

Article

End-to-End Deep Learning Architectures Using 3D Neuroimaging Biomarkers for Early Alzheimer's Diagnosis

Deevyankar Agarwal ^{1,*} , Manuel Alvaro Berbis ² , Teodoro Martín-Noguerol ³, Antonio Luna ³ , Sara Carmen Parrado Garcia ⁴ and Isabel de la Torre-Díez ¹ 

¹ Department of Signal Theory and Communications and Telematics Engineering, University of Valladolid, Paseo de Belén 15, 47011 Valladolid, Spain; isator@tel.uva.es

² Hospital San Juan de Dios, HT Medica, Avda Brillante 106, 14012 Córdoba, Spain; a.berbis@htime.org

³ MRI Unit, Radiology Department, HT Medica, Carmelo Torres No. 2, 23007 Jaén, Spain; t.martin.f@htime.org (T.M.-N.); aluna70@htime.org (A.L.)

⁴ Radiodiagnosis Service, University Clinical Hospital of Valladolid, SACYL, Av. Ramón y Cajal, 3, 47003 Valladolid, Spain; scparrado@saludcastillayleon.es

* Correspondence: deevyankar@gmail.com

Abstract: This study uses magnetic resonance imaging (MRI) data to propose end-to-end learning implementing volumetric convolutional neural network (CNN) models for two binary classification tasks: Alzheimer's disease (AD) vs. cognitively normal (CN) and stable mild cognitive impairment (sMCI) vs. AD. The baseline MP-RAGE T1 MR images of 245 AD patients and 229 with sMCI were obtained from the ADNI dataset, whereas 245 T1 MR images of CN people were obtained from the IXI dataset. All of the images were preprocessed in four steps: N4 bias field correction, denoising, brain extraction, and registration. End-to-end-learning-based deep CNNs were used to discern between different phases of AD. Eight CNN-based architectures were implemented and assessed. The DenseNet264 excelled in both types of classification, with 82.5% accuracy and 87.63% AUC for training and 81.03% accuracy for testing relating to the sMCI vs. AD and 100% accuracy and 100% AUC for training and 99.56% accuracy for testing relating to the AD vs. CN. Deep learning approaches based on CNN and end-to-end learning offer a strong tool for examining minute but complex properties in MR images which could aid in the early detection and prediction of Alzheimer's disease in clinical settings.

Keywords: Alzheimer's disease; CNN; deep learning; end-to-end learning; MCI; MRI; neuroimaging

MSC: 68T073



Citation: Agarwal, D.; Berbis, M.A.; Martín-Noguerol, T.; Luna, A.; Garcia, S.C.P.; de la Torre-Díez, I. End-to-End Deep Learning Architectures Using 3D Neuroimaging Biomarkers for Early Alzheimer's Diagnosis. *Mathematics* **2022**, *10*, 2575. <https://doi.org/10.3390/math10152575>

Academic Editor: Christophe Guyeux

Received: 27 June 2022

Accepted: 21 July 2022

Published: 25 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

AD is the most common form of dementia, and there is currently no proven cure for it. AD and other dementias are estimated to impact at least 50 million individuals worldwide today [1]. Before clinical symptoms appear, AD advances gradually over several years [2]. In 2022, the number of people living with AD in the United States reached 6.5 million. By 2050, some 14 million people are expected to have the disease [3]. It is critical to study novel early diagnosis methods for different kinds of dementia, including AD or mild cognitive impairment (MCI), to ensure the correct treatment and to slow down the progress of the disease. MCI is a state that falls in between normal cognitive function and AD. MCI affects a person's cognitive ability, although they can still go about their regular lives. Moreover, MCI affects nearly one-fifth of those over the age of 65 [3]. In 3 to 5 years, about one third of them will acquire AD [3]. Anatomical and functional brain abnormalities linked to AD can be studied and evaluated by using magnetic resonance (MR) imaging, which is a non-invasive and effective technology. MR imaging is acknowledged as valuable tool for detecting the progression of AD and is routinely used in clinical practice [4,5].

Numerous neuroimaging studies have employed region of interest (ROI) techniques to quantify and detect minor changes linked with AD [6]. Such research relies primarily on past knowledge to drive ROI and feature selection, neglecting brain changes outside the examined region(s) and failing to uncover new knowledge. ML can discover complicated and nuanced patterns of change across MR images and provide a systematic way to construct sophisticated, automated, and objective classification frameworks for processing high-dimensional data. Furthermore, ML algorithms have recently been proven to predict AD better than physicians in some circumstances [7], making it an essential subject of research related to computer-aided diagnosis. While statistical ML approaches, such as the support vector machine (SVM) [8], were first successful in automating the detection of AD, deep learning (DL) methods, such as convolutional neural networks (CNN) and sparse autoencoders [9–19], have lately outpaced statistical methods. In recent years, numerous research activities on neuroimaging-based, computer-aided categorization of AD and its prodromal stage, MCI, have been published [8,20–22]. Due to their inability to extract adaptive features, SVM-based, automated diagnosis models for neuropsychiatric disorders rely on hand-crafted features [8]. The proliferation of DL algorithms for image classification applications was aided by the rising capability of GPUs. DL is a field of machine learning that simulates the human brain's ability to recognize complicated patterns. It uses impulsive learning to learn features, hidden representations, and disease-related patterns from raw neuroimaging data, as well as examine correlations in various regions of MR images [23,24]. The core foundation of DL is an end-to-end learning design concept. The main benefit of end-to-end learning is that it optimizes all phases of the processing pipeline at the same time, possibly resulting in optimal performance [25].

The authors of [9,26] proposed an end-to-end hierarchy for brain MR image analysis, with levels ranging from 1 (none) to 4 (complete). At Level 1, feature extraction and selection are carried out manually. Three-dimensional (3D) volume data are rearranged into 1D vector form for use as input into DL networks such as the restricted Boltzmann machine (RBM) and deep belief network (DBN) [27–29]. At Level 2, 3D data are separated into white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), and hippocampus regions, or turned into 2D slices during preprocessing, and then fed into a DL network such as a CNN. The visual cortex of the brain is what stimulates CNNs. CNNs are the most effective model for image analysis [20,22,30]. They use two-dimensional or three-dimensional pictures as input and extract features by stacking convolutional layers to make greater use of spatial information. The fact that a CNN combines feature extraction and classification is one of its most significant advantages. At Level 3, preprocessed 3D volume data [31] are used as an input into DL networks. The preprocessing of MR images is critical to the efficacy of any quantitative analysis approach. This kind of preprocessing includes procedures such as denoising, bias field correction, brain extraction, registration, normalization, and smoothing that aim to improve image quality and unify geometric and intensity patterns. Level 4 includes directly feeding DL networks with a 3D MR image obtained from a scanner; however, as far as the author is aware, no study has used this level and documented it in the literature.

The majority of known research employed Level 1 [27–29,32–35] or Level 2 [10–18,36–67], the results of which were based on particular software, hyperparameter tuning, and manual noise reduction. Because of these interdependencies [20,22,68], performance evaluations in these pieces of research only employed a portion of the original datasets, avoiding apparent outliers and making a fair performance comparison challenging.

Using ResNet18, Ramzan et al. [36] investigated the efficacy of resting-state functional magnetic resonance imaging (rs-fMRI) for multi-class categorization of AD and its related phases, including CN, SMC, EMCI, MCI, LMCI, and AD. They used a single-channel input to train the network from scratch, as well as an expanded network architecture, to execute transfer learning with and without finetuning. For early diagnosis of AD, Mehmood et al. [37] employed VGG-19 architecture for transfer learning and tissue segmentation on each subject to extract gray matter (GM) tissue from MR images. Abrol et al. [47]

used the GM area to suggest a CNN-based transfer learning scheme, demonstrating that transfer learning is helpful for CNN-based research at Level 2.

To increase the classification accuracy of AD stages, multimodal DL [54–56,58,59,64] approaches have been sought to combine diverse inputs and DL models. In a specific way, multimodal neuroimaging integrates two or more datasets obtained with various imaging devices with the goal of improving our understanding of the structure and function of the brain by leveraging contrasting physical and physiological sensitivities. Multimodal DL approaches are especially challenging to implement at Level 3 learning because of the restrictions of various resolutions, the number of dimensions, inconsistent data, and limited sample sizes [69,70]. Furthermore, we discovered that multimodal DL techniques were only used in research with Level 2 learning. We did not use a multimodal learning approach since our study is centered on Level 3 learning implementation.

Lee et al. [64] used an RNN to predict AD by extracting multimodal characteristics from MRI, cohort data, and CSF data. To integrate and learn feature representation from multimodal neuroimaging data for AD diagnosis, Shi et al. [61] created a multimodal stacked deep polynomial network (MM-SDPN). The MM-SDPN was made up of two-stage SDPNs. Two SDPNs were used to learn high-level MR imaging and PET properties, which were fed into another SDPN to integrate multimodal neuroimaging data for AD stage classification. Lu et al. [58] proposed a novel deep neural network (DNN)-based method that used multi-scale and multimodal knowledge (MMDNN), combining metabolism (FDG-PET) and regional volume (T1-MRI) with a focus on assessing classification accuracy in stable MCI and progressive MCI subjects with known future conversion to probable AD. Song et al. [39] developed an image fusion approach to help AD diagnosis by combining the GM tissue area of brain MR images and FDG-PET images by registration and mask coding to create a new fused modality known as “GM-PET”. The GM region, which is crucial for AD diagnosis, was highlighted in the single composite image, but the contour and metabolic properties of the subject’s brain tissue were preserved. They tested the performance of image fusion methods in binary and multi-classification tasks using 3D CNN and 3D multi-scale CNN.

A limited number of studies used the Level 3 hierarchy. Rieke et al. [71] used MR images to train a 3D CNN for AD vs. CN classification, and they used various visualization techniques to demonstrate that their CNN focused on brain areas linked with AD, specifically the medial temporal lobe. Korolev et al. [72] utilized MR images based on deep 3D CNN to train VoxCNN, which is similar to VGG and ResNet, to categorize different phases of AD. Liu et al. [73] proposed a multimodal DL framework for hippocampus segmentation and AD classification simultaneously based on multi-task CNN and 3D DenseNet by using MR images. Recognizing the benefits of pre-training knowledge, Gao et al. [74] introduced the AD-NET (age-adjusted 3D CNN), with the pre-training model serving two purposes: extracting and transferring features, as well as gaining and transferring knowledge. The knowledge being transmitted in this study was a surrogate biomarker for age that was used to classify MCI converters vs. non-converters on an individual basis. Basaia et al. [75] used two datasets collected using distinct MR protocols and scanners to train, evaluate, and test a 3D CNN in order to cover the complete range of data heterogeneity and provide a less dataset-specific methodology. For AD vs. CN classification, Oh et al. [9] employed convolutional autoencoder (CAE)-based unsupervised learning and supervised transfer learning by transferring gained information from AD versus CN to solve the progressive MCI vs. stable MCI classification task by using 3D MR images. They also used a gradient-based visualization approach to simulate the spatial relevance of the CNN model in detecting the most relevant biomarkers linked to AD and MCI development. The temporal and parietal lobes were identified as crucial classification areas.

Volumetric medical imaging data may be interpreted using 3D CNN models in their original volumetric input form. However, we did not find any comparative analyses of these models for early AD diagnosis nor any implementation of several of them utilizing end-to-end learning at Level 3 during our review of the relevant literature. The purpose

of this study is to use Level 3 end-to-end learning, MR images, and cutting-edge open-source software to determine the best 3D CNN model for classifying the different stages of AD. This study's major contributions and qualities may be described as follows: (1) Use of MR images and Level 3 end-to-end learning, implementing different state-of-the-art, 3D-CNN-based architectures from the DenseNet family [76]: DenseNet121, DenseNet169, DensNet201, and DensNet264 and from the EfficientNets family [77]: EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, and EfficientNet-B3. (2) Comparative analysis of the implementations. (3) Use of Medical Open Network for AI (MONAI) [78] to implement the models and ANTsPyNet [79] to preprocess MR images.

To the best of the authors' knowledge, there has been no previous study that employed MOANI or ANTsPyNet to create a 3D CNN or preprocess neuroimaging data for the early diagnosis of AD. To train the models in the shortest amount of time, the whole implementation was performed in PyTorch GPU. Therefore, the main contribution of this work is a comparative analysis of 3D CNNs for the categorization of different stages of AD utilizing a Level 3 end-to-end learning technique. This article's remainder is organized as follows: The study approach is presented in Section 2, together with datasets, MR image preprocessing, built 3D CNNs, and experimental setup with the algorithm. Section 3 contains the findings: a comparative study of all models that have been built using a rank mechanism and comprehensive performance indicators; a comparison of our best outcomes with published state-of-the-art implemented techniques; and a conclusion section. The Supplementary Data include the URL to the code for researchers.

2. Materials and Methods

The proposed method for encouraging end-to-end learning by applying 3D CNN models for early AD diagnosis is summarized in Figure 1 and explained in further detail in this section. To differentiate AD from CN and AD from sMCI, MR images of patients with AD, CN, and stable MCI were preprocessed using AntsPyNet and then supervised fine-tuning was utilized to generate eight state-of-the-art, 3D-CNN-based classifiers leveraging end-to-end learning and MONAI. Finally, a comparative study was conducted, utilizing the accuracy, AUC, precision, recall, and F1-score metrics of deployed models to identify the best 3D CNN model to help future researchers.

2.1. Datasets and Preprocessing of MR Images

The Alzheimer's Disease Neuroimaging Initiative (ADNI) and Information eXtraction from Images (IXI) datasets, both of which are publicly accessible on the web [80,81], were used in this research. The ADNI intends to find more sensitive and accurate methods for detecting AD early on, as well as biomarkers to track the AD progression. IXI is a collection of over 600 MR images taken from healthy, normal people. The IXI dataset includes participants from three London hospitals: Hammersmith Hospital, Guy's Hospital, and the Institute of Psychiatry. We utilized 719 MP-RAGE T1-weighted, structural MR images downloaded in NIfTI format for this study, which were originally categorized into AD ($n = 245$), CN ($n = 245$), and sMCI ($n = 229$) at baseline. MR images for AD and sMCI were acquired from ADNI, whereas CN MR images were obtained from IXI. Only MCI images that were stable for at least 4 years and up to 15 years, as specified in the ADNI description files, were downloaded. Because the method for recognizing sMCI MR image IDs from ADNI was not found in any article during the literature analysis, we present it here to enable readers to identify patients with sMCI and download them from ADNI. Researchers have to examine the following two CSV files:

1. ADNIMERGE [80]: Can be downloaded from study data -> Test Data -> Data for Challenges. To identify stable MCI, MCI converted to AD, or MCI converted to normal, the factors mentioned in Table 1 may be examined;
2. Diagnostic Summary [ADNI1,GO,2,3] [80]: Can be downloaded from study data -> Assessments -> ALL Diagnosis. The factors mentioned in Table 2 may be examined to identify the various phases of AD.

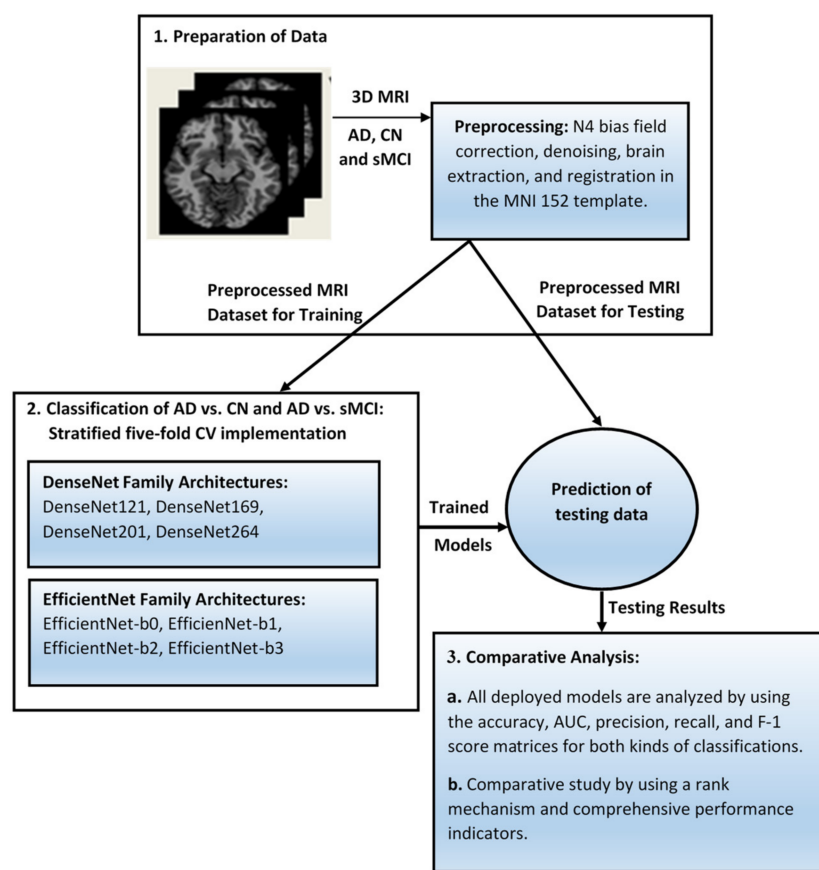


Figure 1. The proposed method's overall architecture.

Table 1. Factors to consider while analyzing ADNIMERGE file.

Name	Description	Value Set
DX.bl	Baseline diagnosis	CN, MCI, AD, EMCI, LMCI, SMC
DX	Current diagnosis status	Same, MCI to AD, AD to MCI, MCI to CN

Table 2. Factors to consider while analyzing a diagnostic summary file. NL: normal control, AD: dementia.

Name	Description	Value Set
DXCHANGE	Which best characterizes the participant's cognitive status change from the previous visit to the current appointment?	1 = Stable: NL; 2 = Stable: MCI; 3 = Stable: Dementia; 4 = Conversion: NL to MCI; 5 = Conversion: MCI to Dementia; 6 = Conversion: NL to Dementia; 7 = Reversion: MCI to NL; 8 = Reversion: Dementia to MCI; 9 = Reversion: Dementia to NL
DIAGNOSIS	Specify diagnostic category.	1 = Cognitively Normal; 5 = Significant Memory Concern; 2 = Early MCI; 3 = Late MCI; 4 = Alzheimer's Disease
DXCURRENT	Current diagnosis.	1 = NL; 2 = MCI; 3 = AD
DXCONV	Has there been a conversion or reversion to NL/MCI?	1 = Yes—Conversion; 2 = Yes—Reversion; 0 = No
DXCONTYP	If yes—conversion, choose type.	1 = Normal Control to MCI; 2 = Normal Control to AD; 3 = MCI to AD

Data are provided for each visit of each patient, and researchers must review the diagnosis and conversion to the next step for each visit. The authors meticulously examined roughly 6000 rows of the csv data described above to obtain information on patients with sMCI. Only 229 of those with sMCI with clear information for each visit remained constant

for at least 4 years and up to 15 years. We did not find the value for DXCHANGE and DX variable for several visits; hence, MR images were ignored. Because our work was centered on end-to-end learning at Level 3, we did not take into account any other factors, such as age, gender, clinical dementia rating (CDR) [82] score, mini-mental state examination (MMSE) [83] score, or the 4 allele of apolipoprotein E (APOE4) [84], which are used in clinical settings [85] and in many studies [17–19,21,22,24,27] to identify the various stages of AD. The downloaded MR images from ADNI often had $256 \times 256 \times 176$ voxels with $1 \text{ mm} \times 1 \text{ mm} \times 1.2 \text{ mm}$ sizes, whereas those from IXI typically had $256 \times 256 \times 256$ voxels with $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$ sizes.

The baseline MRI scans were downloaded in the Neuroimaging Informatics Technology Initiative (NIfTI) [86] format from the ADNI and IXI databases. NIfTI is an upgraded version of the Analyze file format, which was created to be easier to use than DICOM while keeping all of the necessary information. It also has the advantage of storing a volume in a single file, with only a basic header followed by raw data. This allows it to load and process quickly. The ANTsPyNet [79] utilities were used to perform a standard preprocessing procedure on each image. The Advanced Normalization Tools (ANTs) pipeline [87–89] was employed. As illustrated in Figure 2, the implemented preprocessing pipeline included (A) N4 bias correction, (B) denoising, (C) brain extraction, and (D) affine registration to MNI152 template. The N4 bias field correction technique [90] is a widely used approach for correcting low-frequency intensity non-uniformity, often known as a bias or gain field, in MR image data. This strategy relies on a basic parametric model with no tissue classification. N4 bias correction of MR images was performed using the `ants.utils.n4 bias field correction()` [91] function, which was followed by denoising. Denoising's [92] main purpose is to estimate the original image by suppressing noise in a noise-contaminated version of the image. Image noise may be created by a variety of internal and extrinsic factors that are difficult to prevent in real-world settings. As a result, picture denoising is significant in image classification, where recovering the original image content is critical for good results. In our work, denoising was performed in two steps: first, we added different intensities of salt and pepper noise [93,94] to the MR image and then we removed the noise using a spatially adaptive filter initially proposed by Manjon et al. [95] through `ants` utility `ants.denoise_image()` [96], which was followed by brain extraction. Brain extraction was conducted on MR images using AntsPyNet's `brain extraction()` [97] tool, which uses a 3D U-net model called `brainy` [98] and ANTs-based training data to achieve brain extraction. The key advantage of `brainy` is its ability to exploit interslice contextual information [99]. This model obtained a median Dice score of 0.97, a mean of 0.96, a minimum of 0.91, and a maximum of 0.98 on a validation dataset of 99 T1-weighted brain scans and their associated, binarized FreeSurfer segmentations [99]. In three seconds, this model could predict the brain mask for a volume of $256 \times 256 \times 256$, independent of orientation. Predicting the brain mask of each picture took around five seconds with our implementation. It was followed by affine fast registration [100] in the MNI 152 template [101], a universal brain atlas template, utilizing the ANTsPy tool `ants.registration()` [102]. The goal of registration is to eliminate any spatial disparities across subjects in the scanner and to reduce translations and rotations from a standard orientation. This aids the subsequent classification's accuracy. After registration, the dimensions were uniformly rescaled to $182 \times 218 \times 182$ for CNN learning. This registered MR image was used to classify the various stages of AD. In our implementation, the preparation of one MRI scan took around two minutes.

2.2. Deployed 3D CNN

CNNs are gaining prominence as a result of their significant advantages in medical image classification applications [103]. In 2D CNN [35,37–42,44,45,49,63,65] approaches for classifying the different stages of AD, where the 3D MR images are evaluated slice by slice, the anatomical context in directions orthogonal to the 2D plane is completely ignored. While using 3D data as a complete input may improve accuracy [104], the computational complexity and memory cost increase as the number of factors grows. Although many stud-

ies preferred to build their own 2D/3D CNN structures [19,71,74,75,105,106], we did find some that used well-established, pre-trained CNN structures for classifying the different stages of AD, such as deep recurrent neural networks (RNN) [12,34], ResNet [15,36,47,72], CaffeNet and GoogleNet [17], DenseNet [13], and Inception V4 and VGG16 [67]. Most of them, however, adopted Level 2 learning since all of these models only support transfer learning for 2D data.

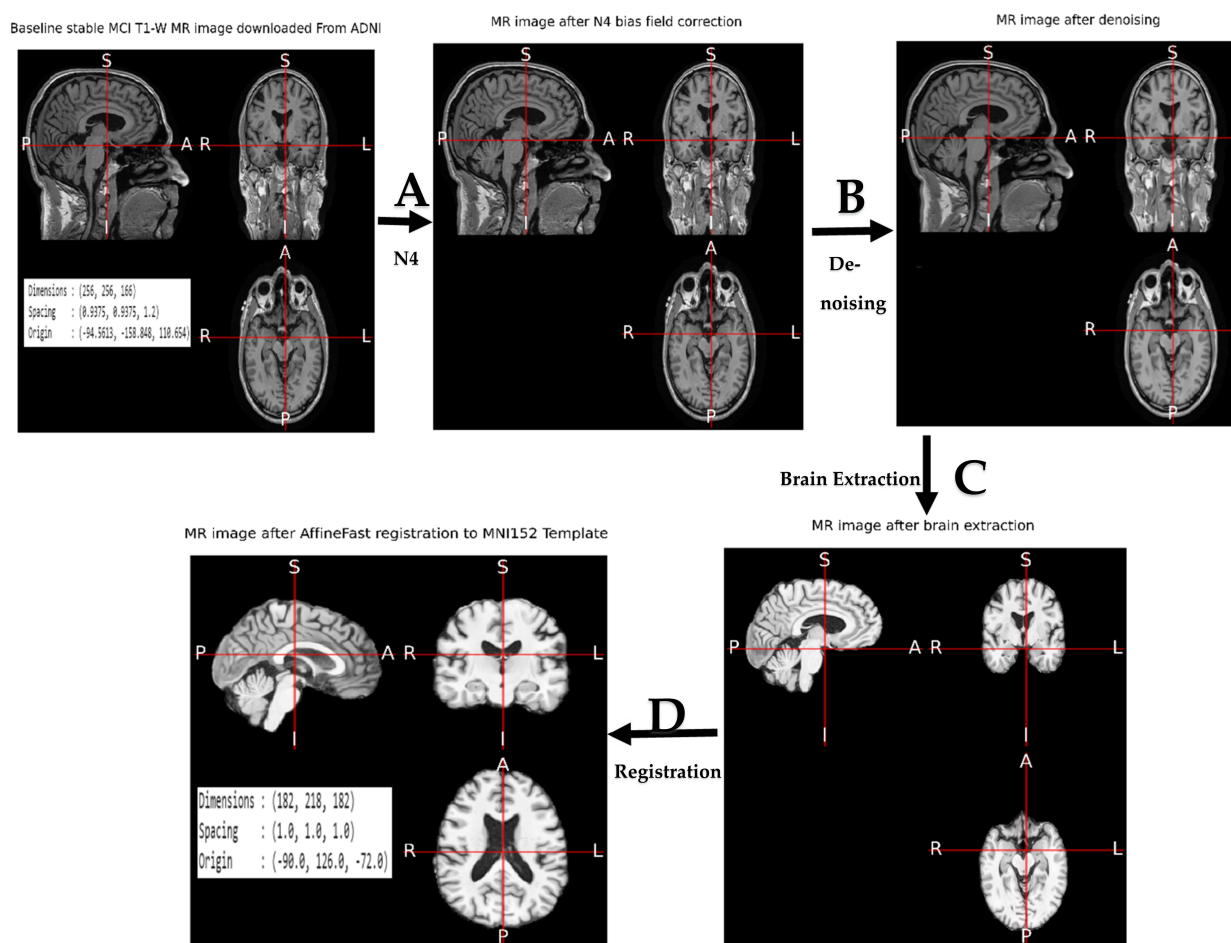


Figure 2. This illustration shows an example of sMCI MR image preprocessing implemented in our study. The original MR picture dimensions were $256 \times 256 \times 166$, but the output MRI dimensions were altered to $182 \times 218 \times 182$. The actions listed were performed in sequence. (A) N4 bias field correction, (B) denoising, (C) 3D U-Net used for brain extraction, (D) affine fast registration in the MNI152 template.

In numerous pieces of research, local TL was used. The idea behind local TL is to utilize the AD vs. CN classifier's finalized weights as the initial weights for sMCI vs. progressive MCI classification. CAE-based unsupervised learning was utilized by Oh et al. [9] to extract sparse representations from 3D MR images of AD and CN subjects, and they used them to classify AD vs. CN using a 3D CNN. The final weights of the CNN used to categorize AD vs. CN were then transferred as the initial weight of the sMCI vs. pMCI classifier. Silvia Basaia et al. [75] employed a 3D CNN without any prior feature engineering and in the face of imaging protocol and scanner erroneousess. However, we were unable to locate any implementation of a number of state-of-the-art, 3D-CNN-based designs that have recently been shown to be extremely effective in other medical data categorization tasks [107–113], specifically, those from the DenseNet [76] and EfficientNets [77] families. Taking all of these things into account, the authors chose to use Level 3 end-to-end learning to implement the

following 3D CNN models for classifying AD vs. CN and sMCI vs. AD patients. All of the models used 3D, processed MRI scans with dimensions of $182 \times 218 \times 182$.

DenseNet: A dense convolutional network (DenseNet) is a feed-forward network that links each layer to the next. DenseNet has $L(L + 1)/2$ direct connections compared to L connections in standard convolutional networks with L layers [76]. All preceding layer feature maps are utilized as inputs for each layer, and its own feature maps are used as inputs for all subsequent layers. DenseNets offer numerous appealing advantages: they solve the vanishing gradient issue, improve feature propagation, enhance feature reuse, and decrease the number of parameters by a significant number. In our investigation, we used 3D DenseNet designs; each design was made up of four DenseBlocks with different numbers of layers. The number of layers in each block, the number of parameters, and the size of the deployed architectures are shown in Table 3.

Table 3. Implemented DenseNet architectures.

CNN	Number of Parameters for 3D Data	Number of Layers in DenseBlocks	Size (in MB)
DenseNet121	11,244,674	[6, 12, 24, 16]	9392.20
DenseNet169	18,546,050	[6, 12, 32, 32]	9891.26
DenseNet201	25,334,658	[6, 12, 48, 32]	10,923.15
DenseNet264	40,251,266	[6, 12, 64, 48]	12,423.07

EfficientNet: EfficientNet [77] is a lightweight model based on the AutoML framework [114] that was used to build a baseline EfficientNet-B0 network and evenly scale up the depth, width, and resolutions using a simpler and effective compound coefficient to enhance EfficientNet models B1–B7. On the ImageNet datasets, these models performed well and outperformed the previous CNN models. EfficientNets are smaller, quicker, and generalize effectively to achieve improved accuracy on other datasets often used for transfer learning. However, they only support transfer learning for 2D data. In the proposed research, end-to-end learning was used to categorize the various stages of AD by using EfficientNet models B0–B3, as shown in Table 4. Due to the direct input of the 3D volume into the model, the increased number of parameters, and the restricted GPU resources and RAM, B4–B7 could not be implemented in the current study.

Table 4. Implemented EfficientNet architectures.

CNN	Number of Parameters for 3D Data	Size (in MB)
EfficientNet-B0	4,690,942	7800.03
EfficientNet-B1	7,449,058	10,222.55
EfficientNet-B2	8,717,764	10,630.40
EfficientNet-B3	12,061,546	14,293.25

2.3. Experimental Setup

The eight 3D CNN architectures were analyzed by applying binary auto-diagnostic problems: (1) AD vs. CN and (2) sMCI vs. AD. The method for evaluating used stratified, five-fold cross-validation, which is detailed in the algorithm provided below.

Step 1: [Preparing Datasets]

- 1.1 Analyze and download baseline T1 MP-RAGE MRI images of AD, CN, and SMCI (stable for at least 4 years, up to 15 years) individuals in NIfTI format: 245 (AD), 245 (CN), and 229 (sMCI).

[Data Sources: ADNI1, ADNI2, ADNI3, ADNIGO, and IXI].

- 1.2 Preprocess downloaded MRIs by using ANTsPyNet.
 - 1.2.1 N4_bias_field_correction.
 - 1.2.2 Denoise_image.
 - 1.2.3 Brain_extraction by using 3D U-Net model brainy.
 - 1.2.4 AffineFast Transformation to register the MRIs in MNI152_T1_1mm_brain template.

Step 2: Set the path for directories based on labels and datasets and then repeat steps 3, 4, and 5 for each dataset.

[Due to Google Colab Pro+'s limited GPU support, we created five different datasets for the classification of AD vs. CN and five separate datasets for the classification of AD vs. sMCI in order to perform a 5-fold, stratified CV by using self-written code.]

Step 3: Specify the path of folders and transformations to create the MONAI dataset.

[In our implementation, we simply used the MONAI load image transformation to read 3D NIfTI images.]

Step 4: Create a PyTorch DataLoader for training, validation, and testing.

[Parameters: MONAI dataset, batch size = 2, number of workers = 2 to load data asynchronously and make multi-processing easier. Set shuffle = True to make batches different each time and increase generalization.]

Step 5: Follow the instructions below eight times for each of the eight models.

[DenseNet121, DenseNet169, DenseNet201, DenseNet264, and EfficientNet-B0 to -B3].

- 5.1 Set device = cuda.
- 5.2 Use Monai.networks.nets to create the model.

[Spatial dims = 3, in channels = 1, out channels = 2, loss function = CrossEntropyLoss (), Adam optimizer with learning rate = 0.0001, AUC metric = ROCAUCMetric]
- 5.3 For fifty training epochs, use the outer loop.
 - 5.3.1 Mini-batches for stochastic gradient descent require an inner loop.
 - 5.3.1.1 Obtain a batch of input from the training data loader.
 - 5.3.1.2 Set the optimizer's gradients to zero.
 - 5.3.1.3 For a given batch of data, the model makes an inference.
 - 5.3.1.4 Compare the set of predictions to the dataset's labels and calculate the loss.
 - 5.3.1.5 The backward gradients over the learning weights are calculated.
 - 5.3.1.6 Use the optimizer to update the model's learning weights for this batch using the observed gradients.
 - 5.3.2 Evaluate the model: inner loop for calculating accuracy and AUC metrics and validating relative loss on a set of data that was not included in the training phase. [Utilize the validation data loader].
 - 5.3.3 Compare the accuracy metric of the current epoch to the accuracy metric of the previous epoch and save the best metric model.
- 5.4 Load the best metric model that was discovered in step 5.3.
- 5.5 Examine the model: [Use test data loader]
 - 5.5.1 Data that were not used in the training or validation processes are classified in this loop.
 - 5.5.2 Calculate the precision, recall, AUC, and F1-score measures.
 - 5.5.3 Make a ROC curve and a confusion matrix.

Step 6: Choose the model that performed best in tests.

ANTsPyNet utilities [79] were used to preprocess all of the MR images. ANTsPyNet is a set of deep learning architectures and applications for basic medical image processing that have been ported to the Python programming language. We compared all of the state-of-the-art models using the same training and testing data, as described in Table 5,

from the ADNI and IXI databases, since fair performance comparisons necessitated the use of the same MRI data.

Table 5. Number of MR images used during implementation work.

Classification	Training	Validation	Testing	Original Dimension	Dimension after Preprocessing
AD vs. CN	AD-160, CN-160	AD-40, CN-40	AD-45, CN-45	$256 \times 256 \times 176$	$182 \times 218 \times 182$
AD vs. sMCI	AD-160, sMCI-160	AD-40, sMCI-40	AD-29, sMCI-29	$256 \times 256 \times 256$	$182 \times 218 \times 182$

All of the work was completed using Google Colab Pro+ [115], which was released in August 2021. It has several important features, including background execution, priority access to faster GPUs, and more RAM. To use it, we must pay USD 49.99 every month. Table 6 shows the resources provided by Colab Pro+ that we employed in our research.

Table 6. Google Colab Pro+ resources used in this research work.

Resources	Option
GPU	CUDA-capable (NVIDIA Deep Learning GPU) (either Tesla V100 or P100). Although Colab Pro+ does not promise support for a particular GPU, it does offer you first choice of what is available. You may achieve a lower quality GPU even with Pro+ if you have a period of high use. The V100 Tensor Core is the most sophisticated GPU created for graphics, high-performance computing (HPC), and AI [116]
RAM	The “High-RAM” runtime option served its purpose by offering 53 GB of RAM and 8 CPU cores
Runtime	24 h. Watch out, since even on Pro+, the runtime disconnects after a while of inactivity
Background execution	Yes
Storage	150 GB

MONAI [78], a freely available, community-supported, PyTorch-based framework for deep learning in healthcare imaging was used to implement all the models in this study. In a native PyTorch paradigm, it provides domain-optimized core features for constructing healthcare imaging training workflows. NVIDIA and King’s College London launched Project MONAI to create an inclusive network of AI researchers for the development and exchange of best practices for AI in healthcare imaging between academic and enterprise researchers. The authors were unable to discover any studies that employed MONAI to determine the different phases of AD. Hyperparameters are an important aspect of neural network training in addition to image preprocessing. Table 7 lists the hyperparameters that were employed in all of the models in this study. Because of the 3D volumetric input and large size of the CNN models, the Adam optimizer [117] was created with a mini-batch size of 2 and an initial learning rate of 1×10^{-4} . The majority of instances in the experiment attained a convergence state within 50 training epochs, which we used as a performance measurement during the cross-validation. Adam was the first “adaptive optimizer” to acquire general acceptance [118]. Instead of using a separate learning rate scheduler, adaptive optimizers include learning rate optimization directly into the optimizer. Adam takes it a step further by controlling the learning rates per weight basis. In other words, it allocates a learning rate to each free variable in the model. The value Adam sets to this learning rate is an optimizer implementation detail that you cannot modify directly. Because of this implementation logic of Adam, the authors did not employ any learning rate schedulers [119] such as ReduceLROnPlateau or EarlyStopping mechanism. In this study, the cross-entropy loss function and ROCAUCMetric were utilized. During the backpropagation process, the output value of a neural network is usually a minimum, considerably below the actual target value. The gradient is frequently relatively low, making it difficult for the neural network to use the data it has to alter the weights and optimize itself. The logarithm of the cross-entropy function enables the network to assess such tiny faults and try to eradicate them. The cross-entropy function allows a CNN to

utilize this change as guidance in the intended direction considerably more effectively than the mean-squared error function does. The ROCAUCMetric indicates how well the model can differentiate across classes. The better the model predicts 0 classes as 0 and 1 classes as 1, the higher the AUC.

Table 7. Network hyperparameters.

Hyperparameter	Option
Number of epochs	50
Batch Size	2
Learning Rate	0.0001
Optimizer	ADAM
Lss Function	CrossEntropyLoss
auc_metric	ROCAUCMetric

3. Results

The performance of eight DL models with similar batch sizes and epochs was examined and compared to gain a complete understanding of how well they performed for the classification of AD vs. CN and AD vs. sMCI. Tables 8 and 9, as well as Figures 3 and 4, illustrate the findings of five measures (precision, recall, F1-score, accuracy, and AUC) for comparing the eight models implemented by using end-to-end learning for both training and testing. For both kinds of classification, the DenseNet-based models outperformed the EfficientNet-based models by a margin of 7 to 14 percent for AD vs. CN classification and 5 to 7 percent for AD vs. sMCI classification during testing. In all forms of classification, DenseNet264 outscored the rest of the DenseNet family. EfficientNet-B0 outperformed the other EfficientNet-based models. During the sMCI vs. AD classification tests, DenseNet201 and EfficientNet-B0 outperformed DenseNet264 by a small margin of 1 to 3 percent for several evaluation matrices. Figures 5 and 6 demonstrate the confusion matrix and ROCAUC of DenseNet264's best fold for both categories of classifications. The confusion matrix and ROCAUC for all other deployed models are given in Supplementary Materials. The first thing we discovered was that stable MCI was noisy during training. This was deduced from the data in Figure 6 and Table 9. This attribute might be because the class is biased and is made up of at least two types of class: those who will acquire AD and those who will stay stable. These might be the factors that make classification challenging. We obtained a maximum accuracy of 82.50% for sMCI vs. AD classification, which has to be improved in future research. We want to underline that, compared to accuracy [62], AUC is seen as a more reliable indicator in the field of medical research. Figures 5 and 6 additionally indicate that the area under each curve tended to be 1.0 in both training and testing for AD vs. CN and 0.90 and 0.79 in training and testing for sMCI vs. AD, demonstrating the classifier's diagnostic capabilities.

Table 8. The findings of the implemented models for AD vs. CN classification.

AD vs. CN (Average of 5-Fold Stratified CV)	Metrics	Dense Net121	Dense Net169	Dense Net201	Dense Net264	B0	B1	B2	B3
Training	Accuracy	99.25	99.50	99.50	100	99.75	98.75	99.00	99.50
	Precision	99.28	99.52	99.52	100	99.75	98.78	99.01	99.51
	Recall	99.25	99.50	99.50	100	99.75	98.75	99.00	99.50
	AUC	100	100	100	100	99.94	98.34	99.52	99.74
	F1-Score	99.25	99.50	99.50	100	99.75	98.75	99.00	99.50
Testing	Accuracy	97.33	92.89	98.22	99.55	90.91	86.91	91.55	85.56
	Precision	97.41	94.35	98.30	99.56	92.90	89.98	93.30	88.84
	Recall	97.33	92.91	98.22	99.55	90.91	86.93	91.55	85.58
	AUC	97.33	92.89	98.22	99.55	90.91	86.91	91.55	85.56
	F1-Score	97.33	92.674	98.22	99.55	90.58	86.40	91.38	85.11

Table 9. The findings of the implemented models for AD vs. sMCI classification.

AD vs. sMCI (Average of 5-Fold Stratified CV)	Metrics	Dense Net121	Dense Net169	Dense Net201	Dense Net264	B0	B1	B2	B3
Training	Accuracy	78.50	81.25	78.75	82.50	80.50	78.25	77.75	72.00
	Precision	78.99	82.83	79.45	84.10	81.44	79.39	79.33	73.99
	Recall	78.48	81.25	78.75	82.50	80.50	78.25	77.75	72.00
	AUC	85.42	86.23	83.20	87.63	81.38	82.59	83.14	73.49
	F1-Score	78.42	80.94	78.63	82.15	80.33	78.06	77.38	70.71
Testing	Accuracy	81.72	79.65	82.06	81.03	81.38	80.69	73.79	74.83
	Precision	82.72	82.83	83.70	83.29	82.91	84.49	79.65	79.22
	Recall	81.72	79.65	82.06	81.03	81.38	80.69	73.80	74.83
	AUC	81.73	79.65	82.06	81.03	81.38	80.69	73.79	74.83
	F1-Score	81.59	79.07	81.84	80.60	80.96	80.00	71.97	72.60

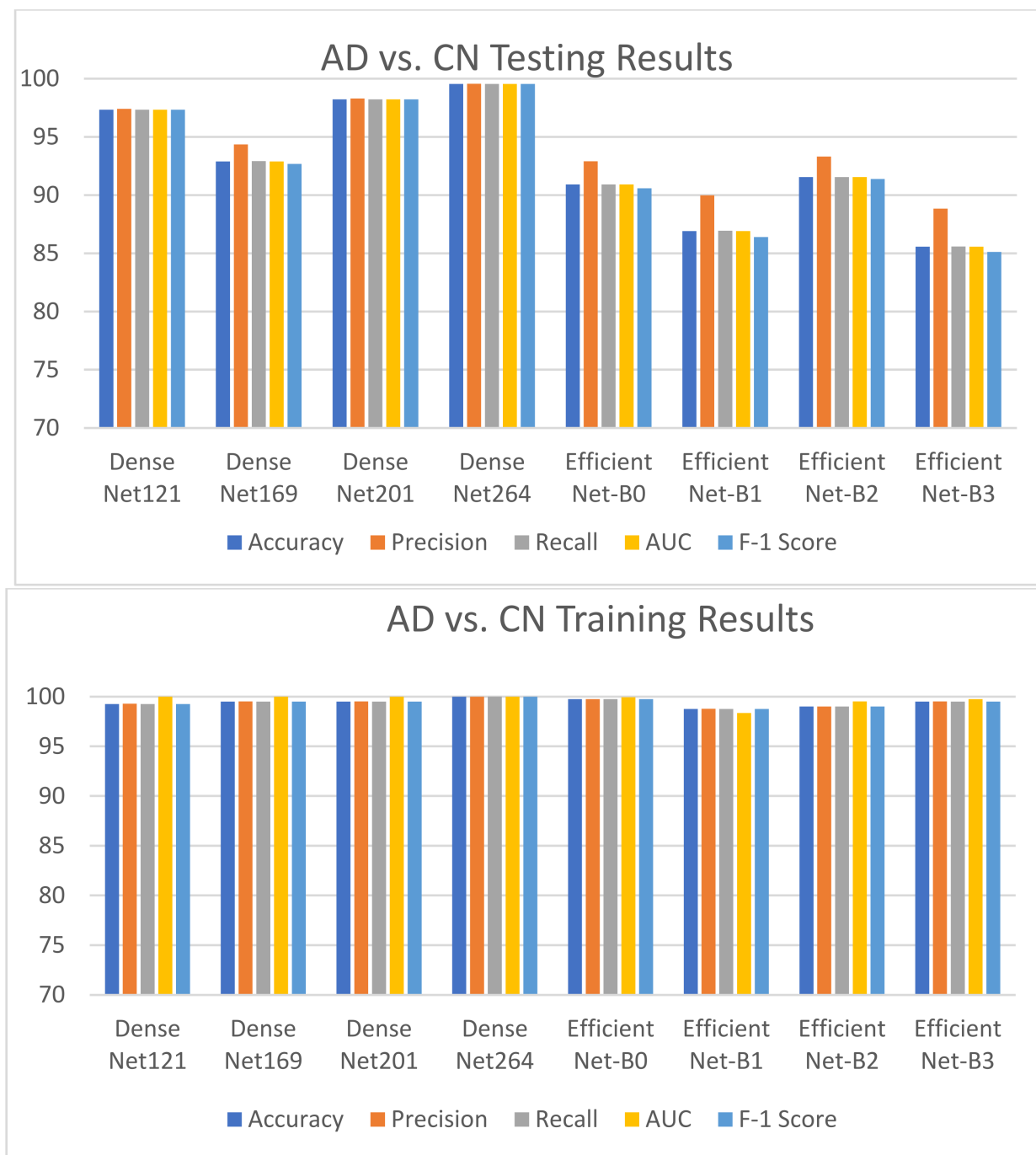


Figure 3. For AD vs. CN classification, the comparison of the results of five evaluation matrices of eight DL models with epoch = 50 and batch size = 2.

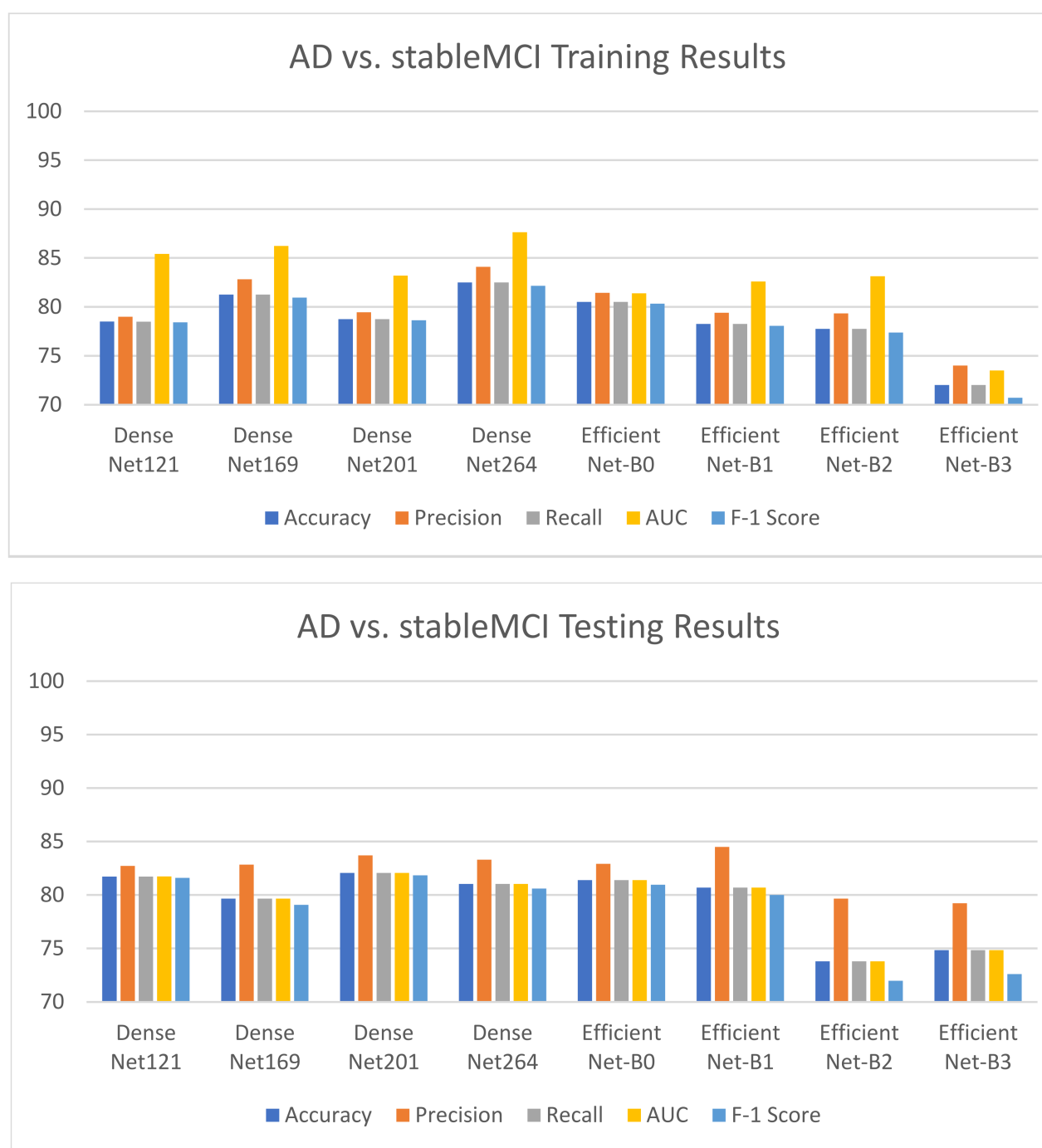


Figure 4. For AD vs. sMCI classification, the comparison of the results of five evaluation matrices of eight DL models with epoch = 50 and batch size = 2.

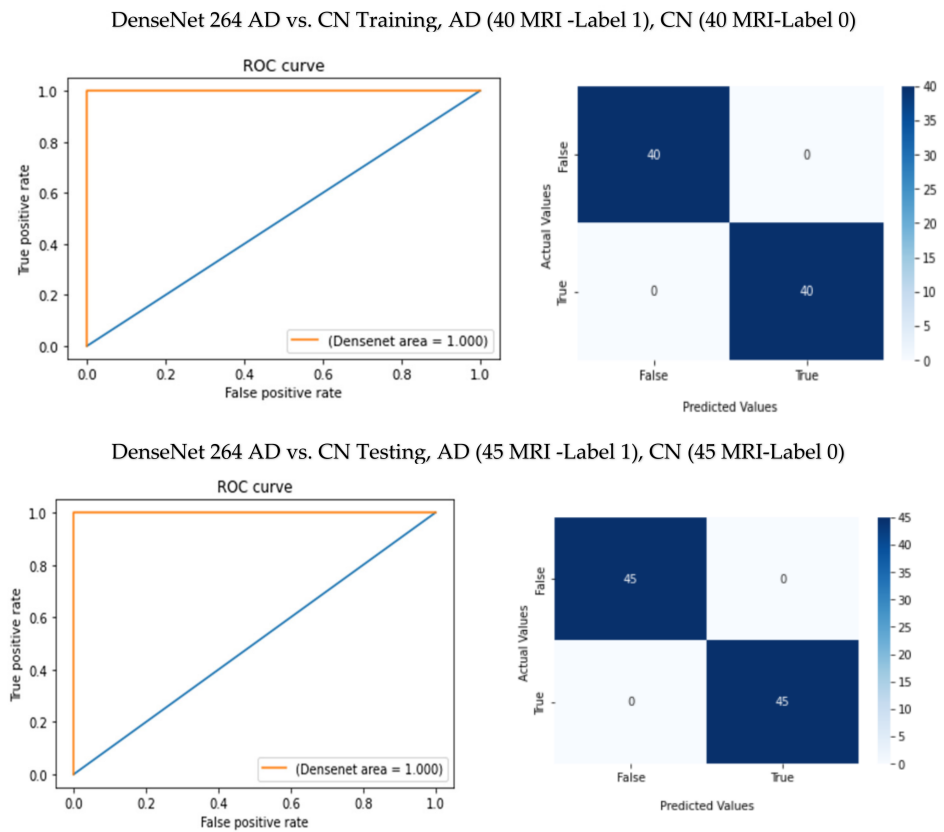


Figure 5. Confusion matrix and ROCAUC for DenseNet264's best fold of AD vs. CN classification.

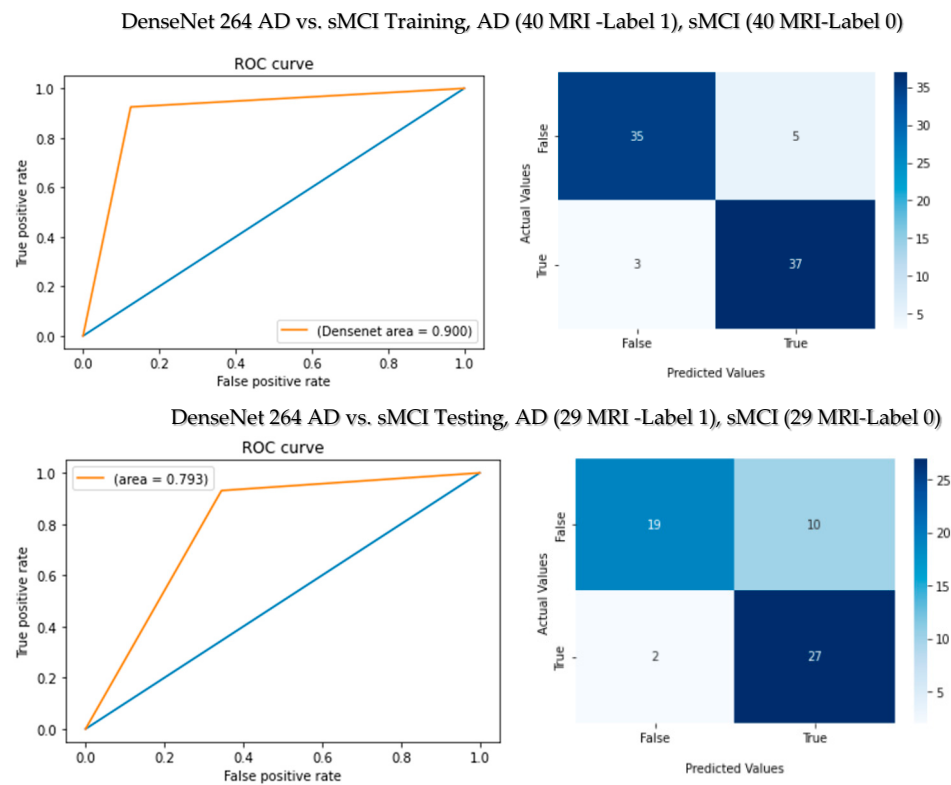


Figure 6. Confusion matrix and ROCAUC for DenseNet264's best fold of AD vs. sMCI classification.

4. Discussion

The following two processes were used to provide a clear and comprehensive description of the outcomes for a relative comparison of implemented models.

4.1. Ranking Mechanism

In this study, the basic ranking approach described by Zorlu et al. [120] was employed. The overall rank of each model was determined individually for the training and testing datasets. Because we had eight models and the best performance index was given the highest rating, the maximum rate value for each performance index was eight (8). After that, each model's overall performance rating was derived by adding its total rank of training and total rank of testing datasets. DenseNet264 obtained the highest-ranking value among all eight DL models for both types of binary classification, as shown in Tables 10 and 11, and was selected as the best model in this research. It may be inferred that DenseNet264 can provide high performance capacity in the early detection of AD. In the AD vs. CN classification, DenseNet264 obtained the best possible score. In AD vs. sMCI Classification training, DenseNet264 outperformed DenseNet201, but DenseNet201 outperformed DenseNet264 in testing. Even in testing, DenseNet121 and EfficientNet-B0 outscored DenseNet264. The overall ranking of DenseNet264 was higher. It was observed that DenseNet201, DenseNet121, and EfficientNet-B0 might be utilized to experiment with more training data in order to build a generalizable DL model for the classification of different phases of AD. For both classes, EfficientNet-B2 and -B3 had the lowest ranking.

Table 10. The acquired rankings of all DL model performance indices for AD vs. CN classification.

DL Model	Stage	Accuracy Rank	Precision Rank	Recall Rank	AUC Rank	F1-Score Rank	Total Rank	Grand Total Rank
DenseNet121	Training	5	4	5	8	5	27	57
	Testing	6	6	6	6	6	30	
DenseNet169	Training	6	6	6	8	6	32	57
	Testing	5	5	5	5	5	25	
DenseNet201	Training	6	6	6	8	6	32	67
	Testing	7	7	7	7	7	35	
DenseNet264	Training	8	8	8	8	8	40	80
	Testing	8	8	8	8	8	40	
EfficientNet-B0	Training	7	7	7	7	7	35	50
	Testing	3	3	3	3	3	15	
EfficientNet-B1	Training	3	2	3	4	3	15	25
	Testing	2	2	2	2	2	10	
EfficientNet-B2	Training	4	3	4	5	4	20	40
	Testing	4	4	4	4	4	20	
EfficientNet-B3	Training	6	5	6	6	6	29	34
	Testing	1	1	1	1	1	5	

4.2. Comprehensive Indicators and Efficiency-Effects Graph

This study looked at how to combine these five measures (precision, recall, F1-score, accuracy, and AUC) to estimate the performance of the eight models in a complete and accurate way. Several of these five indicators, however, are interconnected. The F1-score is a combined indication of accuracy and recall. It was also observed that certain models performed well in terms of recall but badly in terms of accuracy and precision or vice versa, indicating that the models did not function well. As a result, we utilized Yang et al.'s [121]

approach to assess the models' strengths in a more thorough manner. The dispersion and standard deviations (std) of the four indicators (precision, recall, accuracy, and AUC) were computed. First, the four indicators were added up (sum) for each model. Then, we calculated their standard deviation (std). Finally, we added a constant ($k = 0.04$) to the std to avoid making a division by zero mistake when figuring out the comprehensive indicators. To calculate the comprehensive evaluation indicator, we divided the final two numbers ($\text{sum}/(\text{std} + 0.04)$). This process is shown in Tables 12 and 13.

Table 11. The acquired rankings of all DL model performance indices for AD vs. sMCI classification.

DL Model	Stage	Accuracy Rank	Precision Rank	Recall Rank	AUC Rank	F1-Score Rank	Total Rank	Grand Total Rank
DenseNet121	Training	4	2	4	6	4	20	51
	Testing	7	3	7	7	7	31	
DenseNet169	Training	7	7	7	7	7	35	50
	Testing	3	4	2	3	3	15	
DenseNet201	Training	5	5	5	5	5	25	64
	Testing	8	7	8	8	8	39	
DenseNet264	Training	8	8	8	8	8	40	66
	Testing	5	6	5	5	5	26	
EfficientNet-B0	Training	6	6	6	2	6	26	55
	Testing	6	5	6	6	6	29	
EfficientNet-B1	Training	3	4	3	3	3	16	40
	Testing	4	8	4	4	4	24	
EfficientNet-B2	Training	2	3	2	4	2	13	19
	Testing	1	2	1	1	1	6	
EfficientNet-B3	Training	2	1	1	1	1	6	15
	Testing	2	1	2	2	2	9	

Table 12. The acquired comprehensive indicators of all DL model performance indices for AD vs. CN.

Model	Stage	Accuracy	Precision	Recall	AUC	Sum	Std	Std + 0.04	Indicator
Dense Net121	Training	0.99	0.99	0.99	1.00	3.98	0.00	0.04	91.02
	Testing	0.97	0.97	0.97	0.97	3.89	0.00	0.04	96.39
Dense Net169	Training	1.00	1.00	1.00	1.00	3.99	0.00	0.04	93.84
	Testing	0.93	0.94	0.93	0.93	3.73	0.01	0.05	78.92
Dense Net201	Training	1.00	1.00	1.00	1.00	3.99	0.00	0.04	93.84
	Testing	0.98	0.98	0.98	0.98	3.93	0.00	0.04	97.27
Dense Net264	Training	1.00	1.00	1.00	1.00	4.00	0.00	0.04	100.00
	Testing	1.00	1.00	1.00	1.00	3.98	0.00	0.04	99.43
Efficient Net-B0	Training	1.00	1.00	1.00	1.00	3.99	0.00	0.04	97.48
	Testing	0.91	0.93	0.91	0.91	3.66	0.01	0.05	73.20

Table 12. Cont.

Model	Stage	Accuracy	Precision	Recall	AUC	Sum	Std	Std + 0.04	Indicator
Efficient Net-B1	Training	0.99	0.99	0.99	0.98	3.95	0.00	0.04	93.72
	Testing	0.87	0.90	0.87	0.87	3.51	0.02	0.06	63.40
Efficient Net-B2	Training	0.99	0.99	0.99	1.00	3.97	0.00	0.04	93.12
	Testing	0.92	0.93	0.92	0.92	3.68	0.01	0.05	75.48
Efficient Net-B3	Training	1.00	1.00	1.00	1.00	3.98	0.00	0.04	96.70
	Testing	0.86	0.89	0.86	0.86	3.46	0.02	0.06	61.30

Table 13. The acquired comprehensive indicators of all DL model performance indices for AD vs. sMCI.

Model	Stage	Accuracy	Precision	Recall	AUC	Sum	Std	Std + 0.04	Indicator
Dense Net121	Training	0.79	0.79	0.78	0.85	3.21	0.03	0.07	43.49
	Testing	0.82	0.83	0.82	0.82	3.28	0.00	0.04	72.89
Dense Net169	Training	0.81	0.83	0.81	0.86	3.32	0.02	0.06	52.23
	Testing	0.80	0.83	0.80	0.80	3.22	0.02	0.06	57.56
Dense Net201	Training	0.79	0.79	0.79	0.83	3.20	0.02	0.06	52.19
	Testing	0.82	0.84	0.82	0.82	3.30	0.01	0.05	68.44
Dense Net264	Training	0.83	0.84	0.83	0.88	3.37	0.02	0.06	52.46
	Testing	0.81	0.83	0.81	0.81	3.26	0.01	0.05	63.62
Efficient Net-B0	Training	0.81	0.81	0.81	0.81	3.24	0.01	0.05	71.55
	Testing	0.81	0.83	0.81	0.81	3.27	0.01	0.05	68.64
Efficient Net-B1	Training	0.78	0.79	0.78	0.83	3.18	0.02	0.06	52.63
	Testing	0.81	0.84	0.81	0.81	3.27	0.02	0.06	55.35
Efficient Net-B2	Training	0.78	0.79	0.78	0.83	3.18	0.03	0.07	48.60
	Testing	0.74	0.80	0.74	0.74	3.01	0.03	0.07	43.45
Efficient Net-B3	Training	0.72	0.74	0.72	0.73	2.91	0.01	0.05	58.00
	Testing	0.75	0.79	0.75	0.75	3.04	0.02	0.06	49.03

The number of model parameters was also employed as an indicator to assess the models' merits for the image classification task in addition to the comprehensive indicators, which were the most relevant index. As a result, the efficiency-effects plot is presented in Figures 7 and 8, where the horizontal coordinate is the number of model parameters, and the vertical coordinate is the model's comprehensive indicator. The model is better and more efficient if the point representing it is as near as possible to the top-left corner of the efficiency-effects graph. Models in the bottom-right corner have the opposite effect. The EfficientNet-B0 model had the greatest overall metrics and reduced model parameters for training, whereas the DenseNet121 model had the highest testing performance for both types of classifications. DenseNet264 outperformed the others in terms of evaluation matrices for AD vs. CN classification, but it also contained the most parameters, requiring more resources to train. In terms of overall metrics, the DenseNet169 and EfficientNet-B1, -B2, and -B3 models performed moderately. In testing, DenseNet201 did particularly well, almost approaching DenseNet121 and better than DenseNet264 for both types of classification. It may be inferred that having a higher number of model parameters does not always mean improved overall model performance.

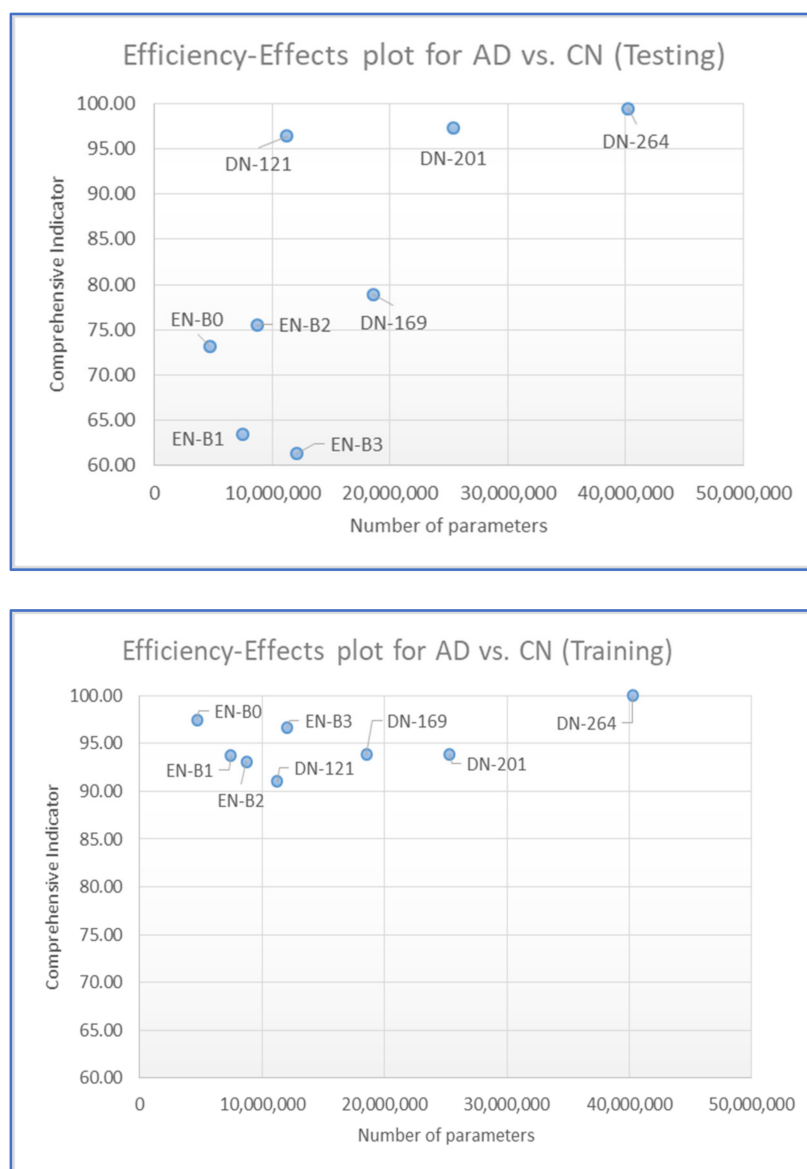


Figure 7. A comparison of the comprehensive performance indicators of eight deployed models for AD vs. CN classification.

4.3. A Comparison with Cutting-Edge Techniques Described in Publications

As indicated in Tables 14 and 15, we compared our classification results to those given in the literature. The approaches that were compared ranged from learning Level 1 to learning Level 3. We also compared our results for stable MCI vs. AD with those for non-converter (stable) MCI vs. (progressive) converter MCI, since converted MCI indicates AD and non-converter MCI indicates stable MCI.

End-to-end learning allowed us to distinguish AD from CN with the maximum degree of accuracy, and, as a result, this classifier may be used in clinical situations after some qualitative analysis. The issue of early auto-diagnosis of MCI patients who are at risk of developing AD from a clinical standpoint is still more important than the AD vs. CN conundrum for successful AD therapy. The categorization of AD and sMCI, on the other hand, is more difficult than that of AD and CN, since the morphological alterations that must be recognized are more subtle. Many of the research findings in Table 15 showed low accuracy of 70 to 80%. Our model, likewise, did well in this categorization, performing best among Level 3 learning classifiers. Only one of Pan et al.'s [54] Level 2 studies outperformed us in sMCI vs. AD classification: by 1.3 percent. They proposed MiSePyNet,

a CNN model for the 18F-FDG PET modality. MiSePyNet was based on the concept of factorized convolution and used separable CNNs, slice- and spatial-wise CNNs, for each view. However, sMCI vs. AD classification accuracy needs to be improved further to aid in clinical settings. This can be achieved by using more training data, using a pre-trained 3D model, and using local transfer learning. We could not find any research that employed the DenseNet264 or EfficientNet family models to classify the various phases of AD.

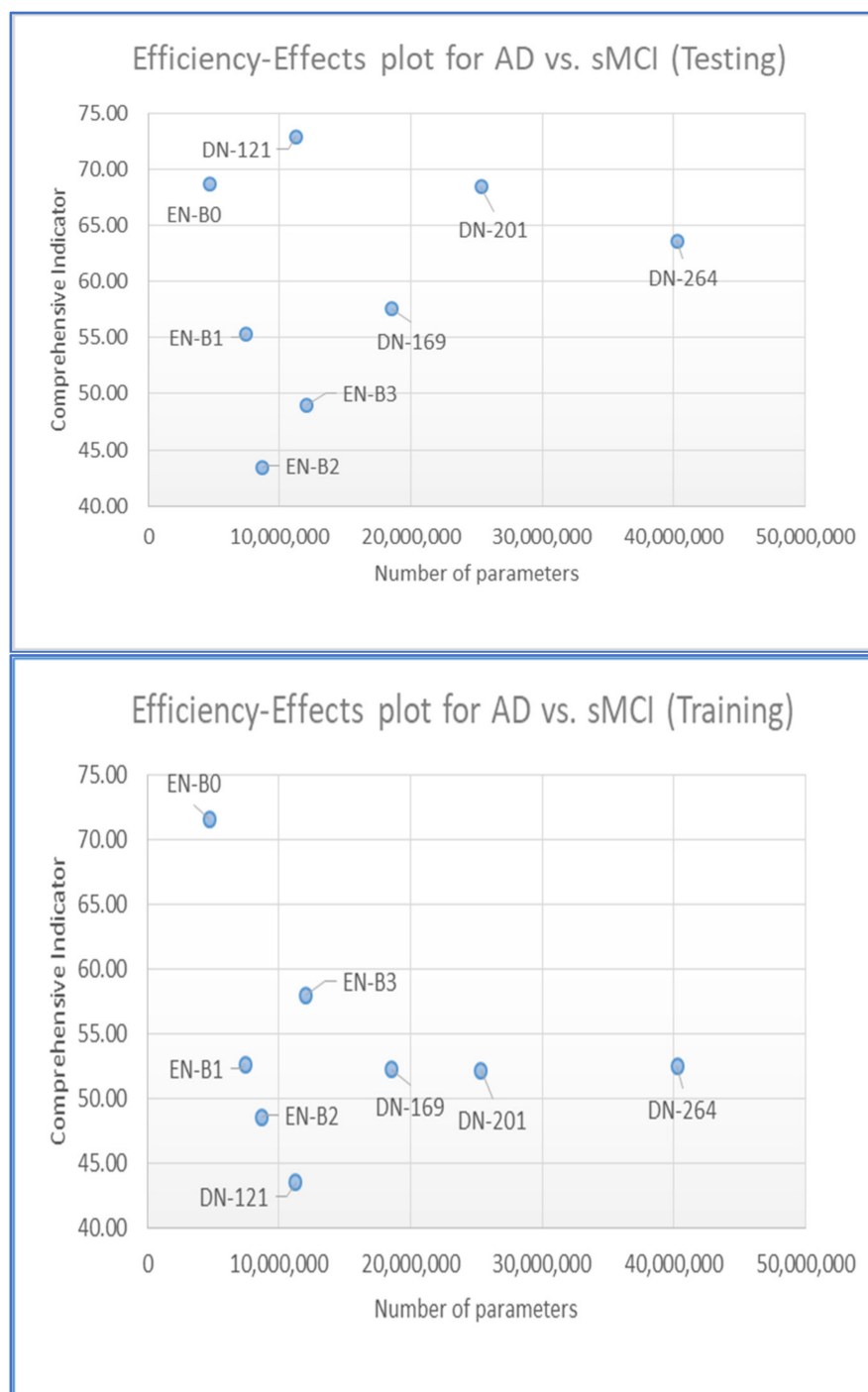


Figure 8. A comparison of the comprehensive performance indicators of eight deployed models for AD vs. sMCI classification.

Table 14. Classification (AD vs. CN) performance of the published state-of-the-art methods.

References	Learning Level/Classifier	Subjects	AD vs. CN					
			Accuracy	Precision	SEN/Recall	AUC	F1-Score	SPE
Toshkhujaev et al. [32]	L1/RBF-SVM	AD-71, CN-171	91.57	-	81.82	-	-	100
Suk et al. [27]	L1/Regression + CNN	AD-186, CN-286	91.02 ± 4.29	-	92.72	92.72	-	89.94
Zhang et al. [10]	L2/CNN	AD-280, CN-275	97.35	-	97.10	99.70	-	97.95
Li et al. [12]	L2/CNN + RNN	AD-194, CN-216	89.10	-	84.6	91.0	-	93.1
Mehmood et al. [50]	L2/VGG-19 (2D TL)	AD-85, CN-75	98.73	-	98.19	-	-	99.09
Tuan et al. [61]	L2/CNN + SVM	CN-98, AD-99	89.00	-	-	-	-	-
Song et al. [60]	L2/3D CNN	CN-126, AD-95	94.11	-	-	-	-	-
Nanni et al. [51]	L2/AlexNet ^P	AD-137, CN-162		-	-	90.8	-	-
	L2/GoogleNet ^P			-	-	89.6	-	-
	L2/ResNet50 ^P			-	-	89.8	-	-
	L2/ResNet101 ^P			-	-	89.9	-	-
	L2/InceptionV3			-	-	88.8	-	-
	L2/3DCNN			-	-	84.1	-	-
A et al. [18]	L2/2D CNN	CN-635, AD-220	96.8	-	94.0	-	-	96.0
Li et al. [46]	L2/CNN	CN-216, AD-194	85.9	-	81.5	88.4	-	89.9
Cui and Liu. [39]	L2/3DCNN	CN-223, AD-192	92.29	-	90.63	96.95	-	93.72
Liu et al. [47]	L2/2DCNN	CN-100, AD-93	93.26	-	92.55	95.68	-	93.94
Xu et al. [64]	L2/SRC	CN-117, AD-113	94.8	-	95.6		-	94.0
Pan et al. [54]	L2/CNN	AD-237, CN-242	93.75	-	91.49	96.87	-	95.92
Shi et al. [59]	L2/MM-SDPN	AD-51, CN-52	97.13 ± 4.44	-	95.93 ± 7.84	-	-	95.93 ± 7.84
Lu et al. [49]	L2/MDNN and TL	CN-304, AD-226	93.58	-	91.54	-	-	95.06
Hon and Khan [43]	L2/InceptionV4	AD-200, CN-100	96.25	-	-	-	-	-
Liu et al. [73]	L3/3D CNN	AD-97, CN-119	88.9	-	86.6	92.5	-	90.8
Oh et al. [9]	L3/CAE + 3DCNN	CN-230, AD-198	86.60 ± 3.66	-	88.55	-	-	84.54
Proposed	L3/DenseNet264	CN-245, AD-245	99.55	99.56	99.55	99.55	99.55	99.55

Abbreviations—P: pertained, MM-SDPN: multimodal stacked deep polynomial networks, MDNN: multistate deep neural network, CAE: convolutional autoencoder, TL: transfer learning, SRC: sparse representation-based classification.

Table 15. Classification (AD vs. sMCI and sMCI vs. pMCI) performance of the published state-of-the-art methods.

References	Learning Level/Classifier	Subjects	AD vs. Stable MCI OR Non-Converter (Stable) MCI vs. (Progressive) Converter MCI					
			Accuracy	Precision	SEN	AUC	F1-Score	SPE
Suk et al. [27]	L1/Regression + CNN	pMCI-167, sMCI-226	74.82 ± 6.80	-	70.93	75.39	-	78.82
Zhang et al. [10]	L2/CNN	pMCI-162, sMCI-251	78.79	-	75.16	86.79	-	82.42
Li et al. [12]	L2/CNN + RNN	pMCI-164, sMCI-233	72.5	-	61.0	74.6	-	82.5
Nanni et al. [45]	L2/AlexNet ^P	sMCI-234, pMCI-240	-	-	-	69.1 ± 1.3	-	-
	L2/GoogleNet ^P		-	-	-	70.0 ± 1.3	-	-
	L2/ResNet50 ^P		-	-	-	70.4 ± 1.0	-	-
	L2/ResNet101 ^P		-	-	-	71.2 ± 1.2	-	-
	L2/InceptionV3 ^P		-	-	-	69.8 ± 3.5	-	-
	L2/3DCNN		-	-	-	61.1 ± 1.0	-	-
Li et al. [52]	L2/CNN	pMCI-164, sMCI-233	71.0	-	59.8	71.9	-	79.0
Cui and Liu. [53]	L2/3DCNN	sMCI-231, pMCI-	75.00	-	73.33	77.70	-	76.19
Xu et al. [56]	L2/SRC	MCI-110	77.8	-	74.10	-	-	81.50
Pan et al. [57]	L2/MiSePyNet	sMCI-360, pMCI-166	83.81	-	75.76	88.89	-	87.50
Shi et al. [61]	L2/MM-SDPN	pMCI-43, sMCI-56	78.88 ± 4.38	-	68.04 ± 9.99	-	-	86.81 ± 9.12
Lu e al. [62]	L2/MDNN and TL	sMCI-409, pMCI-112	81.55	-	73.33	-	-	83.83
Shen et al. [63]	L2/RNN	pMCI-307, sMCI-558	80.00	-	81.00	-	-	80.00
Yang and Liu [66]	L2/SVM	sMCI-270, pMCI-70	78.56	-	91.02	-	-	77.63
Gao et al. [74]	L3/3DCNN	pMCI-168, sMCI-129	76.0	-	77.0	81.0	-	76.0
Oh et al. [9]	L3/CAE + 3DCNN	sMCI-101, pMCI-166	73.95 ± 4.82	-	77.46	-	-	70.71
Proposed	L3/DenseNet264	sMCI-229, AD-229	82.50	84.10	82.50	87.63	82.15	82.50

Abbreviations—P: pertained, MM-SDPN: multimodal stacked deep polynomial networks, MDNN: multistate deep neural network, CAE: convolutional autoencoder, TL: transfer learning, SRC: sparse representation-based classification, MiSePyNet: multi-view separable pyramid network.

5. Conclusions

We show a variety of findings in this section. Even with the endemic challenges of neuroimaging, where training data are few and sample dimensionality is large, end-to-end learning without the use of hand-crafted features is achievable. We also performed an in-depth comparative analysis of eight state-of-the-art models, DenseNet121, DenseNet169, DenseNet201, DenseNet264, and EfficientNet family models from B0 to B3, that were implemented by using 3D MRI input and cutting-edge software such as MONAI and ANTsPyNet on a Pytorch-based GPU setup.

The experimental findings on the ADNI and IXI data showed that our model outperformed current state-of-the-art models in terms of performance and efficiency. The findings of this study may be used to advise researchers in determining the best model to use and understanding the situations in which the models would give better outcomes. A neural network model with more layers or more parameters does not always deliver superior overall performance for a very small data regime. In general, neural networks from the DenseNet family, such as DenseNet121 or DenseNet201 and DenseNet264, and EfficientNet-B0 provide superior results for categorizing the various phases of AD. This research, however, had some drawbacks.

- First, the number of subjects employed for the training and test phases was still small for promoting end-to-end learning. When more data become available in the future, we think this method will help learning models to generalize better than approaches that are made by hand;
- Second, our AD vs. sMCI classification accuracy was still only 82.50 percent, which has to be improved in order to provide better therapy for AD patients. A pre-trained 3D CNN model, as well as an exploratory study into local transfer learning, is required to achieve this goal in the future.

Despite these flaws, to the best of our knowledge, this is the first piece of research to use end-to-end learning with volumetric CNN architecture to compare eight CNN-based 3D models to categorize the various stages of AD without hand-crafted features. In future studies, to find the best network model, it may be necessary to perform a lot of experiments that include network structures, hyperparameters, and other neuroimaging data.

Supplementary Materials: The supporting information can be downloaded at: <https://drive.google.com/drive/folders/1EpeDISfKc7p-DGba1XWdjbp3qpq6PW1?usp=sharing> (accessed on 26 June 2022). Researchers may obtain the preprocessing script for MR images with findings as well as the scripts for all models. It contains two folders: one for the categorization of AD vs. CN and the other for AD vs. sMCI.

Author Contributions: D.A., M.A.B. and T.M.-N. participated in the conception and methodology and implementation of the models and manuscript writing. I.d.I.T.-D., A.L. and S.C.P.G. participated in the review and manuscript writing and data collection and preprocessing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by European Commission and the Ministry of Industry, Energy and Tourism under the project AAL-20125036 named BWetake Care: ICTbased Solution for (Self-) Management of Daily Living.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are openly available in ADNI at <https://adni.loni.usc.edu/> (accessed on 26 June 2022) and in IXI at <https://brain-development.org/ixi-dataset/> (accessed on 26 June 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Patterson, C. *World Alzheimer Report 2018*; Report; Alzheimer's Disease International: London, UK, 2018.
2. Hardy, J. Amyloid, the Presenilins and Alzheimer's Disease. *Trends Neurosci.* **1997**, *20*, 154–159. [CrossRef]
3. Alzheimer's Disease Facts and Figures. Alzheimer's Disease and Dementia. Available online: <https://www.alz.org/alzheimers-dementia/facts-figures> (accessed on 29 April 2022).
4. Klöppel, S.; Stonnington, C.M.; Chu, C.; Draganski, B.; Scahill, R.I.; Rohrer, J.D.; Fox, N.C.; Jack, C.R., Jr.; Ashburner, J.; Frackowiak, R.S.J. Automatic Classification of MR Scans in Alzheimer's Disease. *Brain* **2008**, *131*, 681–689. [CrossRef] [PubMed]
5. Hinrichs, C.; Singh, V.; Mukherjee, L.; Xu, G.; Chung, M.K.; Johnson, S.C. Spatially Augmented LPboosting for AD Classification with Evaluations on the ADNI Dataset. *NeuroImage* **2009**, *48*, 138–149. [CrossRef] [PubMed]
6. Baron, J.C.; Chételat, G.; Desgranges, B.; Perche, G.; Landeau, B.; de la Sayette, V.; Eustache, F. In Vivo Mapping of Gray Matter Loss with Voxel-Based Morphometry in Mild Alzheimer's Disease. *NeuroImage* **2001**, *14*, 298–309. [CrossRef]
7. Klöppel, S.; Stonnington, C.M.; Barnes, J.; Chen, F.; Chu, C.; Good, C.D.; Mader, I.; Mitchell, L.A.; Patel, A.C.; Roberts, C.C.; et al. Accuracy of Dementia Diagnosis—A Direct Comparison between Radiologists and a Computerized Method. *Brain* **2008**, *131*, 2969–2974. [CrossRef]
8. Rathore, S.; Habes, M.; Iftikhar, M.A.; Shacklett, A.; Davatzikos, C. A Review on Neuroimaging-Based Classification Studies and Associated Feature Extraction Methods for Alzheimer's Disease and Its Prodromal Stages. *Neuroimage* **2017**, *155*, 530–548. [CrossRef]
9. Oh, K.; Chung, Y.-C.; Kim, K.W.; Kim, W.-S.; Oh, I.-S. Classification and Visualization of Alzheimer's Disease Using Volumetric Convolutional Neural Network and Transfer Learning. *Sci. Rep.* **2019**, *9*, 18150. [CrossRef]
10. Zhang, J.; Zheng, B.; Gao, A.; Feng, X.; Liang, D.; Long, X. A 3D Densely Connected Convolution Neural Network with Connection-Wise Attention Mechanism for Alzheimer's Disease Classification. *Magn. Reson. Imaging* **2021**, *78*, 119–126. [CrossRef]

11. Mehmood, A.; Maqsood, M.; Bashir, M.; Shuyuan, Y. A Deep Siamese Convolution Neural Network for Multi-Class Classification of Alzheimer Disease. *Brain Sci.* **2020**, *10*, 84. [\[CrossRef\]](#)
12. Li, F.; Liu, M. A Hybrid Convolutional and Recurrent Neural Network for Hippocampus Analysis in Alzheimer's Disease. *J. Neurosci. Methods* **2019**, *323*, 108–118. [\[CrossRef\]](#)
13. Solano-Rojas, B.; Villalón-Fonseca, R. A Low-Cost Three-Dimensional DenseNet Neural Network for Alzheimer's Disease Early Discovery. *Sensors* **2021**, *21*, 1302. [\[CrossRef\]](#)
14. Folego, G.; Weiler, M.; Casseb, R.F.; Pires, R.; Rocha, A. Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI. *Front. Bioeng. Biotechnol.* **2020**, *8*, 534592. [\[CrossRef\]](#)
15. Odusami, M.; Maskeliūnas, R.; Damaševičius, R.; Krilavičius, T. Analysis of Features of Alzheimer's Disease: Detection of Early Stage from Functional Brain Changes in Magnetic Resonance Images Using a Finetuned ResNet18 Network. *Diagnostics* **2021**, *11*, 1071. [\[CrossRef\]](#)
16. Basheera, S.; Sai Ram, M.S. Convolution Neural Network-Based Alzheimer's Disease Classification Using Hybrid Enhanced Independent Component Analysis Based Segmented Gray Matter of T2 Weighted Magnetic Resonance Imaging with Clinical Valuation. *Alzheimers Dement.* **2019**, *5*, 974–986. [\[CrossRef\]](#)
17. Wu, C.; Guo, S.; Hong, Y.; Xiao, B.; Wu, Y.; Zhang, Q.; Alzheimer's Disease Neuroimaging Initiative. Discrimination and Conversion Prediction of Mild Cognitive Impairment Using Convolutional Neural Networks. *Quant. Imaging Med. Surg.* **2018**, *8*, 992–1003. [\[CrossRef\]](#)
18. Ahila, A.M.P.; Hamdi, M.; Bourouis, S.; Rastislav, K.; Mohamed, F. Evaluation of Neuro Images for the Diagnosis of Alzheimer's Disease Using Deep Learning Neural Network. *Front. Public Health* **2022**, *10*, 834032.
19. Gocer, E. Diagnosis of Alzheimer's Disease with Sobolev Gradient-Based Optimization and 3D Convolutional Neural Network. *Int. J. Numer. Methods Biomed. Eng.* **2019**, *35*, e3225. [\[CrossRef\]](#)
20. Sethi, M.; Ahuja, S.; Rani, S.; Koundal, D.; Zaguia, A.; Enbeyle, W. An Exploration: Alzheimer's Disease Classification Based on Convolutional Neural Network. *BioMed Res. Int.* **2022**, *2022*, e8739960. [\[CrossRef\]](#)
21. Ebrahimiaghnavieh, M.A.; Luo, S.; Chiong, R. Deep Learning to Detect Alzheimer's Disease from Neuroimaging: A Systematic Literature Review. *Comput. Methods Programs Biomed.* **2020**, *187*, 105242. [\[CrossRef\]](#)
22. Agarwal, D.; Marques, G.; de la Torre-Díez, I.; Franco Martin, M.A.; García Zapirain, B.; Martín Rodríguez, F. Transfer Learning for Alzheimer's Disease through Neuroimaging Biomarkers: A Systematic Review. *Sensors* **2021**, *21*, 7259. [\[CrossRef\]](#)
23. Shen, D.; Wu, G.; Suk, H.-I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [\[CrossRef\]](#)
24. Plis, S.M.; Hjelm, D.R.; Salakhutdinov, R.; Allen, E.A.; Bockholt, H.J.; Long, J.D.; Johnson, H.J.; Paulsen, J.S.; Turner, J.A.; Calhoun, V.D. Deep Learning for Neuroimaging: A Validation Study. *Front. Neurosci.* **2014**, *8*, 229. [\[CrossRef\]](#)
25. Glasmachers, T. Limits of End-to-End Learning. In Proceedings of the Ninth Asian Conference on Machine Learning; PMLR, Seoul, Korea, 15–17 November 2017; pp. 17–32.
26. Wadekar, S.N.; Schwartz, B.J.; Kannan, S.S.; Mar, M.; Manna, R.K.; Chellapandi, V.; Gonzalez, D.J.; Gamal, A.E. Towards End-to-End Deep Learning for Autonomous Racing: On Data Collection and a Unified Architecture for Steering and Throttle Prediction. *arXiv* **2021**, arXiv:2105.01799.
27. Suk, H.-I.; Lee, S.-W.; Shen, D.; Alzheimer's Disease Neuroimaging Initiative. Deep Ensemble Learning of Sparse Regression Models for Brain Disease Diagnosis. *Med. Image Anal.* **2017**, *37*, 101–113. [\[CrossRef\]](#)
28. Vieira, S.; Pinaya, W.H.L.; Mechelli, A. Using Deep Learning to Investigate the Neuroimaging Correlates of Psychiatric and Neurological Disorders: Methods and Applications. *Neurosci. Biobehav. Rev.* **2017**, *74 Pt A*, 58–75. [\[CrossRef\]](#)
29. Liu, M.; Zhang, J.; Lian, C.; Shen, D. Weakly Supervised Deep Learning for Brain Disease Prognosis Using MRI and Incomplete Clinical Scores. *IEEE Trans. Cybern.* **2020**, *50*, 3381–3392. [\[CrossRef\]](#)
30. 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.* **2020**, *63*, 101694. [\[CrossRef\]](#)
31. Manjón, J.V. MRI Preprocessing. In *Imaging Biomarkers: Development and Clinical Integration*; Martí-Bonmatí, L., Alberich-Bayarri, A., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 53–63. [\[CrossRef\]](#)
32. Toshkhujaev, S.; Lee, K.H.; Choi, K.Y.; Lee, J.J.; Kwon, G.-R.; Gupta, Y.; Lama, R.K. Classification of Alzheimer's Disease and Mild Cognitive Impairment Based on Cortical and Subcortical Features from MRI T1 Brain Images Utilizing Four Different Types of Datasets. *J. Healthc. Eng.* **2020**, *2020*, e3743171. [\[CrossRef\]](#)
33. Ju, R.; Hu, C.; Zhou, P.; Li, Q. Early Diagnosis of Alzheimer's Disease Based on Resting-State Brain Networks and Deep Learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 244–257. [\[CrossRef\]](#)
34. Nguyen, M.; He, T.; An, L.; Alexander, D.C.; Feng, J.; Yeo, B.T.T.; Alzheimer's Disease Neuroimaging Initiative. Predicting Alzheimer's Disease Progression Using Deep Recurrent Neural Networks. *Neuroimage* **2020**, *222*, 117203. [\[CrossRef\]](#)
35. Ramírez, J.; Górriz, J.M.; Ortiz, A.; Martínez-Murcia, F.J.; Segovia, F.; Salas-Gonzalez, D.; Castillo-Barnes, D.; Illán, I.A.; Puntonet, C.G. Alzheimer's Disease Neuroimaging Initiative. Ensemble of Random Forests One vs. Rest Classifiers for MCI and AD Prediction Using ANOVA Cortical and Subcortical Feature Selection and Partial Least Squares. *J. Neurosci. Methods* **2018**, *302*, 47–57. [\[CrossRef\]](#) [\[PubMed\]](#)

36. Ramzan, F.; Khan, M.U.G.; Rehmat, A.; Iqbal, S.; Saba, T.; Rehman, A.; Mehmood, Z. A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State FMRI and Residual Neural Networks. *J. Med. Syst.* **2019**, *44*, 37. [[CrossRef](#)] [[PubMed](#)]
37. Mehmood, A.; Yang, S.; Feng, Z.; Wang, M.; Ahmad, A.S.; Khan, R.; Maqsood, M.; Yaqub, M. A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images. *Neuroscience* **2021**, *460*, 43–52. [[CrossRef](#)] [[PubMed](#)]
38. Tuan, T.A.; Pham, T.B.; Kim, J.Y.; Tavares, J.M.R.S. Alzheimer's Diagnosis Using Deep Learning in Segmenting and Classifying 3D Brain MR Images. *Int. J. Neurosci.* **2020**, *130*, 689–698. [[CrossRef](#)] [[PubMed](#)]
39. Song, J.; Zheng, J.; Li, P.; Lu, X.; Zhu, G.; Shen, P. An Effective Multimodal Image Fusion Method Using MRI and PET for Alzheimer's Disease Diagnosis. *Front. Digit. Health* **2021**, *3*, 637386. [[CrossRef](#)]
40. Odusami, M.; Maskeliūnas, R.; Damaševičius, R. An Intelligent System for Early Recognition of Alzheimer's Disease Using Neuroimaging. *Sensors* **2022**, *22*, 740. [[CrossRef](#)]
41. Bi, X.; Liu, W.; Liu, H.; Shang, Q. Artificial Intelligence-Based MRI Images for Brain in Prediction of Alzheimer's Disease. *J. Healthc. Eng.* **2021**, *2021*, 8198552. [[CrossRef](#)]
42. Puente-Castro, A.; Fernandez-Blanco, E.; Pazos, A.; Munteanu, C.R. Automatic Assessment of Alzheimer's Disease Diagnosis Based on Deep Learning Techniques. *Comput. Biol. Med.* **2020**, *120*, 103764. [[CrossRef](#)]
43. Tufail, A.B.; Ma, Y.-K.; Zhang, Q.-N. Binary Classification of Alzheimer's Disease Using SMRI Imaging Modality and Deep Learning. *J. Digit. Imaging* **2020**, *33*, 1073–1090. [[CrossRef](#)]
44. Herzog, N.J.; Magoulas, G.D. Brain Asymmetry Detection and Machine Learning Classification for Diagnosis of Early Dementia. *Sensors* **2021**, *21*, 778. [[CrossRef](#)]
45. Nanni, L.; Interlenghi, M.; Brahnam, S.; Salvatore, C.; Papa, S.; Nemni, R.; Castiglioni, I. The Alzheimer's Disease Neuroimaging Initiative. Comparison of Transfer Learning and Conventional Machine Learning Applied to Structural Brain MRI for the Early Diagnosis and Prognosis of Alzheimer's Disease. *Front. Neurol.* **2020**, *11*, 576194. [[CrossRef](#)]
46. Jiang, J.; Kang, L.; Huang, J.; Zhang, T. Deep Learning Based Mild Cognitive Impairment Diagnosis Using Structure MR Images. *Neurosci. Lett.* **2020**, *730*, 134791. [[CrossRef](#)]
47. Abrol, A.; Bhattarai, M.; Fedorov, A.; Du, Y.; Plis, S.; Calhoun, V. Deep Residual Learning for Neuroimaging: An Application to Predict Progression to Alzheimer's Disease. *J. Neurosci. Methods* **2020**, *339*, 108701. [[CrossRef](#)]
48. Prakash, D.; Madusanka, N.; Bhattacharjee, S.; Kim, C.-H.; Park, H.-G.; Choi, H.-K. Diagnosing Alzheimer's Disease Based on Multiclass MRI Scans Using Transfer Learning Techniques. *Curr. Med. Imaging* **2021**, *17*, 1460–1472. [[CrossRef](#)]
49. Gupta, Y.; Lee, K.H.; Choi, K.Y.; Lee, J.J.; Kim, B.C.; Kwon, G.R.; The National Research Center for Dementia; Alzheimer's Disease Neuroimaging Initiative. Early Diagnosis of Alzheimer's Disease Using Combined Features from Voxel-Based Morphometry and Cortical, Subcortical, and Hippocampus Regions of MRI T1 Brain Images. *PLoS ONE* **2019**, *14*, e0222446. [[CrossRef](#)]
50. Zeng, A.; Jia, L.; Pan, D.; Song, X. Early prognosis of Alzheimer's disease based on convolutional neural networks and ensemble learning. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi* **2019**, *36*, 711–719. [[CrossRef](#)]
51. Ortiz, A.; Munilla, J.; Górriz, J.M.; Ramírez, J. Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer's Disease. *Int. J. Neural Syst.* **2016**, *26*, 1650025. [[CrossRef](#)]
52. Li, A.; Li, F.; Elahifasae, F.; Liu, M.; Zhang, L. Alzheimer's Disease Neuroimaging Initiative. Hippocampal Shape and Asymmetry Analysis by Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis. *Brain Imaging Behav.* **2021**, *15*, 2330–2339. [[CrossRef](#)]
53. Cui, R.; Liu, M. Hippocampus Analysis by Combination of 3-D DenseNet and Shapes for Alzheimer's Disease Diagnosis. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 2099–2107. [[CrossRef](#)]
54. Zhang, T.; Shi, M. Multi-Modal Neuroimaging Feature Fusion for Diagnosis of Alzheimer's Disease. *J. Neurosci. Methods* **2020**, *341*, 108795. [[CrossRef](#)]
55. Liu, M.; Cheng, D.; Wang, K.; Wang, Y.; Alzheimer's Disease Neuroimaging Initiative. Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis. *Neuroinformatics* **2018**, *16*, 295–308. [[CrossRef](#)]
56. Xu, L.; Wu, X.; Chen, K.; Yao, L. Multi-Modality Sparse Representation-Based Classification for Alzheimer's Disease and Mild Cognitive Impairment. *Comput. Methods Programs Biomed.* **2015**, *122*, 182–190. [[CrossRef](#)]
57. Pan, X.; Phan, T.-L.; Adel, M.; Fossati, C.; Gaidon, T.; Wojak, J.; Guedj, E. Multi-View Separable Pyramid Network for AD Prediction at MCI Stage by 18F-FDG Brain PET Imaging. *IEEE Trans. Med. Imaging* **2021**, *40*, 81–92. [[CrossRef](#)]
58. Lu, D.; Popuri, K.; Ding, G.W.; Balachandar, R.; Beg, M.F.; Alzheimer's Disease Neuroimaging Initiative. Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease Using Structural MR and FDG-PET Images. *Sci. Rep.* **2018**, *8*, 5697. [[CrossRef](#)]
59. Abrol, A.; Fu, Z.; Du, Y.; Calhoun, V.D. Multimodal Data Fusion of Deep Learning and Dynamic Functional Connectivity Features to Predict Alzheimer's Disease Progression. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Berlin, Germany, 23–27 July 2019; Volume 2019, pp. 4409–4413. [[CrossRef](#)]
60. Venugopalan, J.; Tong, L.; Hassanzadeh, H.R.; Wang, M.D. Multimodal Deep Learning Models for Early Detection of Alzheimer's Disease Stage. *Sci. Rep.* **2021**, *11*, 3254. [[CrossRef](#)]
61. Shi, J.; Zheng, X.; Li, Y.; Zhang, Q.; Ying, S. Multimodal Neuroimaging Feature Learning with Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease. *IEEE J. Biomed. Health Inf.* **2018**, *22*, 173–183. [[CrossRef](#)]

62. Lu, D.; Popuri, K.; Ding, G.W.; Balachandar, R.; Beg, M.F.; Alzheimer's Disease Neuroimaging Initiative. Multiscale Deep Neural Network Based Analysis of FDG-PET Images for the Early Diagnosis of Alzheimer's Disease. *Med. Image Anal.* **2018**, *46*, 26–34. [CrossRef] [PubMed]
63. Shen, T.; Jiang, J.; Lu, J.; Wang, M.; Zuo, C.; Yu, Z.; Yan, Z. Predicting Alzheimer Disease from Mild Cognitive Impairment with a Deep Belief Network Based on 18F-FDG-PET Images. *Mol. Imaging* **2019**, *18*, 1536012119877285. [CrossRef] [PubMed]
64. Lee, G.; Nho, K.; Kang, B.; Sohn, K.-A.; Kim, D. Predicting Alzheimer's Disease Progression Using Multi-Modal Deep Learning Approach. *Sci. Rep.* **2019**, *9*, 1952. [CrossRef] [PubMed]
65. Er, F.; Goularas, D. Predicting the Prognosis of MCI Patients Using Longitudinal MRI Data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 1164–1173. [CrossRef] [PubMed]
66. Yang, Z.; Liu, Z. The Risk Prediction of Alzheimer's Disease Based on the Deep Learning Model of Brain 18F-FDG Positron Emission Tomography. *Saudi J. Biol. Sci.* **2020**, *27*, 659–665. [CrossRef]
67. Hon, M.; Khan, N.M. Towards Alzheimer's Disease Classification through Transfer Learning. In Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, USA, 13–16 November 2017; pp. 1166–1169. [CrossRef]
68. Grueso, S.; Viejo-Sobera, R. Machine Learning Methods for Predicting Progression from Mild Cognitive Impairment to Alzheimer's Disease Dementia: A Systematic Review. *Alzheimers Res. Ther.* **2021**, *13*, 162. [CrossRef]
69. Uludağ, K.; Roebroeck, A. General Overview on the Merits of Multimodal Neuroimaging Data Fusion. *NeuroImage* **2014**, *102*, 3–10. [CrossRef]
70. Tulay, E.E.; Metin, B.; Tarhan, N.; Arkan, M.K. Multimodal Neuroimaging: Basic Concepts and Classification of Neuropsychiatric Diseases. *Clin. EEG Neurosci.* **2019**, *50*, 20–33. [CrossRef]
71. Rieke, J.; Eitel, F.; Weygandt, M.; Haynes, J.-D.; Ritter, K. Visualizing Convolutional Networks for MRI-Based Diagnosis of Alzheimer's Disease. *arXiv* **2018**, arXiv:1808.02874. [CrossRef]
72. Korolev, S.; Safiullin, A.; Belyaev, M.; Dodonova, Y. Residual and Plain Convolutional Neural Networks for 3D Brain MRI Classification. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017; pp. 835–838. [CrossRef]
73. Liu, M.; Li, F.; Yan, H.; Wang, K.; Ma, Y.; Alzheimer's Disease Neuroimaging Initiative; Shen, L.; Xu, M. A Multi-Model Deep Convolutional Neural Network for Automatic Hippocampus Segmentation and Classification in Alzheimer's Disease. *Neuroimage* **2020**, *208*, 116459. [CrossRef]
74. Gao, F.; Yoon, H.; Xu, Y.; Goradia, D.; Luo, J.; Wu, T.; Su, Y. AD-NET: Age-Adjust Neural Network for Improved MCI to AD Conversion Prediction. *NeuroImage Clin.* **2020**, *27*, 102290. [CrossRef]
75. Basaia, S.; Agosta, F.; Wagner, L.; Canu, E.; Magnani, G.; Santangelo, R.; Filippi, M. Automated Classification of Alzheimer's Disease and Mild Cognitive Impairment Using a Single MRI and Deep Neural Networks. *NeuroImage Clin.* **2019**, *21*, 101645. [CrossRef]
76. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2018**, arXiv:1608.06993.
77. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* **2020**, arXiv:1905.11946.
78. MONAI—About Us. Available online: <https://monai.io/about.html> (accessed on 5 May 2022).
79. Welcome to ANTsPyNet's Documentation!—ANTsPyNet 0.0.1 Documentation. Available online: <https://antsx.github.io/ANTsPyNet/docs/build/html/index.html> (accessed on 6 May 2022).
80. ADNI | Alzheimer's Disease Neuroimaging Initiative. Available online: <https://adni.loni.usc.edu/> (accessed on 10 May 2022).
81. IXI Dataset—Brain Development. Available online: <https://brain-development.org/ixi-dataset/> (accessed on 10 May 2022).
82. Clinical Dementia Rating—An Overview. ScienceDirect Topics. Available online: <https://www.sciencedirect.com/topics/neuroscience/clinical-dementia-rating> (accessed on 10 May 2022).
83. Mini-Mental State Examination—An Overview. ScienceDirect Topics. Available online: <https://www.sciencedirect.com/topics/medicine-and-dentistry/mini-mental-state-examination> (accessed on 10 May 2022).
84. Emrani, S.; Arain, H.A.; DeMarshall, C.; Nuriel, T. APOE4 Is Associated with Cognitive and Pathological Heterogeneity in Patients with Alzheimer's Disease: A Systematic Review. *Alzheimers Res. Ther.* **2020**, *12*, 141. [CrossRef]
85. Abushakra, S.; Porsteinsson, A.; Vellas, B.; Cummings, J.; Gauthier, S.; Hey, J.A.; Power, A.; Hendrix, S.; Wang, P.; Shen, L.; et al. Clinical benefits of tramiprosate in alzheimer's disease are associated with higher number of apoe4 alleles: The “apoe4 gene-dose effect”. *J. Prev. Alzheimer's Dis.* **2016**, *3*, 219–228. [CrossRef]
86. Moore, C.M. NIFTI (File Format) Radiology Reference Article Radiopaedia.org. Radiopaedia. Available online: <https://radiopaedia.org/articles/nifti-file-format> (accessed on 16 May 2022).
87. Park, B.; Byeon, K.; Park, H. FuNP (Fusion of Neuroimaging Preprocessing) Pipelines: A Fully Automated Preprocessing Software for Functional Magnetic Resonance Imaging. *Front. Neuroinform.* **2019**, *13*, 5. [CrossRef]
88. Tustison, N.J.; Cook, P.A.; Klein, A.; Song, G.; Das, S.R.; Duda, J.T.; Kandel, B.M.; van Strien, N.; Stone, J.R.; Gee, J.C.; et al. Large-Scale Evaluation of ANTs and FreeSurfer Cortical Thickness Measurements. *Neuroimage* **2014**, *99*, 166–179. [CrossRef]
89. Bhagwat, N.; Barry, A.; Dickie, E.W.; Brown, S.T.; Devenyi, G.A.; Hatano, K.; DuPre, E.; Dagher, A.; Chakravarty, M.; Greenwood, C.M.T.; et al. Understanding the Impact of Preprocessing Pipelines on Neuroimaging Cortical Surface Analyses. *GigaScience* **2021**, *10*, gaa155. [CrossRef]

90. Tustison, N.J.; Avants, B.B.; Cook, P.A.; Zheng, Y.; Egan, A.; Yushkevich, P.A.; Gee, J.C. N4ITK: Improved N3 Bias Correction. *IEEE Trans. Med. Imaging* **2010**, *29*, 1310–1320. [CrossRef]
91. Ants. Utils. Bias_Correction—ANTsPy Master Documentation. Available online: https://antspy.readthedocs.io/en/latest/_modules/ants/utils/bias_correction.html (accessed on 15 May 2022).
92. Denoise An Image—Denoiseimage. Available online: <https://antsx.github.io/ANTsRCORE/reference/denoiseImage.html> (accessed on 15 May 2022).
93. Progressive Sprinkles and Salt-and-Pepper Noise. Available online: <https://www.simonwenkel.com/notes/ai/practical/vision/progressive-sprinkles-and-salt-and-pepper-noise.html#salt-and-pepper-noise> (accessed on 16 May 2022).
94. Pepper Noise—An Overview ScienceDirect Topics. Available online: <https://www.sciencedirect.com/topics/engineering/pepper-noise> (accessed on 15 May 2022).
95. Manjón, J.V.; Coupé, P.; Martí-Bonmatí, L.; Collins, D.L.; Robles, M. Adaptive Non-Local Means Denoising of MR Images with Spatially Varying Noise Levels. *J. Magn. Reson. Imaging* **2010**, *31*, 192–203. [CrossRef] [PubMed]
96. Ants. Utils. Denoise_Image—ANTsPy Master Documentation. Available online: https://antspy.readthedocs.io/en/latest/_modules/ants/utils/denoise_image.html (accessed on 16 May 2022).
97. Antspynet. Utilities. Brain_Extraction—ANTsPyNet 0.0.1 Documentation. Available online: https://antsx.github.io/ANTsPyNet/docs/build/html/_modules/antspynet/utilities/brain_extraction.html (accessed on 16 May 2022).
98. Trained Models. Available online: <https://github.com/neuronets/trained-models> (accessed on 16 May 2022).
99. Automated Brain Extraction. Available online: <https://github.com/neuronets/brainy> (accessed on 16 May 2022).
100. Lee, S.; Lee, G.-G.; Jang, E.S.; Kim, W.-Y. Fast Affine Transform for Real-Time Machine Vision Applications. In *Intelligent Computing*; Huang, D.-S., Li, K., Irwin, G.W., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1180–1190. [CrossRef]
101. Fonov, V.; Evans, A.C.; Botteron, K.; Almli, C.R.; McKinstry, R.C.; Collins, D.L. Unbiased Average Age-Appropriate Atlases for Pediatric Studies. *NeuroImage* **2011**, *54*, 313–327. [CrossRef]
102. Registration—ANTsPy Master Documentation. Available online: <https://antspy.readthedocs.io/en/latest/registration.html> (accessed on 17 May 2022).
103. Yadav, S.S.; Jadhav, S.M. Deep Convolutional Neural Network Based Medical Image Classification for Disease Diagnosis. *J. Big Data* **2019**, *6*, 113. [CrossRef]
104. Milletari, F.; Ahmadi, S.-A.; Kroll, C.; Plate, A.; Rozanski, V.; Maiostre, J.; Levin, J.; Dietrich, O.; Ertl-Wagner, B.; Bötzel, K.; et al. Hough-CNN: Deep Learning for Segmentation of Deep Brain Regions in MRI and Ultrasound. *Comput. Vis. Image Underst.* **2017**, *164*, 92–102. [CrossRef]
105. De Luna, A.; Marcia, R.F. Data-Limited Deep Learning Methods for Mild Cognitive Impairment Classification in Alzheimer’s Disease Patients. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Virtual, 1–5 November 2021; Volume 2021, pp. 2641–2646. [CrossRef]
106. Choi, H.; Jin, K.H. Alzheimer’s Disease Neuroimaging Initiative. Predicting Cognitive Decline with Deep Learning of Brain Metabolism and Amyloid Imaging. *Behav. Brain Res.* **2018**, *344*, 103–109. [CrossRef] [PubMed]
107. Pham, T.D. A Comprehensive Study on Classification of COVID-19 on Computed Tomography with Pretrained Convolutional Neural Networks. *Sci. Rep.* **2020**, *10*, 16942. [CrossRef] [PubMed]
108. Automated Medical Diagnosis of COVID-19 through EfficientNet Convolutional Neural Network. *Appl. Soft Comput.* **2020**, *96*, 106691. [CrossRef] [PubMed]
109. Ibrahim, H.A.K. Deep Learning Techniques for Medical Image Classification. Ph.D. Thesis, NOVA Information Management School, Universidade Nova de Lisboa, Lisbon, Portugal, 2021. Available online: <https://run.unl.pt/bitstream/10362/130159/1/D0059.pdf> (accessed on 20 May 2022).
110. Zhang, Y.-D.; Satapathy, S.C.; Zhang, X.; Wang, S.-H. COVID-19 Diagnosis via DenseNet and Optimization of Transfer Learning Setting. *Cogn. Comput.* **2021**, *13*, 1–17. [CrossRef] [PubMed]
111. Shamila Ebenezer, A.; Deepa Kanmani, S.; Sivakumar, M.; Jeba Priya, S. Effect of Image Transformation on EfficientNet Model for COVID-19 CT Image Classification. *Mater. Today Proc.* **2022**, *51*, 2512–2519. [CrossRef]
112. Ali, K.; Shaikh, Z.A.; Khan, A.A.; Laghari, A.A. Multiclass Skin Cancer Classification Using EfficientNets—a First Step towards Preventing Skin Cancer. *Neurosci. Inform.* **2022**, *2*, 100034. [CrossRef]
113. Oloko-Oba, M.; Viriri, S. Ensemble of EfficientNets for the Diagnosis of Tuberculosis. *Comput. Intell. Neurosci.* **2021**, *2021*, 9790894. [CrossRef]
114. MnasNet: Towards Automating the Design of Mobile Machine Learning Models. Google AI Blog. Available online: <http://ai.googleblog.com/2018/08/mnasnet-towards-automating-design-of.html> (accessed on 23 May 2022).
115. Droste, B. Google Colab Pro+: Is it Worth \$49.99? Medium. Available online: <https://towardsdatascience.com/google-colab-pro-is-it-worth-49-99-c542770b8e56> (accessed on 22 May 2022).
116. NVIDIA V100. NVIDIA. Available online: <https://www.nvidia.com/en-us/data-center/v100/> (accessed on 22 May 2022).
117. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980. [CrossRef]
118. Adam Optimizer PyTorch with Examples—Python Guides. 2022. Available online: <https://pythonguides.com/adam-optimizer-pytorch/> (accessed on 23 May 2022).

-
119. Using Learning Rate Scheduler and Early Stopping with PyTorch. DebuggerCafe. 2021. Available online: <https://debuggercafe.com/using-learning-rate-scheduler-and-early-stopping-with-pytorch/> (accessed on 23 May 2022).
 120. Zorlu, K.; Gokceoglu, C.; Ocakoglu, F.; Nefeslioglu, H.A.; Acikalin, S. Prediction of Uniaxial Compressive Strength of Sandstones Using Petrography-Based Models. *Eng. Geol.* **2008**, *96*, 141–158. [[CrossRef](#)]
 121. Yang, Y.; Zhang, L.; Du, M.; Bo, J.; Liu, H.; Ren, L.; Li, X.; Deen, M.J. A Comparative Analysis of Eleven Neural Networks Architectures for Small Datasets of Lung Images of COVID-19 Patients toward Improved Clinical Decisions. *Comput. Biol. Med.* **2021**, *139*, 104887. [[CrossRef](#)]