Assessing the Interpretability of Machine Learning Models in Early Detection of Alzheimer's Disease

Karim Haddada, Mohamed Ibn Khedher, Olfa Jemai, Sarra Iben Khedher, Mounim El Yacoubi

Assessing the Interpretability of Machine Learning Models in Early Detection of Alzheimer's Disease

Karim Haddada*, Mohamed Ibn Khedher†, Olfa Jemai*, Sarra Iben Khedher‡ and Mounim A. El-Yacoubi§
* Research Team in Intelligent Machines (RTIM), National Engineering School of Gabes (ENIG)
University of Gabes
karim.haddada@issatmh.u-monastir.tn, olfa.jemai@isimg.tn
†IRT - SystemX, 2 Bd Thomas Gobert, 91120 Palaiseau, France
mohamed.ibn-khedher@irt-systemx.fr
‡Faculty of Medicine of Monastir, Universty of Monastir, Monastir, 5000, Tunisie
sarra.ibenkhedher@gmail.com
§Samovar, Telecom SudParis, Institut Polytechnique de Paris, 19 place Marguerite Perey, 91120 Palaiseau, France
mounim.el_yacoubi@telecom-sudparis.eu

***Abstract*—Alzheimer's disease (AD) is a chronic and irreversible neurological disorder, making early detection essential for managing its progression. This study investigates the coherence of SHAP values with medical scientific truth. It examines three types of features: clinical, demographic, and FreeSurfer extracted from MRI scans. A set of six ML classifiers are investigated for their interpretability levels. This study is validated on the OASIS-3 dataset with binary classification. The results show that clinical data outperforms the others, with a margin of 14% over FreeSurfer features, the second-best features. In the case of clinical features, the explanations provided by the tree-based classifiers consistently align with medical insights. This comparison was calculated using the Kendall Tau distance.**

***Index Terms*—Early Alzheimer Disease, Interpretability, Machine Learning, SHAP, Explainable AI.**

I. Introduction

Alzheimer's disease primarily affects individuals over the age of 65. It is characterized by a decline in cognitive functions, including memory loss, language difficulties, and impaired judgment. Early detection is crucial for improving the quality of life for those affected [1].

Recent advances in machine learning (ML) have created new opportunities in Alzheimer's research, by analyzing vast amounts of data, e.g., brain scans, medical records, and genetic information, to identify biomarkers associated with the disease [2]. However, the application of ML in Alzheimer's research is not without challenges. The disease's variability, along with differences in data quality and processing methods, can impede the development of accurate models. Moreover, the "black-box" nature of some ML algorithms raises issues regarding their predictions' interpretability. To mitigate these issues, researchers are increasingly turning to explainable AI (XAI) approaches. XAI aims to provide an explanation of ML decisions, enabling clinicians and researchers to understand how predictions are made and identify the most influential factors [3]. XAI calculates an importance score for each input feature influencing the classifier's decision [4].

Previous research has focused on creating highly accurate AI models for detecting AD. However, for these models to be embraced by medical professionals, they must also be understandable and provide clear explanations of their decisions [5]. In other words, the explanation provided by the ML algorithm should align with the doctor's beliefs and/or established medical truths.

This paper evaluates the prediction accuracy and level of interpretability of a set of AI models. In our recent paper [2], we introduced a system for early AD detection based on MRI images using a transfer learning-based CNN. To assess this system interpretability, we identified two strategies for applying XAI methods: one involves computing the importance score per pixel [6], which is not practical in our case as a single pixel cannot determine the outcome independently. Alternatively, dividing the image into regions based on the anatomy of the human brain and then calculating the importance of each region allows us to identify the brain areas responsible for AD. This approach necessitates a segmentation algorithm, and the study's success depends on the quality of this algorithm. Consequently, we have decided to base our study on FreeSurfer [7], a set of features available within the dataset that primarily measures the volume of specific brain regions.

Our contributions encompass the evaluation of a set of AI algorithms using three types of features: FreeSurfer, clinical data, and demographics. We then compare the consistency of AI model explanations generated by SHapley Additive exPlanations (SHAP) with doctor's explanations. To our knowledge, we are the first to propose the Kendall's tau distance for comparing two ordered vectors.

By investigating three features, our approach is ready for a fusion-based method. However, since the features are not synchronous (not taken on the same day), we will leave the investigation of fusion for future work.

The rest of this paper is as follows. Section II presents the related works. Section III details our evaluation pipeline. Section IV, presents results and discussion. Finally, section V outlines our conclusions and future works.

II. Related works

Researchers have evaluated recently AD detection models w.r.t performance and interpretability. Our perspective aims to integrate these two aspects. In this context, we cover the state of the art that, on one hand, focuses on classification approaches that discern whether patients have AD or not, and, on the other hand, delves into XAI methods to gauge how interpretable these models are.

A. Approaches of Early AD Diagnostic

The approaches for AD diagnosis are broadly classified into cognitive and non-cognitive strategies. The first evaluate a patient's mental capacities, such as memory, language, attention, and executive functions, while the second emphasize a more holistic assessment of a patient's health and lifestyle, encompassing their medical and familial history as well as daily habits.

Non-cognitive techniques include the examination of behavior, emotional states, blood-based biomarkers, and neuroimaging being the most popular. A variety of neuroimaging techniques have been utilized, such as MRI, fMRI, sMRI, and PET [8], [9]. To increase diagnostic precision, some research suggests combining different imaging modalities [10], [11].

Generally, two steps are necessary for a neuroimaging-based approach: feature extraction and classification. In first, the image can be used directly or after a feature extraction process. These features can be extracted through techniques such as Bag-of-Features [10] or dimensionality reduction methods like Principal Component Analysis [12]. Additionally, neural networks, particularly CNN [13] and Vision Transformers [14], are used for feature extraction to derive discriminative embedding representations. In the classification step, a variety of ML algorithms are used, including traditional algorithms such as SVM [15] and KNN, as well as deep learning algorithms like CNN [16] and Vision Transformers (ViT).

In addition to neuroimaging, other non-cognitive methods such as behavioral analysis are also utilized. In [17], gait analysis was explored. The authors hypothesized that patients with AD might exhibit unique gait characteristics that differ from those of healthy individuals.

On the other side, cognitive approaches detect AD by evaluating cognitive abilities through standardized tests. These tests assess key areas such as memory, attention, language, and spatial orientation, which are crucial for understanding how individuals acquire, process, and utilize knowledge-fundamental aspects of cognitive function. Traditional cognitive assessments like the Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR) [18], Montreal Cognitive Assessment (MoCA) [19], and the Clock Drawing Test (CDT) [20] are commonly used to identify cognitive impairments. These tests are not only straightforward to administer but also cost-effective, providing clinicians with vital diagnostic information.

B. Interpretability approaches for AD Diagnostic

Interpretability is essential for ML algorithms, particularly in critical fields like healthcare. It allows for comprehensible and explainable predictions, enhancing trust and transparency. Interpretability approaches are generally grouped into three categories: visualization-based, surrogate-based, and intrinsic methods.

Visualization approaches employ graphical representations to illustrate what models have learned from data. They include Grad-CAM (Gradient-weighted Class Activation Mapping), Saliency maps, Gradient Integration, and Layer-Wise Relevance Propagation (LRP). Grad-CAM uses gradient data to spotlight influential regions in images, often applied to neuroimaging with CNNs [21]. It is sometimes combined with Gradient Integration for deeper insights [22]. LRP offers a different approach by allocating "relevance" scores to input features that significantly impact the model's predictions. It has been applied in the AD context using convolutional networks trained from both image and spectral data [23].

Surrogate-based methods entail creating interpretable models to analyze the predictions of a black-box ML model. An explanation for the black-box model is generated by comparing the decisions made by surrogate models with those of the black-box model. It include LIME (Local Interpretable Model-agnostic Explanations) and SHAP. In [24], LIME was used to identify the parts involved in the patient's brain. As input to LIME, a Transfer-based CNN is trained on a neuroimaging dataset. In [25], SHAP was combined with Grad-CAM to interpret the decisions of an XGBoost model trained on a set of features including neuroimaging (MRI and PET) and clinical features.

To align with the state of the art, we have investigated a set of features including clinical data, demographic characteristics, and FreeSurfer features extracted from MRI images. Regarding the interpretability approach, we have selected the SHAP surrogate-based method, which is among the most popular in current state of the art.

III. Proposed approach

Our approach consists of four steps. First, we select a list of features. These inputs are then fed into ML classifiers. Next, an interpretability technique is applied to identify the most important features. To evaluate the classifier, two types of metrics are calculated: one related to performance and the other to the interpretability level. We describe next the details of these different stages. Figure 1 presents the flowchart of our proposed approach.

Fig. 1: Flowchart of our proposed approach

A. Feature Selection

Three types of features are investigated in our paper: clinical features, demographic features, and features extracted from MRI images using FreeSurfer features (i.e. features extracted from the MRI neuroimaging). Based on the identified public dataset, which contains numerous features, it is evident that not all features are of significant importance or influence. Therefore, in collaboration with medical experts, we carefully select a list of the most relevant features for each feature type. These will be detailed subsequently and presented in a ranking format by the medical expert.

Clinical Features include assessments of Memory, Age, MMSE, Judgment (the ability to make sound decisions), and home hobbies (the capacity to engage in activities of daily living). Demographic Features include age (also noted under clinical features), ApoE gene status (which indicates genetic risk factors for Alzheimer's disease), education level, gender, and maternal dementia (whether the patient's mother had dementia). FreeSurfer Features include Intracranial Volume, Subcortical Gray Volume, Total Gray Volume, Cortex Volume, Cortical White Matter Volume, and Supratentorial Volume (the volume of the brain located above the tentorium cerebelli).

B. Classification

In the classification phase, a set of binary ML classifiers was evaluated to differentiate between AD patients and Cognitively Normal (CN) subjects. The models tested included Decision Trees (DT), Random Forests (RF), K-Nearest Neighbors (KNN), eXtreme Gradient Boosting (XGBoost), Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP). For each feature type, the most performant classifier is selected to investigate its interpretability level.

C. Interpretability based on SHAP

This step aims to sort input features according to their importance on the output of the ML classifier. The core of the SHAP method [26] is the assignment, to each feature in a given data sample, of a score, based on the Shapley values, that quantifies its contribution to the model's prediction. The Shapley values ϕ_j for feature j are computed as follows:

$$\phi_j = \sum_{S \subseteq X \setminus j} \frac{|S|!(|X| - |S| - 1)!}{|X|!} \left[f(x_j) - f_S(x_j) \right],$$

where X is the set of input features, $S \subseteq X \setminus j$ is a subset of features that does not include feature j, $f(x_j)$ is the model's prediction for the input sample x_j, and $f_S(x_j)$ is the model's prediction for the input sample x_j with the features in S set to their expected values.

D. Interpretability Evaluation metric

The objective of this step is to evaluate the ranking order of features by their importance using SHAP values and to compare it with a medical expert's assessment. Let $X = \{X_i | 1 \leq i \leq n\}$ be the set of n features ordered by SHAP values and $Y = \{Y_i | 1 \leq i \leq n\}$ be the set ordered by a medical expert. We propose using the Kendall tau distance [27] to measure the similarity.
Formally, observation pairs (X_i, Y_i) and (X_j, Y_j) with $i < j$ are *concordant* if their orderings are consistent; that is, either $X_i > X_j$ and $Y_i > Y_j$ or $X_i < X_j$ and $Y_i < Y_j$. Conversely, if $X_i > X_j$ and $Y_i < Y_j$ or $X_i < X_j$ and $Y_i > Y_j$, the pairs are *discordant*.
The Kendall τ coefficient is then defined as: $\tau = \frac{C-D}{C+D}$ where C is the number of concordant pairs, and D is the number of discordant pairs.

IV. Experimental results

A. Dataset and evaluation Protocol

We validated our approach on the Open Access Series of Imaging Studies (OASIS-3) dataset [28]. This dataset includes details about participants' demographics (713 CN and 346 AD), their clinical history (4473 CN and 1048 AD), and structural information about their brains, derived from MRI scans processed with Freesurfer (1841 CN and 484 AD). To ensure consistency and completeness, we preprocessed the features to eliminate any missing data. Missing values were imputed using the mean of the respective column. Additionally, the CDR values, which range from 0, 0.5, 1, 2, to 3, were categorized into two groups: CN for CDR values of 0, and AD for CDR values of 0.5, 1, 2, and 3.
In our evaluation protocol, we used a 5-fold cross-validation technique. For evaluation, we used accuracy to assess model classification performance and Kendall's Tau distance to measure the interpretability level.

TABLE I: Comparison of Feature Orders and Measures

Feature	SHAP Order	Doctor Order	Measure
Clinical	memory, age, mmse, judgment, homehobb	memory, age, mmse, judgment, homehobb	1
Demographic	Age, APOE , Educ, Gender , momdem	Age, APOE, Educ, Gender, momdem	1
FreeSurfer	IntraCranialVol, CortexVol, TotalGrayVol, CorticalWhiteMatterVol, SubCortGrayVol, SupraTentorialVol	IntraCranialVol, SubCortGrayVol, TotalGrayVol, CortexVol, CorticalWhiteMatterVol, SupraTentorialVol	0.46

TABLE II: ML models Accuracy (%) using different features

Classifier	Clinical	Demographic	FreeSurfer
DT	0.95	0.65	0.77
RF	0.93	0.74	0.81
KNN	0.92	0.72	0.81
SVM	0.90	0.71	0.81
XGboost	0.91	0.74	0.80
MLP	0.89	0.71	0.81

TABLE III: Comparison between SHAP Values and Medical Expert using Kendall's Tau distance.

Model	Feature Type		
	Clinical	Demographic	Freesurfer
DT	1	1	0.06
RF	1	1	0.06
KNN	-0.39	0.19	0.46
SVM	0	0.19	0.46
XGboost	0.39	1	0.06
MLP	0	1	0.06

B. Results and Discussion

Table II shows classifiers' accuracy, while Table III displays the Kendall's Tau similarity between the ranking orders provided by SHAP and a medical expert. These orders are detailed in Table I. For each feature, clinical, demographic, and FreeSurfer, an interpretability analysis is detailed for the most accurate classifier. These are presented respectively in Figures 2, 3, and 4. For each figure, the left side presents the average Shapley value for each input feature across the entire test dataset, while the right side describes the evolution of the Shapley value as the input feature value increases or decreases.

Regarding performance results, we observe that clinical features are more performant than demographic and FreeSurfer features, exceeding 89% in all cases. FreeSurfer are less performant, hovering around 80%. As these features are extracted via a segmentation algorithm, this prompts us to further investigate other segmentation algorithms. Moreover, these results are consistent with the opinion of the medical expert who confirms that they rely on clinical data to identify patients with AD. In terms of classifiers, tree-based classifiers (DT and RF) perform better with clinical data and are also the most interpretable. This coincides with the nature of tree-based classifiers, which are inherently interpretable by design.

The analysis of clinical data shows that for the factors "memory," "judgment," and "homehobb," as their deterioration decreases (depicted in blue on the right side of Figure 2), their impact on the model becomes negative, indicating a lower probability of AD onset. Similarly for age, a decrease in its value is correlated with a reduced risk of AD onset. However, for 'MMSE', an increase in its value (depicted in red) correlates with a negative impact on the model, implying a reduced probability of AD occurrence. These findings are consistent with scientific reality [29].

With demographic features, the four models (DT, RF, XGBoost, and MLP) demonstrated modest performance but offered strong interpretability. However, while KNN and SVM showed comparable performance, they were less interpretable. The analysis of interpretability level (Figure 3) indicates that an increase in age or APOE gene values positively influences the model's output, signifying a heightened risk of AD onset. Likewise, a lower education level is associated with a positive effect on model's predictions, suggesting a higher likelihood of AD development.

Regarding FreeSurfer features, the analysis of interpretability level (see Figure 4) indicates that a decrease in cortical volume, total gray matter volume, and cortical white matter volume has a positive impact on the model's predictions, indicating an increased likelihood of AD onset.

A final observation regarding the trade-off between performance and interpretability levels shows that a high-performing model is not necessarily the most interpretable. Devising more complex models may result in a more effective model, but at the cost of interpretability. The ideal model should be both high-performing and interpretable.

V. Conclusions

In this paper, we have investigated the interpretability level of a set of ML classifiers, in comparison to medical scientific truth. This investigation was conducted using three types of features: clinical, demographic, and FreeSurfer-based. Our comparison strategy involved collecting feature importance using the SHAP technique, on one hand, and by analyzing the scientific truths as dictated by our medical expert collaborator, on the other. The results showed that clinical data are the most performant, and choosing a tree-based classifier yielded a very high level of

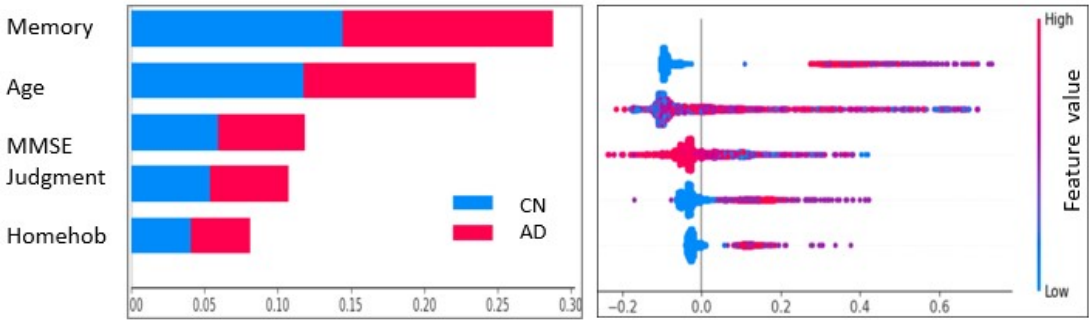

Fig. 2: Analysis of interpretability level for clinical features using the DT classifier. Left: Average feature importance using SHAP. Right: Feature importance for the AD class according to its values

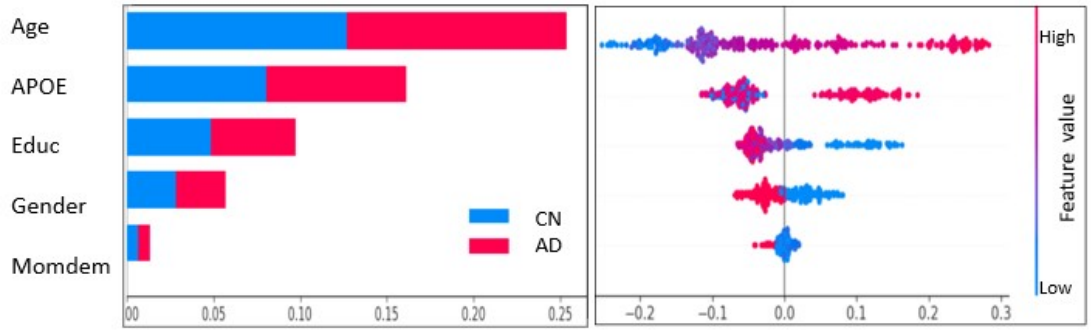

Fig. 3: Analysis of interpretability level for demographic features using the RF classifier. Left: Average feature importance using SHAP. Right: Feature importance for the AD class according to its values

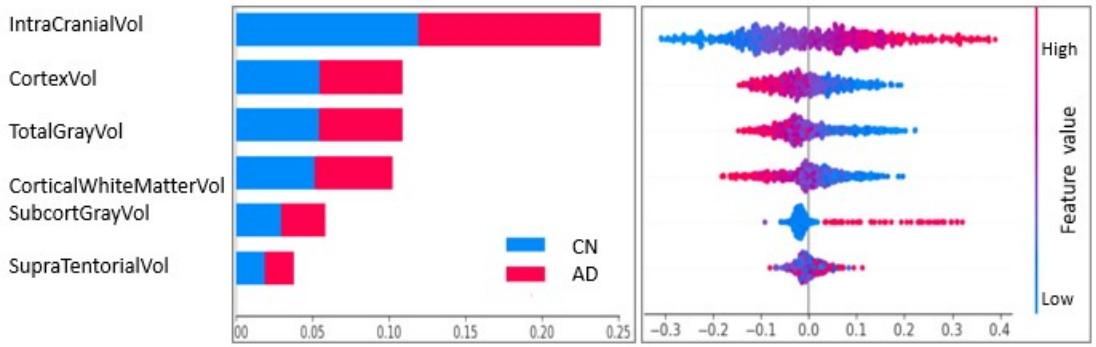

Fig. 4: Analysis of interpretability level for FreeSurfer features using the KNN classifier. Left: Average feature importance using SHAP. Right: Feature importance for the AD class according to its values

interpretability. Our main finding from this investigation was that a trade-off should be taken into account when choosing the best classifier, which should be both high-performing and interpretable.

Future work includes improving the performance of other features, especially FreeSurfer, with the aim of integrating all three types of features to enhance performance. The fusion of multiple features can enhance performance but may also make the model less interpretable due to increased complexity. We plan to test other interpretation methods, such as Integrated Gradients. We will also explore XAI for other modalities harnessed for detecting neurodegenerative diseases, such as speech [30], [31], and handwriting [32], [33], [34], [35], [36].

References

[1] Pradnya Borkar, Vishal Ashok Wankhede, Deepak T Mane, Suresh Limkar, JVN Ramesh, and Samir N Ajani. Deep learning and image processing-based early detection of alzheimer disease in cognitively normal individuals. Soft Computing, pages 1–23, 2023.

[2] 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.

[3] Johannes Allgaier, Lena Mulansky, Rachel Lea Draelos, and Rüdiger Pryss. How does the model make predictions? a systematic literature review on the explainability power of machine learning in healthcare. Artificial Intelligence in Medicine, 143:102616, 2023.

[4] Viswan Vimbi, Noushath Shaffi, and Mufti Mahmud. Interpreting artificial intelligence models: a systematic review on the application of lime and shap in alzheimerâs disease detection. Brain Informatics, 11(1):10, 2024.

[5] Vimbi Viswan, Noushath Shaffi, Mufti Mahmud, Karthikeyan Subramanian, and Faizal Hajamohideen. Explainable artificial intelligence in alzheimerâs disease classification: A systematic review. Cognitive Computation, 16(1):1–44, 2024.

[6] Feng-Lei Fan, Jinjun Xiong, Mengzhou Li, and Ge Wang. On interpretability of artificial neural networks: A survey. IEEE Transactions on Radiation and Plasma Medical Sciences, 5(6):741–760, 2021.

[7] Ales Bartos, David Gregus, Ibrahim Ibrahim, and Jaroslav Tintěra. Brain volumes and their ratios in alzheimer s disease on magnetic resonance imaging segmented using freesurfer 6.0. Psychiatry Research: Neuroimaging, 287:70–74, 2019.

[8] 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.

[9] Iype Cherian, Mahendra Alate, Anil Baburao Desai, MR Prajna, and Devyani Rawat. Early detection of alzheimer's disease using fuzzy c-means clustering and genetic algorithm-based feature selection from pet scans. International Journal of Intelligent Systems and Applications in Engineering, 12(3s):452–463, 2024.

[10] Latifa Houria, Noureddine Belkhamsa, Assia Cherfa, and Yazid Cherfa. Multimodal magnetic resonance imaging for alzheimer's disease diagnosis using hybrid features extraction and ensemble support vector machines. International Journal of Imaging Systems and Technology, 33(2):610–621, 2023.

[11] Yan Tang, Xing Xiong, Gan Tong, Yuan Yang, and Hao Zhang. Multimodal diagnosis model of alzheimerâs disease based on improved transformer. BioMedical Engineering OnLine, 23(1):1–18, 2024.

[12] Princy Matlani. Bilstm-ann: early diagnosis of alzheimerâs disease using hybrid deep learning algorithms. Multimedia Tools and Applications, pages 1–28, 2024.

[13] Ju-Hyeon Noh, Jun-Hyeok Kim, and Hee-Deok Yang. Classification of alzheimerâs progression using fmri data. Sensors, 23(14):6330, 2023.

[14] Uttam Khatri and Goo-Rak Kwon. Rmtnet: Recurrence meet transformer for alzheimerâs disease diagnosis using fdg-pet. January 2024.

[15] Shang Miao, Qun Xu, Weimin Li, Chao Yang, Bin Sheng, Fangyu Liu, Tsigabu T Bezabih, and Xiao Yu. Mmtfn: Multi-modal multi-scale transformer fusion network for alzheimer's disease diagnosis. International Journal of Imaging Systems and Technology, 2024.

[16] Zilun Zhang and Farzad Khalvati. Introducing vision transformer for alzheimer's disease classification task with 3d input. arXiv preprint arXiv:2210.01177, 2022.

[17] Younghoon Jeon, Jaeyong Kang, Byeong C Kim, Kun Ho Lee, Jong-In Song, and Jeonghwan Gwak. Early alzheimerâs disease diagnosis using wearable sensors and multilevel gait assessment: A machine learning ensemble approach. IEEE Sensors Journal, 2023.

[18] Ray-Chang Tzeng, Yu-Wan Yang, Kai-Cheng Hsu, Hsin-Te Chang, and Pai-Yi Chiu. Sum of boxes of the clinical dementia rating scale highly predicts conversion or reversion in predementia stages. Frontiers in Aging Neuroscience, 14:1021792, 2022.

[19] Li Chang Ang, Philip Yap, Sze Yan Tay, Way Inn Koay, and Tau Ming Liew. Examining the validity and utility of montreal cognitive assessment domain scores for early neurocognitive disorders. Journal of the American Medical Directors Association, 24(3):314–320, 2023.

[20] Sophia Lazarova, Denitsa Grigorova, Dessislava Petrova-Antonova, and Alzheimerâs Disease Neuroimaging Initiative. Detection of alzheimerâs disease using logistic regression and clock drawing errors. Brain Sciences, 13(8):1139, 2023.

[21] Sobhana Jahan, Kazi Abu Taher, M Shamim Kaiser, Mufti Mahmud, Md Sazzadur Rahman, ASM Sanwar Hosen, and In-Ho Ra. Explainable ai-based alzheimerâs prediction and management using multimodal data. Plos one, 18(11):e0294253, 2023.

[22] M Lucas, M Lerma, J Furst, and D Raicu. Visual explanations from deep networks via riemann-stieltjes integrated gradient-based localization, 2022.

[23] Marilia Lopes, Raymundo Cassani, Tiago H Falk, et al. Using cnn saliency maps and eeg modulation spectra for improved and more interpretable machine learning-based alzheimerâs disease diagnosis. Computational Intelligence and Neuroscience, 2023, 2023.

[24] Atefe Aghaei, Mohsen Ebrahimi Moghaddam, and Hamed Malek. Interpretable ensemble deep learning model for early detection of alzheimer's disease using local interpretable model-agnostic explanations. International Journal of Imaging Systems and Technology, 32(6):1889–1902, 2022.

[25] Fuliang Yi, Hui Yang, Durong Chen, Yao Qin, Hongjuan Han, Jing Cui, Wenlin Bai, Yifei Ma, Rong Zhang, and Hongmei Yu. Xgboost-shap-based interpretable diagnostic framework for alzheimerâs disease. BMC Medical Informatics and Decision Making, 23(1):137, 2023.

[26] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30, 2017.

[27] Maurice G Kendall. A new measure of rank correlation. Biometrika, 30(1/2):81–93, 1938.

[28] Pamela J LaMontagne, Tammie LS Benzinger, John C Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G Vlassenko, et al. Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. MedRxiv, pages 2019–12, 2019.

[29] Ronald C Petersen, Glenn E Smith, Robert J Ivnik, Emre Kokmen, and Eric G Tangalos. Memory function in very early alzheimer's disease. Neurology, 44(5):867–867, 1994.

[30] S. Mirzaei, M. El Yacoubi, S. Garcia-Salicetti, J. Boudy, C. Kahindo, V. Cristancho-Lacroix, H. KerhervÃ©, and A.-S. Rigaud. Two-stage feature selection of voice parameters for early alzheimerâs disease prediction. IRBM, 39(6):430â435, December 2018.

[31] Holger Frohlich, Noemi Bontridder, Dijana Petrovska-Delacreta, Enrico Glaab, Felix Kluge, Mounim El Yacoubi, Mayca Marin Valero, Jean-Christophe Corvol, Bjoern Eskofier, Jean-Marc Van Gyseghem, Stephane Lehericy, Jurgen Winkler, and Jochen Klucken. Leveraging the potential of digital technology for better individualized treatment of parkinsonâs disease. Frontiers in Neurology, 13, February 2022.

[32] 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.

[33] 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.

[34] Nickson Mwamsojo, Frederic Lehmann, Mounim A. El-Yacoubi, Kamel Merghem, Yann Frignac, Badr-Eddine Benkelfat, and Anne-Sophie Rigaud. Reservoir computing for early stage alzheimerâs disease detection. IEEE Access, 10:59821â59831, 2022.

[35] 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.

[36] Jana Sweidan, Mounim A. El-Yacoubi, and Anne-Sophie Rigaud. Explainability of cnn-based alzheimerâs disease detection from online handwriting. Research Square preprint, April 2024.