International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Predicting Alzheimer's Disease Onset: An Efficient Weighted Probability-Based Deep Ensemble Learning Method (WPBDELM) Using MRI Images

Naveen N.[1]*, Nagaraj G. Cholli[2]

Submitted :13/05/2024 **Revised:** 25/06/2024 **Accepted:** 01/07/2024

Abstract*:* Recent advancements in deep learning have enabled innovative approaches to analyze large datasets, particularly within the healthcare sector. Convolutional neural networks (CNNs) are becoming essential tools for classifying medical imaging data, with increasing attention on their application in neuroscience for Alzheimer's disease (AD) classification. Given that Alzheimer's is the most frequent cause of dementia in the aging population, detection at its early stages is critical. Early diagnosis depends on noninvasive imaging procedures, namely positron emission tomography (PET) and magnetic resonance imaging (MRI). There is much promise for enhancing early and precise AD detection through the analysis of several imaging modalities using CNNs. Furthermore, by combining several models, ensemble learning (EL) can greatly improve the performance of machine learning systems. This work presents an ensemble method for early AD diagnosis using MRI images using deep learning. The proposed method combines six prominent CNNs into an ensemble model, selected through a novel technique called the weighted probability-based deep ensemble learning method (WPBDELM). The study involved collecting and preprocessing data, developing individual and ensemble models, and evaluating them using ADNI data. The evaluation demonstrated high accuracy rates: 98.57% (NC/AD), 98.37% (NC/EMCI), 98.22% (EMCI/LMCI), 99.83% (LMCI/AD), 98.72% (three-way classification), and 98.78% (four-way classification). These results not only exceeded those of most reviewed studies but also were on par with the best-performing methods. Although individual models were outperformed by ensemble methods, there were no discernible differences between the different ensemble techniques. The evaluation outcomes demonstrated that although individual models performed less well in practice, the ensemble approach produced reliable and encouraging results.

Keywords: *Dementia, Alzheimer's disease, Ensemble learning, Deep learning, Mild cognitive impairment, Deep learning, Convolutional neural networks, Transfer learning, Magnetic resonance imaging, Transfer learning*

1. Introduction

Dementia is a group of neurological conditions marked by a steady deterioration in cognitive function. Of all dementia cases, 60 to 80 percent are caused by the most common form, Alzheimer's disease (AD) [1, 2]. It is an irreparable neurological disease that advances over time and is characterized by memory loss, cognitive decline, and difficulties in accomplishing regular tasks [3]. Although the precise reason for AD remains unclear, it is associated with brain changes that begin years before the disease's symptoms become apparent. These variations include the formation of neurofibrillary tangles within neurons, leading to neuronal death, and the buildup of amyloid plaques between nerve cells, which interferes with normal neurotransmitter function [4].

Mild cognitive impairment (MCI), a midway phase situated between normal cognitive function (NC) and Alzheimer's disease (AD), denotes an observable decline in cognitive capabilities without any disruption in an individual's everyday life [5]. Importantly, not all MCI subjects progress toward the development of AD or dementia; however, there is a substantial likelihood but there is a high likelihood of this transition occurring. Consequently, numerous studies have classified MCI as the preliminary phase of AD [6]

The incidence of age-related conditions like Alzheimer's disease has tended to increase within the past years because of the growing obligations of global life [7]. While the death rates of cardiovascular disease and cancer (prostate) have declined over the past two decades, the AD mortality rate has experienced a notable increase of 145%, ranking it as the sixth most fatal disease in the U.S. [8]. Despite the emergence of certain favorable outcomes in recent studies regarding novel pharmaceuticals that target AD, the disease remains devoid of an officially endorsed treatment [9]. As previously mentioned, the identification of a precise means to diagnose AD in its nascent stages can yield numerous advantages, including the cessation or reduction of disease progression, diminished healthcare expenses, and increased quality of life for individuals.

As far as the authors are aware, AD can be diagnosed in three different ways. The first approach, which is widely favored owing to its simplicity and cost-effectiveness, involves specialists employing clinical data, symptoms, and various criteria, such as cognitive assessment scales and questionnaires, to ascertain the presence of AD. However, this particular approach has significant limitations, including susceptibility to subjective factors and suboptimal performance outcomes [7]. The next method involves the measurement of clinical biomarkers, namely, the levels of amyloid-beta (β) and tau proteins, through the examination of cerebrospinal fluid (CSF) or brain autopsy. Although effective, this method is invasive and generally not preferred for routine early diagnosis [8]. The third approach employs neuroimaging techniques, including MRI, fMRI, PET, and DTI to observe the

[1] *Department of Computer Science & Engineering, M. S. Ramaiah University of Applied Sciences, Visvesvaraya Technological University, Belagavi, India. E-mail: naveenn.cs.et@msruas.ac.in*
ORCID iD: https://orcid.org/0000-0001-9146-3251
[2] *Department of Information Science & Engineering, RV College of Engineering, Visvesvaraya Technological University, Belagavi, India. E-mail: nagaraj.cholli@rvce.edu.in*
ORCID iD: https://orcid.org/0000-0001-7409-8272

anatomical structure and functionality of the brain. While this process offers extensive information in a short time frame, the interpretation of the comprehensive details within the images poses a considerable challenge for physicians [8].

Advancements in computational power and the accessibility of publicly accessible Alzheimer's disease datasets have resulted in the utilization of machine learning (ML) methodologies within the realm of early Alzheimer's disease diagnosis [9]. Deep learning (DL), which has garnered considerable attention because of its remarkable achievements in diverse fields and medical image analysis [10], can effectively extract high-level features that have demonstrated greater efficacy than conventional approaches in numerous studies [11]. In recent years, DL has become more prominent in different fields, especially medicine, in which it has become a useful tool for diagnosing AD. For instance, Suk et al. [12] conducted an initial study on using deep learning methods for Alzheimer's diagnosis in 2013, utilizing stacked autoencoders (SAE) for feature extraction and support vector machines (SVM) for classification. We present an overview of several analogous studies on DL implementation in early Alzheimer's diagnosis in the following paragraphs.

2. Related Work

Li et al. [13, 14] focused on diagnosing AD by examining the shape and asymmetry of the hippocampus using cascaded convolutional neural networks (CNNs). Their performance was slightly lower compared to a previous study [15] that accommodated only hippocampal shape features for classification. In contrast, Mehmood et al. [16] and Kang et al. [17] employed a VGG-based 2D-CNN architecture with transfer learning for early AD detection. Kang et al. also integrated a multimodal approach, incorporating MRI and DTI data.

ResNet, which is the highly favored CNN model that is in the scholarly literature, featured in studies conducted by Odusami et al. [18], Jabason et al. [19], Abrol et al. [6], Shanmugam et al. [20], Ramzan et al. [2], and Li et al. [13]. Certain studies have employed ResNet along with other deep learning methods for ensemble approaches [21] or comparative studies ([14, 18, 20, 22]. An attention mechanism-based 3D-ResNet was introduced by Zhang et al. [23] to enhance model interpretability for early AD diagnosis. Other studies, including those by Zhang et al. [23–26], Ji et al. [27], Guan et al. [28], and Liu et al. [29]. VGG and DenseNet were the second and third most prevalent CNN designs, respectively, as the literature indicates. The majority of research using VGGs has used conventional versions, including VGG16 or VGG19; however, certain studies, like those by Zhang et al. [23–26] and Yu et al. [30], have suggested customized variants.

Recent research endeavors have utilized DenseNet, an architecture that has proven highly successful in early AD diagnosis. Li and Liu [13, 31], as well as Liu et al. [32], employed the 3D-DenseNet model in their investigations, aiming for high-level feature extraction and successfully classifying various phases of AD. In their research, Li and Liu [13] employed a patch-based strategy to extract features from various brain regions. In contrast, Liu et al. [32] and Li and Liu [31] focused exclusively on the hippocampus for feature extraction.

Studies by Islam and Zhang [33], Ruiz et al. [34], Wang et al. [35], and others that merged different DenseNet topologies have documented several ensemble techniques. Customized convolutional neural network topologies have also been proposed by several papers examined in this field for the Alzheimer's diagnosis and its prodromal stages, like mild cognitive impairment, including the subtypes of early MCI (EMCI) and late MCI (LMCI). For the automatic classification of AD and MCI subtypes, Basaia et al. [10] suggested a 3D-CNN with 12 convolutional blocks that use a rectified linear unit (ReLU) as the activation layer, a fully connected layer, and a logistic regression layer as the classifier. To achieve the binary classification of prodromal stages of AD, Gorji and Kaabouch [36] built a basic 2D-CNN architecture consisting of three convolution layers, each followed by a max pooling operation, a fully connected layer, and a sigmoid classifier. Gray matter (GM) has received much attention in this study since it has been shown to influence AD's early onset. To detect AD early, Pan et al. [37] used an ensemble technique based on multiple 2D-CNN classifiers. Using single-axial slices of MR images, they produced many base CNN classifiers. By selecting the top classifiers on every axis, they then built an ensemble model.

The literature generally shows positive findings, with multiple studies showcasing the effectiveness of deep learning models in accurately categorizing healthy individuals, Alzheimer's disease patients, and patients with mild cognitive impairment. Consequently, deep learning may prove useful in the early detection of Alzheimer's disease. As such, the rationale for this study is based on two primary factors. Initially, since early diagnosis of Alzheimer's disease directly improves patient outcomes, it is crucial. Early detection of the illness allows for the initiation of treatment, which slows the disease's course and improves the standard of life for those suffering and their family. Second, detection in the early stages can also help reduce healthcare costs.

In Table 1, we will find a summary of the works on using Deep CNN models and Ensemble learning for AD classification. While numerous studies that were reviewed presented promising outcomes, a majority of them

Table 1: Summary of the literature related to AD classification using Deep CNN models and Ensemble learning

Author(s)	Dataset	Modality	Classifier(s)	Classification	Accuracy
Abrol et al. [6]	ADNI	MRI	CNN + 3D ResNet	3-way (NC/MCI/AD)	83.01%
Ramzan et al. [2]	ADNI	fMRI	ResNet-18	4-way (NC/EMCI/LMCI/AD)	97.28%
Odusami et al. [38]	ADNI	MRI	ResNet-18	3-way (NC/MCI/AD)	96.7%
Shanmugam et al. [20]	ADNI	MRI	ResNet + GoogleNet + AlexNet	3-way (NC/MCI/AD)	95.42%
Jabason et al. [19]	ADNI	MRI	Ensemble of DenseNet121, DenseNet169, DenseNet201, and ResNet50	4-way (NC/EMCI/LMCI/AD)	95.23%
Liu et al. [29]	ADNI	MRI	CNN from scratch AlexNet GoogLeNet	3-way (NC/MCI/AD)	78.02% 91.40% 93.02%,
Ruiz et al. [34]	ADNI	MRI	3D DenseNets	4-way (NC/EMCI/LMCI/AD)	96.38%
Pan et al. [37]	ADNI	MRI	2D-CNNs	3-way (NC/MCI/AD)	84.0%
Pexian et al. [39]	ADNI	MRI	Deep Broad Ensemble	3-way (NC/MCI/AD)	91.28%
Tanveer et al. [40]	ADNI	MRI	Deep Transfer Ensemble	4-way (NC/EMCI/LMCI/AD)	93.34%
Mujahid et al. [41]	ADNI	MRI	Ensemble model	4-way (NC/EMCI/MCI/AD)	97.35% (VGG-16 +Efficient Net-B2)
Raza et al. [42]	ADNI	MRI	CNN from scratch	4-way (NC/MCI/LMCI/AD)	97.84%

Savaş [43]	ADNI	MRI	EfficientNet-B0 EfficientNet-B2 EfficientNet-B3	3-way (NC/MCI/AD)	92.98% 94.42%, 97.28%
Rajesh et al. [44]	ADNI	MRI	MobileNetV2 and LSTM	3-way (NC/MCI/AD)	94%
Fathi et al. [45]	ADNI	MRI	Ensemble Model	3-way (NC/MCI/AD) 4-way (NC/EMCI/LMCI/AD)	93.92% 93.88%
Al-Adhaileh [46]	ADNI	MRI	AlexNet ResNet50	3-way (NC/MCI/AD)	94.53% 58.07%
Helaly et al. [47]	ADNI	MRI	2D CNN 3D CNN Fine-tuned VGG-19	4-way (NC/EMCI/LMCI/AD)	93.61% 95.17% 97.0%
Antony et al. [48]	ADNI	MRI	Random forest VGG-16 VGG-19	2-way (AD/NC)	68.0% 81.0% 84.0%
AbdulAzeem et al. [49]	ADNI	MRI	CNN from scratch	2-way (AD/NC)	95.60%
Hazarika et al. [50]	ADNI	MRI	Hybrid model (LeNet + AlexNet)	3-way (NC/MCI/AD)	93.58%
Rizwan et al. [51]	ADNI	MRI	VGG-16 & VGG-19	4-way (NC/EMCI/MCI/AD)	97.89%

failed to tackle all clinically relevant AD classification stages. Moreover, there aren't many thorough comparisons between various ensemble methods and individual CNN classifiers. In this study, we propose an ensemble deep learning-based technique for the early diagnosis of AD based on MR images to address these gaps. The weighted probability-based deep ensemble learning method (WPBDELM), a novel approach, is the basis for six well-known convolutional neural networks in the proposed technique.

The following summarizes the primary contributions of the current study:

1. We present a new ensemble technique called Weighted Probability-Based Deep Ensemble Learning Method (WPBDELM), which is intended to improve the efficacy of individual CNN models in the early detection of Alzheimer's disease.
2. To improve the ensemble method's overall efficiency, a comparison study was used to determine the best hyperparameters and scenarios for integrating individual CNNs.
3. We applied a domain adaptation (DA) based transfer learning approach to significantly boost model performance, surpassing results achieved with other parameter initialization methods
4. We addressed all significant binary and multiclass classification groups, providing a comprehensive evaluation of the proposed approach across diverse diagnostic scenarios.

In light of the aforementioned considerations, the primary aim of this study was to present was to introduce a novel ensemble method named the WPBDELM, which employs various base CNN architectures for early diagnosis of AD. The approach stands out for two key reasons. Firstly, The study differed from previous ones by using six different CNN classifiers, in contrast to earlier studies that employed only one or fewer than three categories of base convolution neural network classifiers or ensemble methods with individual CNN architectures. We carefully selected the quantity and types of base classifiers by comparing popular CNN architectures. Second, the study added a weight factor for every model's prediction, rather than using bagging or simple majority voting in the ensemble technique. This variable represents the accuracy of every disease class, offering a sophisticated method for ensemble decision-making.

3. Proposed Methodology

Detecting AD at an early stage is vital to controlling dementia in the older population since AD is the primary risk factor for dementia in older adults. Early diagnosis is greatly aided by non-invasive brain imaging procedures like magnetic resonance imaging (MRI). With CNNs integrated into the analysis of these imaging modalities, AD can be detected more early and precisely. Furthermore, by merging many models, ensemble learning (EL) has been demonstrated to dramatically improve the performance of machine learning systems. This paper suggests a deep learning-based ensemble method for using MR images to diagnose AD early. The weighted probability-based deep ensemble learning method (WPBDELM), a novel methodology, compares many ensemble scenarios before selecting the ensemble model, which is composed of six well-known CNNs. Data collection, preprocessing, individual and ensemble model creation, and model evaluation using ADNI data were all part of the suggested research. Figure 1 shows an overview of the suggested methodology.

The ADNI1 Complete 2Yr 1.5 T, MP-RAGE sequence standardized dataset, which included patient scans taken at six, twelve, eighteen, and twenty-four months after diagnosis, was used in our study to train our model. All of the sMRI data from the ADNI1 collection are used in this investigation. At baseline, 741 patients in total were included, including 173 AD, 139 EMCI, 156 MCI, 136 LMCI, and 137 NC. Healthy aging controls who have not converted within two years of baseline follow-up visits make up the NC class. The period of supervision for the switch to AD was eighteen months. Individuals with modest cognitive impairments who were nevertheless able to perform daily tasks were kept in the MCI class. Patients who are classified as AD at baseline and show no evidence of reversal after two years of follow-up visits are included in the AD class. The demographic information for the 741 participants (173 AD, 139 EMCI, 156 MCI, 136 LMCI, and 137 NC), whose ages range from 55 to 92 years, is summarized in Table 2.

3.1 Data Preprocessing

The ADNI provides images, such as 3D MRI scans, in NIFTI format. Data from each brain sMRI were normalized through preprocessing. The preprocessing procedure consists of skull-stripping, contrast enhancement, denoising, bias field correction, registration, intensity normalization, and extraction of 2D images from 3D MRI volumes. Skull stripping is the elimination of nonbrain tissue that serves as noise reducing the CNN classification performance. By enhancing pixel value ranges and picture border contrast, contrast enhancement raises an image's quality and brightness. Denoising is the technique of eliminating noise from an image to bring it back to its original

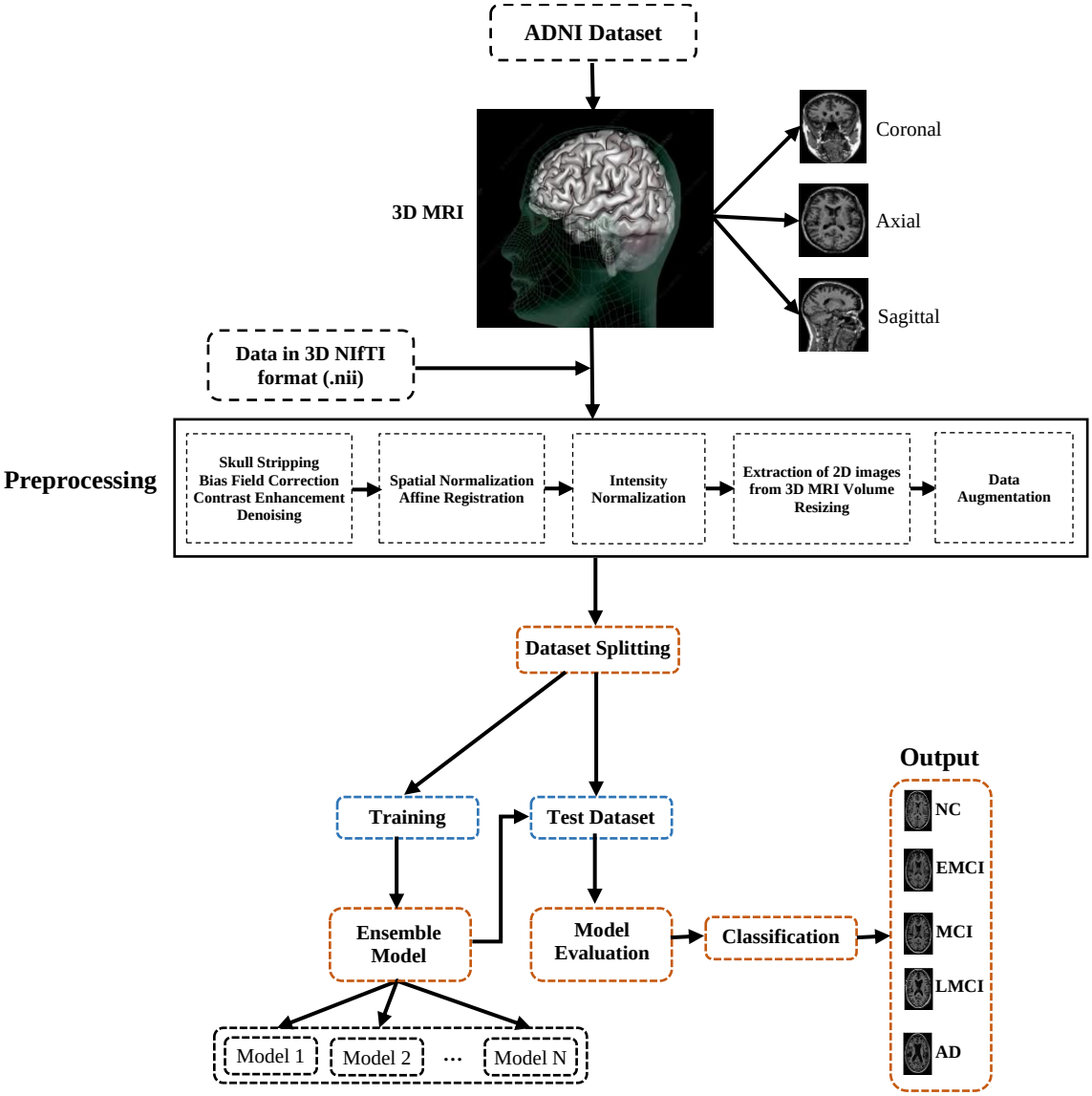

Fig.1: Proposed framework of our study

Table 2: Demographic description of the dataset along with clinical information

Variable	NC	EMCI	MCI	LMCI	AD
Number of participants	137	139	156	136	173
Gender (male: female)	74:63	72:67	81:75	71:65	89:84
Age (year; mean, std)	74.3, 5.4	74.5, 7.2	74.2, 7.1	74.8, 7.3	75.12, 7.56
Weight (kg; mean, std)	73.8, 13.6	77.2,12.7	76.4, 13.7	72.7, 14.3	73.9, 14.6
MMSE (mean, std)	29.18, 0.96	27.19, 1.71	26.39, 1.21	25.47, 1.84	23.2, 2.0
CDR (mean, std)	0.00, 0.00	0.50, 0.00	0.50, 0.00	0.50, 0.00	0.75, 0.25
Total Image Scans	712	667	904	779	1023

quality while keeping crucial information. A 3D Gaussian kernel with voxel size (2×2×2) was used to smooth MRIs. Using the SimpleITK library, the N4 bias field correction technique was applied to fix uneven low-frequency intensity in brain sMRI. We take into account an affine regularization using the ICBM space template, and a bias regularization of 1e−3 with a full bias width at half maximum of 60mm cut-off. The next step was to normalize the MRI signal intensity for each image. This involved dividing the original value of each voxel by the image's initial maximal value, which produced a value between 0 and 1. We used the med2image Python package (version 1.1.2) to take 2D slices out of the 3D MRI scans that had already been processed. The conversion procedure comprised taking 3D MRI volumes and producing 2D slices in lossless PNG format. Each 3D MRI scan, with dimensions of 181×217×181, was re-sliced into three 2D image sets, corresponding to the sagittal, coronal, and transverse orientations, to meet the requirements of CNN training, validation, and testing. The X, Y, and Z axes, which are perpendicular to the corresponding planes, are in line with these orientations. Generally speaking, the brain's center contains more information than its periphery. The slices in the middle have a higher information entropy than the remaining ones. Consequently, not every slice will be utilized in the training process. Sagittal, coronal, and axial view slices all provide complementing data. The middle slices, or the 80th and 85th slices in each orientation, are known to include important anatomical information, thus we extracted slices at intervals of five. The number of 2D images produced by this method was as follows: 4,272 for NC, 4,002 for EMCI, 5,424 for MCI, 4,674 for LMCI, and 6,138 for AD. The center and spatial resolution of the scaled image stayed the same, but each 2D slice was reformatted to a uniform dimension of 224×224 using edge padding and zero filling, making the 2D slice squared. The photos

were now comparable in size to those in the ImageNet dataset—a popular resource for transfer learning—thanks to this normalization. The deep learning system might then employ these produced 2D slices for testing, validation, and training.

3.2 Class Imbalance

We noticed that the dataset we collected had a significant imbalance in the instance count across different AD categories. We implemented two resampling techniques to tackle the issue: oversampling and undersampling. We duplicated instances from underrepresented classes, namely, MCI, EMCI, and LMCI in the oversampling approach. Conversely, in the undersampling approach, we removed instances from overrepresented classes, namely, AD and NC. After applying these resampling methods, we ensured that all the AD classes contained 5,250 MR images, which led to an expanded dataset of 26,250 images.

3.3 Data Augmentation

Deep neural networks (DNNs) enhance model performance, especially when trained with larger datasets. Data augmentation is a technique that involves generating more images to expand the original dataset. Introducing modifications to the images boosts a model's ability to classify more accurately. As a result, the model can become more generalized and less prone to overfitting.

Effective training of deep neural networks requires a vast dataset. However, when the dataset is not sufficiently large, such as in the case of MRI, data augmentation techniques can significantly diversify the data for neural network training. There are various data augmentation techniques including translation, gamma correction, scaling, random affine transformation, random noise addition, and rotation. In our study, we compiled a dataset of images from 741 patients. Unfortunately, the dataset was insufficient for practical deep neural network training and optimal performance. To address this issue, we used image rotation (90°, 180°, and 270°), flipping/reflection (both horizontal and vertical flipping), and zooming in and out data augmentation techniques to expand the dataset. The original training dataset was supplemented with augmented data to provide a suitably big sample size. Data augmentation was also employed to address the original dataset's imbalance, with the number of augmented slices generated varying by class. As a result of implementing these techniques, we enlarged the dataset from 26,250 to 52,500 images. Subsequently, the dataset was divided into three sets for training (70%), validation (15%), and testing (15%). A summary of the dataset used in our research is given in Table 3.

Table 3: Description of the training, validation, and test datasets

Class Label	NC	EMCI	MCI	LMCI	AD	Total
Image count	10500	10500	10500	10500	10500	52500
Train set	7350	7350	7350	7350	7350	36750
Validation set	1575	1575	1575	1575	1575	7875
Test set	1575	1575	1575	1575	1575	7875

3.4 Cross-Validation

Cross-validation is a technique used to assess the performance of a model by splitting the dataset into three subsets. One subset is used for testing the model, while the other subsets are used for training and validation. A 10-fold stratified cross-validation is frequently used for classification tasks to make sure that the distribution of classes in every fold is consistent with the distribution of the entire dataset. This strategy helps in identifying the optimal hyperparameters for the classification model. Training and testing subsets are distinct sets, meaning they do not share any sample, to prevent double dipping. The findings are calculated as the average of 10 evaluations over the course of 10 folds, and this process is repeated for each of the 10 folds. Estimating the generalization error is the primary goal of cross-validation, which guarantees that comparable outcomes will be produced on fresh unseen data (i.e., the least generalization error). Since the models created during the training phase are computed using k-1 folds rather than all the training sets, in practice, this inaccuracy will always lead to an overestimation of the true prediction error. This overestimation decreases as k grows and is dependent on the classifier's learning curve slope.

3.5 The Proposed Model

The preprocessing steps involved normalizing intensities, resizing all slices to a default dimension of 256x256, and selecting informative slices. A slice-based method was then applied by the suggested deep learning-based ensemble technique for early Alzheimer's disease detection. This approach is favored for its simplicity, low complexity, and alignment with the study's end-to-end structure. Furthermore, Furthermore, the transformation of 3D images into 2D slices, the dataset was expanded, addressing concerns about overfitting and enhancing model generalizability.

The method proposed in this study utilizes CNN architectures, with a focus on combining multiple classifiers for enhanced AD detection. ix base classifiers were selected, drawing inspiration from established CNN architectures, which have demonstrated strong performance in previous studies. These architectures include DenseNet12, DenseNet201, DenseNet169, VGG-19, ResNet-50, and Inception-ResNet V2. By altering the last few layers of the architecture, each classifier was customized to the current study setting.

DenseNets utilize a foundational structure known as the dense block, where each layer is directly linked to every other layer. This design improves information flow and addresses the vanishing gradient problem. Standard DenseNet architectures generally consist of four dense blocks, five transition layers, one fully connected layer, and a Softmax layer for classification. The number of layers and structure of the dense blocks differ throughout DenseNet versions, such as DenseNet201 and DenseNet169. The proposed DenseNet architectures for this study are illustrated in Fig. 2.

We changed the final layer of the conventional DenseNet architecture, as shown in Fig. 2. The last layer now consists of a Softmax layer, a dropout layer with a dropout rate of 0.3, a fully connected layer with 32 neurons, and a batch normalization layer.

ResNet was used to reduce the vanishing gradient issue and speed up model convergence by using shortcut connections between layers, much like DenseNet. Our study's modified ResNet design, which was modeled after ResNet-50, has four stages with three, four, six, and three residual blocks in each step. To further enhance the architecture's performance, additional layers similar to those found in the final layers of DenseNet were incorporated at the end of the model (refer to Figure. 3). This configuration aims to leverage the strengths of ResNet while incorporating elements of DenseNet to optimize the model's effectiveness for our specific application. The arrangement of residual blocks at different stages, as shown in Fig. 3, is consistent in terms of the number of layers but differs in terms of the number of kernels.

Combining the best features of two important developments in deep learning, the Inception architecture and residual connections (ResNet), Inception-ResNet V2 is a potent and incredibly effective deep convolutional neural network architecture. The Inception-ResNet V2 model merges the multi-scale processing capability of the Inception architecture with the identity mapping features of residual networks. The Inception modules enable the network to analyze visual information at multiple scales simultaneously, capturing both fine and coarse details from the input images. Conversely, residual connections enable gradients to pass through the network directly during backpropagation, which helps to lessen the effects of the vanishing gradient issue. This feature makes it easier to train considerably deeper networks—which are essential for identifying the complex patterns required for precise categorization. We modified the Inception-ResNet V2 standard version as an additional unique architecture that adds to the ensemble model. Fig. 4 shows the Inception-ResNet's reduced structure.

VGG19 was the last individual architecture in the ensemble model that was suggested. The modified architecture, as shown in Fig. 5, only uses the first six layers of the conventional version to reduce the number of parameters and computational costs, accelerate convergence, and lessen the possibility of overfitting. In addition, the model included two dropout layers, two fully linked layers, two batch normalization layers, and one Softmax layer.

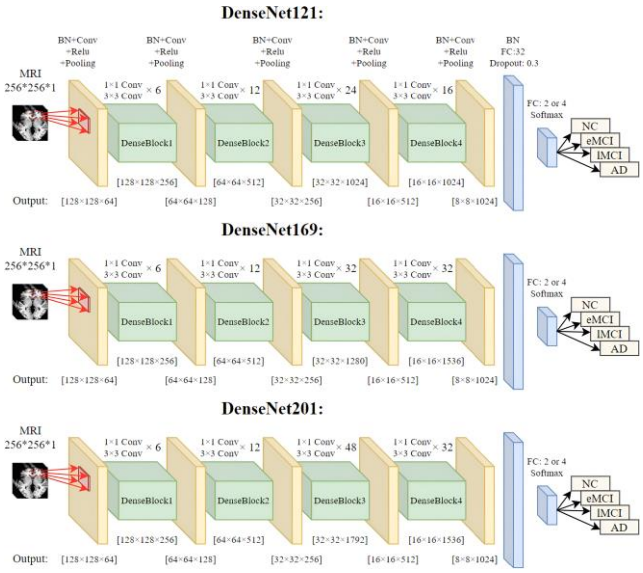

Fig.2. Proposed DenseNet architecture

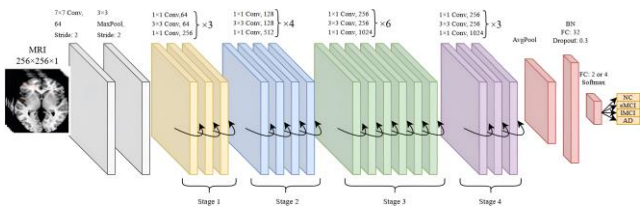

Fig.3. Proposed ResNet architecture

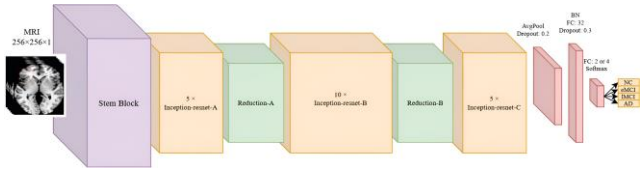

Fig.4. Proposed Inception-ResNet architecture

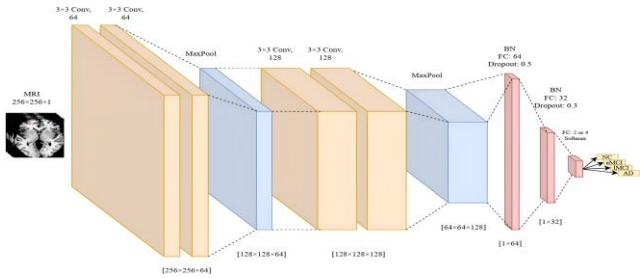

Fig.5. Proposed VGG architecture

3.6 Ensemble Learning

We suggest an ensemble method that is divided into two stages. Using the same training and test datasets, we first trained and assessed each base classifier. To integrate these classifiers in the second phase, we used an ensemble method based on weighted probability. Figure 6 depicts our model's flow.

In the first phase, each classifier's accuracy served as its weight in the final ensemble model. Essentially, classifiers with higher accuracy had a greater influence on the overall ensemble model. The probabilistic value of each class in each classifier was multiplied by its associated weight to get the output of the model. The sum of these weighted probabilities was then processed through a Softmax function, yielding the class with the highest probability as the final output. This step ensured that the ensemble method effectively combined the strengths of the individual models, further enhancing the overall classification accuracy. These steps can be summarized as follows:

$$C_q = \sum_{p=1}^{6} w_p \times \alpha_q^p \; where \; p = 1, \dots, 6 \; and \; q = 1, .., 4 \quad (1)$$

The classifier index p is used to represent each classifier, whereas i is used to represent each class. The weight (accuracy) of the pth classifier is represented by w_p , and the probability value of the qth class in the pth classifier is represented by α_q^p. Additionally, S_q represents the weighted probabilities sum for the qth class. The output of the model is derived from:

$$S = softmax(C_q) \; where \; q = 1, \dots, 4 \quad (2)$$

$$E = argMax(S_q) \; where \; q = 1, \dots, 4 \quad (3)$$

where S stands for the Softmax function's outcome and E for the ensemble method's final output.

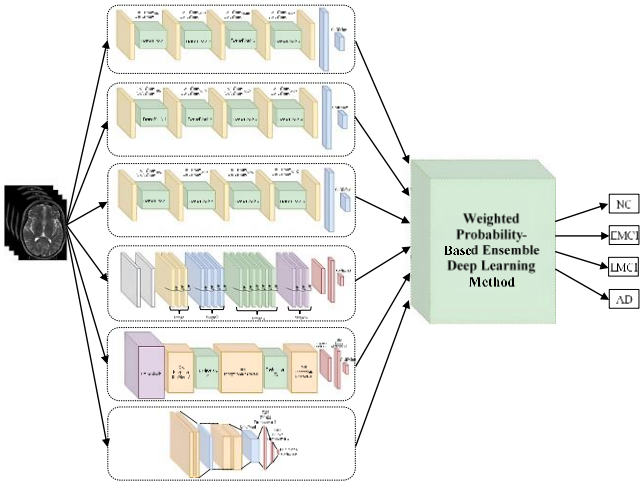

Fig.6. The general framework of the proposed model

3.7 Transfer Learning with Fine Tuning

To improve the efficiency of our proposed model and speed up the training process, we employed a method of transfer learning based on domain adaptation, along with fine-tuning. The source and destination datasets are in the same domain even if they differ from one another. Our approach involved initially training the deep model using a randomized initialization method for the NC/AD binary classification group. The remaining classification groups then employed the parameters from this initial model, utilizing a two-stage transfer and fine-tuning procedure as follows. The pre-trained NC/AD classification model's initial convolutional blocks and layers were locked throughout the transfer phase to avoid additional training, allowing only the custom layers to be trained. The model was retrained using the new categorization group at a learning rate of 0.001, yielding the transferred version of the model. All of the transferred version's layers and convolutional blocks were unlocked at a lower learning rate of 0.0001 during the next step, known as tuning. To get the final, refined (fine-tuned) version, the model was retrained once again.

3.8 Performance Analysis

Upon completing the training phase, the models underwent evaluation using different performance metrics such as accuracy, sensitivity, and specificity. These calculations were based on Eq. (4) – (6). To ensure a fair evaluation, the data was split into training (70%), validation (15%), and testing (15%) sets.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (4)$$
$$Sensitivity = True\ Positive\ Rate = Recall = TP/(TP + FN) \quad (5)$$

$$Specificity = True\ Negative\ Rate = TN/(TN + FP) \quad (6)$$

Where,
True positive (TP): Positive occurrences that were accurately predicted.
True Negative (TN): Negative occurrences that were accurately predicted.
False positive (FP): Positive occurrences that were incorrectly predicted (Type I error).
False Negative (FN): Negative occurrences that were incorrectly predicted (Type II mistake).

We used the ROC curve, in addition to the performance metrics mentioned above, to assess and compare how well our ensemble model performed compared to the individual models.

4. Experimental Results

We conducted a comprehensive evaluation and comparison of the ensemble model and its individual components, examining their performance on a range of binary and multiclass classification groups such as NC/AD, NC/EMCI, EMCI/LMCI, LMCI/AD, and 4-way (NC/EMCI/LMCI/AD).

4.1 Hyperparameter selection

After extensively exploring our data and studying the relevant literature, we confidently determined the ideal values for our hyperparameters. As shown in Table 4, we selected the appropriate hyperparameters.

Table 4: Selection of hyperparameter values

Parameter	Value
Learning rate	0.001, 0.0001
Loss function	Cross-entropy
Optimization Algorithm	Adam
Batch size	32, 64
Epoch count	50,100,200,300,350

Applied in deep learning models are several optimization algorithms like root mean squared propagation (RMSProp), Adadelta, stochastic gradient descent (SGD), Adagrad, and Adam, the current literature favors Adam and SGD [52]. Therefore, in our study, we opted for Adam because it remains the cheapest computational choice compared with other algorithms. Furthermore, it has been demonstrated that cross-entropy performs better in classification models with a Softmax output layer, even though the mean square error (MSE) and cross-entropy are the most often utilized loss functions in the existing literature. During the initial training phase, we used the Keras framework's default learning rate of 0.001. Subsequently, we had to lower the learning rate to 0.0001 for fine-tuning the parameters. We originally set the batch size to 64 to expedite training, but after running into an out-of-memory (OOM) problem, we reduced it to 32. The difficulty of the categorization groups was taken into consideration when determining the number of training epochs by exploratory means. As a result, more epochs were needed for groups that presented a greater challenge. As we delved further into the exploration of optimal model training, we noticed a marked difference. When fine-tuning, the model consistently reached its optimal solution in a fraction of the time needed during initial training. The discovery that fewer epochs were needed for successful results prompted us to delve deeper into the optimal number of epochs for training our model. In our comprehensive analysis, we utilized the dependable DenseNet-121 as the foundation for our research. We then extended our findings to other architectures, yielding broader and more widely applicable conclusions. These findings are illustrated in Fig. 7.

Figure 7 shows that the random parameter initialization strategy takes longer to converge, typically between 200 and 350 epochs, whereas the fine-tuned models converge within 100 epochs. When considering the various classification groups, it is evident that certain groups, such as EMCI/LCMI and 4-way classification, pose a greater challenge and therefore take longer to converge, requiring a greater number of epochs.

In our quest for the best results, we delved even deeper into our analysis. We rigorously investigated various hyperparameters, including the selection of dropout layer numbers and values, the incorporation of batch normalization layers, and the choice of base classifiers and their numbers within our ensemble model. Furthermore, we specifically utilized the NC/AD classification group and implemented the esteemed DenseNet121 as our base classifier. Our examination of these variables was thoroughly conducted in various scenarios, as shown in Fig. 8.

Figure 8 demonstrates that the model performs exceptionally well when a single dropout layer with a value of 0.3 and one batch normalization layer is used compared with the other configurations. We also evaluated prominent CNN models and found that DenseNet-based models outperformed others, as shown in Fig. 8d. On the basis of these results, we defined and evaluated different combinations for merging the individual CNN classifiers to obtain the optimal ensemble model. The performance in different scenarios graphically is shown in Figure 9.

Fig. 9 displays the highest-performing scenarios among the seven defined options, with E6 and E7 achieving an impressive 98.57% accuracy. Considering this, we opt for the E6 scenario, which combines DenseNet121, DenseNet201, DenseNet169, ResNet-50, VGG-19, and Inception-ResNet V2, as our ultimate ensemble model because of its efficient computational cost compared with that of E7.

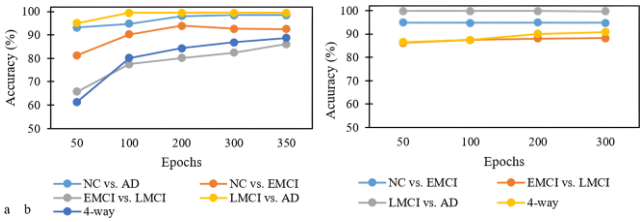

Fig.7. Analyzing the required epoch count using the DenseNet121 architecture - a. Attained accuracy for various epochs in the random parameters initialization approach. b. Attained accuracy for various epochs in fine-tuned models

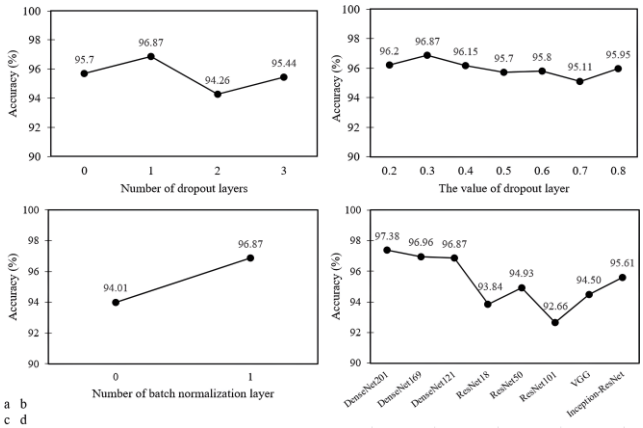

Fig.8. Hyperparameter exploration and analysis: a. assessing the accuracy of the model with different dropout layer counts b. Assessing model accuracy with different dropout layer values. c. batch normalization layer's effect on model prediction. d. accuracy comparison between several well-known CNN designs.

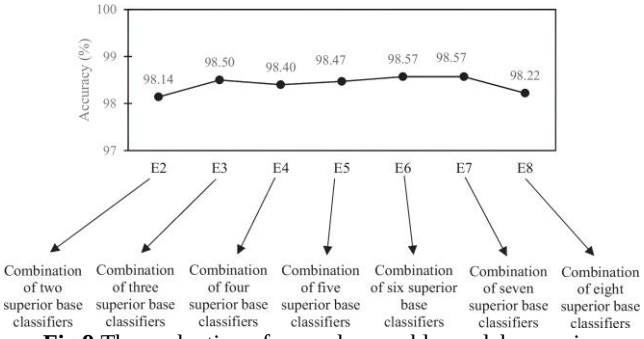

Fig.9 The evaluation of several ensemble model scenarios

4.2 Performance evaluation

4.1.1 Binary and Four-Way Classification

Once the appropriate hyperparameters were chosen, we proceeded to train and evaluate the base classifiers and the ensemble method. Our focus was on early detection of AD, so we created a range of binary and multiclass classification groups. The results for both the binary and four-way classification models are listed in Table 5.

According to Table 5, the ensemble methods (listed in the last three rows) outperformed the individual models in terms of performance. When different ensemble techniques were analyzed, there were no major discrepancies in the classification results. However, the WPBDELM model was marginally better than other widely used ensemble methods across various classification groups. For a visual representation, Figure 10 displays the ROC plot for both individual models and the proposed ensemble approach in every classification group.

Figure 10 illustrates the difficulties encountered in classifying the EMCI/LMCI group, where the combined model outperforms the single classifiers. Keep in mind that ROC curves cannot be directly applied to multiclass classification models; therefore, we drew separate ROC curves for each class in the 4-way classification model. Our suggested ensemble technique showed satisfactory results in identifying all disease categories in the 4-way classification group, with particular success in detecting AD patients and NCs (refer to Fig. 10e).

Table 5. Performance of Binary and Four-Way Classification on the Test Set

Method	NC vs. AD			NC vs. EMCI			EMCI vs. LMCI			LMCI vs. AD			4-way
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spc	Acc
DenseNet201	97.38	96.59	98.43	98.83	98.79	99.68	97.03	98.36	95.61	92.56	91.17	93.96	94.87
DenseNet169	96.96	97.19	96.67	98.74	98.55	99.66	97.81	99.32	95.02	94.65	93.76	95.57	94.29
DenseNet121	96.88	97.78	95.68	98.57	98.70	98.69	98.19	99.13	**96.02**	94.28	93.18	95.09	93.48
ResNet50	94.94	94.67	95.29	97.98	97.96	98.43	96.15	98.64	92.02	94.30	92.98	95.71	94.55
VGG19	94.51	93.19	96.27	97.98	97.22	99.64	97.03	99.27	92.62	94.32	96.18	92.43	93.99
Inception-ResNet	95.61	95.11	96.27	97.89	97.96	98.06	96.24	98.23	93.62	91.88	91.31	92.45	93.80
MVEM	98.31	97.63	98.22	99.16	98.40	99.68	98.29	99.46	94.86	97.77	95.54	98.36	97.93
PBEM	98.50	98.72	98.15	99.13	98.70	99.74	98.34	99.51	94.91	98.07	97.73	98.42	97.83
WPBDELM (Proposed Method)	**98.57**	**98.81**	**98.24**	**99.23**	**99.67**	**99.82**	**98.49**	**99.63**	95.02	**98.27**	**98.13**	**98.42**	**98.78**

Note: PBEM – probability-based ensemble method, MVEM – majority-voting ensemble method, WPBDELM – weighted probability-based ensemble deep learning method

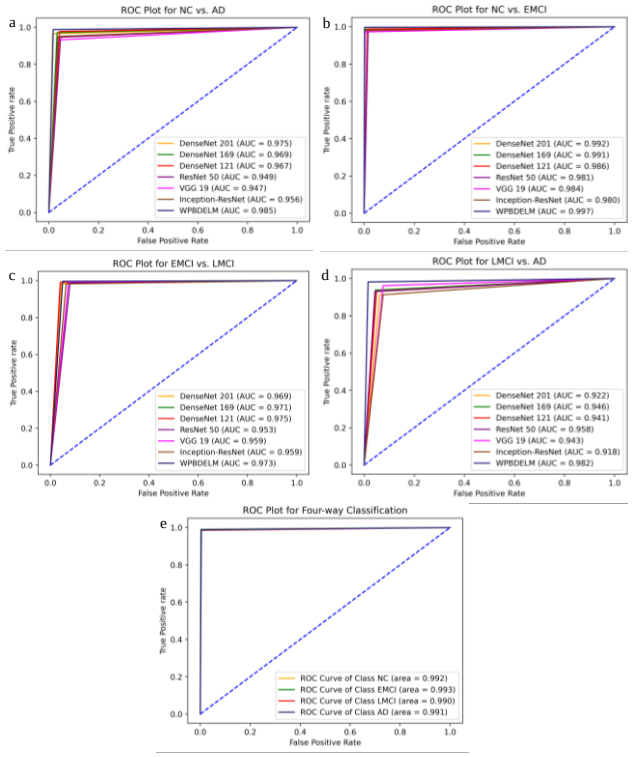

Fig.10. ROC curve for categorization group a. NC/AD b. NC/EMCI c. EMCI/LMCI d. LMCI/AD e. 4-way classification

4.1.2 Three-Way Classification

An ADNI data-driven three-way classification model has been trained and evaluated. The performance results are detailed in Table 6 and Fig. 11, which show that the suggested ensemble framework (WPBDEM) outperformed the individual architectures.

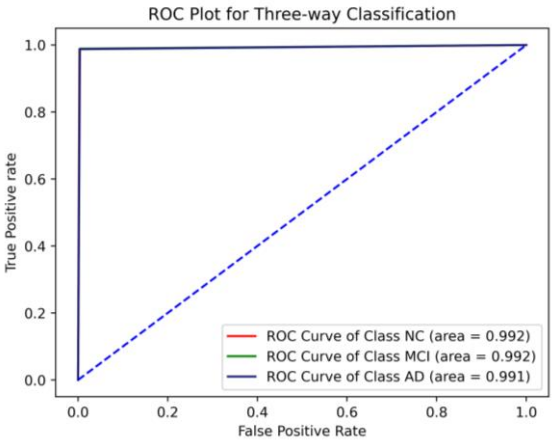

Fig.11. ROC plot for three-way classification

Table 6. Performance of Three-Way Classification on the Test Set

	DenseNet201	DenseNet169	DenseNet121	ResNet50	VGG	Inception-ResNet	WPBEM
Evaluation	97.14	96.91	96.81	97.12	96.29	94.92	**98.72**

5. Discussion

This study aimed to present an ensemble learning model for Alzheimer's disease early detection that makes use of deep learning techniques. The process involved collecting and preprocessing data, forming individual models, and evaluating them against ADNI data. This innovative method was chosen following a comprehensive comparative study of several ensemble scenarios. The final ensemble model was formed by selecting the top six individual CNN-based classifiers. As previously indicated, the model's performance was assessed using critical metrics including accuracy, sensitivity, and specificity, placing it on par with current cutting-edge techniques. Table 5 outlines the comparison of our suggested model and other relevant studies based on the assessment of ADNI data.

In our comprehensive study, we address all significant classification groups, including both binary and multiclass classifications. While we did not include NC/LMCI and EMCI/AD in our analysis because of their lower clinical significance for early AD diagnosis, we did thoroughly cover other MCI subgroups (namely, EMCI and LMCI). Upon reviewing Table 5, which enumerates research utilizing MR images as input data, we discovered that not all classification groups were covered by any of the examined studies. While Yue et al. [53], Basaia et al. [10], and Mehmood et al. [16] covered all binary classification groups in their research, they failed to consider the multiclass classification group, which is just as significant.

According to Table 5, our performance results surpass those of the majority of the reviewed papers and are on par with those of others. Our proposed model particularly excels in four-way classification, outperforming other methods. While Yue et al. [53] produced remarkable results in several binary classification groups, multiclass classification was not addressed in their work.

The ensemble approach has been widely utilized in various studies, with majority voting (MVEM) and probability-based methods (PBEM) being the most commonly used approaches [5, 19, 33, 54, 55]. Notably, MVEM has been prominent in the literature. In addition, the PBEM has also been incorporated into studies conducted by Ruiz et al. [34] and Wang et al. [35]. Inspired by the PBEM, a new method called the WPBDELM was introduced in the present study and compared with other approaches. Despite previous research consistently supporting the effectiveness of ensemble methods over individual architectures, minimal differences are observed among different ensemble approaches according to a comparative analysis [41]. In most binary classification groups, the WPBDELM yields equal or slightly better results than the other approaches do, whereas the MVEM outperforms the other methods in four-way classification (refer to Table 3 for details). A groundbreaking ensemble technique was introduced in a recent study by Ma et al. [39], referred to as the deep-broad ensemble, which combines 3D-residual convolutional blocks and a broad learning system. This innovative approach has shown remarkable performance improvements over individual methods. The proposed approach offers a notable advantage by removing the need for costly hardware and lengthy training times.

6. Conclusion and Future Work

In recent years, the number of elderly people has increased, resulting in various repercussions, such as the increasing prevalence of age-related illnesses like Alzheimer's. As a result, researchers have turned their attention toward finding ways to improve the early detection and diagnosis of these diseases. Currently, machine learning and deep learning techniques are being heavily researched due to their potential to accurately identify these conditions. Despite impressive diagnostic results, addressing some critical concerns surrounding deep learning methods, including the suggested framework, is crucial.

Table 5. Comparison of our proposed model with similar studies in the literature

Author (s)	Method	NC vs. AD			NC vs. EMCI			EMCI vs. LMCI			LMCI vs. AD			4–way
		Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc
Mujahid et al. [41]	2D CNN Ensemble	–	–	–	–	–	–	–	–	–	–	–	–	97.35
Raza et al. [42]	CNN from scratch	–	–	–	–	–	–	–	–	–	–	–	–	97.84
Helaly et al.[47]	2D–CNN+3D–CNN	–	–	–	–	–	–	–	–	–	–	–	–	97.0
Li et al. [56]	2D–CNN+3D–CNN	86.01	82.36	90.19		–	–	72.34	60.28	78.35	–	–	–	–
Mehmood et al. [57]	2D–CNN (VGG)	97.52	97.98	98.99	87.36	87.01	87.01	82.26	81.01	82.61	83.02	82.31	83.01	–
Zhang et al. [24]	3D–CNN (ResAttNet)	92.4	90.98	91.29	–	–	–	83.41	82.12	81.27	–	–	–	–
Ruiz et al. [34]	3D–DenseNet Ensemble	–	–	–	–	–	–	–	–	–	–	–	–	83.33
Kang et al. [17]	2D–CNN (VGG)	–	–	–	93.92	97.43	93.02	–	–	–	–	–	–	–
Pan et al. [58]	2D–CNN	83.98	–	–	–	–	–	63.54	–	–	–	–	–	–
Abrol et al. [6]	3D–CNN (ResNet)	92.05	–	–	–	–	–	78.18	–	–	–	–	–	–
Basaia et al. [10]	3D–CNN	97.21	97.98	97.95	77.24	76.17	78.18	74.91	75.08	75.17	76.31	75.21	78.14	–
Gorji and Kaabouch [36]	2D–CNN	–	–	–	94.06	91.26	97.91	92.96	92.04	95.32	–	–	–	–
Yue et al. [59]	2D–CNN	97.96	98.01	–	98.97	98.95	–	97.98	98.02	–	98.97	99.34	–	–
Lu et al. [54]	SAE+DNN	85.08	81.34	92.08	–	–	–	83.02	80.17	84.35	–	–	–	–
Liu et al. [60]	2D–CNN+3D–CNN	94.04	93.06	94.04	65.74	64.21	68.47	–	–	–	–	–	–	–
Islam and Zhang [33]	2D–DenseNet Ensemble	–	–	–	–	–	–	–	–	–	–	–	–	93.18
Shi et al. [61]	Stacked DPN	96.24	96.07	98.67	–	–	–	79.18	69.07	87.56	–	–	–	57
Suk et al. [62]	2D–CNN	92.18	93.08	90.14	–	–	–	75.14	71.14	83.47	–	–	–	–
Ortiz et al. [63]	Deep Belief Network	90.78	86	94	80	60	90	–	–	–	–	–	–	–
Proposed Model	**Ensemble of six 2D–CNNs with WPBDEM**	**98.57**	**98.81**	**98.24**	**98.37**	**99.63**	**94.97**	**98.22**	**94.08**	**95.36**	**99.83**	**99.70**	**100**	**98.78**

Numerous studies on deep learning have used well-known datasets like ADNI and OASIS datasets to assess their frameworks. While these datasets are valuable for research purposes, their practical applicability as automatic CAD systems need further investigation. To overcome the situation, we conducted a thorough evaluation process, we measured the accuracy of trained models using the ADNI dataset as part of a comprehensive validation process. The outcomes showed that individual models perform poorly in real-world situations. However, the diagnostic outcomes of our recommended ensemble method were encouraging. To validate the model's generalizability, it is necessary to conduct additional experiments using a variety of larger datasets.

This study aimed to present a DL framework that utilizes MR images. However, there is potential for adaptation and retraining with different neuroimaging data, such as PET and fMRI data. Additionally, exploring the potential for developing multimodal

and modality-independent frameworks should be considered. Additional research is necessary to assess the efficacy of alternative deep learning methods, including the recently released CNN architectures, in the early identification of Alzheimer's disease.

Acknowledgments: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the editor and anonymous reviewers for their comments, which helped improve the quality of this work.

Funding Statement: The author(s) received no specific funding for this study.

Authorship: The authors confirm contribution to the paper as follows: study conception and design: N.N.; data collection: NAVEEN N; analysis and interpretation of results: Naveen N, Nagaraj J Cholli; draft manuscript preparation: Naveen N; writing-review and editing: Naveen N, Nagaraj J Cholli; visualization: Naveen N; supervision: Nagaraj J Cholli.; All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data utilized for our study were sourced from the ADNI database, which is accessible at http://adni.loni.usc.edu/.

Conflict of interest: The authors declare that they have no conflicts of interest.

References

[1]. Jain R, Jain N, Aggarwal A, Hemanth DJ (2019) Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. Cogn Syst Res 57:147–159. https://doi.org/10.1016/j.cogsys.2018.12.015

[2]. Ramzan F, Khan MUG, Rehmat A, et al (2020) 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 44:1–16. https://doi.org/10.1007/s10916-019-1475-2

[3]. Jin D, Zhou B, Han Y, et al (2020) Generalizable, Reproducible, and Neuroscientifically Interpretable Imaging Biomarkers for Alzheimer's Disease. Adv Sci 7:. https://doi.org/10.1002/ADVS.202000675

[4]. Janghel RR, Rathore YK (2021) Deep Convolution Neural Network Based System for Early Diagnosis of Alzheimer's Disease. IRBM 42:258–267. https://doi.org/10.1016/J.IRBM.2020.06.006

[5]. Sarraf S, Desouza DD, Anderson JAEE, Saverino C (2019) MCADNNet: Recognizing stages of cognitive impairment through efficient convolutional fMRI and MRI neural network topology models. IEEE Access 7:155584–155600. https://doi.org/10.1109/access.2019.2949577

[6]. Abrol A, Bhattarai M, Fedorov A, et al (2020) Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease. J Neurosci Methods 339:. https://doi.org/10.1016/J.JNEUMETH.2020.108701

[7]. Sun H, Wang A, Wang W, Liu C (2021) An improved deep residual network prediction model for the early diagnosis of Alzheimer's disease. Sensors 21:. https://doi.org/10.3390/S21124182

[8]. Gui J, Sun Z, Wen Y, et al (2023) A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. IEEE Trans Knowl Data Eng 35:3313–3332. https://doi.org/10.1109/TKDE.2021.3130191

[9]. Pellegrini E, Ballerini L, Hernandez M del CV, et al (2018) Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. Alzheimer's Dement Diagnosis, Assess Dis Monit 10:519–535. https://doi.org/10.1016/J.DADM.2018.07.004

[10]. Basaia S, Agosta F, Wagner L, et al (2019) Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. NeuroImage Clin 21:101645. https://doi.org/10.1016/j.nicl.2018.101645

[11]. Liu S, Liu S, Cai W, et al (2015) Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease. IEEE Trans Biomed Eng 62:1132–1140. https://doi.org/10.1109/TBME.2014.2372011

[12]. Suk H Il, Shen D (2013) Deep learning-based feature representation for AD/MCI classification. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 8150 LNCS:583–590. https://doi.org/10.1007/978-3-642-40763-5_72/COVER

[13]. Li F, Liu M (2018) Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. Comput Med Imaging Graph 70:101–110. https://doi.org/10.1016/J.COMPMEDIMAG.2018.09.009

[14]. Li Y, Ding W, Wang X, et al (2021) Alzheimer's Disease Classification Model Based on MED-3D Transfer Learning. ACM Int Conf Proceeding Ser 394–398. https://doi.org/10.1145/3500931.3500999

[15]. Cui R, Liu M (2019) Hippocampus analysis by combination of 3-D densenet and shapes for Alzheimer's disease diagnosis. IEEE J Biomed Heal Informatics 23:2099–2107. https://doi.org/10.1109/jbhi.2018.2882392

[16]. Mehmood A, yang S, feng Z, et al (2021) A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images. Neuroscience 460:43–52. https://doi.org/10.1016/J.NEUROSCIENCE.2021.01.002

[17]. Kang L, Jiang J, Huang J, Zhang T (2020) Identifying Early Mild Cognitive Impairment by Multi-Modality MRI-Based Deep Learning. Front Aging Neurosci 12:. https://doi.org/10.3389/FNAGI.2020.00206

[18]. Odusami M, Maskeliūnas R, Damaševičius R, Krilavičius T (2021) 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 11:. https://doi.org/10.3390/DIAGNOSTICS11061071

[19]. Jabason E, Ahmad MO, Swamy MNS (2019) Classification of Alzheimer's Disease from MRI Data Using an Ensemble of Hybrid Deep Convolutional Neural Networks. Midwest Symp Circuits Syst 2019-August:481–484. https://doi.org/10.1109/MWSCAS.2019.8884939

[20]. Shanmugam JV, Duraisamy B, Simon BC, Bhaskaran P (2022) Alzheimer's disease classification using pre-trained deep networks. Biomed Signal Process Control 71:. https://doi.org/10.1016/J.BSPC.2021.103217

[21]. Jabason E, Omair Ahmad M, Swamy MNS (2019) Hybrid Feature Fusion Using RNN and Pre-trained CNN for Classification of Alzheimer's Disease (Poster). FUSION 2019 - 22nd Int Conf Inf Fusion. https://doi.org/10.23919/FUSION43075.2019.9011301

[22]. L S, V S, Ravi V, et al (2023) Deep learning-based approach for multi-stage diagnosis of Alzheimer's disease. Multimed Tools Appl 83:16799–16822. https://doi.org/10.1007/s11042-023-16026-0

[23]. Zhang X, Han L, Zhu W, et al (2022) An Explainable 3D Residual Self-Attention Deep Neural Network for Joint Atrophy Localization and Alzheimer's Disease Diagnosis Using Structural MRI. IEEE J Biomed Heal Informatics 26:5289–5297. https://doi.org/10.1109/JBHI.2021.3066832

[24]. Zhang J, Zheng B, Gao A, et al (2021) A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification. Magn Reson Imaging 78:119–126. https://doi.org/10.1016/J.MRI.2021.02.001

[25]. Zhang Y, Teng Q, Liu YYYY, et al (2022) Diagnosis of Alzheimer's disease based on regional attention with sMRI gray matter slices. J Neurosci Methods 365:109376. https://doi.org/10.1016/j.jneumeth.2021.109376
[26]. Zhang P, Lin S, Qiao J, Tu Y (2021) Diagnosis of alzheimer's disease with ensemble learning classifier and 3D convolutional neural network. Sensors 21:. https://doi.org/10.3390/S21227634
[27]. Ji H, Liu Z, Yan WQ, Klette R (2020) Early Diagnosis of Alzheimer's Disease Based on Selective Kernel Network with Spatial Attention. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 12047 LNCS:503–515. https://doi.org/10.1007/978-3-030-41299-9_39
[28]. Guan H, Wang C, Cheng J, et al (2022) A parallel attention-augmented bilinear network for early magnetic resonance imaging-based diagnosis of Alzheimer's disease. Hum Brain Mapp 43:760–772. https://doi.org/10.1002/hbm.25685
[29]. Liu Z, Lu H, Pan X, et al (2022) Diagnosis of Alzheimer's disease via an attention-based multi-scale convolutional neural network. Knowledge-Based Syst 238:. https://doi.org/10.1016/J.KNOSYS.2021.107942
[30]. Yu X, Peng B, Shi J, et al (2019) 3D Convolutional Networks Based Automatic Diagnosis of Alzheimer's Disease Using Structural MRI. Proc - 2019 12th Int Congr Image Signal Process Biomed Eng Informatics, CISP-BMEI 2019. https://doi.org/10.1109/CISP-BMEI48845.2019.8965827
[31]. Li F, Liu M (2019) A hybrid Convolutional and Recurrent Neural Network for Hippocampus Analysis in Alzheimer's Disease. J Neurosci Methods 323:108–118. https://doi.org/10.1016/J.JNEUMETH.2019.05.006
[32]. Liu M, Li F, Yan H, et al (2020) A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. Neuroimage 208:. https://doi.org/10.1016/J.NEUROIMAGE.2019.116459
[33]. Islam J, Zhang Y (2018) Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. Brain Informatics 5:1–14. https://doi.org/10.1186/S40708-018-0080-3/FIGURES/10
[34]. Ruiz J, Mahmud M, Modasshir M, Shamim Kaiser M (2020) 3D DenseNet Ensemble in 4-Way Classification of Alzheimer's Disease. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 12241 LNAI:85–96. https://doi.org/10.1007/978-3-030-59277-6_8
[35]. Wang H, Shen Y, Wang S, et al (2019) Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease. Neurocomputing 333:145–156. https://doi.org/10.1016/J.NEUCOM.2018.12.018
[36]. 36. Gorji HT, Kaabouch N (2019) A deep learning approach for diagnosis of mild cognitive impairment based on MRI images. Brain Sci 9:1–14. https://doi.org/10.3390/brainsci9090217
[37]. Pan D, Zeng A, Jia L, et al (2020) Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning. Front Neurosci 14:501050. https://doi.org/10.3389/FNINS.2020.00259/BIBTEX
[38]. Odusami M, Maskeliūnas R, Damaševičius R (2022) An Intelligent System for Early Recognition of Alzheimer’s Disease Using Neuroimaging. Sensors 2022, Vol 22, Page 740 22:740. https://doi.org/10.3390/S22030740
[39]. Ma P, Wang J, Zhou Z, et al (2023) Development and validation of a deep-broad ensemble model for early detection of Alzheimer's disease. Front Neurosci 17:. https://doi.org/10.3389/fnins.2023.1137557
[40]. Tanveer M, Rashid AH, Ganaie MA, et al (2022) Classification of Alzheimer's Disease Using Ensemble of Deep Neural Networks Trained Through Transfer Learning. IEEE J Biomed Heal Informatics 26:1453–1463. https://doi.org/10.1109/JBHI.2021.3083274
[41]. Mujahid M, Rehman A, Alam T, et al (2023) An Efficient Ensemble Approach for Alzheimer's Disease Detection Using an Adaptive Synthetic Technique and Deep Learning. Diagnostics 2023, Vol 13, Page 2489 13:2489. https://doi.org/10.3390/DIAGNOSTICS13152489
[42]. Raza N, Naseer A, Tamoor M, Zafar K (2023) Alzheimer Disease Classification through Transfer Learning Approach. Diagnostics 13:
[43]. Savaş S (2022) Detecting the Stages of Alzheimer's Disease with Pre-trained Deep Learning Architectures. Arab J Sci Eng 47:2201–2218. https://doi.org/10.1007/s13369-021-06131-3
[44]. Rajesh Khanna M (2023) Multi-level classification of Alzheimer disease using DCNN and ensemble deep learning techniques. Signal, Image Video Process 17:3603–3611. https://doi.org/10.1007/S11760-023-02586-Z
[45]. Fathi S, Ahmadi A, Dehnad A, et al (2023) A Deep Learning-Based Ensemble Method for Early Diagnosis of Alzheimer's Disease using MRI Images. Neuroinformatics 22:89–105. https://doi.org/10.1007/S12021-023-09646-2/TABLES/5
[46]. Al-Adhaileh MH (2022) Diagnosis and classification of Alzheimer's disease by using a convolution neural network algorithm. Soft Comput 26:7751–7762. https://doi.org/10.1007/s00500-022-06762-0
[47]. Helaly HA, Badawy M, Haikal AY (2022) Deep Learning Approach for Early Detection of Alzheimer's Disease. Cognit Comput 14:1711–1727. https://doi.org/10.1007/s12559-021-09946-2
[48]. Antony F, Anita HB, George JA (2023) Classification on Alzheimer's Disease MRI Images with VGG-16 and VGG-19. Smart Innov Syst Technol 312:199–207
[49]. AbdulAzeem Y, Bahgat WM, Badawy M (2021) A CNN based framework for classification of Alzheimer's disease. 33:10415–10428
[50]. Hazarika RA, Maji AK, Kandar D, et al (2023) An Approach for Classification of Alzheimer's Disease Using Deep Neural Network and Brain Magnetic Resonance Imaging (MRI). Electron 12:676
[51]. Khan R, Akbar S, Mehmood A, et al (2022) A transfer learning approach for multiclass classification of Alzheimer's disease using MRI images. Front Neurosci 16:. https://doi.org/10.3389/FNINS.2022.1050777
[52]. Ebrahimighahnavieh MA, Luo S, Chiong R (2020) Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. Comput Methods Programs Biomed 187:105242. https://doi.org/10.1016/j.cmpb.2019.105242
[53]. Yue L, Gong X, Li J, et al (2019) Hierarchical feature extraction for early Alzheimer's disease diagnosis. IEEE Access 7:93752–93760. https://doi.org/10.1109/access.2019.2926288
[54]. Lu D, Popuri K, Ding GW, et al (2018) Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images. Sci Rep 8:. https://doi.org/10.1038/S41598-018-22871-Z
[55]. Zheng C, Xia Y, Chen Y, et al (2018) Early Diagnosis of Alzheimer's Disease by Ensemble Deep Learning Using FDG-PET. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 11266 LNCS:614–622. https://doi.org/10.1007/978-3-030-02698-1_53

[56]. Li A, Li F, Elahifasaee F, et al (2021) Hippocampal shape and asymmetry analysis by cascaded convolutional neural networks for Alzheimer's disease diagnosis. Brain Imaging Behav 15:2330–2339. https://doi.org/10.1007/s11682-020-00427-y

[57]. Mehmood A, Abugabah A, AlZubi AA, Sanzogni L (2022) Early Diagnosis of Alzheimer's Disease Based on Convolutional Neural Networks. Comput Syst Sci Eng 43:305–315. https://doi.org/10.32604/csse.2022.018520

[58]. 58. Pan D, Zeng A, Jia L, et al (2020) Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning. Front Neurosci 14:501050

[59]. Yue L, Gong X, Li J, et al (2019) Hierarchical feature extraction for early Alzheimer's disease diagnosis. IEEE Access 7:93752–93760. https://doi.org/10.1109/ACCESS.2019.2926288

[60]. Liu M, Cheng D, Wang K, Wang Y (2018) Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis. Neuroinformatics 16:295–308. https://doi.org/10.1007/S12021-018-9370-4

[61]. Shi J, Zheng X, Li Y, et al (2018) Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. IEEE J Biomed Heal Informatics 22:173–183. https://doi.org/10.1109/jbhi.2017.2655720

[62]. Suk H Il, Lee SW, Shen D (2017) Deep ensemble learning of sparse regression models for brain disease diagnosis. Med Image Anal 37:101–113. https://doi.org/10.1016/j.media.2017.01.008

[63]. Ortiz A, Munilla J, Górriz JM, Ramírez J (2016) Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. Int J Neural Syst 26:1650023–1650025. https://doi.org/10.1142/s0129065716500258