Transfer Learning for Predicting Conversion from Normal Cognition to Cognitive Impairment

Early First Draft

Justin Bardwell

Supervisors: Ghulam Mubashar Hassan, Farzaneh Salami, Naveed Akhtar

20 September 2021

Contents

1 Introduction **3**

1.1 Alzheimer's Disease 3

1.2 Deep Learning in Medical Imaging 3

1.3 Available Data 4

1.4 Machine Learning Classifiers for Alzheimer's Disease Detection 5

1.5 Convolutional Neural Networks for Alzheimer's Disease Detection . . 6

1.6 Predicting Cognitive Decline 8

1.7 Conclusion 9

2 Project Objectives **9**

3 Materials and Methods **9**

3.1 Required Hardware and Software 9

3.2 Obtaining Data 10

3.3 Selecting Subjects 10

3.4 Evaluation Metrics 10

3.5 Experiment 1 - Voxel-based vs. Patch-based approaches 11

3.6 Experiment 2 - CNN Architectures 11

3.7 Experiment 3 - Inclusion of Clinical Data 11

4 Results **12**

4.1 Experiment 1 - Voxel-based vs. Patch-based approaches 12

4.2 Experiment 2 - CNN Architectures 12

4.3 Experiment 3 - Inclusion of Clinical Data 12

1 Introduction

1.1 Alzheimer's Disease

Alzheimer's disease (AD) is a progressive neurodegenerative disease which destroys memory and other important cognitive functions. It is the most common form of dementia affecting up to 70% of those diagnosed [1]. In the United States alone, it is estimated that there are over 6 million people living with AD [2]. Dementia, including AD, has been the second leading cause of death in Australia since 2013 [3]. Currently, there is no cure for AD.

At the present time, there is no single test that can be used to identify AD and the diagnosis can only be confirmed by examination of the subject's brain tissue after death. Clinical diagnosis of AD usually involves a combination of psychiatric assessment, cerebrospinal fluid testing (CSF), APOE genotyping, positron emission tomography (PET) and magnetic resonance imaging (MRI). However, the expert use of these tests only leads to a 77% diagnostic accuracy for clinical diagnosis of AD [4].

Mild cognitive impairment (MCI) is a condition in which people experience significant memory loss but do not yet meet the criteria for a clinical diagnosis of AD [5]. Davis et al. [6] estimated that, at age 65, 8% of people will progress from normal cognition to MCI and 22% of patients with MCI will progress to a clinical diagnosis of AD annually. For a cohort of 100 cognitively normal patients at age 65, it was found that a 20% reduction in the progression rate from normal cognition to MCI would avoid 5.7 cases of MCI and 5.6 cases of AD in the future.

Non-pharmacological interventions such as diet, exercise and cognitive exercise have been shown to have an influence on reducing the incidence of development of MCI and dementia [7]. As such, early intervention should lead to increased life expectancy and less time spent in severe AD health states.

1.2 Deep Learning in Medical Imaging

Recently, deep learning has been used to solve many complex medical problems such as detecting Multiple Sclerosis, Alzheimer's disease and various types of cancer [8]. The interest in deep learning for medical imaging has largely been due to the ability of convolutional neural networks (CNNs) to learn useful representations of complex images.

Lee et al. [9] found that radiological examinations' retrospective error rate is approximately 30%. There are a range of factors that can cause errors to be made but

it was found that errors were mainly due to fatigue and radiologists' inherent biases while performing diagnoses. These diagnostic errors are estimated to account for up to 80,000 annual deaths in the US. As such, the use of automated systems for diagnosis is a growing field of interest. Due to the prevalence of errors in radiological examinations, deep learning methods are being heavily researched and utilised to improve practices.

MRI is a medical imaging technique which uses strong magnetic fields and radio waves to take highly detailed cross-sectional pictures of the anatomy. It is a commonly used technique for diagnosing brain injuries and diseases and is one of the most useful tools a physician can use to assist in the diagnosis of AD. Out of the currently available medical imaging techniques, MRI has been found to perform the most effectively in detection of AD using deep learning [10].

1.3 Available Data

In 2010, the first phase of the Alzheimer's Disease Neuroimaging Initiative (ADNI) was completed and it provided a compilation of MRI and PET scans of 819 elderly subjects [11]. At baseline, 229 of these participants were classified as cognitively normal, 398 had mild cognitive impairment and 192 were AD patients. Since the initial compilation of these subjects, there have been several phases of new participant recruitments and the amount of images available under ADNI has been steadily increasing. Ebrahimighahnavieh et al. [10] found that of the studies into Alzheimer's detection using deep learning, approximately 90% of them used data from ADNI either by itelf or in combination with data from other sources.

The Open Access Series of Imaging Studies (OASIS) is another neuroimaging dataset which is focused on subjects' cognitive decline. OASIS-3 was released in 2018 and is an openly available dataset containing MRI and PET imaging for 1,098 subjects [12]. 850 of the participants entered the study as cognitively normal (CN) while there were 248 participants who entered with some form of cognitive impairment. Throughout the study, 245 of the patients who were initially cognitively normal had converted to a state of cognitive impairment. Over the course of the study, there were a total of 2,168 MRI scans produced. As the OASIS-3 images were obtained over a period of more than 10 years and on a range of different scanners, there were several file types produced. A single standard format has been provided and all data files have been converted to NifTI format files.

The OASIS-3 dataset also includes related clinical data as well as post-processed outputs and regional segmentations of the brain. Clinicians assessed the participants and provided a dementia diagnosis which included categories of "cognitively

normal",“AD dementia” and “vascular dementia”.

The Clinical Dementia Rating (CDR) is a commonly used measure in longitudinal studies of AD [13]. Patients' CDR is based on their impairment in memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. A CDR of 0 corresponds to a cognitively normal patient, 0.5 represents very mild dementia, 1 represents mild dementia, 2 represents moderate dementia and 3 represents severe dementia.

The groups of patients in OASIS-3 have been defined as “Stable Controls”, “Converters” and “Dementia at aging”. Stable controls are individuals who begun with a CDR score of 0 and remained with that score. Converters are those subjects who started with CDR of 0 and progressed to a CDR greater than 0. Subjects defined as dementia at aging were those who initially had a CDR greater than 0.

In both the ADNI and OASIS datasets, subjects completed clinical assessment protocols in line with the National Alzheimer Coordinating Center Uniform Data Set. The main difference between the OASIS-3 and ADNI datasets is that the majority of the OASIS-3 participants were initially categorised as being cognitively normal and their potential decline was followed through longitudinal progression. ADNI, however, primarily enrolled patients who already had some form of dementia or MCI. Thus, OASIS-3 has the potential to greatly improve early detection methods of MCI and AD.

1.4 Machine Learning Classifiers for Alzheimer's Disease Detection

Dukart et al. [14] provided some of the first research into using machine learning techniques to detect AD using MRI data. The study used a 50% split of AD patients and healthy control subjects taken from a 56 subject sample of the ADNI dataset. The volume of several regions of interest of the brain were extracted from the MRI data. A support vector machine (SVM) was fit to this volumetric data and was able to distinguish between AD patients and healthy controls with an accuracy rate of 80.4%.

Gray et al. [15] explored the ability of a random forest (RF) to classify AD using MRI volumes. The results from this study suggested that the RF was able to predict AD with a higher degree of accuracy in comparison to previous studies which used SVMs. A subset of the ADNI data was used and included 37 AD patients, 75 MCI patients and 35 healthy controls. Only using the MRI data from AD patients and healthy controls, the random forest model was able to detect AD patients with

82.5% accuracy. Gray et al. also tested the RF model's ability to classify between MCI patients and healthy controls. This task proved to be more challenging and the model only achieved an accuracy of 67.3%.

1.5 Convolutional Neural Networks for Alzheimer's Disease Detection

The first study to explore the use of deep learning for AD classification was performed by Suk and Shen in 2013 [16]. Unlike previous methods which used simple features extracted from MRI such as brain tissue volumes, the deep learning method implemented here was proposed to be able to extract more complicated patterns from the data. A stacked auto-encoder (SAE) with three hidden layers was utilised and the output layer of the auto-encoder is used to represent the class label of the input data. This implementation of the SAE acts as a classifier and was able to detect between healthy controls and AD patients with 85.7% accuracy. The SAE could also distinguish MCI patients from healthy controls with 70.6% accuracy. The results produced with this deep learning method provided better results than previous methods which considered only low-level features extracted from MRI.

Gunawardena et al. [17] compared the use of SVMs and CNNs in the detection of AD from MRI data. The same data was used for both experiments and consisted of 1615 images from the ADNI dataset. As seen in Figure 1, the CNN model was constructed using two convolution layers, a pooling layer and a fully connected layer. The SVM used in this study predicted AD with an accuracy of 84.4% while the CNN model had an accuracy of 96.0%. This increase in performance is significant and confirms that deep learning methods can greatly improve the ability to automatically detect AD from MRI.

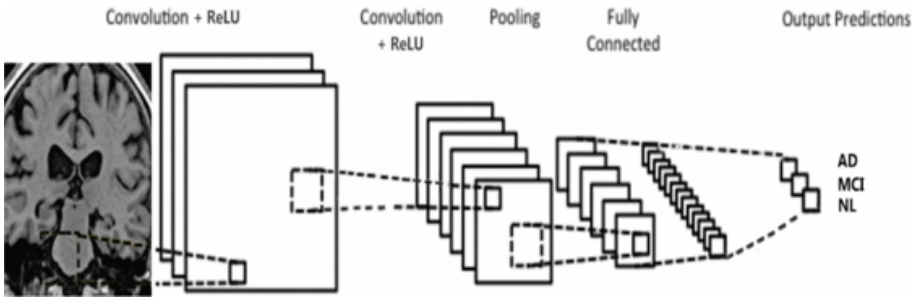

Figure 1: Simple CNN to predict AD from two-dimensional MRI images (Source: Gunawardena et al., 2017)

Farooq et al. [18] investigated the AD detection performance of well-known CNN models such as GoogLeNet and VGG as well as three ResNet models with vary-

ing amounts of layers. 355 MRI volumes from the ADNI dataset were used, with each scan being broken down into many 2D slices. Binary classification between cognitively normal patients and patients with AD or MCI was performed effectively by these models. ResNet-18 and GoogLeNet were the two best performing models with both achieving accuracy rates of over 99%. As seen in Figure 2, the performance of these models exceeded most other state-of-the-art technologies present in the literature.

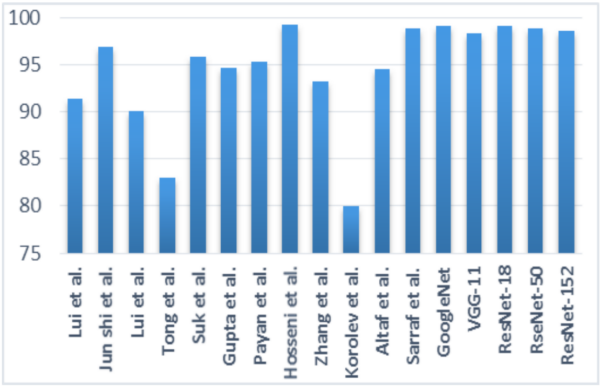

Figure 2: Binary classification accuracy comparison (Source: Farooq et al., 2017)

Although three-dimensional neural networks are more computationally intensive than their 2D counterparts, the ability of 3D CNNs to extract discriminative features from MRI data is superior [19]. The two methods for dealing with 3D images are voxel-based and patch-based approaches. A voxel-based approach uses voxel intensity values from the whole MRI while a patch-based approach breaks down the whole image into several small three-dimensional cubes. It has been found that patch-based approaches are able to learn from the whole brain with significantly reduced dimensionality in comparison to voxel-based approaches [20]. Figure 3 shows the architecture of a successful patch-based CNN that is able to capture subtle brain changes while also being significantly less computationally intensive than approaches that attempt to use image data from the whole brain [21].

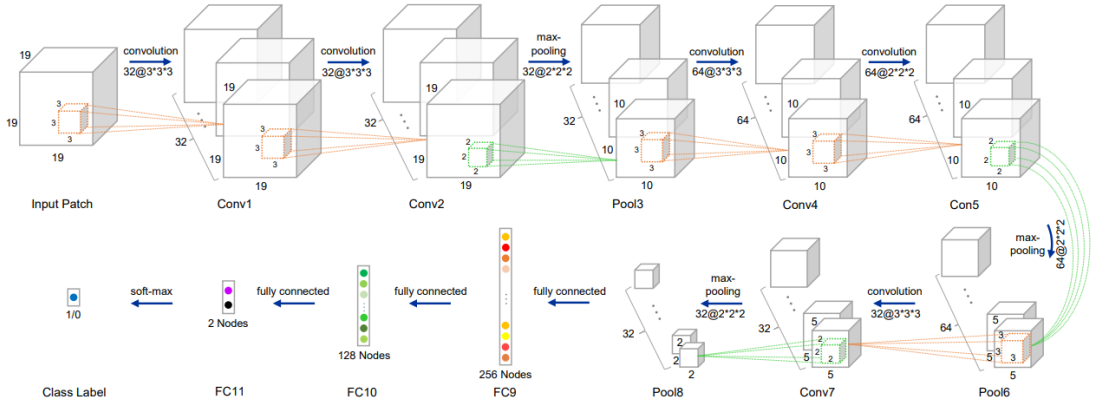

Figure 3: Binary classification accuracy comparison (Source: Liu et al., 2018)

1.6 Predicting Cognitive Decline

Early diagnosis of AD and MCI due to AD could help slow or halt patients' cognitive decline[22]. The methods of distinguishing between CN and MCI patients as well as the prediction of a subject's conversion from MCI to AD and CN to MCI/AD are more valuable compared to other classification approaches [10].

Shen et al. [23] used a 2D CNN to extract features from the MRI scans of 165 MCI patients recorded in ADNI. After feature extraction, a support vector machine was used to classify these features and predict whether the patients would be diagnosed with AD within 12 months. The classification accuracy using an RBF kernel was 92.3%.

Albert et al. [24] investigated the ability to predict individuals' progression from normal cognition to MCI. Their study used baseline data from 224 cognitively normal subjects, of which 75% had a first degree relative with dementia. At the end of the 5 year study, 46 of these subjects had progressed to having MCI. Albert et al found that "age", "APOE4", "CSF p-tau", "right hippocampus volume" and "right entorhinal cortex thickness" were significant for the purposes of prediction of a patient's progression from normal cognition to MCI. This study's use of MRI scans was limited to measuring the brain's right hippocampus volume and entorhinal cortex thickness and using these measurements in their model. For the prediction of progression from normal cognition to MCI, the single-modal use of the MRI domain technique in this study provided an AUC of 0.740. When combined with other demographic and genetic variables the AUC improved to 0.849.

1.7 Conclusion

The use of deep learning in medical imaging has become more prevalent in recent years. CNNs have proven to be highly effective at detecting AD from MRI data. The state-of-the-art methods are currently able to detect AD with accuracy rates that likely exceed the ability of radiologists to correctly diagnose AD using MRI. CNNs can also be used to predict the cognitive decline of a patient from MCI to a clinical diagnosis of AD. To date, there has been limited research into predicting individuals' progression from normal cognition to MCI, meaning there are substantial opportunities to improve this practice.

The release of the OASIS-3 dataset in 2018 has opened up the potential to greatly improve the current methods of predicting cognitive decline from normal cognition to MCI and AD. The methods used in the current literature are limited to traditional machine learning models. As such, there are opportunities for research into the improvements that deep learning can provide in relation to our ability to predict cognitively normal adults' decline to MCI and AD.

2 Project Objectives

The main objective of this research project is to explore the performance of transfer learning based 3D CNNs for the task of predicting conversion from healthy cognition to cognitive impairment using brain MRI image data.

The experiments to be performed will address the following points of interest:

1. Compare the performance of a 3D CNN using data from a voxel-based approach versus a patch-based approach.
2. Explore any improvements in model performance that can be attained by utilising different CNN architectures.
3. Evaluate whether the inclusion of clinical data can assist in our prediction of conversion from healthy cognition to cognitive impairment.

3 Materials and Methods

3.1 Required Hardware and Software

All work will be performed in a Google Colab environment which provides access to GPUs. Implementation of the experiments will be performed using Python 3.7 and

TensorFlow 2.6.

3.2 Obtaining Data

For each subject in OASIS-3, we have access to T1-weighted MRI images. We will use the post-processed MRI images from OASIS-3 that come in the form of FreeSurfer files. These files each contain a 3D image of the brain with the skull stripped and having a shape of $256 \times 256 \times 256$ voxels.

Other clinical data for each subject such as gender, age and APOE genotype is also present and will be explored further in Experiment 3.

3.3 Selecting Subjects

There are two classes of subjects that we are interested in for our experiments. These are subjects who remained cognitively normal for the foreseeable future and subjects who converted from normal cognition to cognitive impairment. For our subjects that remained cognitively normal throughout the study, we will select those who have received a clinical assessment with a CDR of 0 at least 3000 days after their initial scan.

For the converters, we select subjects who initially had a CDR of 0 but at a later assessment received a CDR of 0.5 or higher. As the scans of each subject are not usually taken on the date of their clinical assessment, we must also take steps to ensure that the scans we select are taken at a point in time when the subject is still likely to be cognitively normal. To do this, we will select scans that have occurred within 365 days of receiving a CDR score of 0 in an assessment and receive a CDR rating of 0.5 or above within the next 1000 days after the scan. We also ensure that the date of the scan was closer to their CDR 0 assessment compared to their CDR ≥ 0.5 assessment.

A 50/50 split of subjects in each class will be obtained before any models are fit to the data.

3.4 Evaluation Metrics

The evaluation metrics to be used are accuracy, sensitivity, specificity, area under the receiver operating characteristic curve.

Our evaluation metrics will be applied to a test set made up of 20% of the total data.

3.5 Experiment 1 - Voxel-based vs. Patch-based approaches

In our first experiment, we will compare the performance of a voxel-based model to a patch-based model. To do this we will be using a 3D ResNet-50 model initialised using ImageNet weights. We will first use a voxel-based method where the whole brain image is used as the input to the model. We will compare the accuracy of this model to a model using the same images but split into $3 \times 3 \times 3$ cubes. Each patch will have a 50% overlap with it's neighbouring cube. This patch-based method will result in 27 individual patches for each image being used. A stacked model will be applied by using the outputs from each patch based model as features to feed in to a logistic regression model for us to make our final prediction.

Hypothesis: The patch-based model will be able to provide more accurate predictions than the voxel-based model.

3.6 Experiment 2 - CNN Architectures

We will compare the performance of two of the historically best performing CNN architectures for medical imaging problems, ResNet-50 and DenseNet-121. We will also test the performance of other variants of ResNet such as SE-ResNet and SE-ResNeXt. An ensemble of all the models chosen will also be tested to determine whether this can provide any additional improvements to our models.

Hypothesis: There will be differences in model accuracy between each model. An ensemble method using all of the CNN architectures tested will provide the greatest accuracy for our predictions.

3.7 Experiment 3 - Inclusion of Clinical Data

A logistic regression model will be fit to the clinical data which is made up of subjects' gender, age and APOE genotype. The accuracy of this simple model will be compared to our CNN based models.

Then, we will insert this clinical data as additional features for our previously used stacked patch-based model to test whether this can improve the model's accuracy.

Hypothesis: The clinical data on its own will not be able to predict cognitive decline as well as the image-based CNN models. However, the use of this clinical data as part of a stacked model will improve accuracy.

4 Results

4.1 Experiment 1 - Voxel-based vs. Patch-based approaches

Preliminary results:

Voxel-based approach where the whole brain's image is used provided an accuracy of 65%. Individual patches provided results ranging from 50% accuracy to 85% accuracy. We expect the stacked model of all 27 patches to produce accuracy in excess of 85%.

Final results for the first 2 experiments are nearing completion. Time taken to run models has been slowing our progress.

4.2 Experiment 2 - CNN Architectures

Preliminary results:

DenseNet-121 provided comparable results to ResNet-50. However, SE-ResNet-50 appears to have provided improved results compared to ResNet-50.

4.3 Experiment 3 - Inclusion of Clinical Data

Yet to test this experiment.

References

[1] World Health Organization, "Global action plan on the public health response to dementia 2017 - 2025," 2017.

[2] Alzheimer's Association, "2021 Alzheimer's disease facts and figures," 2021.

[3] Australian Bureau of Statistics, "Causes of death, Australia, 2019," 2020.

[4] M. N. Sabbagh, L.-F. Lue, D. Fayard, and J. Shi, "Increasing precision of clinical diagnosis of Alzheimer's disease using a combined algorithm incorporating clinical and novel biomarker data," *Neurology and therapy*, vol. 6, no. 1, pp. 83–95, 2017.

[5] R. C. Petersen, R. Doody, A. Kurz, R. C. Mohs, J. C. Morris, P. V. Rabins, K. Ritchie, M. Rossor, L. Thal, and B. Winblad, "Current concepts in mild cognitive impairment," *Archives of neurology*, vol. 58, no. 12, pp. 1985–1992, 2001.

[6] M. Davis, T. O'Connell, S. Johnson, S. Cline, E. Merikle, F. Martenyi, and K. Simpson, "Estimating Alzheimer's disease progression rates from normal cognition through mild cognitive impairment and stages of dementia," *Current Alzheimer Research*, vol. 15, no. 8, pp. 777–788, 2018.

[7] S. Yao, Y. Liu, X. Zheng, Y. Zhang, S. Cui, C. Tang, L. Lu, and N. Xu, "Do nonpharmacological interventions prevent cognitive decline? a systematic review and meta-analysis," *Translational psychiatry*, vol. 10, no. 1, pp. 1–11, 2020.

[8] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.

[9] C. S. Lee, P. G. Nagy, S. J. Weaver, and D. E. Newman-Toker, "Cognitive and system factors contributing to diagnostic errors in radiology," *American Journal of Roentgenology*, vol. 201, no. 3, pp. 611–617, 2013.

[10] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," *Computer methods and programs in biomedicine*, vol. 187, p. 105242, 2020.

[11] C. R. Jack Jr, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, and C. Ward, "The alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *Journal of Magnetic*

Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 27, no. 4, pp. 685–691, 2008.

[12] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, and A. Vlassenko, "OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease," *MedRxiv*, 2019.

[13] J. Morris, "Current vision and scoring rules the clinical dementia rating (CDR)," *Neurology*, vol. 43, pp. 2412–2414, 1993.

[14] J. Dukart, K. Mueller, H. Barthel, A. Villringer, O. Sabri, M. L. Schroeter, A. D. N. 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*, vol. 212, no. 3, pp. 230–236, 2013.

[15] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, A. D. N. Initiative *et al.*, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, 2013.

[16] H.-I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 583–590.

[17] K. Gunawardena, R. Rajapakse, and N. Kodikara, "Applying convolutional neural networks for pre-detection of Alzheimer's disease from structural mri data," in *2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*. IEEE, 2017, pp. 1–7.

[18] A. Farooq, S. Anwar, M. Awais, and M. Alnowami, "Artificial intelligence based smart diagnosis of Alzheimer's disease and mild cognitive impairment," in *2017 International Smart cities conference (ISC2)*. IEEE, 2017, pp. 1–4.

[19] E. Yagis, L. Citi, S. Diciotti, C. Marzi, S. W. Atnafu, and A. G. S. De Herrera, "3D convolutional neural networks for diagnosis of Alzheimer's disease via structural MRI," in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2020, pp. 65–70.

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

[21] M. Liu, J. Zhang, D. Nie, P.-T. Yap, and D. Shen, "Anatomical landmark based deep feature representation for MR images in brain disease diagnosis," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1476–1485, 2018.

[22] K. M. Langa and D. A. Levine, "The diagnosis and management of mild cognitive impairment: a clinical review," *Jama*, vol. 312, no. 23, pp. 2551–2561, 2014.

[23] T. Shen, J. Jiang, Y. Li, P. Wu, C. Zuo, and Z. Yan, "Decision supporting model for one-year conversion probability from MCI to AD using CNN and SVM," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 738–741.

[24] M. Albert, Y. Zhu, A. Moghekar, S. Mori, M. I. Miller, A. Soldan, C. Pettigrew, O. Selnes, S. Li, and M.-C. Wang, "Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years," *Brain*, vol. 141, no. 3, pp. 877–887, 2018.