Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Alzheimer's disease classification using deep learning: A novel small-data approach.

RAJA HASEEB [1†], TAE-MIN CHOI[2], AND JONG-HWAN KIM, JR.[3]
[1]Korea Advanced Institute of Science and Technology, Daejeon, South Korea (rajahaseeb147@kaist.ac.kr)
[2]Korea Advanced Institute of Science and Technology, Daejeon, South Korea (e-mail: author@lamar.colostate.edu)
[3]Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA
[†]These authors contributed equally to this work

Corresponding author: Tae-min Choi (e-mail: author@ boulder.nist.gov).

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

ABSTRACT Alzheimer's disease (AD) is a progressive neurodegenerative disease that causes cognitive impairment among elder people. It is the most common cause of dementia. Early and timely diagnosis of AD can help with the overall treatment process of the patients. Classifying AD patients from mild cognitive impairment (MCI) and cognitive normal (CN) patients is very important for AD diagnosis. Recently use of machine learning techniques and deep learning-based algorithms have become a popular choice for the AD classification task. However, most of the existing approaches depend upon large data sets or suffer from data leakage. In this work, we present a novel framework for AD classification in a small data regime. Our approach is based on three main steps: 1) PGGAN based medical image generation to deal with data scarcity; 2) Model pretraining using SimCLR framework; 3) Train the pretrained model for the final classification task. We used ResNet-18 in combination with CBAM (Convolutional Block Attention Module). The CBAM module enhances the useful features in the images, which helps in better training of the model. We performed a clinical evaluation of our model on novel test data set. We achieved an accuracy of 83% for the AD vs. CN classification task using only a few slices for the training process. We also compared our model with the previous approaches and found our results to be comparable even with such small data. Results indicate the effectiveness of our proposed framework and can help in the early diagnosis of AD patients.

INDEX TERMS Alzheimer's disease classification, Alzheimer's disease diagnosis, convolutional neural network (CNN), machine learning, deep learning, contrastive learning, CBAM, generative adversarial network (GAN).

I. INTRODUCTION

Alzheimer's disease (AD), also referred to simply as Alzheimer's, is a progressive neurodegenerative disease. It is the most common cause of dementia in the world, affecting 1 out of 9 people over the age of 65 years [1]. In 2018, 5.7 million Americans were reported to have AD. This number can rise to 14 million by the year 2050 [1]. AD causes a loss of connection between nerve cells in elders. AD results in brain shrinkage, cause enlargement of brain ventricles and decrease the size of the hippocampus. Alzheimer's disease progressively causes cognitive impairment, commonly associated with early memory loss, affecting a person's ability to think and perform common day-to-day tasks. The available treatments for AD at the present can only slow the progression at the best, and there is no validated cure for the patients that already have AD. Therefore, early diagnosis of AD is crucial for the timely treatment of patients and can help abate the progression of the disease. Based on this, the concept of Mild Cognitive Impairment (MCI) was introduced as the prodromal form of AD [2]. It describes people having mild symptoms of brain impairment. The MCI patients are still able to perform daily activities up to some extent. However, their ability to do so declines with time as the disease progresses and the patients in this phase have high

chances of progressing into dementia [1], [3], [4], [5]. As a result, patients are classified into three main categories i.e. AD class representing the patients with the disease, Cognitive Normal (CN) people which have normal brain function, and the MCI patients which represents the stage in between the CN and AD patients.

Consequently, physicians have been trying to apply various clinical methodologies to classify and diagnose AD patients. Cerebrospinal fluid (CSF) concentration in the brain is reported to indicate the presence of AD. As the disease progresses, the level of norepinephrine increase in the CSF. A ventricular puncture is used for the collection of CSF; the physician collects CSF directly from one of the brain ventricles by making a hole in the skull [6]. This process can be arduous and can cause bleeding in the brain. A lot of focus has been put on the development of medical imaging techniques in recent years. Neuroimaging techniques like Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Functional MRI, and Computerized Tomography (CT) are widely used to diagnose functional and structural changes in the brain. MRI modality is commonly used for the diagnosis of AD, as it is easily accessible and provides good results. MRI sequences are usually T1 or T2-weighted scans. It is evident from research that the progression from CN to MCI to AD results in certain changes in the brain structure. A healthy brain can be distinguished from a diseased brain by using morphological changes in the brain structure, volume, and texture via medical imaging [7], [8]. Changes in hippocampus size can be an early indicator of AD and are useful for AD classification [9]. The area of the brain affected by AD can be estimated by observing changes in the white matter (WM) [10]. Gray matter (GM) can help in analyzing AD [11]. However, all these methods for diagnosis require specialists to analyze the neuroimaging sequences and make a decision. This process can be costly, laborious, time-consuming, and sometimes even prone to human errors. Therefore, there is a need for an automated way for the classification of AD patients which takes less time and effort, is reliable, less costly. Such a process can help the practitioners speed up the diagnosis process and treat the patients at an early stage.

Deep learning [12], [13], [14], [15], [16] is the latest paradigm in machine learning which involves high-level abstraction of features and has shown tremendous success across various domains and applications like classification [17], [18], [19], [20], [21], object detection [22], [23], object tracking [24], depth estimation [25], [26], semantic and instance segmentation [27], [28], [29], [30], speech recognition [31], [32], natural language processing (NLP) [33], [34], [35] and so on. Numerous deep learning architectures have been developed in recent years including convolutional neural networks (CNNs) [36], [17], recurrent neural networks (RNNs) [37], recursive neural networks [38], auto-encoders [39] and Generative Adversarial Networks (GANs) [40]. Such discoveries aroused the interest of data scientists and researchers in the field of machine learning and also adapting

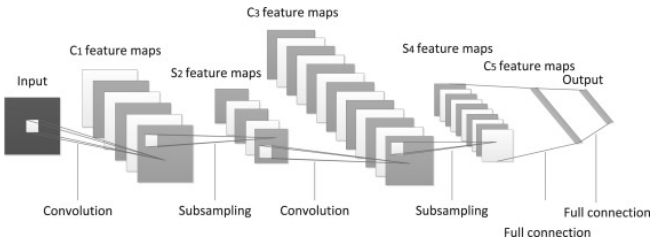

FIGURE 1. A schematic diagram of standard CNN [16].

it to the medical domain. Deep learning caught the eye of many healthcare practitioners, and it did not take long for researchers to realize yet another opportunity for the application of deep learning. The healthcare sector is not small anymore. Due to the increase in the number of patients and medical devices, the amount of medical data is quite large. Therefore, it makes medical diagnosis and analysis of the patient result more challenging. Such rapid growth of medical data requires tedious efforts by doctors and medical experts and are usually susceptible to human error, and can have various interpretations among different experts. There is a need for efficient and effective tools and methods to help speed up the diagnosis process by proper management and processing of this large data while maintaining high levels of accuracy and performance [41].

Soon enough deep neural networks found their application in the medical and healthcare sector for diagnosis of various diseases [42]. This lead to the use of deep learning methods for the recognition of Alzheimer's Disease (AD). Many scientists and researchers started applying deep learning models to medical biomarkers and imaging data [43], [44], [45]. As expected, deep learning showed high performance in AD diagnosis like every other field. Modern deep learning algorithms like stacked auto-encoders (SAE), residual networks, and Deep Boltzmann Machine (DBM) outperformed previously used traditional machine learning algorithms like SVMs and random forest.

However, despite this tremendous success of deep learning techniques across multiple domains, one important issue still needs to be addressed, and researchers and data scientists are making continuous efforts to address that issue i.e. the need for a large training data set for deep learning models. Although deep learning models show promising results with a very high level of accuracy, this requires a very careful training process along with a large number of training examples. It is a common belief among all the researchers that the more data you feed into your network, the better the performance will be. So you need a sufficiently large volume of data to train a successful machine learning algorithm [14], [46]. Data with high-class imbalance or insufficient data can lead to the poor performance of the algorithm [47]. In the case of common real-world images, this is not a problem as we have enough training data available from various resources. However, this issue becomes more prominent when we talk about domains like medical and healthcare.

Despite the rapid growth in medical records and data, there are many cases when training data is not sufficient. One main reason is that patient data is well protected by the patient data laws. Also, it is expensive and time-consuming to obtain more medical data especially annotated ones since you need a medical expert for that purpose. The imaging and data retrieval standards also differ from country to country and even from one hospital to another, which makes the process even more complicated. In recent years, some big hospitals and authorities made some medical data available to the public in an anonymous way to further drive deep learning research. However, still, this data is not big enough to properly train the current deep learning models. Also, the associated annotation varies considerably due to differences in data acquisition systems and different annotators. Therefore, in clinical settings, usually small data sets are obtained. Creating large data sets is usually time-consuming and might not be feasible especially in the case of some rare diseases or new imaging modalities [48]. For example, in the case of the recent COVID-19 infection, the training data is still an issue for many researchers [49] and practitioners are in continuous efforts to obtain more data through clinical assessments, which is going to take quite some time because the infection is novel and trials are still in early stages.

Recently there have been several attempts to work with small medical data. Researchers have employed methods used in other domains in the healthcare sector. Such work can help mitigate the need for huge medical data required for training the algorithm and can help reduce the time and cost of the annotation process. This research presents a novel approach to deal with medical data scarcity for the case of Alzheimer's Disease (AD) diagnosis and classification. A Progressive GAN (PGGAN) [50] based data augmentation approach is combined with a recent simple framework for contrastive learning of visual representations (SimCLR) [51] and Convolutional block attention module (Cbam) [52] to achieve better results on AD vs. CN vs. MCI classification using small data set. The details of the algorithm and methodology are explained in the later sections.

II. RELATED WORD

In the recent few years, computer-aided diagnosis (CAD) has been utilized to assist physicians. Many researchers and data scientists have been putting efforts to develop CAD systems to diagnose AD. Most of the practitioners rely on Mini-Mental State Examination and other physical tests to verify the stage of AD [7] [53]. Clinically, AD classification is done by developing different biomarkers and by collecting different parameters to test the AD stage [54]. Machine learning has been deployed in recent CAD systems to examine and analyze the pattern in the medical data. Machine learning has shown promising results with better accuracies based on the features extracted from single and multi-modal brain images [55].

There have been various attempts to classify CN, AD, and MCI patients using image data, cognitive features, and other medical biomarkers. A random forest-based approach is used for AD classification using CSF measurement, genetic information, shape, voxel intensity and brain volume as feature values as features to the network [56]. The dimensionality of these features is reduced using PCA and classification is performed using particle swarm optimization (PSO) and support vector machine (SVM) [57]. Brain tissues like WM, GM, and hippocampus are affected as AD progresses. The learning vector quantization approach was used on brain MRI images to segment WM and GM. then SVM was used to perform classification. Texture changes in GM and WM are also a good indicator to distinguish AD, MCI, and CN patients. First-order statistical features are extracted from the histogram to measure texture changes. Then GLCM and Gabor filters are used to extract second-order statistical parameters [58]. These extracted features are then used for patient classification using SVM [59] and KNN [58].

Hybrid features generated by the combination of volume and texture information such as GM volume along with texture features are used to perform AD classification using SVM-Random Fourier Expression (SVM-FRE) [60]. A combination of features extracted from clinical data and segmented brain images is used for multi-class classification of AD from CN and MCI [61]. The classification accuracy usually increases with the number of features. However, this makes the overall training process more complicated and less accurate. To select important features that have more discriminative properties, greedy score, and kernel-based discriminative method is used [62]. According to Fan *et al.* [63] positron emission tomography (PET) scans provide better features and classification accuracy compared to structural MRI (sMRI). This fact was supported by Dukart *et al.* [64] and it was reported that compared to sMRI, fluorodeoxyglucose-PET (FDG-PET) features are more discriminative. Better accuracy results are reported in [65] for AD vs. CN classification with PET images (100%) instead of single-photon emission computed tomography (SPECT) images (97.5%). A similar finding is reported in [66] for CN vs. AD where SPECT image accuracy (94.5%) was lower compared to PET images (96.67%). A novel automatic classification and tissue segmentation method based on Independent Component Analysis (ICA) and SVM classifier was proposed in [67] where the input to the network is a combination of T1, T2, and proton density (PD) scans. Hojatti *et al.* [68] utilized resting-state functional MRI (rs-fMRI) to classify MCI converters (MCIc) and MCI non-converters (MCInc) by observing changes in brain connectivity. Connectivity information from fMRI data was used in [69].

Vemuri *et al.* [70] observed improved classification for CN vs. AD by combining sMRI scan features with genetic and demographic information of the patients. To detect subtle changes in the gray matter (WM), a refined parcellation method is proposed in [66]. Histogram of regions of interests (ROIs) was used for feature selection to classify AD and CN patients by Magnin et al. [71]. Shape deformation features of the hippocampus region were found to be better

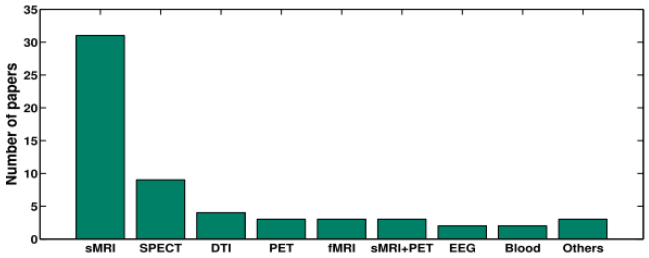

FIGURE 2. Plot showing data and image modalities used with SVM for Alzheimer's in various research [89].

than volumetric features to discriminate AD, MCI, and CN patients [72]. To extract useful voxels as features (VAF) from PECT images, fisher discriminate ratio (FDR) is used [73]. To discriminate early AD from CN, normalized mean square error (NMSE) features are used as input to the SVM classifier [74]. Adjacent voxels are clustered together for classification of AD, MCI, and CN patients [75]. To make data linearly separable, a gaussian mixture model (GMM)-based feature extraction method is proposed in [66]. A comparison of principal component analysis (PCA) and non-negative matrix factorization (NMF)-based features with SVM is made in [76] to classify AD patients from CN. It is reported that NMF-based features give better performance compared to PCA. Oritz *et al.* [77] used sparse inverse covariance estimation (SICE) with SVM to extract most discriminative features from sMRI and PET data. Volume-based features are found to be more useful compared to voxel-based morphometry (VBM) [78] for AD vs. CN. Plocharsky *et al.* used morphological features of brain regions to perform AD vs. CN classification [79]. Tangaro *et al.* [80] used a fuzzy-based classes approach for hippocampus volume to classify MCIc vs. MCIn and AD vs. CN.

Various works used wavelet-based features. Zhang *et al.* found 3D discrete wavelet transform (DWT) and SVM to be useful for classification of AD, MCI and CN patients [81]. DWT features are used by Chaplot et al. [82]. Unsupervised segmentation of sMRI brain images is performed by Oritz *et al.* [83] via self-organizing maps (SOMs) for classification of AD vs. CN. However, the results of the SVM-based approach using T2-weighted images are reported to be better than the SOM-based method [82]. SVM in combination with ICA is used in various works for AD vs. CN classification [84] [85]. SVM-based classifier with EEG recordings is also used for AD vs. CN classification [86]. Word comprehension activity was used to take EEG readings of patients for the classification of MCIc, MCInc and CN patients [87]. A linear SVM is used to classify patients based on speech patterns data of subjects [88].

To classify CN with pathologically confirmed cases of AD, a linear SVM is used by Kloppel *et al.* [11]. Multi-class classification of AD, CN, and MCI patients is performed via polynomial kernel [90]. PCA features with the polynomial kernel are found to be useful in AD diagnosis [81]. Lahmiri *et al.* [91] used cognitive test scores and volumetric features with a polynomial kernel to classify AD vs. CN patients. For high dimensional data, it is reported that the linear kernel outperforms RBF and polynomial kernels in terms of classification accuracy [84] [73].

Kernel ensembles are also adopted by some researchers. Alam *et al.* [92] used multiple kernel SVM for AD vs. MCI vs. CN classification. A combination of RBF, linear and polynomial kernel is used by Kamath *et al.* [67]. A multi-modal (features from MRI and genetic data) and multiple SVM kernel [93] approach is presented for the classification of subjects. SPECT images of Alzheimer's and normal subjects are used as input to a contiguous SVM (CSVM) classifier for more spatial information [94]. CSVM is reported to build a more robust classifier by using the voxel connectivity information. A twin support vector machine (TWSVM)-based approach is used by Zhang *et al.* for AD vs. CN classification [81], while Xu *et al.* used structural least squares twin support vector machine (S-LSTSVM). Linear SVM in combination with genetic algorithm (GA) is used for AD vs. CN and progressive MCI (pMCI) vs. stable MCI (sMCI) classification [95].

A temporally structured SVM (TS-SVM)-based approach is present for early diagnosis of AD cite. MCI converters and non-converters are classified based on the longitudinal MR images. A random forest robust SVM (RF-RSVM) is proposed by Lu *et al.* for MCI vs. CN classification using FDG-PET images. A combination of LDA, PCA, and dual-tree complex wavelet transform (DTCWT) is used to classify AD vs. CN with TWSVM [96]. A switching delayed particle swarm optimization SVM (SDPSO-SVM) is proposed for optimization of SVM parameters for AD classification [97]. Features from rs-fMRI are used for Ad vs. CN classification using random SVM clusters []. This reduces the complexity of the SVM algorithm [98]. In this approach, training is done on randomly chosen features and samples from the data set, resulting in less computational complexity. Improved prediction accuracy is reported in various research for AD classification using ensemble-based SVM [65] [99] [100].

Artificial neural networks (ANNs) can model complex patterns of non-linear data and are highly used for machine learning (ML) applications. Many researchers adopted ANNs for AD classification. A standard ANN is shown in Figure 3. EEG signals are used to classify Ad vs. CN with ANN [101] [102]. A comparison between different variants of ANN by Savio *et al.* [103] and they found that learning vector quantization neural network (LVQNN) shows best results. Long et al. [104] utilized an unsupervised learning approach for AD vs. CN classification. A comparison is made between various algorithms like random forest (RF), decision tree, bagging, multi-layer perceptron (MLP) and ANN for AD classification [105]. Mahmood *et al.* used ANN in combination with PCA to reduce the complexity of high dimensional data [106]. ANN is used for AD vs. CN classification [107]. A bidirectional gated recurrent neural network (BGRU) is cascaded with an MLP for AD classification [108]. AD subjects are classified based on sequential auditory data by

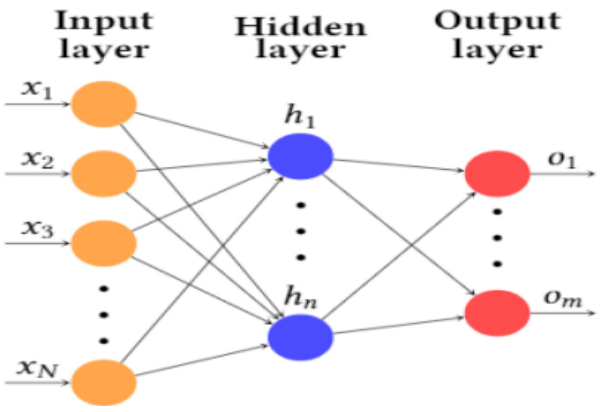

FIGURE 3. Standard Neural Network Architecture [89].

using an LSTM (long short-term memory) network [109]. To discriminate AD subjects from CN, Kar *et al.* [110] used an ANN with a fuzzy-based approach on features from DTI images.

All these approaches require some kind of feature selection (FS) method to extract meaningful features from medical data. To extract features from EEG signals, visibility graphs (VG) are used [111]. To classify subjects, Rodrigues *et al.* utilized WT and SIFT (short-time Fourier transform) feature to train an ANN. As mentioned earlier, AD results in tissue loss in WM and GM of the brain. This causes shrinkage in the volume of WM and GM. Also, the CSF volume increase, and ventricles enlarge. A classification framework is proposed by Yang *et al.* [112]that combines CSF volumetric features and shape features of WM, GM, and ventricles. Cortical features in MR images also show signs of AD. Cortical thickness data is used to train a robust incremental classifier [113]. To classify MCI and AD subjects, data from neuropsychological tests, education, and age is used [114]. A combination of shape and volumetric features is used by Yang *et al.* [57]. A feed-forward neural network (FFNN) is used along with a hybrid of an artificial bee colony (ABC) and PSO [115].

Deep learning has become very popular in recent years across various domains including the medical sector as well. Deep neural networks have more hidden layers which make it possible for the network to learn high dimensional non-linear relations in the data. A deep convolutional neural network (DCNN) contains hierarchical layers which are connected and operate directly on the image data to extract feature maps. This eliminates the need for hardcore feature engineering and the network automatically extracts features. This is the main advantage of deep learning over traditional approaches. Due to its tremendous success across different domains, deep learning also found its way into the medical and healthcare sector.

In medical image analysis, deep learning (DL) has shown good results and success [116]. Suk *et al.* reported improved accuracy for AD vs. MCI classification by applying a stacked auto-encoder (SAE) [117]. A deep Boltzmann machine (DBM) is used for AD vs. MCI vs. CN classification using patches from PET and MRI data [45]. To perform AD classification, an entropy-based convolutional neural network is used to extract radiomic features from 3D MRI images [118]. A multi-scale deep neural network is proposed based on multi-modal data from FDG-PET and MRI images by Lu *et al.* [119]. Segmented gray matter is divided into patches which are then used for feature extraction. Diagnosis of AD and its stages is performed using a deep CNN on MRI scans [120]. 3D-CNN is reported to perform better than SAE on AD classification task [121]. To classify CN, AD, MCI patients, an SAE-based classifier is used on top of DL-based latent features [122]. A multi-modal deep learning architecture based on SAE is presented for the AD diagnosis [43]. To classify AD, MCI, and CN subjects, Hosseini *et al.* [123] presented a deeply supervised adaptive 3D-CNN (DSA-3DCNN) using MRI data. The network was pretrained by a 3D convolutional auto-encoder (3D-CAE). High classification accuracy is obtained by training a DBN-SVM classifier on extracted 3D patches [124].

Multiple GRU architectures are used to extract features from multi-model patient data like MRI, cognitive performance, CSF volumes, and demographic information [125]. Segmented GM voxels from T2-weighted MRI are enhanced via hybrid enhanced independent component analysis [2]. A CNN is then used on these voxels to classify AD, MCI, and CN patients. FreeSurfer [126] is used to compute various features including cortical thickness measurements from segment gray matter [127]. Various ML algorithms are then trained with non-linear SVM showing the best performance results. A deep CNN model is combined with sparse regression models known as deep ensemble sparse regression network (DeepESRNet) for AD diagnosis [128]. A deep learning model was trained on fMRI data for the first time by Sarraf *et al.* [129]. A combination of deep CNN and sparse regression networks is used to classify AD patients [130]. A multi-modal stacked deep polynomial networks (MM-SDPN) algorithm is introduced by Shi *et al.* [131]. Data from multiple modalities i.e. PET and MRI are used for feature extraction. Bilinear shake fusion is used to combine features extracted from a CNN train on MRI images, and a fully-connected (FC) network trained on clinical data (APOE genotype, cognitive scores, age, and gender) [132]. ROI features like cortical thickness, volume, and surface areas are extracted from DTI, PET, and MRI images using the FreeSurfer tool to predict AD using an LSTM network [133]. Small-sized patches of the hippocampus region were extracted from structural MRI (sMRI) images. Patches from sagittal, axial, and coronal views were combined and fed into CNN ensembles [134].

Basaia *et al.* utilized T1-weighted 3D sMRI data to build a DL network that was able to differentiate AD, CN and MCI patients [135]. EEG is used to calculate power spectral density (PSD) of brain states which is then used to train a DL network [136] on PSD spectrograms to classify dementia stages. FDG-PET data is used to train a multi-scale

deep neural network (MDNN) [137]. The network showed good discriminative properties for AD diagnosis based on data from a single modality. Multiple cluster dense CNNs (DenseNets) are used by Li *et al.*. 3D patches were extracted from T1-weighted sMRI images and K-means clustering was used to cluster the patches [138]. This process eliminated the need for rigid registration and segmentation of MRI data for AD diagnosis via DenseNets. A VGG16 [139] network pretrained on Imagenet is used in a transfer learning approach to classify AD patients with sMRI slices [140]. Khan *et al.* presented another transfer learning approach for AD prediction based on VGG19 architecture [141]. Most informative slices were selected from the patient's MRI data based on the entropy values. To handle class imbalance, a data augmentation approach is presented [142]. Training is done on main MRI view and 3D views by using AlexNet architecture pretrained on Imagenet data. To make parameter selection more efficient, grouped and 3D separable convolutions based deep network are used with T1-weighted sMRI images to extract descriptive features [143] for AD classifications. To enhance performance for AD vs MCI diagnosis on sMRI images and reduce overfitting, Whang et al. [144] proposed an ensemble of 3D-DenseNets.

A. CONTRIBUTION OF THIS RESEARCH

In this work, we present a novel approach to classify AD patients and to deal with medical data scarcity. The key contributions of this work are listed below.

- **A small-data approach.** Dealing with medical data scarcity specifically for the case of AD diagnosis. We use a small data set approach and achieve good results using only a few MRI slices.
- **Dealing with class imbalance.**ALso removed imbalance among classes usign data augmentation.
- **Novel medical image synthesis.**We present a novel PGGAN-based approach for medical image synthesis in combination with traditional data augmentation.
- **Applciation of SimCLR to AD classification.** We present application of SimCLR [51] and CBAM [52] to the case of AD classification for the first time.
- **Unbiased Evaluation.** To correctly evaluate our model without data leakage, we prepared separate training, validation, and test sets on subject-level. This results in effective hyper-parameter optimization and evaluation of our approach.
- **An analysis of various approaches.** We also present an analysis of some previous approaches and different architectures. Results indicate the presence of data leakage and overfitting in these different approaches.

This research aims to design an automated model with improved performance for AD diagnosis specifically for the small data set settings. Although there have been some approaches to deal with medical data scarcity. However, none of them was focused on the AD classification task, specifically using GAN-based image synthesis. Contrastive representation learning is still novel in the medical sector. There hasn't been any work based on contrastive learning for the classification of Alzheimer's Disease (AD).

III. PROPOSED METHOD

The section describes the overall pipeline and architecture of our proposed framework.

A. DATA PREPARATION

The first setup was to select and prepare the data set. We collected T2-weighted MRI scans from the ADNI (Alzheimer's Disease Neuroimaging Initiative) website. Data for 246 patients (82 AD, 82 MCI, and 82 CN) is used for training. Additional data from 54 patients were kept separate for testing purposes. This test data is further divided into two parts i.e. validation and test data. To clear evaluate our model, validation data is selected from test data instead of training data. Validation data consisted of 115 slices (51 for AD, 64 for CN) and test data consisted of slices from 10 patients for each category. Demographic information of the training data is shown in Table 1.

In our case, we only used axial view slices and the dimensions of each slice are 256x256. Initially, each patient had around 48 slices. Manual data cleaning was performed and only good slices with visible brain regions were selected. The slices with no or less information were removed from the data.

B. DATA AUGMENTATION

In this research, we presented a small-data approach. One common problem faced when training with limited data is the overfitting. To deal with this, we use a novel GAN architecture for image synthesis. However, training a GAN also requires a large amount of data, which brings us to traditional data augmentation techniques. We performed traditional data augmentation on our data including small rotations, shear, zoom, width, and height shift. The resulting number of slices is 2435, 2429, and 2389 for AD, CN, and MCI respectively.

C. GAN-BASED IMAGE GENERATION

After the data augmentation, we generated whole-brain MRI images using GAN. There are various GAN architectures available. We decided to implement three architectures i.e. DCGAN [145], RaLSGAN [146] and PGGAN [147]. DCGAN is hard-coded for 64x64 images. Since we wanted to keep our image size 256x256, we tried to train DCGAN on 256x256 images. However, the training of the network is not stable for higher resolutions.

A comparison was made between RaLSGAN and PGGAN for medical image synthesis based on the FID (Fréchet Inception Distance) score. The FID score is a measure based on the similarity between real and generated images and is commonly used to evaluate GAN models. The lower the FID score, the better the model. In our case, PGGAN outperformed RaLSGAN for image synthesis. Moreover, Progressive Growing of GANs (PGGANs) with the Wasserstein

TABLE 1. Demographic representation of training MRI images

Data source	Research group	Total subjects	Sex		Age (years)	Total MRI volumes	Image slices	Imaging protocol
			M	F				
ADNI	AD	82	46	44	55-93	82	406	Axial,
	CN	82	46	46	70-96	82	347	2D,
	MCI	82	46	46	58-96	82	279	1.5-Tesla

TABLE 2. Total training data after PGGAN-based image generation

Data source	Research group	MRI volumes		
		Before Augmentation	After Augmentation	After PGGAN
ADNI	AD	406	2435	4870
	CN	347	2429	4870
	MCI	279	2389	4870

loss [148] using gradient penalty generated very realistic and high-resolution images. The FID score for PGGAN-WP for our case is around 40 while for RaLSGAN it is around 140. A comparison of generated images by the two models is shown in Figure 4. The final training data set consisted of 4870 slices for each category. Table 2 shows the number of training samples after PGGAN-based image generation.

D. COMPARISON OF VARIOUS ARCHITECTURES

As a first step, a detailed analysis was made on the AD vs. CN classification accuracy using different architectures. These architectures include Custom CNN architecture, VGG16 [139], inception model [149], residual attention network [150], CBAM [52] architectures and Multi-scale CNN. We also evaluate pretrained models on ImageNet data. Since the validation data was unseen and novel, all the models were evaluated based on the validation accuracy. Table 3 shows a comparison of various architectures. Testing results are reported based on the majority voting decision for the patients.

All these architectures showed higher accuracy during training. However, the validation accuracy is not very high. Models were also used to predict the patients using the test data. The testing accuracy of all the models was nearly the same as the validation accuracy. Majority voting was used to make and categorize the patients. If a patient has more slices with AD prediction, then the patient was classified as AD. Otherwise, the patient belongs to CN. The best performance was achieved in some cases with 2-3 miss-classifications per category.

Based on these results, we can conclude that these models are not very promising for small medical data classification. As mentioned earlier, transfer learning from real-world data is also not very useful in the case of medical data. Most of the models seem to overfit due to the small size of training data. Raghu *et al.* evaluated transfer learning for small medical data regime. It has been reported that for medical image tasks, transfer learning does not significantly affect the model performance [151]. Models trained from scratch showed comparable performance to the standard ImageNet architectures. The reason can be that these large ImageNet models might be over-parametrized for very small data sets. Therefore, there is a need for a better model and approach that can achieve robust performance for small medical data.

E. PROPOSED ARCHITECTURE

In this research, we propose a novel method for AD classification in small data regime. Proposed methodology consists of three main parts.

- PGGAN-based image generation
- Model pre-training with SimCLR
- Training the pre-trainined model for classification task

We used ResNet-18 [152] with CBAM [52] as our base model. After the PGGAN-based image generation, the base model is first pretrained via SimCLR [51] to learn discriminative representations. The new train this final model again for the classification task. Details and working of the above architectures are mentioned in the next sections.

1) PGGAN

The Progressive-Growing GAN architecture [147] release from NVIDIA has shown impressive performance in GAN-based image synthesis. Classical GAN architectures struggle to produce good images even for low resolution like 32x32 or 128x128. However, PGGAN can output good quality high-resolution images of up to 1024x1024.

Typically, a GAN consists of two parts i.e. a generator and a discriminator (aka critic). The generator takes random noise as input and produces an image sample. Ideally, this image sample should be indistinguishable from the training distribution. The discriminator network acts as a critic that assesses the generated images. The goal of the generator is to fool the discriminator network by generating as real images as possible, while the discriminator penalizes the generator

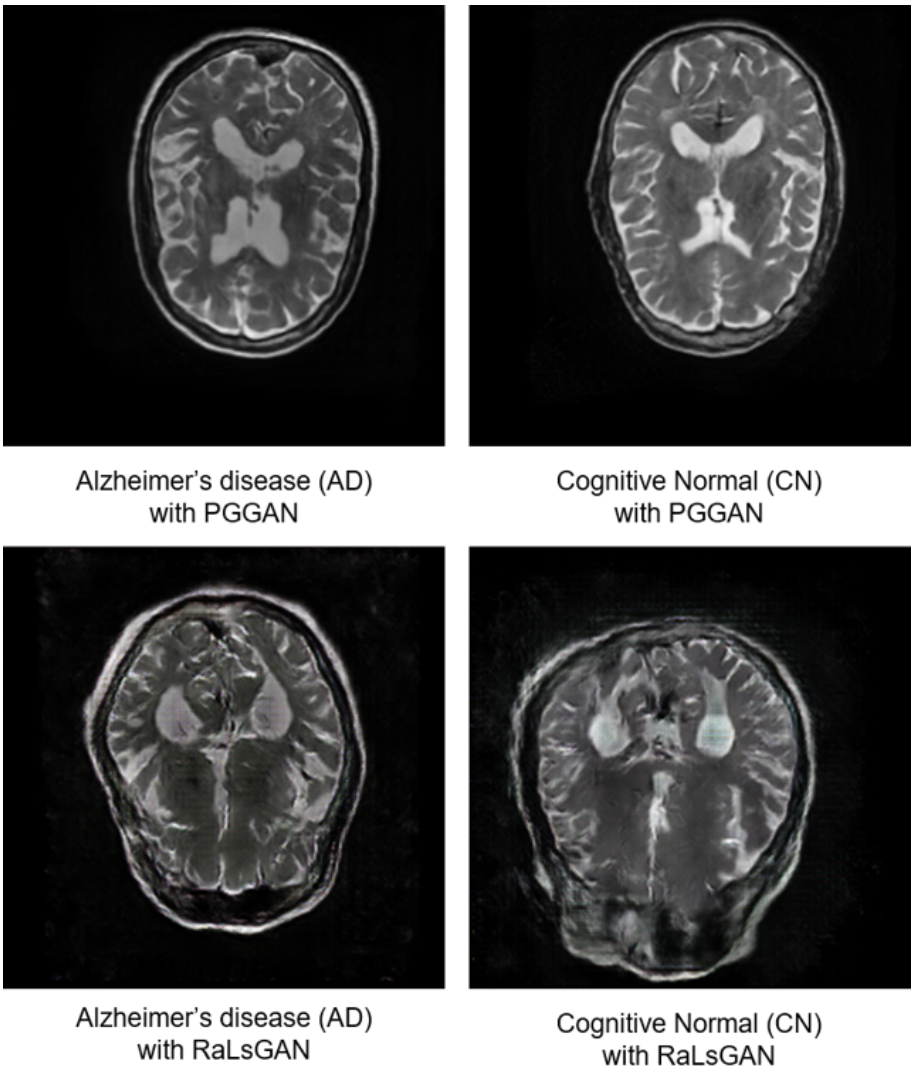

FIGURE 4. Comparison of MRI images generated by PGGAN-WP and RaLSGAN

if it fails to do so. Loss is calculated for the discriminator classification, which is then backpropagated through the network for gradient calculation. These gradients are then used to update the generator's parameters.

In PGGAN, the main difference is in the training methodology. It utilizes a multi-scale architecture where the training starts at low-resolution and then the resolution is increased by adding new layers to the network in a progressive manner as shown in Figure. The original data set with real training sampled is first classified into different resolutions like 4^2, 8^2 up to 1024^2. The generator first produces 4^2 images until it reaches some equilibrium point. Then the training proceeds with 8^2 images up to 1024^2. This strategy improves the training stability. The main reason is the when we go from a latent space z to high-resolution like 1024^2, there is an enormous amount of variance present in the space. As has been the trend in previous GAN research, generating 128^2 RGB ImageNet images is much harder than low-resolution images such as 28^2 MNIST images. Figure 5 shows the proposed multi-scale architecture.

This study adopts the PGGAN architecture to generate 256x256 brain MRI images using an ADNI data set. In our case, we limited the resolution up to 256^2. We train and generate AD, CN, and MCI images separately. The details of the model used are shown in Figure 6.

2) Convolutional Block Attention Module

Attention modules are used in computer vision to make the model learn and focus more on the important information, rather than learning background information. The idea of attention was first introduced in the domain of Natural Language Processing (NLP), introduced by Google Brain [153] in NeurIPS 2017. Recently, there have been many works that adopted the idea of the attention to computer vision tasks.

A typical attention module generates a mask of the input feature map using a simple 2D-convolutional layer, multi-

TABLE 3. Comparison of various architectures for AD vs. CN classification

Accuracy	Custom CNN	ResNet-18	Inception pretrained	Residual Attention Network	CBAM Inception V3	Multi-scale CNN	DenseNet
Train (%)	98.04%	99%	99.8%	85%	99%	100%	97.4%
Validation(%)	72.91%	75%	70%	76.67%	70.8%	70.83%	72.32%
Testing (#)	15/20	16/20	13/20	12/20	15/20	12/20	15/20

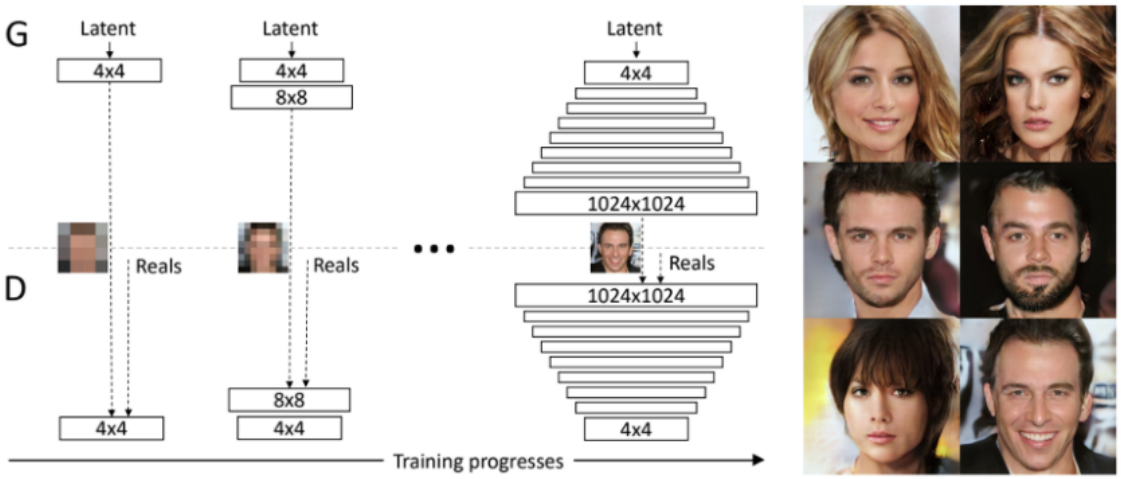

FIGURE 5. Diagram of the Multi-Scale Architecture used in Progressively-Growing GANs as proposed in the original paper [147].

Generator	Activation	Output Shape
Latent vector	–	512 × 1 × 1
Conv 4 × 4	LReLU	512 × 4 × 4
Conv 3 × 3	LReLU	512 × 4 × 4
Upsample	–	512 × 8 × 8
Conv 3 × 3	LReLU	512 × 8 × 8
Conv 3 × 3	LReLU	512 × 8 × 8
Upsample	–	512 × 16 × 16
Conv 3 × 3	LReLU	256 × 16 × 16
Conv 3 × 3	LReLU	256 × 16 × 16
Upsample	–	256 × 32 × 32
Conv 3 × 3	LReLU	128 × 32 × 32
Conv 3 × 3	LReLU	128 × 32 × 32
Upsample	–	128 × 64 × 64
Conv 3 × 3	LReLU	64 × 64 × 64
Conv 3 × 3	LReLU	64 × 64 × 64
Upsample	–	64 × 128 × 128
Conv 3 × 3	LReLU	32 × 128 × 128
Conv 3 × 3	LReLU	32 × 128 × 128
Upsample	–	32 × 256 × 256
Conv 3 × 3	LReLU	16 × 256 × 256
Conv 3 × 3	LReLU	16 × 256 × 256
Conv 1 × 1	Linear	1 × 256 × 256

Discriminator	Activation	Output Shape
Input image	–	1 × 256 × 256
Conv 1 × 1	LReLU	16 × 256 × 256
Conv 3 × 3	LReLU	16 × 256 × 256
Conv 3 × 3	LReLU	32 × 256 × 256
Downsample	–	32 × 128 × 128
Conv 3 × 3	LReLU	32 × 128 × 128
Conv 3 × 3	LReLU	64 × 128 × 128
Downsample	–	64 × 64 × 64
Conv 3 × 3	LReLU	64 × 64 × 64
Conv 3 × 3	LReLU	128 × 64 × 64
Downsample	–	128 × 32 × 32
Conv 3 × 3	LReLU	128 × 32 × 32
Conv 3 × 3	LReLU	256 × 32 × 32
Downsample	–	256 × 16 × 16
Conv 3 × 3	LReLU	256 × 16 × 16
Conv 3 × 3	LReLU	512 × 16 × 16
Downsample	–	512 × 8 × 8
Conv 3 × 3	LReLU	512 × 8 × 8
Conv 3 × 3	LReLU	512 × 8 × 8
Downsample	–	512 × 4 × 4
Minibatch stddev	–	513 × 4 × 4
Conv 3 × 3	LReLU	512 × 4 × 4
Conv 4 × 4	LReLU	512 × 1 × 1
Fully-connected	Linear	1 × 1 × 1

FIGURE 6. PGGAN architecture details for generator/discriminator used for the MRI synthesis [147].

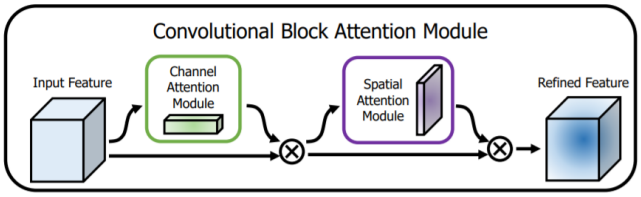

FIGURE 7. The overview of CBAM [52].

layer perceptron (MLP), and a sigmoid function at the end. Convolutional Block Attention Module or CBAM [52] is a recently proposed attention module for CNNs. Provided an input feature map, it computes the attention maps along two dimensions i.e. channel and spatial. These inferred attention maps are then multiplied with the input feature map to further refine the features. The intuition behind this idea is that blind attachment of an attention module can result in a 3D attention map which can be computationally expensive. This work [52] proposes the idea to compute channel attention and spatial attention separately using two sequential sub-modules called the Channel Attention Module (CAM) and the Spatial Attention Module (SAM). Results indicate that the proposed method achieves a similar effect with much fewer parameters. Convolutional Block Attention Module has shown consistent improvements in classification and detection tasks over the previous state-of-the-art models. The general structure of the CBAM module is shown in Figure 7.

Two pooling methods i.e. average and max pooling, are used at the same time to compute channel-wise attention for the given input feature map. This results in two Cx1x1 vectors, one produced by max-pooling and the other by average pooling. These are then passed through a simple bottleneck dense layer and then combined with a summation. The sigmoid function is applied at the end to obtain a Cx1x1 vector which shows the importance of each channel in the original feature map. Figure 8 shows the channel attention sub-module. This channel attention vector is applied to the

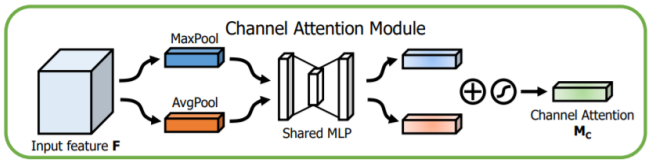

FIGURE 8. Channel attention sub-module [52].

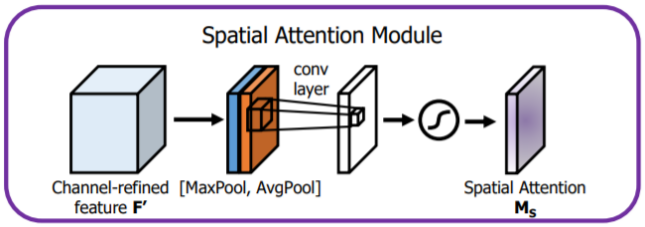

FIGURE 9. Spatial attention sub-module [52].

input feature in a pointwise manner, which creates a new vector $\mathbf{F}'$ which is shaped the same as the original input feature map $\mathbf{F}$.

After channel dimension, the next step is to process features in width and height dimensions. From the channel attention module, we obtain a CxHxW map. In the spatial attention module, average and max pooling are applied pointwise which results in two 1xHxW features. A 7x7 convolution is applied after concatenating the two features. In the end, the sigmoid function is applied to get a 1xHxW shaped feature, which is called a spatial attention map. This spatial attention map is applied to $\mathbf{F}'$ pointwise, resulting in a CxHxW vector and we get the final output of CBAM. Figure 9 shows the channel attention sub-module.

3) SimCLR

Numerous self-supervised methods have been presented in recent years, each better than the previous ones. However, their performance was still lower than the supervised learning methods. Cheng *et al.* changed this by presenting their research paper "SimCLR: a simple framework for contrastive learning of visual representations" [51]. The SimCLR paper not only improves upon the previous state-of-the-art self-supervised learning methods but also shows better performance for ImageNet classification than the supervised learning methods when scaling up the architecture.

The basic intuition behind contrastive learning is to teach a machine how to distinguish between similar and dissimilar things. SimCLR blends contrastive learning with some novel ideas to learn the visual representation without human supervision. The idea of SimCLR framework is rather very simple. Given an input image, random transformations are applied to get two augmented versions of the image x_i and x_j. Representation is then obtained for these augmented images by passing them through an encoder network. These encoded vector are represented as h_i and h_j. Then a non-linear fully connected layer is applied to get representations z. The objective is to maximize the similarity between these two representations z_i and z_j for augmented versions of the same image. Figure 10 shows working of SimCLR framework.

In their research [51], the authors used a combination of random transformations including crop, flip, color jitter, and grayscale. A random transformation function (T) is applied to get a pair of two images for each input image in the batch. Hence for the two-class case and batch size of 2 (one from each class), we have 4 images after augmentation. To get the encodings h_i and h_j, the authors used ResNet-50 [21] architecture. These encoded representations are then passed through a series of **Dense -> ReLU -> Dense** non-linear transformation layers to obtain representations z_i and z_j. In the paper, the non-linear projection part is referred to as the projection head and is denoted by $g(.)$. Cosine similarity shown in equation 3.1, is calculated between the representations z_i and z_j. The similarity of the augmented images belonging to the same class will be higher compared to the similarity between images from different classes.

$$s_{i,j} = \frac{z_i^T z_j}{||z_i||||z_j||} \tag{1}$$

SimCLR used contrastive loss called NT-Xent loss. The augmented pair in the batch are taken one by one and the probability of the two images being similar is calculated by applying the softmax function. The loss is calculated by taking the negative log of the above calculation (equation 3.2). Loss is computed for the same pair a second time, by interchanging the positions of the images in the pair. Finally, we compute loss over all the pairs in the batch of size N=2 and take an average.

$$L = \frac{1}{2N}\sum_{k=1}^{2N}[l(2k-1,2k) + l(2k,2k-1)] \tag{2}$$

The encoder and projected representation are improved over time based on the above loss values placing similar images closer to each other and dissimilar images apart. Once the training is finished on the contrastive learning task, this pretrained model can be used in a transfer learning fashion to perform a downstream task. For this, the representations from the encoder are used instead of representations obtained from the projection head.

In our case, we used random cropped versions of the input images as transformations, and the downstream task is the AD vs. CN vs. MCI classification. The ResNet-18 model with CBAM is pretrained using the SimCLR framework and is then applied to the classification task.

IV. EXPERIMENTS

We then performed experiments using our approach. Training and testing are done for two-category cases (AD vs. CN) and three-category cases (AD vs. CN vs. MCI). Our training details and results are mentioned in the next sections.

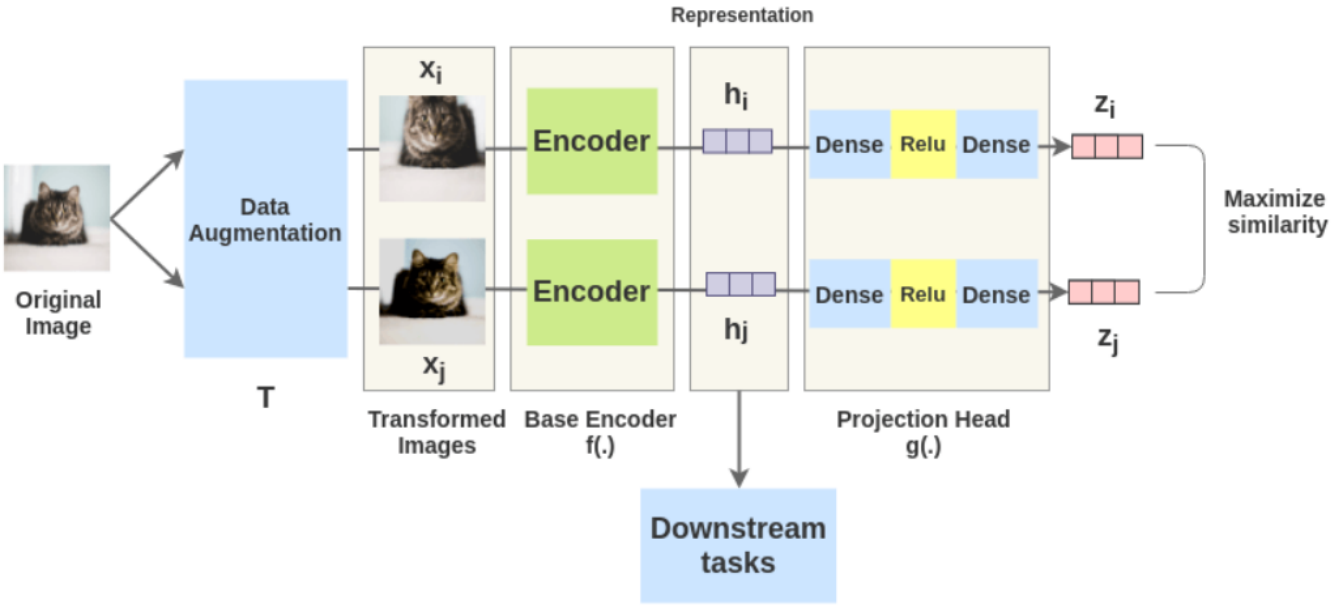

FIGURE 10. Working of SimCLR framework [51].

A. TRAINING DETAILS

1) System and environment

We used PyTorch modules to design and train our proposed model. All the training is performed using NVIDIA RTX 2080 Ti GPU.

2) Data Preprocessing

The original images obtained from the ADNI repository are 256x256 DICOM images. These images are converted into PNG format using a python library named pydicom [154]. The resulting images are then center-cropped to 224x224 and image pixels are normalized between 0 and 1. As mentioned earlier, images with no or little information are removed manually and only informative slices are used.

3) SimCLR

Next we pretrain our model using SimCLR [51] framework. Model is trained for 1000 epochs with a batch size of 128. SGD optimizer is used with a learning rate of 0.05, momentum value of 0.9, and weight decay regularization of 0.0004. We used 80%:20% split to divide our data into training and validation sets.

4) Classification task

Next, we load our pre-trained model for the classification task. Model is trained for 100 epochs with a batch size of 128. SGD optimizer is used with a learning rate of 0.001 which is decayed after a certain number of epochs. Momentum value of 0.9 and weight decay regularization of 0.0004. As mentioned earlier, separate data sets are prepared and used for training, validation, and testing the model.

B. RESULTS

We trained our ResNet-18 [21] mode for comparison. We also compare results with and without CBAM and SimCLR architectures to highlight their effect. We also compared our model to a custom-designed CNN (five convolutional layers and five fully connected layers). Classification results are shown in the table. Here we only mention validation and test accuracies. Training accuracy can easily be very high and does not convey much meaningful information in terms of the actual model performance. Table 4 shows results of the proposed framework for AD vs. CN classification task.

As seen in the results shown in the table, ResNet-18 architecture combined with CBAM and SimCLR provides better results. When training from scratch, the model seems to overfit and the training process is unstable. However, after pretraining, the model training on classification tasks is more stable and the resulting accuracy also goes up to 75%. The model performance is further refined when combined with CBAM architecture since the model learns to focus and refine important features for the classification task. This shows that both SimCLR and CBAM help the model in achieving considerable performance gains on medical data.

In further experiments, the model was trained for the three-way classification task. However, the model validation and testing accuracies ranged between 55-60%. Here model shows lower performance. The main reason is that MCI class is very hard to distinguish from AD and CN classes. Also, the amount of data is small which can result in poor training of the model. We also investigated the effect of PGGAN-based medical image generation. The model shows 3-4 misclassifications during the testing phase when trained on data with only traditional data augmentation. Adding PGGAN data increases the amount and also a variety of data

TABLE 4. AD vs. CN classification results of proposed framework

Accuracy	Custom CNN	ResNet-18 (scratch)	ResNet-18 (SimCLR)	ResNet-18 + CBAM (SimCLR)
Validation	75%	70%	75%	81%
Test(per slices)	77%	69%	75%	83%
Test(per patient)	9/10, 7/10	6/10, 8/10	8/10, 8/10	9/10, 9/10

which can help the model learn more generalized features.

C. COMPARISON WITH EXISTING METHODS

Finally, we compare our results with some other deep learning-based approaches in recent years for the AD classification task. However, the comparison is not completely fair since most of these past approaches use large data sets or suffer from data leakage and also lack proper evaluation methods.

Aderghal *et al.* achieved an accuracy of 84% percent for the AD vs. CN classification task using the ROI-based approach. Accuracy of 85% is achieved using a 3D subject-level approach [155]. Cheng *et al.* reported an accuracy of 87% for AD vs. CN using a 3D patch-level approach. Korolev *et al.* [156] and Li *et al.* [157] obtained accuracies of 80% and 88% using 3D subject-level approaches. Valliani and Soni [] reported accuracy of 81% for AD vs. CN classification and 57% for three-way classification using 2D slices. Senanayake *et al.* achieved 76% accuracy using 3D subject-level approach for AD vs. CN classification. Compared with our approach, some of these methods have higher accuracy. These methods are free from data leakage. However, the amount of training data was large.

An accuracy of 91% is achieved using ROI-based approach by Aderghal *et al.* [158]. However, their approach suffers from late-split and the absence of an independent test set. Similar limitations are observed in [135] [159] [160] [161] where reported accuracies are very high but there was no proper test set and evaluation was biased. Liu *et al.* [162] achieved 85% accuracy for AD vs. CN task using 3D patches. However, their approach suffers from biased transfer learning (see Section 2.3.1). Farooq *et al.* achieved a very high accuracy of 99% for AD vs. CN vs. MCI classification using 2D slices. However, their data split method was wrong and resulted in a biased evaluation of the model. Vu *et al.* reported 86% and 80% accuracy for AD vs. CN and Ad vs. CN vs. MCI classification respectively using 3D data. However, this approach also suffers from a wrong data split. Accuracy of 95% is reported by Wu *et al.* for three-way classification using 2D slices but there was no independent test set or proper data split. Approaches by Hosseini *et al.* [163] and Wang *et al.* [144] also suffer from these limitations.

In our approach, we avoided all kinds of data leakage and provided an unbiased evaluation of our model which is very important for clinical applications. Our model achieves comparable results to the past approaches even with a very small data set.

V. CONCLUDING REMARKS

We have proposed a novel framework for the AD classification task. The proposed architecture comprises a baseline ResNet-18 model in combination with CBAM which helps the model to refine the important discriminative features and put less focus on the background features. The whole model is pretrained using SimCLR [51] that enables the model to learn useful representations in a self-supervised manner. To deal with the issue of data scarcity, we presented a novel PGGAN-WGP based medical image synthesis scheme which generates very realistic MRI images.

We compare our results with the previous approaches. We evaluated our model on separate validation and test data without any kind of data leakage as seen in many previous approaches [164]. Our proposed model achieves an accuracy of 83% on the test data set with one misclassification per category for the task of AD vs. CN classification. These results are comparable to previous approaches despite the small data set. In the case of three-way classification, model achieves an accuracy of up to 60% on the test data. The proposed model shows promising results and can help in early diagnosis which can result in timely treatment of AD patients.

A. FUTURE DIRECTION

There is still room for improvement in the model performance, especially for the case of AD vs. CN vs. MCI classification. As part of future work, we will test our model with more and more large data sets and compare the results. This can result in better training and make the model more generalizable. We also plan to look into better data preprocessing methods like segmentation and better slice selection approaches. Improving the data can help the model performance. Taking help from medical experts can help in this process.

Another future direction is to investigate the performance of 3D CNN architecture instead of 2D CNN with our approach. We can also extend this framework for the diagnosis of other diseases. We can also test this framework for other modalities or multiple modalities and compare the results. Another possible application can be explainable AI in which we can interpret how our model is making decisions during the evaluation process. Therefore, there are a lot of possible paths for future work and further work in this direction can

help with the early diagnosis of various diseases, especially in small data regimes.

REFERENCES

[1] Alzheimer's Association. 2015 alzheimer's disease facts and figures. Alzheimer's & Dementia, 11(3):332–384, 2015.

[2] Shaik Basheera and M Satya Sai Ram. 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. Alzheimer's & Dementia: Translational Research & Clinical Interventions, 5:974–986, 2019.

[3] Marilyn S Albert, Steven T DeKosky, Dennis Dickson, Bruno Dubois, Howard H Feldman, Nick C Fox, Anthony Gamst, David M Holtzman, William J Jagust, Ronald C Petersen, et al. The diagnosis of mild cognitive impairment due to alzheimer's disease: recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. Alzheimer's & dementia, 7(3):270–279, 2011.

[4] Clifford R Jack Jr, Marilyn S Albert, David S Knopman, Guy M McKhann, Reisa A Sperling, Maria C Carrillo, Bill Thies, and Creighton H Phelps. Introduction to the recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. Alzheimer's & dementia, 7(3):257–262, 2011.

[5] Reisa A Sperling, Paul S Aisen, Laurel A Beckett, David A Bennett, Suzanne Craft, Anne M Fagan, Takeshi Iwatsubo, Clifford R Jack Jr, Jeffrey Kaye, Thomas J Montine, et al. Toward defining the preclinical stages of alzheimer's disease: recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. Alzheimer's & dementia, 7(3):280–292, 2011.

[6] Rachael Elrod, Elaine R Peskind, Lillian DiGiacomo, Kayla I Brodkin, Richard C Veith, and Murray A Raskind. Effects of alzheimer's disease severity on cerebrospinal fluid norepinephrine concentration. The American journal of psychiatry, 154(1):25–30, 1997.

[7] Gali H Weissberger, Jessica V Strong, Kayla B Stefanidis, Mathew J Summers, Mark W Bondi, and Nikki H Stricker. Diagnostic accuracy of memory measures in alzheimer's dementia and mild cognitive impairment: a systematic review and meta-analysis. Neuropsychology review, 27(4):354–388, 2017.

[8] Sara Lorio, Ferath Kherif, Anne Ruef, Lester Melie-Garcia, Richard Frackowiak, John Ashburner, Gunther Helms, Antoine Lutti, and Bodgan Draganski. Neurobiological origin of spurious brain morphological changes: A quantitative mri study. Human brain mapping, 37(5):1801–1815, 2016.

[9] Stanisław Adaszewski, Juergen Dukart, Ferath Kherif, Richard Frackowiak, Bogdan Draganski, Alzheimer's Disease Neuroimaging Initiative, et al. How early can we predict alzheimer's disease using computational anatomy? Neurobiology of aging, 34(12):2815–2826, 2013.

[10] Tommaso Maggipinto, Roberto Bellotti, Nicola Amoroso, Domenico Diacono, Giacinto Donvito, Eufemia Lella, Alfonso Monaco, Marzia Antonella Scelsi, Sabina Tangaro, Alzheimer's Disease Neuroimaging Initiative, et al. Dti measurements for alzheimer's classification. Physics in Medicine & Biology, 62(6):2361, 2017.

[11] Stefan Klöppel, Cynthia M Stonnington, Carlton Chu, Bogdan Draganski, Rachael I Scahill, Jonathan D Rohrer, Nick C Fox, Clifford R Jack Jr, John Ashburner, and Richard SJ Frackowiak. Automatic classification of mr scans in alzheimer's disease. Brain, 131(3):681–689, 2008.

[12] Yichuan Tang. Deep learning using linear support vector machines. arXiv preprint arXiv:1306.0239, 2013.

[13] Jürgen Schmidhuber. Deep learning in neural networks: An overview. Neural networks, 61:85–117, 2015.

[14] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. nature, 521(7553):436–444, 2015.

[15] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. Deep learning, volume 1. MIT press Cambridge, 2016.

[16] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E Alsaadi. A survey of deep neural network architectures and their applications. Neurocomputing, 234:11–26, 2017.

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25:1097–1105, 2012.

[18] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[19] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1–9, 2015.

[20] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning, pages 6105–6114. PMLR, 2019.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[22] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788, 2016.

[23] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497, 2015.

[24] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems, 30(11):3212–3232, 2019.

[25] Hamid Laga, Laurent Valentin Jospin, Farid Boussaid, and Mohammed Bennamoun. A survey on deep learning techniques for stereo-based depth estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

[26] Fayao Liu, Chunhua Shen, and Guosheng Lin. Deep convolutional neural fields for depth estimation from a single image. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5162–5170, 2015.

[27] Liang-Chieh Chen, Alexander Hermans, George Papandreou, Florian Schroff, Peng Wang, and Hartwig Adam. Masklab: Instance segmentation by refining object detection with semantic and direction features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4013–4022, 2018.

[28] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431–3440, 2015.

[29] Shervin Minaee, Yuri Y Boykov, Fatih Porikli, Antonio J Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[30] Abdul Mueed Hafiz and Ghulam Mohiuddin Bhat. A survey on instance segmentation: state of the art. International Journal of Multimedia Information Retrieval, pages 1–19, 2020.

[31] Dario Amodei, Sundaram Ananthanarayanan, Rishita Anubhai, Jingliang Bai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Qiang Cheng, Guoliang Chen, et al. Deep speech 2: End-to-end speech recognition in english and mandarin. In International conference on machine learning, pages 173–182. PMLR, 2016.

[32] Alex Graves and Navdeep Jaitly. Towards end-to-end speech recognition with recurrent neural networks. In International conference on machine learning, pages 1764–1772. PMLR, 2014.

[33] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning, pages 160–167, 2008.

[34] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing, pages 1631–1642, 2013.

[35] Yoav Goldberg. A primer on neural network models for natural language processing. Journal of Artificial Intelligence Research, 57:345–420, 2016.

[36] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

[37] Zachary C Lipton, John Berkowitz, and Charles Elkan. A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019, 2015.

[38] Richard Socher, Cliff Chiung-Yu Lin, Andrew Y Ng, and Christopher D Manning. Parsing natural scenes and natural language with recursive neural networks. In ICML, 2011.

[39] Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. Recurrent neural network based language model. In Eleventh annual conference of the international speech communication association, 2010.

[40] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. arXiv preprint arXiv:1406.2661, 2014.

[41] Wullianallur Raghupathi and Viju Raghupathi. Big data analytics in healthcare: promise and potential. Health information science and systems, 2(1):1–10, 2014.

[42] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics, 19(6):1236–1246, 2018.

[43] Siqi Liu, Sidong Liu, Weidong Cai, Hangyu Che, Sonia Pujol, Ron Kikinis, Dagan Feng, Michael J Fulham, et al. Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease. IEEE Transactions on Biomedical Engineering, 62(4):1132–1140, 2014.

[44] Siqi Liu, Sidong Liu, Weidong Cai, Sonia Pujol, Ron Kikinis, and Dagan Feng. Early diagnosis of alzheimer's disease with deep learning. In 2014 IEEE 11th international symposium on biomedical imaging (ISBI), pages 1015–1018. IEEE, 2014.

[45] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. NeuroImage, 101:569–582, 2014.

[46] Pedro Domingos. A few useful things to know about machine learning. Communications of the ACM, 55(10):78–87, 2012.

[47] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging, 35(5):1285–1298, 2016.

[48] Alfonso Medela, Artzai Picon, Cristina L Saratxaga, Oihana Belar, Virginia Cabezón, Riccardo Cicchi, Roberto Bilbao, and Ben Glover. Few shot learning in histopathological images: reducing the need of labeled data on biological datasets. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pages 1860–1864. IEEE, 2019.

[49] Yujin Oh, Sangjoon Park, and Jong Chul Ye. Deep learning covid-19 features on cxr using limited training data sets. IEEE Transactions on Medical Imaging, 39(8):2688–2700, 2020.

[50] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196, 2017.

[51] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In International conference on machine learning, pages 1597–1607. PMLR, 2020.

[52] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), pages 3–19, 2018.

[53] Shannon Zofia Klekociuk, Jeffery J Summers, James C Vickers, and Mathew J Summers. Reducing false positive diagnoses in mild cognitive impairment: the importance of comprehensive neuropsychological assessment. European journal of neurology, 21(10):1330–e83, 2014.

[54] Katarzyna Adamczuk, Jolien Schaeverbeke, Hugo MJ Vanderstichele, Johan Lilja, Natalie Nelissen, Koen Van Laere, Patrick Dupont, Kelly Hilven, Koen Poesen, and Rik Vandenberghe. Diagnostic value of cerebrospinal fluid aβ ratios in preclinical alzheimer's disease. Alzheimer's research & therapy, 7(1):1–11, 2015.

[55] Dinggang Shen, Chong-Yaw Wee, Daoqiang Zhang, Luping Zhou, and Pew-Thian Yap. Machine learning techniques for ad/mci diagnosis and prognosis. In Machine learning in healthcare informatics, pages 147–179. Springer, 2014.

[56] Katherine R Gray, Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative, et al. Random forest-based similarity measures for multi-modal classification of alzheimer's disease. NeuroImage, 65:167–175, 2013.

[57] Shih-Ting Yang, Jiann-Der Lee, Tzyh-Chyang Chang, Chung-Hsien Huang, Jiun-Jie Wang, Wen-Chuin Hsu, Hsiao-Lung Chan, Yau-Yau Wai, and Kuan-Yi Li. Discrimination between alzheimer's disease and mild cognitive impairment using som and pso-svm. Computational and mathematical methods in medicine, 2013, 2013.

[58] Wed Kadhim Oleiwi. Alzheimer disease diagnosis using the k-means, glcm and k_nn. Journal of University of Babylon for Pure and Applied Sciences, 26(2):57–65, 2018.

[59] Akhila Reghu and Reby John. Glcm based feature extraction of neurodegenerative disease for regional brain patterns. Int J Eng Res Gen Sci, 4:319–323, 2016.

[60] Zhe Xiao, Yi Ding, Tian Lan, Cong Zhang, Chuanji Luo, and Zhiguang Qin. Brain mr image classification for alzheimer's disease diagnosis based on multifeature fusion. Computational and mathematical methods in medicine, 2017, 2017.

[61] Tooba Altaf, Syed Muhammad Anwar, Nadia Gul, Muhammad Nadeem Majeed, and Muhammad Majid. Multi-class alzheimer's disease classification using image and clinical features. Biomedical Signal Processing and Control, 43:64–74, 2018.

[62] Ramesh Kumar Lama, Jeonghwan Gwak, Jeong-Seon Park, and Sang-Woong Lee. Diagnosis of alzheimer's disease based on structural mri images using a regularized extreme learning machine and pca features. Journal of healthcare engineering, 2017, 2017.

[63] Yong Fan, Susan M Resnick, Xiaoying Wu, and Christos Davatzikos. Structural and functional biomarkers of prodromal alzheimer's disease: a high-dimensional pattern classification study. Neuroimage, 41(2):277–285, 2008.

[64] Juergen Dukart, Karsten Mueller, Henryk Barthel, Arno Villringer, Osama Sabri, Matthias Leopold Schroeter, Alzheimer's Disease Neuroimaging Initiative, et al. Meta-analysis based svm classification enables accurate detection of alzheimer's disease across different clinical centers using fdg-pet and mri. Psychiatry Research: Neuroimaging, 212(3):230–236, 2013.

[65] Ignacio Álvarez, Míriam López, Juan Manuel Górriz, Javier Ramírez, Diego Salas-Gonzalez, Carlos García Puntonet, and Fermín Segovia. Automatic classification system for the diagnosis of alzheimer disease using component-based svm aggregations. In International Conference on Neural Information Processing, pages 402–409. Springer, 2008.

[66] F Segovia, JM Górriz, J Ramírez, D Salas-González, I Álvarez, M López, R Chaves, and P Padilla. Classification of functional brain images using a gmm-based multi-variate approach. Neuroscience Letters, 474(1):58–62, 2010.

[67] Rupali S Kamathe and Kalyani R Joshi. A novel method based on independent component analysis for brain mr image tissue classification into csf, wm and gm for atrophy detection in alzheimer's disease. Biomedical Signal Processing and Control, 40:41–48, 2018.

[68] Seyed Hani Hojjati, Ata Ebrahimzadeh, Ali Khazaee, Abbas Babajani-Feremi, Alzheimer's Disease Neuroimaging Initiative, et al. Predicting conversion from mci to ad using resting-state fmri, graph theoretical approach and svm. Journal of neuroscience methods, 282:69–80, 2017.

[69] Jinhua Sheng, Bocheng Wang, Qiao Zhang, Qingqiang Liu, Yangjie Ma, Weixiang Liu, Meiling Shao, and Bin Chen. A novel joint hcpmmp method for automatically classifying alzheimer's and different stage mci patients. Behavioural brain research, 365:210–221, 2019.

[70] Prashanthi Vemuri, Jeffrey L Gunter, Matthew L Senjem, Jennifer L Whitwell, Kejal Kantarci, David S Knopman, Bradley F Boeve, Ronald C Petersen, and Clifford R Jack Jr. Alzheimer's disease diagnosis in individual subjects using structural mr images: validation studies. Neuroimage, 39(3):1186–1197, 2008.

[71] Benoît Magnin, Lilia Mesrob, Serge Kinkingnéhun, Mélanie Pélégrini-Issac, Olivier Colliot, Marie Sarazin, Bruno Dubois, Stéphane Lehéricy, and Habib Benali. Support vector machine-based classification of alzheimer's disease from whole-brain anatomical mri. Neuroradiology, 51(2):73–83, 2009.

[72] Emilie Gerardin, Gaël Chételat, Marie Chupin, Rémi Cuingnet, Béatrice Desgranges, Ho-Sung Kim, Marc Niethammer, Bruno Dubois, Stéphane Lehéricy, Line Garnero, et al. Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage, 47(4):1476–1486, 2009.

[73] Javier Ramírez, Juan Manuel Górriz, Míriam López, Diego Salas-Gonzalez, Ignacio Álvarez, Fermín Segovia, and Carlos García Puntonet. Early detection of the alzheimer disease combining feature selection and kernel machines. In International Conference on Neural Information Processing, pages 410–417. Springer, 2008.

[74] R Chaves, J Ramírez, JM Górriz, M López, D Salas-Gonzalez, I Alvarez, and F Segovia. Svm-based computer-aided diagnosis of the alzheimer's

disease using t-test nmse feature selection with feature correlation weighting. Neuroscience letters, 461(3):293–297, 2009.

[75] Claudia Plant, Stefan J Teipel, Annahita Oswald, Christian Böhm, Thomas Meindl, Janaina Mourao-Miranda, Arun W Bokde, Harald Hampel, and Michael Ewers. Automated detection of brain atrophy patterns based on mri for the prediction of alzheimer's disease. Neuroimage, 50(1):162–174, 2010.

[76] P Padilla, JM Górriz, J Ramírez, EW Lang, R Chaves, F Segovia, M López, D Salas-González, and I Álvarez. Analysis of spect brain images for the diagnosis of alzheimer's disease based on nmf for feature extraction. Neuroscience letters, 479(3):192–196, 2010.

[77] Andrés Ortiz, Jorge Munilla, Ignacio Álvarez-Illán, Juan M Górriz, Javier Ramírez, Alzheimer's Disease Neuroimaging Initiative, et al. Exploratory graphical models of functional and structural connectivity patterns for alzheimer's disease diagnosis. Frontiers in computational neuroscience, 9:132, 2015.

[78] Daniel Schmitter, Alexis Roche, Bénédicte Maréchal, Delphine Ribes, Ahmed Abdulkadir, Meritxell Bach-Cuadra, Alessandro Daducci, Cristina Granziera, Stefan Klöppel, Philippe Maeder, et al. An evaluation of volume-based morphometry for prediction of mild cognitive impairment and alzheimer's disease. NeuroImage: Clinical, 7:7–17, 2015.

[79] Maciej Plocharski, Lasse Riis Østergaard, Alzheimer's Disease Neuroimaging Initiative, et al. Extraction of sulcal medial surface and classification of alzheimer's disease using sulcal features. Computer methods and programs in biomedicine, 133:35–44, 2016.

[80] Sabina Tangaro, Annarita Fanizzi, Nicola Amoroso, Roberto Bellotti, Alzheimer's Disease Neuroimaging Initiative, et al. A fuzzy-based system reveals alzheimer's disease onset in subjects with mild cognitive impairment. Physica Medica, 38:36–44, 2017.

[81] Yudong Zhang, Shuihua Wang, Preetha Phillips, Zhengchao Dong, Genlin Ji, and Jiquan Yang. Detection of alzheimer's disease and mild cognitive impairment based on structural volumetric mr images using 3d-dwt and wta-ksvm trained by psotvac. Biomedical Signal Processing and Control, 21:58–73, 2015.

[82] Sandeep Chaplot, Lalit M Patnaik, and NR Jagannathan. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomedical signal processing and control, 1(1):86–92, 2006.

[83] Andrés Ortiz, Juan M Górriz, Javier Ramírez, Francisco Jesús Martínez-Murcia, Alzheimer's Disease Neuroimaging Initiative, et al. Lvq-svm based cad tool applied to structural mri for the diagnosis of the alzheimer's disease. Pattern Recognition Letters, 34(14):1725–1733, 2013.

[84] Laila Khedher, Ignacio A Illán, Juan M Górriz, Javier Ramírez, Abdelbasset Brahim, and Anke Meyer-Baese. Independent component analysis-support vector machine-based computer-aided diagnosis system for alzheimer's with visual support. International journal of neural systems, 27(03):1650050, 2017.

[85] Laila Khedher, Javier Ramírez, Juan Manuel Górriz, Abdelbasset Brahim, and IA Illán. Independent component analysis-based classification of alzheimer's disease from segmented mri data. In International Work-Conference on the Interplay between Natural and Artificial Computation, pages 78–87. Springer, 2015.

[86] NN Kulkarni and VK Bairagi. Extracting salient features for eeg-based diagnosis of alzheimer's disease using support vector machine classifier. IETE Journal of Research, 63(1):11–22, 2017.

[87] Ali Mazaheri, Katrien Segaert, John Olichney, Jin-Chen Yang, Yu-Qiong Niu, Kim Shapiro, and Howard Bowman. Eeg oscillations during word processing predict mci conversion to alzheimer's disease. NeuroImage: Clinical, 17:188–197, 2018.

[88] Gábor Gosztolya, Veronika Vincze, László Tóth, Magdolna Pákáski, János Kálmán, and Ildikó Hoffmann. Identifying mild cognitive impairment and mild alzheimer's disease based on spontaneous speech using asr and linguistic features. Computer Speech & Language, 53:181–197, 2019.

[89] M Tanveer, B Richhariya, RU Khan, AH Rashid, P Khanna, M Prasad, and CT Lin. Machine learning techniques for the diagnosis of alzheimer's disease: A review. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 16(1s):1–35, 2020.

[90] Salim Lahmiri and Mounir Boukadoum. New approach for automatic classification of alzheimer's disease, mild cognitive impairment and healthy brain magnetic resonance images. Healthcare technology letters, 1(1):32–36, 2014.

[91] Salim Lahmiri and Amir Shmuel. Performance of machine learning methods applied to structural mri and adas cognitive scores in diagnosing alzheimer's disease. Biomedical Signal Processing and Control, 52:414–419, 2019.

[92] Saruar Alam, Goo-Rak Kwon, and Alzheimer's Disease Neuroimaging Initiative. Alzheimer disease classification using kpca, lda, and multi-kernel learning svm. International Journal of Imaging Systems and Technology, 27(2):133–143, 2017.

[93] Jialin Peng, Xiaofeng Zhu, Ye Wang, Le An, and Dinggang Shen. Structured sparsity regularized multiple kernel learning for alzheimer's disease diagnosis. Pattern recognition, 88:370–382, 2019.

[94] Glenn Fung and Jonathan Stoeckel. Svm feature selection for classification of spect images of alzheimer's disease using spatial information. Knowledge and Information Systems, 11(2):243–258, 2007.

[95] Iman Beheshti, Hasan Demirel, Hiroshi Matsuda, Alzheimer's Disease Neuroimaging Initiative, et al. Classification of alzheimer's disease and prediction of mild cognitive impairment-to-alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. Computers in biology and medicine, 83:109–119, 2017.

[96] Saruar Alam, Goo-Rak Kwon, Ji-In Kim, and Chun-Su Park. Twin svm-based classification of alzheimer's disease using complex dual-tree wavelet principal coefficients and lda. Journal of healthcare engineering, 2017, 2017.

[97] Nianyin Zeng, Hong Qiu, Zidong Wang, Weibo Liu, Hong Zhang, and Yurong Li. A new switching-delayed-pso-based optimized svm algorithm for diagnosis of alzheimer's disease. Neurocomputing, 320:195–202, 2018.

[98] Xia-an Bi, Qing Shu, Qi Sun, and Qian Xu. Random support vector machine cluster analysis of resting-state fmri in alzheimer's disease. PloS one, 13(3):e0194479, 2018.

[99] IA Illán, JM Górriz, MM López, Javier Ramírez, Diego Salas-Gonzalez, Fermín Segovia, Rosa Chaves, and Carlos García Puntonet. Computer aided diagnosis of alzheimer's disease using component based svm. Applied Soft Computing, 11(2):2376–2382, 2011.

[100] Diego Salas-Gonzalez, Juan Manuel Górriz, Javier Ramírez, M López, I Alvarez, F Segovia, R Chaves, and CG Puntonet. Computer-aided diagnosis of alzheimer's disease using support vector machines and classification trees. Physics in Medicine & Biology, 55(10):2807, 2010.

[101] Helder Frederico da Silva Lopes, Jair M Abe, and Renato Anghinah. Application of paraconsistent artificial neural networks as a method of aid in the diagnosis of alzheimer disease. Journal of medical systems, 34(6):1073–1081, 2010.

[102] Ziad Sankari and Hojjat Adeli. Probabilistic neural networks for diagnosis of alzheimer's disease using conventional and wavelet coherence. Journal of neuroscience methods, 197(1):165–170, 2011.

[103] Alexandre Savio, Maite García-Sebastián, Carmen Hernández, Manuel Graña, and Jorge Villanúa. Classification results of artificial neural networks for alzheimer's disease detection. In International Conference on Intelligent Data Engineering and Automated Learning, pages 641–648. Springer, 2009.

[104] Xiaojing Long and Chris Wyatt. An automatic unsupervised classification of mr images in alzheimer's disease. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 2910–2917. IEEE, 2010.

[105] Sandhya Joshi, Deepa Shenoy, GG Vibhudendra Simha, PL Rrashmi, KR Venugopal, and LM Patnaik. Classification of alzheimer's disease and parkinson's disease by using machine learning and neural network methods. In 2010 Second International Conference on Machine Learning and Computing, pages 218–222. IEEE, 2010.

[106] Rigel Mahmood and Bishad Ghimire. Automatic detection and classification of alzheimer's disease from mri scans using principal component analysis and artificial neural networks. In 2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP), pages 133–137. IEEE, 2013.

[107] B Al-Naami, N Gharaibeh, and A AlRazzaq Kheshman. Automated detection of alzheimer disease using region growing technique and artificial neural network. World Acad. Sci. Eng. Technol. Int. J. Biomed. Biol. Eng, 7(5), 2013.

[108] Ruoxuan Cui, Manhua Liu, and Gang Li. Longitudinal analysis for alzheimer's disease diagnosis using rnn. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pages 1398–1401. IEEE, 2018.

[109] Julian Fritsch, Sebastian Wankerl, and Elmar Nöth. Automatic diagnosis of alzheimer's disease using neural network language models. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5841–5845. IEEE, 2019.

[110] Subrata Kar and D Dutta Majumder. A novel approach of diffusion tensor visualization based neuro fuzzy classification system for early detection of alzheimer's disease. Journal of Alzheimer's disease reports, 3(1):1–18, 2019.

[111] Mehran Ahmadlou, Hojjat Adeli, and Anahita Adeli. New diagnostic eeg markers of the alzheimer's disease using visibility graph. Journal of neural transmission, 117(9):1099–1109, 2010.

[112] Shih-Ting Yang, Jiann-Der Lee, Chung-Hsien Huang, Jiun-Jie Wang, Wen-Chuin Hsu, and Yau-Yau Wai. Computer-aided diagnosis of alzheimer's disease using multiple features with artificial neural network. In Pacific Rim International Conference on Artificial Intelligence, pages 699–705. Springer, 2010.

[113] Youngsang Cho, Joon-Kyung Seong, Yong Jeong, Sung Yong Shin, Alzheimer's Disease Neuroimaging Initiative, et al. Individual subject classification for alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage, 59(3):2217–2230, 2012.

[114] María Quintana, Joan Guàrdia, Gonzalo Sánchez-Benavides, Miguel Aguilar, José Luis Molinuevo, Alfredo Robles, María Sagrario Barquero, Carmen Antúnez, Carlos Martínez-Parra, Anna Frank-García, et al. Using artificial neural networks in clinical neuropsychology: High performance in mild cognitive impairment and alzheimer's disease. Journal of Clinical and Experimental Neuropsychology, 34(2):195–208, 2012.

[115] Shuihua Wang, Yudong Zhang, Zhengchao Dong, Sidan Du, Genlin Ji, Jie Yan, Jiquan Yang, Qiong Wang, Chunmei Feng, and Preetha Phillips. Feed-forward neural network optimized by hybridization of pso and abc for abnormal brain detection. International Journal of Imaging Systems and Technology, 25(2):153–164, 2015.

[116] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. Annual review of biomedical engineering, 19:221–248, 2017.

[117] Heung-Il Suk and Dinggang Shen. Deep learning-based feature representation for ad/mci classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 583–590. Springer, 2013.

[118] Ahmad Chaddad, Christian Desrosiers, and Tamim Niazi. Deep radiomic analysis of mri related to alzheimer's disease. IEEE Access, 6:58213–58221, 2018.

[119] Donghuan Lu, Karteek Popuri, Gavin Weiguang Ding, Rakesh Balachandar, and Mirza Faisal Beg. Multimodal and multiscale deep neural networks for the early diagnosis of alzheimer's disease using structural mr and fdg-pet images. Scientific reports, 8(1):1–13, 2018.

[120] Ammarah Farooq, SyedMuhammad Anwar, Muhammad Awais, and Saad Rehman. A deep cnn based multi-class classification of alzheimer's disease using mri. In 2017 IEEE International Conference on Imaging systems and techniques (IST), pages 1–6. IEEE, 2017.

[121] Adrien Payan and Giovanni Montana. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. arXiv preprint arXiv:1502.02506, 2015.

[122] Heung-Il Suk, Seong-Whan Lee, and Dinggang Shen. Latent feature representation with stacked auto-encoder for ad/mci diagnosis. Brain Structure and Function, 220(2):841–859, 2015.

[123] Ehsan Hosseini-Asl, Georgy Gimel'farb, and Ayman El-Baz. Alzheimer's disease diagnostics by a deeply supervised adaptable 3d convolutional network. arXiv preprint arXiv:1607.00556, 2016.

[124] Andres Ortiz, Jorge Munilla, Juan M Gorriz, and Javier Ramirez. Ensembles of deep learning architectures for the early diagnosis of the alzheimer's disease. International journal of neural systems, 26(07):1650025, 2016.

[125] Garam Lee, Kwangsik Nho, Byungkon Kang, Kyung-Ah Sohn, and Dokyoon Kim. Predicting alzheimer's disease progression using multi-modal deep learning approach. Scientific reports, 9(1):1–12, 2019.

[126] Bruce Fischl, Martin I Sereno, and Anders M Dale. Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system. Neuroimage, 9(2):195–207, 1999.

[127] VP Subramanyam Rallabandi, Ketki Tulpule, Mahanandeeshwar Gattu, Alzheimer's Disease Neuroimaging Initiative, et al. Automatic classification of cognitively normal, mild cognitive impairment and alzheimer's disease using structural mri analysis. Informatics in Medicine Unlocked, 18:100305, 2020.

[128] Heung-Il Suk and Dinggang Shen. Deep ensemble sparse regression network for alzheimer's disease diagnosis. In International Workshop on Machine Learning in Medical Imaging, pages 113–121. Springer, 2016.

[129] Saman Sarraf, Ghassem Tofighi, Alzheimer's Disease Neuroimaging Initiative, et al. Deepad: Alzheimer's disease classification via deep convolutional neural networks using mri and fmri. BioRxiv, page 070441, 2016.

[130] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Deep ensemble learning of sparse regression models for brain disease diagnosis. Medical image analysis, 37:101–113, 2017.

[131] Jun Shi, Xiao Zheng, Yan Li, Qi Zhang, and Shihui Ying. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease. IEEE journal of biomedical and health informatics, 22(1):173–183, 2017.

[132] Tsubasa Goto, Caihua Wang, Yuanzhong Li, and Yukihiro Tsuboshita. Multi-modal deep learning for predicting progression of alzheimer's disease using bi-linear shake fusion. In Medical Imaging 2020: Computer-Aided Diagnosis, volume 11314, page 113141X. International Society for Optics and Photonics, 2020.

[133] Xin Hong, Rongjie Lin, Chenhui Yang, Nianyin Zeng, Chunting Cai, Jin Gou, and Jane Yang. Predicting alzheimer's disease using lstm. IEEE Access, 7:80893–80901, 2019.

[134] Samsuddin Ahmed, Kyu Yeong Choi, Jang Jae Lee, Byeong C Kim, Goo-Rak Kwon, Kun Ho Lee, and Ho Yub Jung. Ensembles of patch-based classifiers for diagnosis of alzheimer diseases. IEEE Access, 7:73373–73383, 2019.

[135] Silvia Basaia, Federica Agosta, Luca Wagner, Elisa Canu, Giuseppe Magnani, Roberto Santangelo, Massimo Filippi, Alzheimer's Disease Neuroimaging Initiative, et al. Automated classification of alzheimer's disease and mild cognitive impairment using a single mri and deep neural networks. NeuroImage: Clinical, 21:101645, 2019.

[136] Cosimo Ieracitano, Nadia Mammone, Alessia Bramanti, Amir Hussain, and Francesco C Morabito. A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of eeg recordings. Neurocomputing, 323:96–107, 2019.

[137] Donghuan Lu, Karteek Popuri, Gavin Weiguang Ding, Rakesh Balachandar, Mirza Faisal Beg, Alzheimer's Disease Neuroimaging Initiative, et al. Multiscale deep neural network based analysis of fdg-pet images for the early diagnosis of alzheimer's disease. Medical image analysis, 46:26–34, 2018.

[138] Fan Li, Manhua Liu, Alzheimer's Disease Neuroimaging Initiative, et al. Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. Computerized Medical Imaging and Graphics, 70:101–110, 2018.

[139] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[140] Rachna Jain, Nikita Jain, Akshay Aggarwal, and D Jude Hemanth. Convolutional neural network based alzheimer's disease classification from magnetic resonance brain images. Cognitive Systems Research, 57:147–159, 2019.

[141] Naimul Mefraz Khan, Nabila Abraham, and Marcia Hon. Transfer learning with intelligent training data selection for prediction of alzheimer's disease. IEEE Access, 7:72726–72735, 2019.

[142] Sitara Afzal, Muazzam Maqsood, Faria Nazir, Umair Khan, Farhan Aadil, Khalid M Awan, Irfan Mehmood, and Oh-Young Song. A data augmentation-based framework to handle class imbalance problem for alzheimer's stage detection. IEEE Access, 7:115528–115539, 2019.

[143] Simeon Spasov, Luca Passamonti, Andrea Duggento, Pietro Liò, Nicola Toschi, Alzheimer's Disease Neuroimaging Initiative, et al. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to alzheimer's disease. Neuroimage, 189:276–287, 2019.

[144] Hongfei Wang, Yanyan Shen, Shuqiang Wang, Tengfei Xiao, Liming Deng, Xiangyu Wang, and Xinyan Zhao. Ensemble of 3d densely connected convolutional network for diagnosis of mild cognitive impairment and alzheimer's disease. Neurocomputing, 333:145–156, 2019.

[145] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2016.

[146] Alexia Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard gan, 2018.

[147] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation, 2018.

[148] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028, 2017.
[149] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014.
[150] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification, 2017.
[151] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding transfer learning for medical imaging, 2019.
[152] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
[153] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
[154] pydicom. pydicom. https://github.com/pydicom/pydicom, 2020. Accessed: 2021-04-19.
[155] Danni Cheng and Manhua Liu. Cnns based multi-modality classification for ad diagnosis. In 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pages 1–5. IEEE, 2017.
[156] Sergey Korolev, Amir Safiullin, Mikhail Belyaev, and Yulia Dodonova. Residual and plain convolutional neural networks for 3d brain mri classification. In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pages 835–838. IEEE, 2017.
[157] Fan Li, Danni Cheng, and Manhua Liu. Alzheimer's disease classification based on combination of multi-model convolutional networks. In 2017 IEEE International Conference on Imaging Systems and Techniques (IST), pages 1–5. IEEE, 2017.
[158] Karim Aderghal, Jenny Benois-Pineau, Karim Afdel, and Catheline Gwenaëlle. Fuseme: Classification of smri images by fusion of deep cnns in 2d+ ε projections. In Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, pages 1–7, 2017.
[159] Jyoti Islam and Yanqing Zhang. A novel deep learning based multi-class classification method for alzheimer's disease detection using brain mri data. In International conference on brain informatics, pages 213–222. Springer, 2017.
[160] Arwa Mohammed Taqi, Ahmed Awad, Fadwa Al-Azzo, and Mariofanna Milanova. The impact of multi-optimizers and data augmentation on tensorflow convolutional neural network performance. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pages 140–145. IEEE, 2018.
[161] Shui-Hua Wang, Preetha Phillips, Yuxiu Sui, Bin Liu, Ming Yang, and Hong Cheng. Classification of alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. Journal of medical systems, 42(5):1–11, 2018.
[162] Mingxia Liu, Jun Zhang, Ehsan Adeli, and Dinggang Shen. Landmark-based deep multi-instance learning for brain disease diagnosis. Medical image analysis, 43:157–168, 2018.
[163] Ehsan Hosseini Asl, Mohammed Ghazal, Ali Mahmoud, Ali Aslantas, Ahmed Shalaby, Manual Casanova, Gregory Barnes, Georgy Gimel'farb, Robert Keynton, and Ayman El Baz. Alzheimer's disease diagnostics by a 3d deeply supervised adaptable convolutional network. FRONTIERS IN BIOSCIENCE-LANDMARK, 23:584–596, 2018.
[164] Junhao Wen, Elina Thibeau-Sutre, Mauricio Diaz-Melo, Jorge Samper-González, Alexandre Routier, Simona Bottani, Didier Dormont, Stanley Durrleman, Ninon Burgos, Olivier Colliot, et al. Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation. Medical image analysis, 63:101694, 2020.

RAJA HASEEB received the B.S. degree in Electrical Engineering for Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan in 2018. He is currently pursuing in M.S. in Electrical Engineering at Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. He is also working as a researcher in Robot Intelligence Technology (RIT) Lab at KAIST.

His research interests include Computer Vision (CV) applications, AI and deep learning for healthcare sector and medical diagnosis.

SECOND B. AUTHOR was born in Greenwich Village, New York, NY, USA in 1977. He received the B.S. and M.S. degrees in aerospace engineering from the University of Virginia, Charlottesville, in 2001 and the Ph.D. degree in mechanical engineering from Drexel University, Philadelphia, PA, in 2008.

From 2001 to 2004, he was a Research Assistant with the Princeton Plasma Physics Laboratory. Since 2009, he has been an Assistant Professor with the Mechanical Engineering Department, Texas A&M University, College Station. He is the author of three books, more than 150 articles, and more than 70 inventions. His research interests include high-pressure and high-density nonthermal plasma discharge processes and applications, microscale plasma discharges, discharges in liquids, spectroscopic diagnostics, plasma propulsion, and innovation plasma applications. He is an Associate Editor of the journal *Earth, Moon, Planets*, and holds two patents.

Dr. Author was a recipient of the International Association of Geomagnetism and Aeronomy Young Scientist Award for Excellence in 2008, and the IEEE Electromagnetic Compatibility Society Best Symposium Paper Award in 2011.

JONG-HWAN KIM (M'87) m (F'09) received the Ph.D. degree in electronics engineering from Seoul National University, Seoul, South Korea, in 1987. Since 1988, he has been with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, where he is leading the Robot Intelligence Technology Laboratory as a KT Endowed Chair Professor. He is the Director of KoYoung-KAIST AI Joint Research Center and Machine Intelligence and Robotics Multi-Sponsored Research and Education Platform, KAIST. He has authored 5 books and 5 edited books, 2 journal special issues, and around 400 refereed papers in technical journals and conference proceedings. His current research interests include intelligence technology, machine intelligence learning, and AI robots