnas, R. Damaševi ius, and S. Misra, "ResD hybrid model based on ResNet18 and DenseNet121 for early Alzheimer disease classification," in *Proc. Int. Conf. Intell. Syst. Design Appl.* Cham, Switzerland: Springer, 2022, pp. 296–305.
[49] B. Solano-Rojas and R. Villalón-Fonseca, "A low-cost three-dimensional DenseNet neural network for Alzheimer's disease early discovery," *Sensors*, vol. 21, no. 4, p. 1302, Feb. 2021.
[50] J. Ruiz, M. Mahmud, M. Modasshir, and M. S. Kaiser, "3D DenseNet ensemble in 4-way classification of Alzheimer's disease," in *Proc. Int. Conf. Brain Informat.* Cham, Switzerland: Springer, 2020, pp. 85–96.
[51] B. Solano-Rojas, R. Villalón-Fonseca, and G. Marín-Raventós, "Alzheimer's disease early detection using a low cost three-dimensional DenseNet-121 architecture," in *Proc. Int. Conf. Smart Homes Health Telematics*. Cham, Switzerland: Springer, 2020, pp. 3–15.
[52] A. Ebrahimi, S. Luo, and R. Chiong, "Deep sequence modelling for Alzheimer's disease detection using MRI," *Comput. Biol. Med.*, vol. 134, Jul. 2021, Art. no. 104537.
[53] L. Fulton, D. Dolezel, J. Harrop, Y. Yan, and C. Fulton, "Classification of Alzheimer's disease with and without imagery using gradient boosted machines and ResNet-50," *Brain Sci.*, vol. 9, no. 9, p. 212, Aug. 2019.
[54] D. Jin, J. Xu, K. Zhao, F. Hu, Z. Yang, B. Liu, T. Jiang, and Y. Liu, "Attention-based 3D convolutional network for Alzheimer's disease diagnosis and biomarkers exploration," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1047–1051.
[55] M. Jiang, B. Yan, Y. Li, J. Zhang, T. Li, and W. Ke, "Image classification of Alzheimer's disease based on external-attention mechanism and fully convolutional network," *Brain Sci.*, vol. 12, no. 3, p. 319, Feb. 2022. [Online]. Available: https://www.mdpi.com/2076-3425/12/3/319
[56] S.-H. Wang, Q. Zhou, M. Yang, and Y.-D. Zhang, "ADVIAN: Alzheimer's disease VGG-inspired attention network based on convolutional block attention module and multiple way data augmentation," *Frontiers Aging Neurosci.*, vol. 13, p. 313, Jun. 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnagi.2021.687456
[57] X. Huo, C.-M. Own, Y. Zhou, N. Wu, and J. Sun, "Multistage diagnosis of Alzheimer's disease based on slice attention network," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2022, pp. 255–266.
[58] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 835–838.
[59] X. Chen, A. Diaz-Pinto, N. Ravikumar, and A. Frangi, "Deep learning in medical image registration," *Prog. Biomed. Eng.*, vol. 3, no. 1, 2021, Art. no. 012003.
[60] A. Khademi, B. Reiche, J. DiGregorio, G. Arezza, and A. R. Moody, "Whole volume brain extraction for multi-centre, multi-disease FLAIR MRI datasets," *Magn. Reson. Imag.*, vol. 66, pp. 116–130, Feb. 2020.
[61] J. E. Iglesias, C. Y. Liu, P. M. Thompson, and Z. Tu, "Robust brain extraction across datasets and comparison with publicly available methods," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1617–1634, Sep. 2011.
[62] *Adni_Go_Procedures_Manual_06102011.pdf*. Accessed: Jul. 29, 2022. [Online]. Available: https://adni.loni.usc.edu
[63] F. Alam, S. U. Rahman, M. Hassan, and A. Khalil, "An investigation towards issues and challenges in medical image registration," *J. Postgraduate Med. Inst.*, vol. 31, no. 3, pp. 1–10, Jul. 2017.
[64] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. van der Walt, M. Descoteaux, I. Nimmo-Smith, and D. Contributors, "Dipy, a library for the analysis of diffusion MRI data," *Frontiers Neuroinform.*, vol. 8, pp. 1–8, Feb. 2014.
[65] A. C. Evans, D. L. Collins, S. R. Mills, E. D. Brown, R. L. Kelly, and T. M. Peters, "3D statistical neuroanatomical models from 305 MRI volumes," in *Proc. IEEE Conf. Rec. Nucl. Sci. Symp. Med. Imag. Conf.*, vol. 3, Nov. 1993, pp. 1813–1817.
[66] H. Zunair, A. Rahman, N. Mohammed, and J. P. Cohen, "Uniformizing techniques to process CT scans with 3D CNNs for tuberculosis prediction," in *Proc. Int. Workshop Predictive Intell. Med.* Cham, Switzerland: Springer, 2020, pp. 156–168.
[67] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
[68] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, J. Zhang, J. Santamaría, Y. Duan, and S. R. Oleiwi, "Towards a better understanding of transfer learning for medical imaging: A case study," *Appl. Sci.*, vol. 10, no. 13, p. 4523, Aug. 2020.
[69] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[70] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[71] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lucic, and C. Schmid, "ViViT: A video vision transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 6836–6846.
[72] G. Bertasius, H. Wang, and L. Torresani, "Is space-time attention all you need for video understanding?" in *Proc. ICML*, 2021, vol. 2, no. 3, p. 4.
[73] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[74] *NITRC: Robust Brain Extraction (ROBEX): Tool/Resource Info*. Accessed: Jan. 8, 2022. [Online]. Available: https://www.nitrc.org/projects/robex/
[75] B. J. Soher, B. M. Dale, and E. M. Merkle, "A review of MR physics: 3 T versus 1.5 T," *Magn. Reson. Imag. Clinics North Amer.*, vol. 15, no. 3, pp. 277–290, Aug. 2007.
[76] L. N. Tanenbaum, "Clinical 3 T MR imaging: Mastering the challenges," *Magn. Reson. Imag. Clinics North Amer.*, vol. 14, no. 1, pp. 1–15, Feb. 2006.
[77] M. Raju and V. P. Gopi, "Multi-class classification of Alzheimer's disease using 3DCNN features and multilayer perceptron," in *Proc. 6th Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2021, pp. 368–373.
[78] B. T. Wyman et al., "Standardization of analysis sets for reporting results from ADNI MRI data," *Alzheimer's Dementia*, vol. 9, no. 3, pp. 332–337, 2013.

AYA GAMAL was born in Cairo, Egypt, in 1997. She received the bachelor's degree in systems and biomedical engineering from Cairo University, Giza, Cairo, Egypt, in 2020. She is currently pursuing the master's degree in informatics program with the Information Technology and Computer Science School, Nile University, Giza.

She is currently a Research Assistant with the Medical Imaging and Image Processing Research Group, Center for Informatics Science, Nile University. She is passionate about machine learning and deep learning aspects mainly in medical imaging field. Medical imaging tasks motivate her to work on since graduation, as she participated in many projects. Her research interest includes in designing deep networks to build reliable medical imaging solutions that can assist the physicians in the future.

MUSTAFA ELATTAR received the bachelor's degree in systems and biomedical engineering from Cairo University, Giza, Cairo, Egypt, the master's degree in communication and information technology, in 2010, after conducting image analysis research for cardiac imaging, and the Ph.D. degree in biomedical engineering and physics, medicine from the Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, after developing preoperative planning framework for transcatheter aortic valve implantation. He joined the Medical Imaging and Image Processing Research Group, Nile University, Giza, as a Research Assistant. He joined The Netherlands Cancer Institute (NKI) as a Postdoctoral Fellow, where he conducted research for image-guided radiotherapy. After that, he joined Nile University as an Assistant Professor at the Information Technology and Computer Science School. He is also leading the Medical Imaging and Image Processing Research Group, with focusing on incorporating deep neural networks in 2-D and 3-D medical image analysis contexts. On the professional side, he worked at the Research and Development Division, Diagnosoft Inc., 3mensio B.V., PieMedical N.V., and Myocardial Solutions Inc. In August 2018, he founded and currently leading Intixel Company S.A.E., which focuses on serving medical imaging solution firms with its turnkey artificial intelligence solutions that achieve their business needs.

SAHAR SELIM received the B.Sc. and M.Sc. degrees from the Faculty of Computer Science and Information Systems, Ain Shams University, and the Ph.D. degree in brain–computer interface (BCI) from the Faculty of Computers and Information, Cairo University, Giza, Egypt.

She is currently an Assistant Professor with the Information Technology and Computer Science School, Nile University, Giza. She is also a member with the Medical Imaging and Image Processing Research Group, Center for Informatics Sciences, Nile University. She contributed to several research projects. She is interested in applying machine learning techniques in medical applications. She has a solid technical background with a strong interest in machine learning and artificial intelligence. Her research interests include machine learning and deep learning, brain–computer interface, medical imaging, neurological diseases diagnosis, and software engineering.