

Convolutional Autoencoder based Deep Learning Approach for Alzheimer’s Disease Diagnosis using Brain MRI

Ekin Yagis, Alba G. Seco De Herrera, Luca Citi
Computer Science and Electrical Engineering (CSEE)
University of Essex
Colchester, United Kingdom
{e.yagis, alba.garcia, lciti}@essex.ac.uk

Abstract—Rapid and accurate diagnosis of Alzheimer’s disease (AD) is critical for patient treatment, especially in the early stages of the disease. While computer-assisted diagnosis based on neuroimaging holds vast potential for helping clinicians detect disease sooner, there are still some technical hurdles to overcome. This study presents an end-to-end disease detection approach using convolutional autoencoders by integrating supervised prediction and unsupervised representation. The 2D neural network is based upon a pre-trained 2D convolutional autoencoder to capture latent representations in structural brain magnetic resonance imaging (MRI) scans. Experiments on the OASIS brain MRI dataset revealed that the model outperforms a number of traditional classifiers in terms of accuracy using a single slice.

Index Terms—Alzheimer’s Disease, Deep Learning, Image Classification, Autoencoder, MRI, Neuroimaging

I. INTRODUCTION

Alzheimer’s disease (AD) is the most common neurodegenerative disease which progressively impairs cognitive functions [1]. Clinical and neurological evaluations are an integral component of an accurate diagnosis considering the complex nature of disease pathology [2], [3]. The integration of neuroimaging can provide essential additional information and therefore contribute to a more accurate and earlier AD diagnosis. An emerging approach to support AD diagnosis is the computer-assisted diagnosis of AD based on neuroimage analysis [4], [5]. In neurodegenerative diseases research, the clinical understanding of neuroimaging scans can be complex, as brain modifications can be challenging to discern from those due to healthy aging. Especially in the early stages of an illness, detecting disease-related changes from magnetic resonance imaging (MRI) scans could be extremely problematic. Thus, in the last few years, there has been a research interest in modeling the deviation of brain structure due to neurodegeneration [6].

Among those, deep learning-based approaches quickly stand out as they automatically discover discriminative features in the training data collection even when the raw data is used as input [7]. Here, in medical image analysis, one of the biggest challenges is the high dimensionality of the input [8]. For instance, even though there are only several hundred MRIs

in the Open Access Series of Imaging Studies (OASIS), each image has more than six million dimensions ($176 \times 176 \times 208$) [9].

In this work, to parse neuroanatomical alterations in AD, we proposed an end-to-end deep learning approach based on deep convolutional autoencoders (CAE) using MRI [10], [11]. An autoencoder is an artificial neural network built to recreate its input. Deep CAEs consist of two parts. The first component, which is the encoding function of the model, learns how to compress the original input in a latent representation. The second part, known as the decoder, learns to recreate the input data as near as possible to the original using the latent representation. In this work, a 26-layer deep CAE model has been used to retrieve a lower-dimensional representation of the data, which contains all the important information needed to describe the original data point. Then, those latent space representations extracted from brain MRI data are used to differentiate subjects with AD and mild cognitive impairment (MCI) from healthy controls (HC). We are also searching for cerebral atrophy patterns to discover the early changes in the brain characterising AD. To this end, we visualised the intermediate activations across different convolutional layers to understand why the model makes certain decisions.

The rest of the paper is organised as follows: The following section II briefly describes the related work in the literature. In section III, the detailed methodology of the proposed framework is presented. Section IV explains evaluation methods and section V presents the obtained results and following by discussion. Finally Section VI discusses the results and Section VII concludes this work.

II. RELATED WORKS

A set of 2D slices extracted from the MRI volume was used as input to 2D CNN architectures in various studies for the purpose of AD diagnosis [12]–[19]. Among those, not many studies explored the possibility of integrating CAE into their framework to learn an efficient representation of data. Martinez et al. [20] proposed a deep CAE architecture to extract data-driven features and stated that in the case

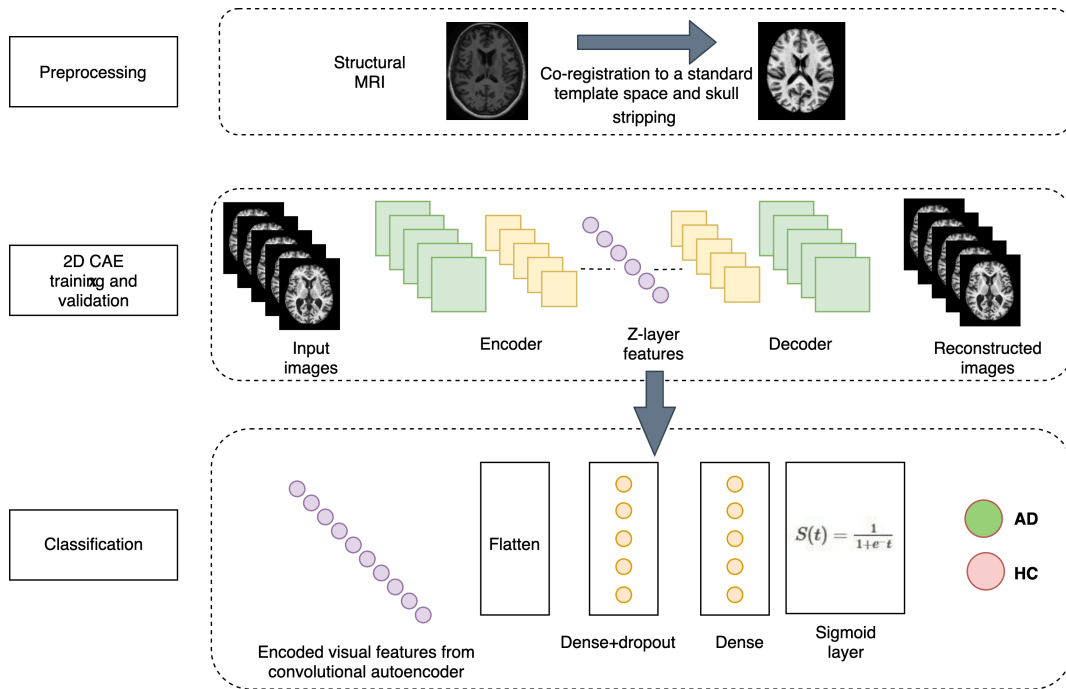


Fig. 1. Overview of the proposed method.

of neuropsychological assessment variables like the Mini-Mental State Exam (MMSE) or the AD Assessment Scale (ADAS11) ratings, imaging-derived markers could forecast clinical variables with correlations above 0.6 [21], [22]. In 2020, Oh et al. [23] used volumetric CAE-based unsupervised learning for the AD vs. HC classification task, then applied supervised transfer learning to solve the progressive mild cognitive impairment (pMCI) vs. stable mild cognitive impairment (sMCI) classification task. Basu et al [24] proposed a model which consists of a 3D convolutional variational autoencoder and a Multi-Layer Perceptron (MLP) to predict the likelihood of the next disease label. Lastly, in 2021 Ferri et al. [25] presented an ANN with stacked autoencoders to differentiate AD and HC using resting-state electroencephalogram (rsEEG), MRI, and rsEEG + MRI features.

III. PROPOSED METHODOLOGY

We propose an end-to-end AD diagnostic framework that extracts latent representations for each class from a brain MRI with a 2D-CAE, then performs classification with a stacked CNN. The methodology is structured by two main components: 2D CAE training/validation for latent space representation and disease classification using latent representation, as shown in Figure 1. In section III-A we present our data selection procedure together with the preprocessing steps. CAE architecture and training strategy is illustrated in section III-B, followed by our classification approach in section III-C. Finally, visualisation of activations is described in section III-D.

A. MRI preprocessing

The publicly available pre-processed version of OASIS data (gain-field corrected, brain masked, and co-registration) (Han et al., 2018) has been used in the experiments. In that version, an atlas-registration-based method was used to create the OASIS brain masks (Marcus et al., 2007). The Talairach and Tournoux atlases were also used for co-registration of each volume. The data matrix size of each pre-processed T1-weighted volume was $176 \times 208 \times 176$, and the voxel size was $1mm \times 1mm \times 1mm$ (Han et al., 2018). From these volumes, we selected the middle axial slice (the 106th) as input for our models. In the work of Mendoza-Léon et al. [26], it has been shown that this axial location corresponds to the anatomical slice, which has a higher degree of disease-associated information due to its high individual content-based image retrieval performance. When the disease label was used as the criteria of interest, the performance results are evaluated by mean average precision values for axial plane [27]. This finding was interpreted as an indication of a higher degree of disease-related knowledge, making them good candidates for a single-slice classification method. However, it should be noted that the number of the selected slice is heavily dependent on the dataset/atlas used. Mendoza-Leon et al. [26] have been used the same dataset; therefore, we followed the same procedure while selecting our single slice candidate. Before feeding the network, the MRI slices in the dataset are normalised in the range [0,1] to receive an unvaried contrast and intensity range.

B. Convolutional autoencoder

Autoencoder is an unsupervised artificial neural network that consists of two parts: an encoder and a decoder. While the encoder tries to learn efficient representations of the input in a reduced dimension, the decoder part of the network reconstructs the input as close to the original as possible using latent representation coming from the encoding part. In other words, an autoencoder aims to learn an approximation to the identity function by minimizing the reconstruction error between input and output. In this work, the Mean Squared Error (MSE) is used as reconstruction error between the input image x and the reconstructed image at the output $\hat{x}_i = g(f(x_i))$:

$$\mathcal{L} = \frac{1}{N} \sum_i (x_i - g(f(x_i)))^2 \quad (1)$$

Autoencoders are mainly used for data dimensionality reduction and image denoising as well as learning latent representations that can be used to generate novel data samples.

The number of convolutional layers, filter size of convolutional layers, and convolutional kernel size are the three main hyperparameters in the CAE. In our model, the encoder has three convolution blocks, where each block has a convolution layer (a kernel size of 3×3) followed by a batch normalization layer. After the first and second convolution blocks, a max-pooling layer (a kernel size of 2×2) is used to downsample the output features of the convolutions. Whereas in the decoder, there are two convolution blocks with convolutional layers (a kernel size of 3×3) with ReLU activations and batch normalization layers. Here, upsampling layers (a kernel size of 2×2) are used after the first and second convolution blocks. Moreover, batch normalization is used to standardise the layer's input for each mini-batch and stabilise the learning process. The details of the network can be seen in Figure 2.

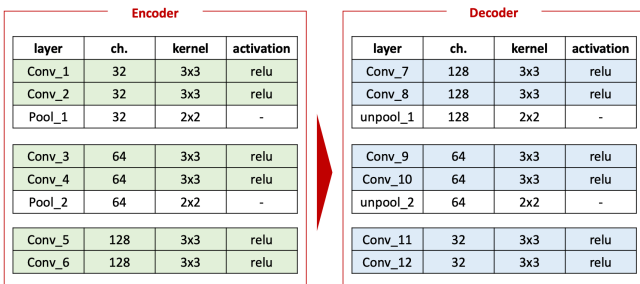


Fig. 2. Detailed architecture of the proposed convolutional autoencoder.

The given model is fit to the training data for 400 epochs with a batch size of 32. Out of 200 subjects, we picked 140 (70 AD, 70 HC) for training the autoencoder. 20% of the training data (28 subjects) were used as validation to control model generalisation, and to interrupt training when generalization stops improving. Thus, train data shape is (112, 176, 176, 1) whereas validation data shape is (28, 176, 176, 1). The remaining 60 subjects have been chosen to use in the

subsequent experiments for testing the full model with unseen patients. We used a first-order gradient-based optimization algorithm called Adam with adaptive learning rates (alpha = 0.0001, beta1 = 0.9, beta2 = 0.999).

C. Classification model

In the classification part, we used the exact same encoder architecture that we used in the convolutional autoencoder model. After the last convolutional layer of the encoder, there is a ‘flatten’ layer in which the two-dimensional matrix of features is flattened into a vector that can be fed into dense layers. Flatten layer is followed by two dense layers with 256 and 128 nodes, respectively. A dropout of 0.2 was added to the first dense layer together with ReLU activation. In the output layer, the sigmoid activation function has been used. In the training process, we first started by freezing the first 15 layers coming from the pre-trained autoencoder and train only the dense layers. Then, we fine-tune all the layers in the second stage. The model has been trained for 400 epochs each time with a batch size of 32 using Adam optimizer with a 0.001 learning rate. Binary cross entropy has been used as a loss function. As overfitting was a big concern for us due to our relatively complex model with a small dataset, dropout regularization has been implemented to prevent the network from overfitting.

D. Visualisation

In the literature, several methods for comprehending and visualising convolutional networks have been created mostly because to interpret the learned features in a neural network [28]–[30]. In the context of our research, the activations or, in other words, feature maps of the network during the forward pass have been used. Feature maps are created by applying filters to the input image or the feature map output of the previous layers. The internal representations for the input for each of the layers in the model are shown by visualisation.

The effect of applying the filters in the first convolutional layer, as seen in Figure 3, is a variety of representations of the axial brain image with various features illuminated. Some draw attention to shapes, while some concentrate on the background or the foreground. The feature maps closest to the model’s input catch a lot of fine detail in the picture, while the feature maps reveal less and less detail as we go further into the model.

IV. EVALUATION

This section reports a detailed description of the dataset together with the range of evaluation methods. In the subsection IV-A, we describe the data collection and dataset used in the experiments in more detail, whereas in the subsection IV-B model training and validation strategies are presented.

A. OASIS dataset description

In this study, we use the Open Access Series of Imaging Studies (OASIS) dataset¹, which is publicly available [9]. The

¹<https://www.oasis-brains.org/>

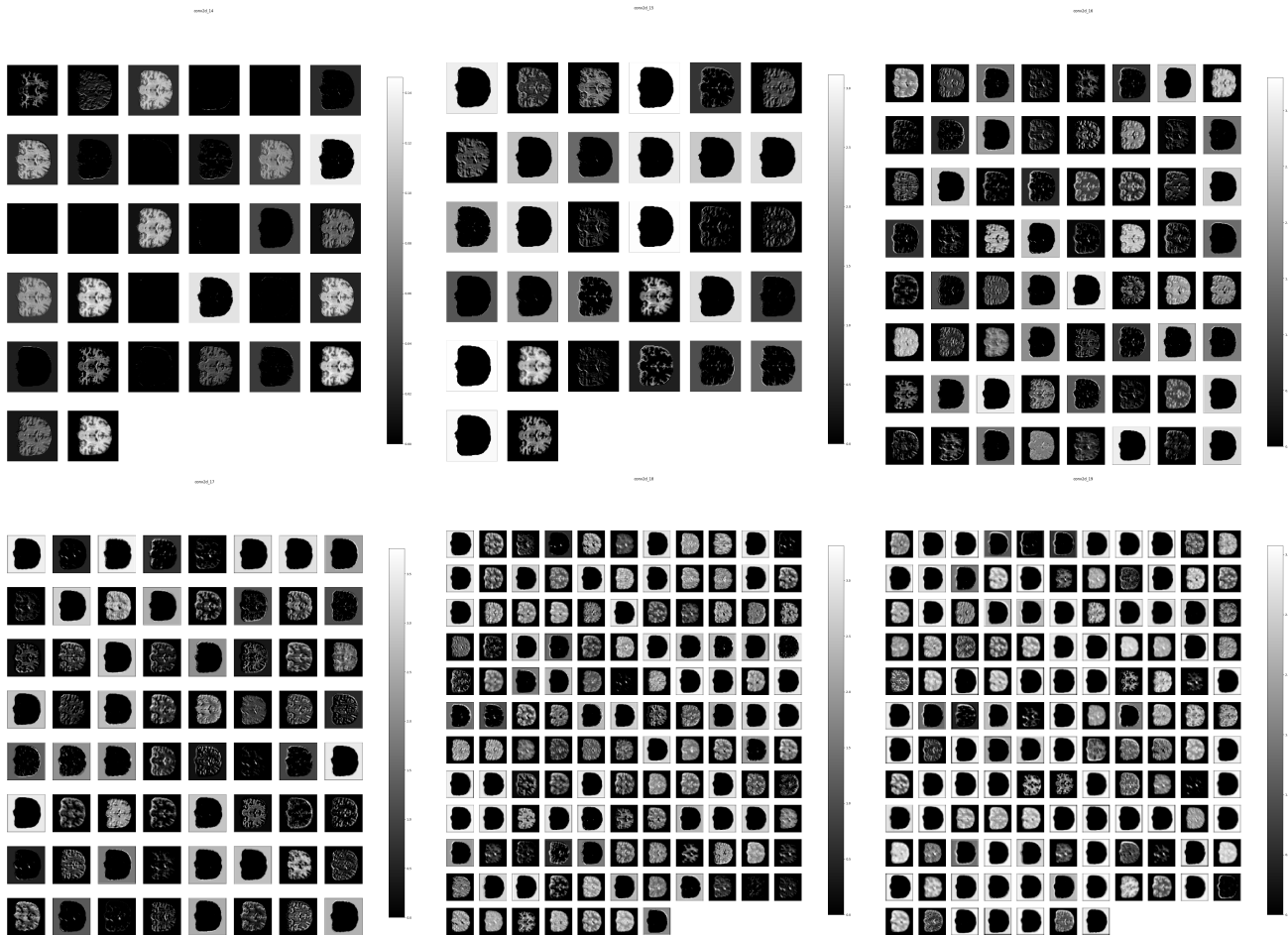


Fig. 3. Visualisation results of selected convolutional layer feature maps. First row, from top to bottom: first, second and third convolutional layers. Second row, from top to bottom: fourth, fifth and sixth convolutional layers.

T1-weighted images of 100 AD patients and 100 HCs have been selected from the OASIS-1 study – a cross-sectional cohort of the OASIS brain MRI dataset (Marcus et al., 2007). In the dataset, there was no substantial difference in age ($p = 0.15$ at t-test), but there was a significant (borderline) difference in gender ($p = 0.04$ at χ^2 -test) between the two classes. The clinical characteristics of the subjects included in this study are summarised in Table I. T1-weighted images were acquired on a 1.5 T MR scanner (Vision, Siemens, Erlangen, Germany) in the sagittal plane using a Magnetization Prepared Rapid Gradient Echo (MPRAGE) series.

The global Clinical Dementia Rating (CDR) score derived from individual CDR ratings is used in OASIS-1 to assess the diagnosis of AD as well as the seriousness of the disorder. On the Clinical Dementia Rating (CDR) scale, MCI is staged at the 0.5 mark. In the scope of our experiments, HCs had the clinical dementia rating (CDR) scores 0, while scores of 0.5 (very mild), 1 (mild), 2 (moderate), and 3 (severe) were all labeled as AD, making classification task more challenging compared to AD vs. HC.

TABLE I
DEMOGRAPHIC FEATURES OF SUBJECTS BELONGING TO OASIS DATASET.

OASIS	Patients	Healthy controls
Number of subjects	100	100
Age (range, years)	62-96	59-94
Age (mean \pm SD, years)	76.70 \pm 7.10	75.50 \pm 9.10
Gender (women/men)	59/41	73/27

B. Model training and validation

The MSE (see Equation 1) has been used as an evaluation measure to show how well the AE is capable of reconstructing unseen images. Moreover, in the scope of this experiment, the peak signal-to-noise ratio (PSNR) has been calculated as a quality measurement between the original and a reconstructed image. A measure of image quality is required when comparing reconstructed outcomes. MSE and PSNR ratio are two widely used metrics. One drawback of MSE is that it is highly dependent on image intensity scaling. By scaling the MSE according to the image range, PSNR prevents this problem, and it is calculated as:

$$\text{PSNR} = 10 \log_{10} \frac{\text{Max}_I^2}{\text{MSE}} \quad (2)$$

where Max_I is the maximum pixel value.

To measure the prediction performance of the model, accuracy and F1 score have been used as evaluation metrics.

The model has been evaluated using an independent test set. The average accuracy is obtained by repeating the full experimentation five times. The architecture was built using Keras (v2.3.1) with TensorFlow backend (v2.0.0) [31], [32]. The training and validation of CNN models were performed on a workstation equipped with an NVidia RTX2080 GPU. The average computational time for model training was 3.2 hours.

V. EXPERIMENTAL RESULTS

In this section, the detailed performance of the end-to-end stacked autoencoder model is reported together with the disease prediction performance.

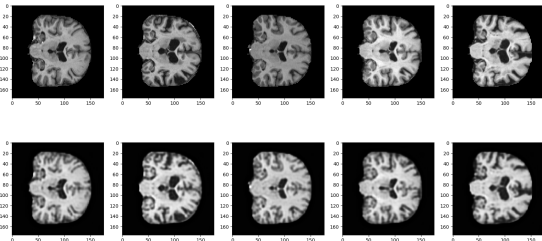


Fig. 4. Sample test images (above) and reconstruction of test images (below).

1) *Reconstruction capability*: The image quality of the restored image increases as the PSNR grows. In our experiments, we achieved 33.34 dB PSNR of reconstructed validation images. The visualisation of sample test images and reconstructed test images can be seen in Figure 4.

2) *Performance of the classification*: The mean percentage \pm standard deviation of accuracy, over 400 iterations are reported in Table II. The autoencoder-based classification model achieves 74.66 ± 0.01 with an F1 score of 0.66.

TABLE II
CLASSIFICATION PERFORMANCE ON THE TEST SET. THE MEAN PERCENTAGE \pm STANDARD DEVIATION OF ACCURACY AND F1 SCORE ARE LISTED.

Model	Parameters	Accuracy	F1- score
CAE+NN	723,905	$74.66 \pm 0.01\%$	0.66

VI. DISCUSSION

We used a single 2D slice of MRI volume for each subject in our framework, which provided many functional benefits. First, it reduced the computational time and resources drastically as the processing complexity, and memory bandwidth demands of 2D CNN models are smaller than 3D CNN models. Second, by using 2D MRI slices, clinical researchers

could take advantage of the most recent CNN architectures, which are often implemented in 2D due to the availability of large 2D image datasets such as ImageNet, CIFAR, and so on.

Given the scale of the networks and the small amount of data used, overfitting was a major concern for us. In order to prevent the network from overfitting, the dropout method is implemented. Moreover, during the experiments, we realised that after epoch 400, the error on training data kept decreasing, whereas validation loss started to increase to a considerable value. Thus, early stopping is employed to avoid overfitting during training.

MCI patients are considered to be at a higher risk of dementia, so they are included in our research. MCI is clinically challenging to distinguish from cognitively stable HC. Thus, in most AD diagnosis frameworks, they are not added to the dataset, which explains the higher accuracy reported in previous works [20].

Compared to other methods where multiple MRI slices are used with well-known deep 2D CNN architectures such as VGG16 and ResNet50, the proposed method outperforms in terms of diagnosis accuracy [18] (In that work, VGG16 model achieved 64.3% whereas ResNet50 obtained 67.1% using transfer learning). The experimental results show that when latent representations are learned in a way that promotes sparsity, classification accuracy improves [33].

VII. CONCLUSION

An early and precise AD diagnosis is needed, and an automated diagnostic tool helped by neuroimaging data analysis would offer a more detailed and effective solution, as well as potentially increase diagnostic precision. In this paper, we presented a CAE-based deep learning method for classifying AD+MCI vs. HC subjects using single 2D brain MRI slices. Experimental results on the test set demonstrated the effectiveness of the proposed method in the classification of AD+MCI vs. HC. As far as classification accuracy is concerned, our approach outperforms the conventional approach where deep CNNs use whole MRI slices as input instead of the latent representations [18]. By integrating supervised prediction and unsupervised representation together, the model achieves 74.66% classification accuracy using only one MRI slice for each subject. As future work, we will extend our current model to 3D to extract volumetric latent representation for MCI and HC classification while trying to expand our datasets and perform a deeper analysis.

ACKNOWLEDGMENT

This work was funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824153 “POTION”.

OASIS data were provided by OASIS: Cross-Sectional: Principal Investigators: D. Marcus, R. Buckner, J. Csernansky J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382.

REFERENCES

- [1] S. Hague, S. Klaffke, and O. Bandmann, "Neurodegenerative disorders: Parkinson's disease and huntington's disease," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, no. 8, pp. 1058–1063, 2005.
- [2] M. Pievani, W. de Haan, T. Wu, W. W. Seeley, and G. B. Frisoni, "Functional network disruption in the degenerative dementias," *The Lancet Neurology*, vol. 10, no. 9, pp. 829–843, 2011.
- [3] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter, "The brain's default network: anatomy, function, and relevance to disease," 2008.
- [4] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick, "Detection of prodromal alzheimer's disease via pattern classification of magnetic resonance imaging," *Neurobiology of aging*, vol. 29, no. 4, pp. 514–523, 2008.
- [5] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative, *et al.*, "Deep ensemble learning of sparse regression models for brain disease diagnosis," *Medical image analysis*, vol. 37, pp. 101–113, 2017.
- [6] M. Habes, R. Pomponio, H. Shou, J. Doshi, E. Mamourian, G. Erus, I. Nasrallah, L. J. Launer, T. Rashid, M. Bilgel, *et al.*, "The brain chart of aging: Machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the istaging consortium of 10,216 harmonized mr scans," *Alzheimer's & Dementia*, vol. 17, no. 1, pp. 89–102, 2021.
- [7] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.
- [8] J. Weese and C. Lorenz, "Four challenges in medical image analysis from an industrial perspective," 2016.
- [9] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults," *Journal of cognitive neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.
- [10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," tech. rep., California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [11] Y. Le Cun, "Learning process in an asymmetric threshold network," in *Disordered systems and biological organization*, pp. 233–240, Springer, 1986.
- [12] A. Farooq, S. Anwar, M. Awais, and S. 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)*, pp. 1–6, IEEE, 2017.
- [13] 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)*, pp. 1–7, IEEE, 2017.
- [14] M. Hon and N. M. Khan, "Towards alzheimer's disease classification through transfer learning," in *2017 IEEE International conference on bioinformatics and biomedicine (BIBM)*, pp. 1166–1169, IEEE, 2017.
- [15] 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, p. 2, 2018.
- [16] A. Valliani and A. Soni, "Deep residual nets for improved alzheimer's diagnosis," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 615–615, 2017.
- [17] S.-H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. 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*, vol. 42, no. 5, p. 85, 2018.
- [18] E. Yagis, A. G. S. De Herrera, and L. Citi, "Generalization performance of deep learning models in neurodegenerative disease classification," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1692–1698, IEEE, 2019.
- [19] T. Jo, K. Nho, and A. J. Saykin, "Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data," *Frontiers in aging neuroscience*, vol. 11, p. 220, 2019.
- [20] F. J. Martinez-Murcia, A. Ortiz, J.-M. Gorriz, J. Ramirez, and D. Castillo-Barnes, "Studying the manifold structure of alzheimer's disease: A deep learning approach using convolutional autoencoders," *IEEE journal of biomedical and health informatics*, vol. 24, no. 1, pp. 17–26, 2019.
- [21] W. G. Rosen, R. C. Mohs, and K. L. Davis, "A new rating scale for alzheimer's disease," *The American journal of psychiatry*, 1984.
- [22] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "'mini-mental state': a practical method for grading the cognitive state of patients for the clinician," *Journal of psychiatric research*, vol. 12, no. 3, pp. 189–198, 1975.
- [23] K. Oh, Y.-C. Chung, K. W. Kim, W.-S. Kim, and I.-S. Oh, "Classification and visualization of alzheimer's disease using volumetric convolutional neural network and transfer learning," *Scientific Reports*, vol. 9, no. 1, pp. 1–16, 2019.
- [24] S. Basu, K. Wagstyl, A. Zandifar, D. L. Collins, A. Romero, and D. Precup, "Early prediction of alzheimer's disease progression using variational autoencoders," in *MICCAI (4)*, pp. 205–213, 2019.
- [25] R. Ferri, C. Babiloni, V. Karami, A. I. Triggiani, F. Carducci, G. Noce, R. Lizio, M. T. Pascarelli, A. Soricelli, F. Amenta, *et al.*, "Stacked autoencoders as new models for an accurate alzheimer's disease classification support using resting-state eeg and mri measurements," *Clinical Neurophysiology*, vol. 132, no. 1, pp. 232–245, 2021.
- [26] R. Mendoza-Léon, J. Puentes, L. F. Uriza, and M. H. Hoyos, "Single-slice alzheimer's disease classification and disease regional analysis with supervised switching autoencoders," *Computers in biology and medicine*, vol. 116, p. 103527, 2020.
- [27] R. A. M. Leon, J. Puentes, F. A. González, and M. H. Hoyos, "Empirical evaluation of general-purpose image features for pathology-oriented image retrieval of alzheimer disease cases," in *CARS 2016: 30th International Congress on Computer Assisted Radiology and Surgery In: International Journal of Computer Assisted Radiology and Surgery*, vol. 11, pp. S39–S40, Springer, 2016.
- [28] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.
- [29] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*, pp. 818–833, Springer, 2014.
- [30] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.
- [31] F. Chollet *et al.*, "keras," 2015.
- [32] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pp. 265–283, 2016.
- [33] A. Makhzani and B. Frey, "K-sparse autoencoders," *arXiv preprint arXiv:1312.5663*, 2013.