



**HAL**  
open science

## Improving alzheimer's diagnosis using vision transformers and transfer learning

Marwa Zaabi, Mohamed Ibn Khedher, Mounim El Yacoubi

► **To cite this version:**

Marwa Zaabi, Mohamed Ibn Khedher, Mounim El Yacoubi. Improving alzheimer's diagnosis using vision transformers and transfer learning. 16th International Conference on Human System Interaction (HSI), Jul 2024, Paris, France. 10.1109/HSI61632.2024.10613527 . hal-04621338

**HAL Id: hal-04621338**

**<https://hal.science/hal-04621338v1>**

Submitted on 24 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# Improving Alzheimer’s Diagnosis using Vision Transformers and Transfer Learning

Marwa Zaabi\*, Mohamed Ibn Khedher† and Mounim A. El-Yacoubi‡

\*CEM Laboratory, ENIS, Sfax University, Tunisia

marwa.zaabi@isimg.tn

†IRT - SystemX, 2 Bd Thomas Gobert, 91120 Palaiseau, France

mohamed.ibn-khedher@irt-systemx.fr

‡Samovar, Telecom SudParis, Institut Polytechnique de Paris, 19 place Marguerite Perey, 91120 Palaiseau, France

mounim.el\_yacoubi@telecom-sudparis.eu

**Abstract**—Alzheimer’s disease is a neurodegenerative disorder defined by memory loss and primarily affects older individuals. Currently, there is no definitive cure available. Although medications are accessible, they only serve to slow the progression of the disease. In this paper, we propose the use of Vision Transformers and Transfer Learning for Alzheimer’s classification. Our approach leverages the temporal aspect of the transformer to model the correlation between different image patches. Transfer learning enables us to mitigate the issue of insufficient available data. Our method has been validated on the OASIS dataset, which consists of 250 brain scans. The results demonstrate that transfer learning with Transformer models surpasses the performance of transfer learning with CNN models by 4% and exceeds traditional CNN models without transfer learning by 8%. Two types of Transformers were tested: ViT-B16 and ViT-B32. The results are comparable, with ViT-B32 outperforming ViT-B16 by 1%.

**Index Terms**—Alzheimer’s disease, Vision Transformer, Transfer Learning.

## I. INTRODUCTION

The timely detection of Alzheimer’s disease (AD) is essential for effective interventions, and neuroimaging stands out as a particularly promising method for this purpose. AD, a neurodegenerative disorder, progressively damages brain connections, leading to memory loss and cognitive decline. While a cure for AD remains elusive, prompt treatment upon diagnosis can greatly mitigate symptom severity, offering relief to patients and alleviating their suffering. Undoubtedly, the primary risk factor for AD is advancing age, typically after the age of 65.

Recently, the integration of Artificial Intelligence (AI) has been pivotal in steering healthcare towards a preventive model, enabling the foresight and early detection of a spectrum of conditions, thus streamlining treatment timelines. This is particularly pertinent for neurodegenerative diseases, where AI’s role is significant in aiding early diagnosis and intervention, crucial for conditions such as AD. AI-powered innovations are also creating supportive technologies, including mobile apps, wearable devices, and interactive chatbots, to provide essential assistance to patients.

An effective approach for early AD detection based on AI consists of two important choices. The first choice concerns

the data modality used, which can include Speech, Gait, Handwriting, or Neuroimaging data. The second choice involves selecting the AI model, which can be based on Deep Learning or traditional algorithms.

Several modalities offer complementary insights for detecting and monitoring AD progression. Neuroimaging, particularly Magnetic Resonance Imaging (MRI), offers through visuals of brain structures, which helps in identifying anatomical changes and atrophy typical of AD [1], [2]. Clinical and neuropsychological data including memory tests and cognitive assessments reveal deficits typical of AD. Analyses of gait and movement can also indicate alterations in physical actions, while changes in speech and language usage might suggest cognitive decline. Lastly, changes in handwriting, such as the loss of writing skills or alterations in letter shaping, may also indicate cognitive decline [3], [4], [5], [6].

Machine Learning (ML) algorithms can be broadly categorized into two groups: traditional ML algorithms and deep learning (DL) algorithms. Traditional ML algorithms, including Support Vector Machines, XGBoost, and Random Forests are often preferred when dealing with pre-processed data, where handcrafted features have been extracted for classification. In contrast, DL algorithms have recently gained traction due to their capability to autonomously acquire feature representations from raw data, thereby often yielding superior performance. Prominent examples include Convolutional Neural Networks [7], Vision Transformers [8] and Autoencoders [9]. These models excel in handling complex data patterns, which is particularly beneficial for interpreting the subtle and intricate changes associated with AD.

Deep Neural Networks (DNNs) capable of acquiring intricate patterns and structures from data. However, neural networks trained from the ground up may encounter challenges in generalizing effectively with limited training data. Transfer learning enables leveraging knowledge acquired from one task to improve generalization on another task. It involves transferring the acquired weights from a network trained on task *A* to improve performance on a new task *B*. The basic concept involves utilizing a model’s knowledge acquired from a task rich in labeled training data to aid in a new task with limited data [1].

Our approach relies on two key strategies. On one hand, we take advantage of the potential of neuroimaging techniques, particularly MRI, which are non-invasive methods capable of detecting structural changes in the brain. On the other hand, we leverage the potential of transformers [10], which are new DNNs capable of modeling the correlations between different areas of the image. We combine the transformer with transfer learning techniques to overcome the limitations posed by the small size of the dataset.

The primary contribution of our paper is the strategic extraction of Regions Of Interest (ROIs), containing the hippocampal areas, rather than utilizing entire images. Following this extraction, we evaluated two architectures of transformers, ViT-B16 and ViT-B32, for classification.

This paper is structured as follows: Section II provides an overview of the pertinent literature. In Section III, we describe our suggested methodology. Section IV expounds on the results achieved and provides a discussion of our findings. Lastly, Section V outlines our conclusions and proposes possible avenues for further research.

## II. RELATED WORK

An approach to early AD detection generally consists of two parts. The first part involves the modalities used. This stage includes selecting the modality and extracting the features. This extraction can be automated through DL or handcrafted features. The second part concerns the classifiers. The choice of classifier may depend on the modality selected. In the following, we present a state-of-the-art review of both parts.

### A. Modalities and Feature Extraction

As mentioned in the introduction, a variety of modalities are utilized for analysis, including neuroimaging, speech, gait, handwriting, and clinical data.

Neuroimaging includes MRI scans, which offer detailed images of the brain's anatomy and is instrumental in identifying structural changes such as atrophy in the hippocampus and other brain regions associated with AD [11]. Positron Emission Tomography (PET) scans are instrumental in observing brain metabolism and the presence of amyloid plaques, a hallmark of the disease [12]. Cerebrospinal Fluid (CSF) analysis, though more invasive, identifies specific biomarkers such as tau and beta-amyloid proteins [13]. In most Neuroimaging-based approaches, whole images are utilized. However, there are instances where the images are segmented to extract local features. Among these local features, techniques such as FreeSurfer [14] are used for detailed segmentation.

Another modality for Early AD detection is handwriting analysis. In collaboration with Broca Hospital in Paris, the authors of [15] propose characterizing AD by analyzing online handwritten cursive loops. This study models the full dynamics of loop velocity trajectories in an unsupervised manner. To find clusters, the authors use dynamic time warping as a dissimilarity metric and temporal clustering with K-medoids.

Concerning speech analysis, the authors explore the potential of using vocal patterns to detect early signs of AD.

They focus on analyzing variations in speech features such as pitch, rate, and pauses, which can be indicative of cognitive impairment [16], [17]. In [18], the authors divided the audio recordings into segments and extracted spectrograms from each segment. These features were subsequently combined into a single input vector for algorithmic classification.

### B. Classification based DL algorithms

Classifiers can be divided into DL and traditional ML algorithms. This section will focus on DL-based classifiers due to their superior ability to handle complex and high-dimensional data, such as medical images. We identify three main DL applications. The first is for extracting features and classification. This is the classic use where DL is applied both to extract features and to perform classification [19]. In the second, DL first serves to extract discriminant features, called embeddings, which are then input to a shallow ML model to perform classification [20]. In the third, DL model training is based on Transfer Learning, which uses big datasets to leverage pre-trained models, to specifically address the challenges posed by AD detection [1].

In terms of architectures, CNN has gained significant traction in the realm of AD diagnosis. In [21], a 3D CNN approach was employed on PET scans, while the same modality was utilized with a 2D CNN [22]. Subsequent advancements were made by [23], who enhanced the architecture of CNN and introduced a multi-stream CNN for classification. Lian et al. [24] employed a deep learning technique to identify AD using structural MRI brain data. To recognize the AD, a fully deep learning architecture structure with a hierarchical foundation was developed. In order to distinguish the AD image from the non-AD image, the built internally generated characteristics were categorized by the fully deep learning architectural structure. In a different study, Yigit [25] divided slice data to simplify 3D to 2D using structural magnetic resonance imaging (SMRI) as the input. After that, they used CNN to identify AD. Their model achieved an 80% classification accuracy on average.

Several other studies have explored different CNN architectures tailored for AD diagnosis. Hussain et al. [26] introduced a multiple layer CNN model specifically designed for AD and compared its effectiveness against established architectures (InceptionV3, Xception, MobilenetV2, and VGG). Similarly, Erdogmus and Kabakus [27] proposed another CNN architecture, achieving a binary classification accuracy of 90.4%. This study adds to the growing body of research aiming to improve diagnostic accuracy. Manimurugan [28] employed fine-tuning techniques to train the VGG-19 architecture, achieving an impressive accuracy of 95.82%. This showcases the effectiveness of fine-tuning in optimizing model performance for AD diagnosis. In [29], the authors proposed the enhancing of VGG16 through the integration of a convolutional block attention module. The authors of [30] modified the densely connected neural network by incorporating an attention mechanism to delve deeper into feature extraction.

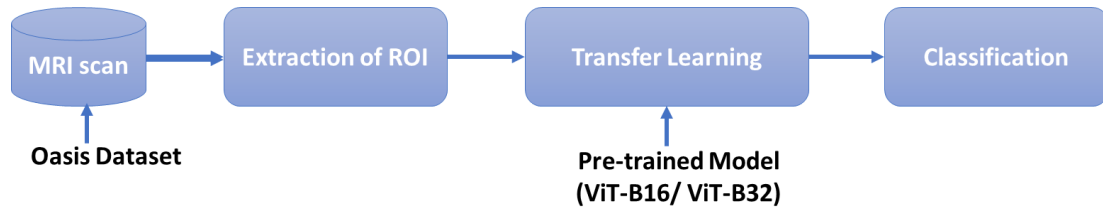


Fig. 1: Flowchart of our proposed approach

Recent studies have tackled the limitations of CNNs by incorporating transformers to extract global features and create long-range connections between image features. The ViT architecture was applied to PET scans in [31]. In [32], a 3D deformable self-attention module was incorporated by the authors into the ViT architecture. Adopting a fusion-based strategy, the authors of [33] combined the strengths of ViT, 2D CNN, and 3D CNN, enabling the effective capture of global dependencies, which resulted in improved model generalization.

To distinguish our research from the existing state of the art, we have selected MRI as input modality for its high diagnostic precision and non-invasive nature. The choice of transformers as classifiers is motivated by their capacity to identify intricate, far-reaching relationships in the data. To counteract the limitation of small dataset sizes, we have incorporated transfer learning strategies, enhancing model generalization by utilizing knowledge from larger datasets.

### III. METHOD

Our methodology comprises three primary phases (Figure 1). First, we extract the region of interest, containing the hippocampus, to focus on the most relevant area for analysis. This targeted step ensures that we are extracting the most critical part of the brain related to AD. Next, we use a pre-trained Transformer model, which we fine-tune to adapt to the specifics of our dataset. Fine-tuning enables the model to accurately adapt to the specific features of our dataset, thereby improving its performance. Finally, we proceed to the classification stage, where the model distinguishes between patients having AD or not, using usual metrics such as accuracy. We describe next each of these steps in detail.

#### A. Extraction of Regions of Interest

The hippocampus is a small region in the brain primarily associated with memory and spatial navigation. The hippocampus is one of the first regions to suffer damage following AD; early symptoms typically include memory loss and disorientation. Our motivation is to target the most relevant regions of the image rather than using the entire image.

First, all images are normalized to the same size. Then, each image is divided into patches of 32x32 pixels. Visually, we have identified that the hippocampus corresponds precisely to patches numbered from 43 to 46.

Figure 2 illustrates these three steps. An image is first divided into square patches of 32x32 pixels. Subsequently, the patches

corresponding to the hippocampus are extracted and serve as the input for the classifier. By focusing on these specific patches, we can efficiently extract the regions that include the hippocampus, leading to more accurate classification.

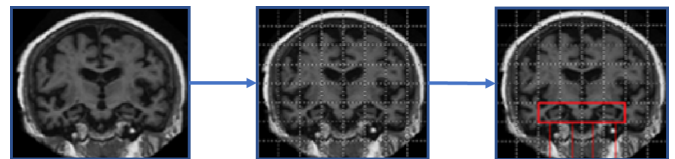


Fig. 2: Extraction of ROI. From left to right: Original image, Segmentation into patches, Extracting hippocampus regions.

#### B. Vision Transformer

The Vision Transformer (ViT), a specialized variant of the transformer model, is tailored for computer vision tasks. At its core lie fundamental components: patch embedding, transformer encoder, and a multilayer perceptron (MLP) classifier. Crucially, the self-attention layer within the Vision Transformer addresses a key drawback of traditional CNNs by dynamically adjusting feature importance. This mechanism empowers the network to better grasp the pivotal connections and interdependencies among image patches, thus enhancing overall performance.

The input image is divided into a grid of smaller sub-image patches by the ViT. These patches are then flattened and processed through a feedforward layer. This layer treats each patch individually, projecting its features and consolidating them into a large-sized feature vector known as patch embedding [34].

Next, position embeddings are incorporated into the sequence of image patches through linear addition. This step preserves positional information by providing details on the relative or absolute positions of the patches within the sequence. Subsequently, the patches embedded are processed through within a vision encoder, which includes both a Multi-Head Self-Attention (MSA) layer and a MLP layer [34].

The MSA layer partitions the input into multiple heads to extract intricate features from the image, thereby generating a comprehensive global representation. This layer repeatedly employs self-attention within each head. The outputs from every heads are then merged through concatenation and forwarded into the MLP for classification.

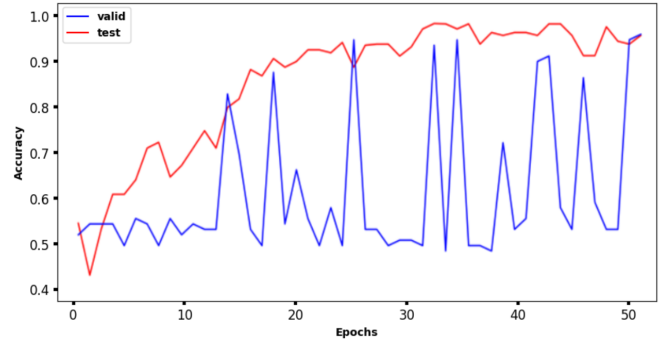
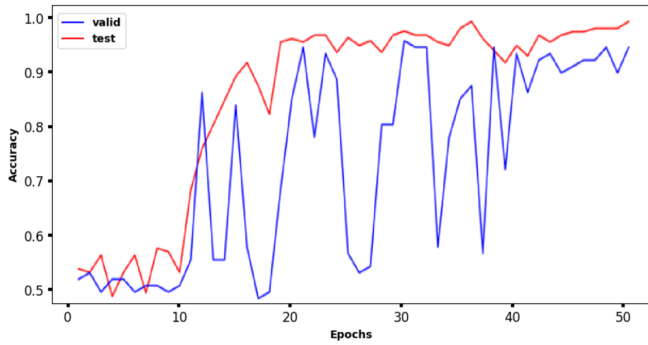


Fig. 3: Evolution of Accuracy on Training and Validation Sets Over Epochs: Left on ViT-B16 and Right on ViT-B32

### C. Transfer Learning based on pretrained ViT

Transfer learning using pretrained Vision Transformers (ViTs) has proven highly effective in various tasks. By leveraging the knowledge embedded in ViTs, which have been trained on extensive image datasets, models can quickly adapt to new, specific tasks with significantly less data and computational resources. Using the pre-trained model's learnt features reduces the requirement to create a new model from scratch. This approach enhances performance, reduces training time, and mitigates the need for large, labeled datasets, making it a powerful technique for applications such as image classification. In this study, we examine two transfer learning schemes, based on the ViT B16 and ViT B32 models.

1) *ViT-B16*: It is a pre-trained ViT that was trained using more than a million images from the ImageNet dataset. It excels in generating comprehensive and resilient image representations, proving valuable across diverse computer vision applications. The "B16" designation in the model signifies its size, with "16" indicating the patch size used to split input images. The ViT-B16 model consists of 16 transformer blocks, with each input image divided into non-overlapping patches with size 16x16 pixels [35]. Regarding our architecture, the base is the ViT-B16, outputting a 768-dimensional vector as its output. The base is followed by a set of layers, especially dense layers. We added three dense layers with 128, 64, and 32 neurons respectively, followed by a single-neuron output layer.

2) *ViT-B32*: It is similar to ViT-B16, both of which are pre-trained on the ImageNet dataset. It uses a patch size of 32x32 pixels to divide the image. Architecturally, it consists the base, which is the ViT-B-32, followed by the same series of additional layers as those used for the ViT-B32.

## IV. EXPERIMENTAL RESULTS

### A. Dataset

The OASIS (Open Access Series of Imaging Studies) dataset [36] is used for evaluating our approach. OASIS contains 250 MRI neuroimaging scans, of which 130 depict brains affected by AD and 120 depict brains not affected by AD. For the evaluation protocol, 20% of the data is used for testing, while the remaining 80% is used for training.

TABLE I: Comparison of classification rates

Approach	Model	Description	Accuracy (%)
[2]	CNN	Classic CNN for classification	88.10
[2]	CNN	Transfer Learning on Alexnet	92.86
[20]	CNN	CNN combined with SVM	94.44
Our approach	Transformer	Transfer Learning on ViT-B32	96.25
	Transformer	Transfer Learning on ViT-B16	95.00

The OASIS dataset was collected from participants spanning a wide range of ages and various health conditions. As an evaluation metric, we used the usual metric accuracy.

### B. Results and Discussion

In this section, we detail the results of our method based on Transformer learned via Transfer Learning on the OASIS dataset in terms of accuracy. Table I shows the comparison of our approach with the state of the art. Figure 3 show the progression of accuracy on the training and validation datasets during the first 50 epochs of training.

In this study, three observations can be drawn regarding the application of transfer learning and Transformer models. First, the application of transfer learning enhanced the performance of CNN models by 4%, demonstrating the importance of this method in improving existing models. Second, when comparing transfer-based approaches, Transformer models outperformed CNN models by 4%, highlighting their superior capability to model the correlation between different image patches. Finally, regarding the settings of the Transformers, optimizing hyperparameters showed that changing the batch size from 16 to 32 results in a performance gain of 1%, a modest but noteworthy improvement in critical fields such as healthcare.

## V. CONCLUSION

In this work, we have investigated the power of combining Transformers and Transfer Learning to enhance AD classification performance. As a preliminary step, we proposed focusing

solely on the most crucial regions in the brain that contain the Hippocampus rather than using the entire image. Our investigation convincingly demonstrated that the use of Transformers significantly outperforms CNNs. This outcome underscores the capability of Transformers to effectively model the correlations among different patches of brain images, which is critical for capturing nuanced patterns associated with AD.

In the future work, we aim to extend the validation of our method on a large number of datasets, which will help in establishing the robustness and generalizability of our model. Additionally, we will delve deeper into optimizing the parameters of the transformer to further enhance its accuracy and efficiency [37], [38], [39], [40], [41].

## REFERENCES

- [1] Karim Haddada, Mohamed Ibn Khedher, and Olfa Jemai. Comparative study of deep learning architectures for early alzheimer detection. In *2023 International Conference on Cyberworlds (CW)*, pages 185–192. IEEE, 2023.
- [2] Marwa Zaabi, Nadia Smaoui, Houda Derbel, and Walid Hariri. Alzheimer’s disease detection using convolutional neural networks and transfer learning based methods. In *2020 17th International Multi-Conference on Systems, Signals & Devices (SSD)*, pages 939–943. IEEE, 2020.
- [3] Mounim A. El-Yacoubi, Sonia Garcia-Salicetti, Christian Kahindo, Anne-Sophie Rigaud, and Victoria Cristancho-Lacroix. From aging to early-stage alzheimer’s: Uncovering handwriting multimodal behaviors by semi-supervised learning and sequential representation learning. *Pattern Recognition*, 86, 2019.
- [4] Nickson Mwamsojo, Frederic Lehmann, Mounim A. El-Yacoubi, Kamel Merghem, Yann Frignac, Badr-Eddine Benkelfat, and Anne-Sophie Rigaud. Reservoir computing for early stage alzheimer’s disease detection. *IEEE Access*, 10:59821–59831, 2022.
- [5] Quang Dao, Mounim A. El-Yacoubi, and Anne-Sophie Rigaud. Detection of alzheimer disease on online handwriting using 1d convolutional neural network. *IEEE Access*, 11:2148–2155, 2023.
- [6] Jana Sweidan, Mounim A. El-Yacoubi, and Anne-Sophie Rigaud. Explainability of cnn-based alzheimer’s disease detection from online handwriting. *Research Square preprint*, April 2024.
- [7] Ahmad Waleed Salehi, Preety Baglat, Brij Bhushan Sharma, Gaurav Gupta, and Ankita Upadhy. A cnn model: earlier diagnosis and classification of alzheimer disease using mri. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pages 156–161. IEEE, 2020.
- [8] Andreas Miltiadous, Emmanouil Gionanidis, Katerina D Tzimourta, Nikolaos Giannakeas, and Alexandros T Tzallas. Dice-net: a novel convolution-transformer architecture for alzheimer detection in eeg signals. *IEEE Access*, 2023.
- [9] Pham Minh Tuan, Trong-Le Phan, Mouloud Adel, Eric Guedj, and Nguyen Linh Trung. Autoencoder-based feature ranking for alzheimer disease classification using pet image. *Machine Learning with Applications*, 6:100184, 2021.
- [10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [11] Shagun Sharma, Kalpna Guleria, Sunita Tiwari, and Sushil Kumar. A deep learning based convolutional neural network model with vgg16 feature extractor for the detection of alzheimer disease using mri scans. *Measurement: Sensors*, 24:100506, 2022.
- [12] Taeho Jo, Kwangsik Nho, Shannon L Risacher, Andrew J Saykin, and Alzheimer’s Neuroimaging Initiative. Deep learning detection of informative features in tau pet for alzheimer’s disease classification. *BMC bioinformatics*, 21:1–13, 2020.
- [13] Elena Ryzhikova, Nicole M Ralbovsky, Vitali Sikirzhitski, Oleksandr Kazakov, Lenka Halamkova, Joseph Quinn, Earl A Zimmerman, and Igor K Lednev. Raman spectroscopy and machine learning for biomedical applications: Alzheimer’s disease diagnosis based on the analysis of cerebrospinal fluid. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 248:119188, 2021.
- [14] Donatas Sederevičius, Didac Vidal-Piñeiro, Øystein Sørensen, Koen van Leemput, Juan Eugenio Iglesias, Adrian V Dalca, Douglas N Greve, Bruce Fischl, Atle Bjørnerud, Kristine B Walhovd, et al. Reliability and sensitivity of two whole-brain segmentation approaches included in freesurfer–aseg and samseg. *Neuroimage*, 237:118113, 2021.
- [15] Christian Kahindo, Mounim El Yacoubi, Sonia Garcia-Salicetti, Anne-Sophie Rigaud, and Victoria Cristancho-Lacroix. Characterizing early-stage alzheimer through spatiotemporal dynamics of handwriting. *IEEE Signal Processing Letters*, PP:1–1, 01 2018.
- [16] Jane Marshall, Jo Atkinson, Alice Thacker, and Bencie Woll. Is speech and language therapy meeting the needs of language minorities? the case of deaf people with neurological impairments. *International Journal of Language & Communication Disorders*, 38(1):85–94, 2003.
- [17] Karmele López-de Ipiña, Jesús B Alonso, Nora Barroso, Marcos Faundez-Zanuy, Miriam Ecay, Jordi Solé-Casals, Carlos M Travieso, Ainara Estanga, and Aitzol Ezeiza. New approaches for alzheimer’s disease diagnosis based on automatic spontaneous speech analysis and emotional temperature. In *International Workshop on Ambient Assisted Living*, pages 407–414. Springer, 2012.
- [18] Lin Liu, Shenghui Zhao, Haibao Chen, and Aiguo Wang. A new machine learning method for identifying alzheimer’s disease. *Simulation Modelling Practice and Theory*, 99:102023, 2020.
- [19] Qiuhui Chen, Qiang Fu, Hao Bai, and Yi Hong. Longformer: Longitudinal transformer for alzheimer’s disease classification with structural mris. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3575–3584, January 2024.
- [20] Amira Ben Rabeh, Faouzi Benzarti, and Hamid Amiri. Cnn-svm for prediction alzheimer disease in early step. In *2023 International Conference on Control, Automation and Diagnosis (ICCAD)*, pages 1–6. IEEE, 2023.
- [21] K.T. Oh M.S. Byun D. Yi J.H. Lee Y.K. Kim B.S. Ye M.J. Yun D.Y. Lee S. Kim, P. Lee and Y. Jeong. Deep learning-based amyloid pet positivity classification model in the alzheimer’s disease continuum by using 2-[18f]fdg pet. *EJNMMI*, 56, 2021.
- [22] M.T.B. Othman I. Ullah R.A. Khan Y.-K. Ma D. Adhikari A.U. Rehman M. Shafiq H. Hamam A.B. Tufail, N. Anwar. Early-stage alzheimer’s disease categorization using pet neuroimaging modality and convolutional neural networks in the 2d and 3d domains. *Advances in IoMT for Healthcare Systems*, 12:4609, 2022.
- [23] M.M. Dehshibi M. Ashtari-Majlan, A. Seifi. A multi-stream convolutional neural network for classification of progressive mci in alzheimer’s disease using structural mri images. *IEEE J. Biomed. Health Inform.*, 26:3918–3926, 2022.
- [24] J. Zhang D. Shen C. Lian, M. Liu. Hierarchical fully convolutional network for joint atrophy localization and alzheimer’s disease diagnosis using structural mri. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42:880–893, 2020.
- [25] Z. Isik A. Yigit. Applying deep learning models to structural mri for stage prediction of alzheimer’s disease. *Turkish Journal of Electrical Engineering and Computer Sciences*, 28:196–210, 2020.
- [26] S. Z. Hassan T. H. Azmi M. A. Rahman E. Hussain, M. Hasan and M. Z. Parvez. Deep learning based binary classification for alzheimer’s disease detection using brain mri images. *IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 1115–1120, 2020.
- [27] P. Erdogmus and A. T. Kabakus. The promise of convolutional neural networks for the early diagnosis of the alzheimer’s disease. *Engineering Applications of Artificial Intelligence*, 123:106254, 2023.
- [28] S.Manimurugan. Classification of alzheimer’s disease from mri images using cnn based pre-trained vgg-19 model. *Journal of Computational Science and Intelligent Technologies*, 2:34–41, 2020.
- [29] M. Yang Y.-D. Zhang S.-H. Wang, Q. Zhou. Advian: Alzheimer’s disease vgg inspired attention network based on convolutional block attention module and multiple way data augmentation. *Front. Aging Neurosci.*, 13, 2021.
- [30] A. Gao X. Feng-D. Liang X. Long J. Zhang, B. Zheng. A 3d densely connected convolution neural network with connection-wise attention mechanism for alzheimer’s disease classification. *Magn. Reson. Imaging*, 78:119–126, 2021.
- [31] Y. Zhang S. Khanal-A.-L. Lin N. Jacobs X. Xing, G. Liang. Vadvit: vision transformer on multi-modality pet images for alzheimer disease diagnosis. *IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4, 2022.
- [32] P. Xu Z. Chen-W. Li X. Yuan G. Zhong C.-M. Pun Z. Huang Q. Zhao, G. Huang. Ida-net: inheritable deformable attention network of structural

- mri for alzheimer's disease diagnosis, biomed. *Signal Process. Control*, 84:104787, 2023.
- [33] D. Hwang J. Jang. M3t: three-dimensional medical image classifier using multiplane and multi-slice transformer. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, page 20686–20697, 2022.
- [34] M. Tmar F. Gargouri R. Kadri, B. Bouaziz. Efficient multimodel method based on transformers and coatnet for alzheimer's diagnosis. *Digital Signal Processing*, 143:104229, 2023.
- [35] A.A. Asiri et al. Advancing brain tumor classification through fine-tuned vision transformers: A comparative study of pre-trained models. *Medical Imaging and Sensing Technologies*, 23:7913, 2023.
- [36] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy 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*, 19(9):1498–1507, 2007.
- [37] Hatem Ibn-Khedher, Mohamed Ibn Khedher, and Makhlof Hadji. Mathematical programming approach for adversarial attack modelling. In *ICAART (2)*, pages 343–350, 2021.
- [38] Mohamed Ibn Khedher, Hatem Ibn-Khedher, and Makhlof Hadji. Dynamic and scalable deep neural network verification algorithm. In *ICAART (2)*, pages 1122–1130, 2021.
- [39] Omar El Mellouki, Mohamed Ibn Khedher, and Mounim A El-Yacoubi. Abstract layer for leakyrelu for neural network verification based on abstract interpretation. *IEEE Access*, 11:33401–33413, 2023.
- [40] Mohamed Ibn Khedher, Houda Jmila, and Mounim El Yacoubi. On the formal evaluation of the robustness of neural networks and its pivotal relevance for ai-based safety-critical domains. *International Journal of Network Dynamics and Intelligence*, page 100018, 2023.
- [41] Houda Jmila and Mohamed Ibn Khedher. Adversarial machine learning for network intrusion detection: A comparative study. *Comput. Networks*, 214:109073, 2022.