



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: <https://doi.org/10.22214/ijraset.2026.82473>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Explainable Deep Learning for Alzheimer's Diagnosis: A Comprehensive Review of Models, Methods, and Clinical Relevance

Shrutha Prakash

St Claret College Autonomous, India

Abstract—Alzheimer's Disease (AD) is a progressive neurodegenerative condition mostly affecting the elderly, causing cognitive decline. It is an active field of research as currently there is no approved cure. Current breakthroughs in deep learning have been promising for AD detection from neuroimaging and clinical information. The "black-box" limitation of such models is a concern regarding trust and clinical uptake. This research investigates the potential of integrating deep learning with Explainable AI (XAI) to resolve these issues. It outlines a systematic review of the literature from 2015 to 2024, organizing studies into four categories: general reviews, models that lack explainability, models with the integration of partial or complete XAI, and explainability methods. The article compares methods, datasets (ADNI, NACC, AIBL), and main results, emphasizing advances in visual interpretability yet pointing towards persisting issues like a deficiency of standard metrics in evaluation, sparse clinical validation, and limited generalizability. Several studies lack rule-based explanations or fail to quantify the reliability of model decisions and there is a necessity to integrate more longitudinal, multimodal, and demographically diversified data in AD research. This review highlights the importance of standardised, clinically validated explainability frameworks to connect research with real-world usage. It seeks to direct future efforts in the creation of responsible, interpretable AI tools for healthcare, providing beneficial insights for researchers, clinicians, and policymakers.

Index Terms—Explainable AI, Alzheimer's Disease, Deep Learning, Medical Imaging, Model Interpretability

I. INTRODUCTION

Alzheimer's disease (AD) is one of the world's most pressing public health challenges that is characterised by early but subtle alterations in the brain much before to the onset of noticeable memory impairment. Alzheimer's Disease (AD) by nature is a gradual and the predominant source of dementia. It principally afflicts people over 65 years and the symptoms progressively deteriorate over the years. In advanced stages, they will be sufficiently severe to meddle in daily activities of affected people. For diagnosis, specialists frequently depend upon an assemblage of techniques and resources including patient's clinical background, psychological condition analyses, somatic plus neural scrutinies, diagnostic assays, MRI or PET scans. These investigations have found that pathological changes commence far prior to symptoms becoming noticeable such as memory loss or confusion. Accordingly, the prompt ascertainment of AD is requisite [49].

There is not a currently established cure for Alzheimer's; however, understanding the origins and causes has improved significantly. Contemporary advancements in artificial intelligence (AI), especially in Machine Learning (ML) and Deep Learning (DL) have emerged as ground-breaking tools for the early diagnosis of AD. These methods aid in-depth analysis, identifying subtle and previously overlooked changes in brain structure and functions linked to Alzheimer's Disease. Most often these models operate as "black boxes," exhibiting a lack of transparency in their decision-making processes leading to apprehensions in clinical settings. For advancement of trust and transparency, Explainable AI (XAI) [53] is scrutinized and assimilated in healthcare applications, presenting visual perceptions into model decisions.

There is an increasing demand for Explainable AI (XAI) [53] frameworks capable of delivering clinically interpretable insights with high diagnostic accuracy to close the gap between performance and trust. SHAP (Shapley Additive Explanations) [4], LIME (Local Interpretable Model-Agnostic Explanations) [3], Grad-CAM (Gradient-weighted Class Activation Mapping) [9], Integrated Gradients (IG) [55], DeepLIFT [7] and Layer-wise Relevance Propagation (LRP) [1] not only confirm model choices but also enable regulatory approval and ethical AI deployment in healthcare.

Objectives and Purpose of the Review

Accurately detecting early stages of AD (Mild Cognitive Impairment (MCI) vs stable MCI vs progressive MCI) is a complex clinical challenge. The need for accurate and reliable frameworks to provide clinically relevant insights is growing with the increased application of AI in healthcare. This review covers literature published between 2015 and 2024 and seeks to:

- Investigate the recent AI-based approaches applied in early detection of AD.
- Study the various explainability methods providing in-sights into major biomarkers in the context of healthcare applications.
- Compare the various XAI methods across different studies involving imaging and clinical data for their clinical utility.
- Map emerging trends and limitations in existing literature.

II. LITERATURE REVIEW

The reviewed studies have been divided into four distinct clusters according to the methodology used: (i) review studies related to AD research, (ii) studies that use machine learning or deep learning models without incorporating explainability, (iii) studies that integrate XAI methods within ML/DL frameworks, and (iv) studies focusing on the novel explainable AI (XAI) techniques.

A. Review studies in AD research

Numerous studies have examined the implementation of DL techniques for AD diagnosis, emphasising both benefits and limitations. While Yan et al. [13] investigate approaches such as Transfer Learning, Ensemble Learning and architectures like CNNs, RNNs, GCNs, Autoencoders and highlight the impact of pre-processing, class imbalance, and performance evaluation on standard datasets, explainability and comparative analysis have been overlooked. Surveys by Khojaste-Sarakhsi et al. and Z. Yao et al. [28], [36] address advanced strategies, including attention and graph-based networks, but offer a restricted evaluation of specific XAI methodologies.

[36] emphasises MRI-based neuroimaging and highlights LRP and counterfactuals for trust; nonetheless, issues pertaining to clinical validation and standardisation persist. Murad et al. [39] group XAI methodologies into intrinsic, visualization-based, and distillation categories, and suggest the use of fuzzy logic to address ambiguity in clinical data. These studies collectively emphasise the necessity for enhanced integration of XAI and real-world validation in Alzheimer's disease diagnosis.

B. Studies Using ML or DL Models Without XAI Integration

CNNs, ensemble methods, and lightweight architectures have demonstrated significant improvement in diagnostic accuracy but they fall short in offering interpretable decisions, limiting their transparency and clinical adoption. Ensemble techniques have been used by researchers to boost model performance in AD diagnosis. Islam and Zhan et al. [5] used an ensemble of deep CNNs on the OASIS dataset, reporting improved accuracy over traditional models. Loddoo et al. [20] combined multiple architectures in a deep ensemble pipeline and achieved high accuracy across OASIS, ADNI, and Kaggle datasets. A "Clinical Decision Support System (CDSS)" introduced by Salami et al. [22] incorporated MRI and clinical data through an ensemble of CNNs, demonstrating improved diagnosis accuracy. Yet in spite of their efficiency, these models lack transparency, as they fail to provide interpretation for their predictions, hence restricting their reliability and practical use.

Liu et al. [15] introduced a lightweight model aimed at efficiency and portability, demonstrating robust performance with MRI data; nevertheless, it exhibited challenges in generalisation to external datasets such as ADNI and lacked interpretability. Shanmugam et al. [18] used T1-weighted MRI images and leveraged transfer learning on pre-trained models, with ResNet-18 providing the best outcome. Han et al. [24] developed a lightweight 3D DL framework for differentiating AD and MCI. While it highlighted automation and efficiency it too lacked the integration of XAI methods, limiting insights into model decisions and affecting clinical transparency.

A 2023 study by Menagadevi et al. [31] proposes an automated technique that integrates white matter segmentation and machine learning for the diagnosis of Alzheimer's disease. Improved image processing and feature extraction result in highly accurate performance. Arafa et al. [41] assessed two methodologies: a custom CNN developed from scratch and a transfer learning model employing VGG16 for MRI-based categorisation. Fouad et al. [38] developed a fully automated system utilising EEG signals, employing ResNet-50 to identify significant patterns in cerebral activity. Although the models demonstrate accuracy, they lack insight into feature contributions, hence limiting their transparency and potential for clinical use. Another fully automated deep learning pipeline by Inan et al. [43] uses a slice-selection strategy to enhance computational efficiency and diagnostic accuracy. Model was validated on ADNI and OASIS datasets supporting its clinical relevance. However, it struggled to distinguish cognitively normal individuals from early-stage MCI.

C. Studies Integrating Explainability

Integrating XAI into deep learning frameworks in healthcare is gaining focus in the recent studies. Lee et al. [8] integrated different analysis methods into their framework to predefine parts of the brain known to be affected in AD. The approach heavily relies on the correctness of these predefined regions. Nigri et al. [10] took an alternative approach via Swap Test method that compares MRI patches with reference scans and highlights key regions. Though it improved interpretability, the accuracy depends on the choice of the correct reference image. F. Zhang et al. [26] used Grad-CAM++ to produce attention maps from CNN models and identify disease-relevant areas. It was observed that the method had limitations in differentiating stable from progressive MCI. Other studies—like those by Shojaei et al. [35], Z. Yao et al. [37], Nguyen et al. [32], and D. Wang et al. [33]—examined SHAP, voxel grading, and quantitative heatmap validation to enhance model transparency. Most common limitations encountered were small datasets or lack of resolution. More advanced models, such as MAXNet [30], MPS-FFA [34], and LA-GMF [47], utilized attention mechanisms and graph-based fusion to better visualization, attaining better explanations while balancing model complexity and training requirements.

Researchers have also explored the usage of EEG and longitudinal data. Sidulova et al. [17] applied LIME to get insights into CNN predictions based on EEG signals to identify key brain regions. Viton et al. [11] made use of heatmaps in a CNN-based framework to visualize features significant for predicting hospital mortality, presenting interesting inferences for temporal medical data.

There is an increased focus on using multimodal data and electronic health records (EHR) for AD diagnosis. Abuhmed et al. [14] used MRI, PET, cognitive scores, and genetics in their model and added interpretability through fuzzy logic and decision trees. SHAP was used to explain model predictions by Bogdanovic et al. [23]. Though the dataset bias was a concern, the model predictions were based on lifestyle and cognitive features, revealing novel insights into AD risk. Rule-based methods were used with SHAP and LIME by Alatrany et al. [42] to make predictions more transparent, while visual explanations with the help of saliency maps in ensemble CNNs were used by Mahmud et al. [40]. Zhu et al.'s TRADE model [45], examined EHRs from over a million patients to predict AD risk and used integrated gradients for explainability.

Graph-based models effectively represent relationships between different brain regions. L. Wang et al. [25] formulated a GNN model that uses self-attention to demonstrate the interactions between different parts of the brain. Zen et al.

[27] developed a system that allocates dynamic weights to features and samples, enabling the model focus on the most relevant information. Effectiveness of these approaches rely significantly on the data quality and labelling.

Among attention-based frameworks, the AXIAL model introduced by Lozupone et al. [48] is notable because of its clinically validated and interpretable outcomes. The model utilizes 2D CNNs for processing 3D MRI images and uses a soft attention mechanism to produce voxel-level maps for localizing disease-relevant regions. It employs double transfer learning from AD vs CN classification to the more intricate pMCI versus sMCI differentiation. The model reveals 17 brain areas, including the hippocampus, parahippocampus, and amygdala, that strongly align with radiologist verdicts, making it interpretable, replicable and clinically relevant.

These studies demonstrate the increasing awareness and effort to make AI in healthcare not only accurate but also reliable. Despite the progress that has been made many frameworks still lack clinical validation, rule-based reasoning, or generalization across populations. These remain important areas for future research.

D. XAI Techniques

The black-box nature of deep models has inspired the development and adaptation of explainability methods like Grad-CAM, LRP, LIME and SHAP. These methods enable researchers to get insights about the most influential features in the context of model's prediction, thereby making the models more interpretable and trustworthy in clinical settings. Some studies also employ attention mechanisms. Recent research has also included hybrid approaches that combine visual heatmaps with quantitative attributions and graph-based embeddings to better capture structural representations, hence improving accuracy and interpretability, in addition to these pixel-level explanation techniques. Table I provides a comprehensive overview of the XAI techniques explored in this study.

III. GAPS IDENTIFIED FROM THE LITERATURE & DISCUSSION

Considering the growing use of deep learning and machine learning frameworks for Alzheimer's diagnosis, certain challenges remain. Studies make improving predictive performance the priority but overlook the explainability aspect, which is essential for clinical credibility.

Despite the use of visual explanations, an accepted standard for assessing their reliability or effectiveness remains absent. Heatmaps, though common, seldom offer explicit, rule-based justification for predicted outcomes. Furthermore, a majority of approaches are evaluated on a singular dataset, such as ADNI, which restrains their applicability across diverse settings or patient demographics. Clinical validation for the predictions and explanations produced by these models is available for only a few studies conducted. As early identification is crucial, a few approaches strive to differentiate between MCI subtypes, such as progressing and stable MCI, which might offer better treatment when identified accurately.

Effective comparison of the studies is challenging due to the use of diverse architectures, datasets, and explainability approaches, together with the frequent inconsistency in result reporting. This lack of consistency complicates the ability to ascertain which strategies are most effective.

A. Deep Learning in AD Research

AD research studies utilize deep learning widely especially with MRI and PET data. 2D and 3D CNN variants are commonly used to detect the structural changes in brain. Transfer learning and ensemble approaches improve the prediction performance especially when the availability of data is limited. Transformers and attention-based models have shown promise in predicting the AD risk and are applied increasingly to handle sequential data such as EHRs or identify important features in images. Another commonly considered approach is the utilization of Graph Neural Networks (GNNs) to map and understand connectivity patterns (both local and global) in brain fMRI data. Other approaches that have shown promising results are - usage of autoencoders for identifying important features, GANs for generating synthetic data and hybrid models using different methods and multimodal data.

B. XAI Techniques for AD Diagnosis

Explainability methods explored in this study have been broadly classified as:

- **Visualization Techniques:** The techniques provide visual perceptions for model decisions. Gradient-based techniques such as Grad-CAM and its variations, LRP, Saliency maps highlight most influential areas of the image regions, approaches such as [48] take advantage of attention-based explainability to enable interpretability in localized diagnosis. [30] uses Dual Attention Module and High-resolution Activation Mapping (HAM) for additional visual interpretability. Occlusion-based techniques such as the Occlusion Test and Swap Test [10] measure feature importance by occluding portions of the input in a systematic manner and measuring its effect on the model's prediction.
- **Feature-based Methods:** LIME, SHAP provide importance scores to the input features to account for their contribution to the model's prediction. Causal Analysis models such as ACE (Average Causal Effect) may be employed to determine the causal relevance of features in CNN predictions. Rule-based techniques such as Class Association Rule Mining (CAR) and SIRUS [42] derive human-readable rules from the model predictions, giving global interpretability. Anchors also offer explanations as if-then rules.

TABLE I
XAI METHODS REVIEWED

Ref	XAI Method	Key Findings
[2]	Class Activation Map (CAM)	Class Activation Maps are generated using global average pooling and linear classification. CAM allows for the visualization of class-specific units in CNNs, providing insights into which features and image regions are most discriminative for different classes.
[1]	Layer-wise Relevance Propagation	Provides pixel-level attribution by redistributing prediction scores backward through the network and can provide meaningful explanations for the decisions of non-linear image classifiers. The generated heatmaps highlight image regions that are relevant for a particular classification decision, even for complex models like deep neural networks. While visual assessment is important, there is a need for more robust, quantitative methods to evaluate the quality of pixel-wise explanations.

[3]	LIME	Perturbs input features and trains local surrogate models. SP-LIME, by selecting a diverse and representative set of explanations, provides a valuable global understanding of the model's behaviour and helps users identify potential issues and limitations.
[4]	SHAP	SHAP provides a unique solution among additive feature attribution methods that guarantees local accuracy, missingness, and consistency. Kernel SHAP is more sample-efficient and accurate than Shapley sampling and LIME.
[7]	DeepLIFT	Tracks differences from a reference input across layers. The choice of the DeepLIFT rule (Linear, Rescale, Reveal Cancel) can impact performance, and the optimal choice may depend on the network architecture and the task.
[9]	Grad-CAM	Grad-CAM generalizes the CAM technique, making gradient-based localization applicable to a much wider range of CNN architectures. It produces class-discriminative visual explanations using gradient information. It can be effectively used to diagnose failure modes in CNNs, understand vulnerabilities to adversarial attacks, and identify biases in training datasets. Combining Grad-CAM with high-resolution gradient-based methods like Guided Backpropagation yields Guided Grad-CAM, which provides both class-discriminative and fine-grained visual explanations.
[46]	A Hybrid Approach Combining GradCAM and LRP for CNN Interpretability	An innovative technique to improve the interpretability of CNNs by producing clearer and more precise visual explanations. The combined method offers superior interpretability and comprehensible results compared to standalone GradCAM, LRP, and combinations of GradCAM with other visualization techniques.
[44]	Gaussian-Class Activation Mapping Explainer (G-CAM)	Gaussian-weighted Class Activation Maps for object localization. G-CAM achieves state-of-the-art performance in speed and accuracy for object detection explainability. Especially effective for tiny object detection where other methods struggle. Reduces saliency noise, improves explanation trust, and offers real-time applicability.

- Example-based Methods: These methods explain a prediction by relating it to similar or contrasting examples. Methods involve counterfactual explanations, Contrastive Explanation Method (CEM)[12], and employing prototypes. MAXNet's [30] Prediction-basis Creation and Retrieval (PCR) module employs a reference set of relevant labeled samples to supply evidence for model predictions.

IV. CONCLUSION AND FUTURE WORK

This review highlights the increasing overlap between AI and healthcare research, particularly in the context of Alzheimer's Disease. Recent advancements in Alzheimer's Disease (AD) diagnosis show deep learning models' promise when applied to datasets such as ADNI, AIBL, and NACC. These models have been achieving outstanding precision in diagnostics. Explainability is stressed now to improve model transparency for all. This emphasis helps in building trust. Still, there are challenges that persist despite all of these developments. Since many heatmap-based outputs do still fail to provide actionable rule-based perceptions, the field now lacks standardized benchmarks for any evaluating of explainability methods. Small or imbalanced datasets generate concern, plus underrepresentation of diverse patient groups and limited external validation generate concern.

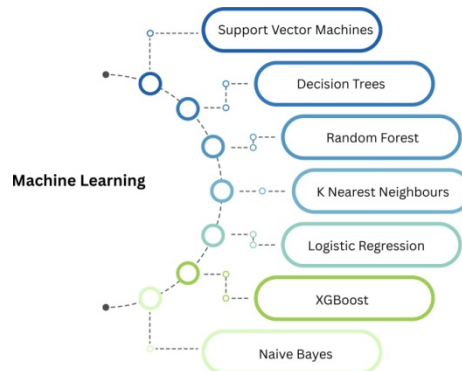


Fig.1.MachineLearningTechniquesUsedInADReasearch

Real-world adoption is further obstructed via reporting model choice inconsistencies as well as clinical validation absence. These issues must be addressed by standardization, included by a broader dataset, and integrated effectively by collaboration with clinical stakeholders for explainable AI in AD diagnosis.

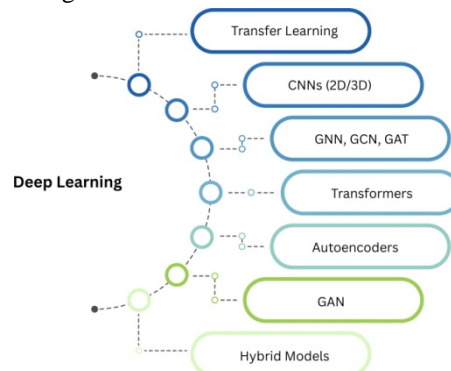


Fig.2.DeepLearningTechniquesUsedInADReasearch

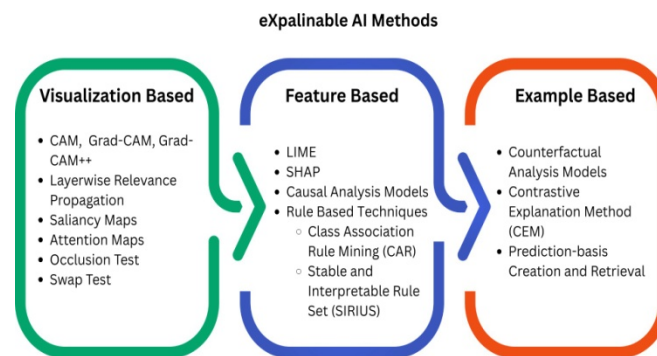


Fig.3.XAITechniquesUsedInADReasearch

REFERENCES

- [1] S.Bach,A.Binder,G.Montavon,F.Klauschen,K.-R.Mu"ller,andW.Samek,“OnPixel-WiseExplanationsforNon LinearClassifierDecisionsbyLayer-WiseRelevancePropagation,”PLoSOne,vol.10,no.7,p.e0130140,Jul.2015,doi: https://doi.org/10.1371/journal.pone.0130140.1371/journal.pone.0130140.
- [2] B.Zhou,A.Khosla,A.Lapedriza,A.Oliva,anda.Torralba,“Learning Deep Features for DiscriminativeLocalization,”arXiv preprint arXiv:1512.04150, Dec. 2015, doi:https://doi.org/10.48550/arXiv.1512.04150.1512.04150.48550/arXiv.1512.04150.
- [3] M.T.Ribeiro,S.Singh,andC.Guestrin,““WhyShouldITrustYou?":ExplainingthePredictionsofAnyClas-sifier,” arXiv preprint arXiv:1602.04938, Aug. 2016, doi:https://doi.org/10.48550/arXiv.1602.04938.10.48550/arXiv.1602.04938.
- [4] S.LundbergandS.-I.Lee,“AUnifiedApproachtoInterpretingModel Predictions,” arXiv preprint arXiv:1705.07874, Nov. 2017, doi:https://doi.org/10.48550/arXiv.1705.07874.10.48550/arXiv.1705.07874.
- [5] J. Islam and Y. Zhang, “Brain MRI analysis for Alzheimer’s diseasediagnosis using an ensemble system of deep convolutional neuralnetworks,” Brain Informatics, vol. 5, no. 2, p. 2, May 2018, doi:https://doi.org/10.1186/s40708-018-0080-3.10.1186/s40708-018-0080-3.

- [6] R.Zhu,F.Dornaika,andY.Ruichek,“Learningadiscriminantgraph-based embedding with feature selection for imagecategorization,” *Neural Networks*, vol. 111, pp. 35–46, Mar. 2019, doi:<https://doi.org/10.1016/j.neunet.2018.12.008>.
- [7] A.Shrikumar,P.Greenside,andA.Kundaje,“LearningImportantFeaturesThroughPropagatingActivationDiffer-ences,” *arXiv preprint arXiv:1704.02685*, Oct. 2019, doi:<https://doi.org/10.48550/arXiv.1704.02685>.
- [8] E. Lee, J.-S. Choi, M. Kim, and H.-I. Suk, “Toward an interpretableAlzheimer’s disease diagnostic model with regional abnormality repre-sentation via deep learning,” *NeuroImage*, vol. 202, p. 116113, Nov.2019, doi: 10.1016/j.neuroimage.2019.116113.
- [9] R.R.Selvaraju,M.Cogswell,A.Das,R.Vedantam,D.Parikh,and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks viaGradient-based Localization,” *Int J Comput Vis*, vol. 128, no. 2, pp.336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.
- [10] E.Nigri,N.Ziviani,F.Cappabianco,A.Antunes,andA.Veloso, “Explainable Deep CNNs for MRI-Based Diagnosis ofAlzheimer’s Disease,” Apr. 25, 2020, arXiv:arXiv:2004.12204. doi:10.48550/arXiv.2004.12204.
- [11] F. Viton, M. Elbattah, J.-L. Gue’rin, and G. Dequen, “Heatmaps forVisual Explainability of CNN-Based Predictions for Multivariate TimeSeries with Application to Healthcare,” in *2020 IEEE InternationalConference on Healthcare Informatics (ICHI)*, Nov. 2020, pp. 1–8. doi:10.1109/ICHI48887.2020.9374393.
- [12] D. Dave, H. Naik, S. Singhal, and P. Patel, “Explainable AI meetsHealthcare: A Study on Heart Disease Dataset,” Nov. 06, 2020, arXiv:arXiv:2011.03195. doi: 10.48550/arXiv.2011.03195.
- [13] K. Yang and E. A. Mohammed, “A Review of Artificial IntelligenceTechnologies for Early Prediction of Alzheimer’s Disease,” Dec. 22,2020, arXiv:arXiv:2101.01781. doi: 10.48550/arXiv.2101.01781.
- [14] T.Abuhmed,S.El-Sappagh,andJ.M.Alonso,“Robusthy-brid deep learning models for Alzheimer’s progression detection,”*Knowledge-Based Systems*, vol. 213, p. 106688, Feb. 2021, doi:10.1016/j.knsys.2020.106688.
- [15] J. Liu, M. Li, Y. Luo, S. Yang, W. Li, and Y. Bi, “Alzheimer’s diseasedetection using depthwise separable convolutional neural networks,”*Computer Methods and Programs in Biomedicine*, vol. 203, p. 106032,May 2021, doi: 10.1016/j.cmpb.2021.106032.
- [16] A.Yadu,P.K.Suhas,andN.Sinha,“ClassSpecificInterpretabilityin CNN Using Causal Analysis,” in *2021 IEEE International Con-ference on Image Processing (ICIP)*, Sep. 2021, pp. 3702–3706. doi:10.1109/ICIP42928.2021.9506118.
- [17] M. Sidulova, N. Nehme, and C. H. Park, “Towards Explainable ImageAnalysis for Alzheimer’s Disease and Mild Cognitive Impairment Di-agnosis,”in*2021IEEEAppliedImageryPatternRecognitionWorkshop(AIPR)*, Oct. 2021, pp. 1–6. doi: 10.1109/AIPR52630.2021.9762082.
- [18] J. V. Shanmugam, B. Duraisamy, B. C. Simon, and P. Bhaskaran,“Alzheimer’s disease classification using pre-trained deep networks,”*Biomedical Signal Processing and Control*, vol. 71, p. 103217, Jan.2022, doi: 10.1016/j.bspc.2021.103217.
- [19] A. Raza, K. P. Tran, L. Koehl, and S. Li, “Designing ECG monitoringhealthcaresystemwithfederatedtransferlearningandexplainableAI,” *Knowledge-Based Systems*, vol. 236, p. 107763, Jan. 2022, doi:10.1016/j.knsys.2021.107763.
- [20] A. Loddo, S. Buttau, and C. Di Ruberto, “Deep learning based pipelinesfor Alzheimer’s disease diagnosis: A comparative study and a noveldeep-ensemble method,” *Computers in Biology and Medicine*, vol. 141, p.105032, Feb.2022,doi:10.1016/j.compbiomed.2021.105032.
- [21] Z. Liu, H. Lu, X. Pan, M. Xu, R. Lan, and X. Luo, “Diagnosis ofAlzheimer’s disease via an attention-based multi-scale convolutionalneural network,” *Knowledge-Based Systems*, vol. 238, p. 107942, Feb.2022, doi: 10.1016/j.knsys.2021.107942.
- [22] F. Salami, A. Bozorgi-Amiri, G. M. Hassan, R. Tavakkoli-Moghaddam, and A. Datta, “Designing a clinical decision support system forAlzheimer’s diagnosis on OASIS-3 data set,” *Biomedical SignalProcessing and Control*, vol. 74, p. 103527, Apr. 2022, doi:10.1016/j.bspc.2022.103527.
- [23] B. Bogdanovic, T. Eftimov, and M. Simjanoska, “In-depth insights intoAlzheimer’s disease by using explainable machine learning approach,”*SciRep*, vol. 12, no. 1, p.6508, Apr.2022, doi:10.1038/s41598-022- 10202-2.
- [24] R. Han, Z. Liu, and C. L. P. Chen, “Multi-scale 3D convolution feature-based Broad Learning System for Alzheimer’s Disease diagnosis viaMRI images,” *Applied Soft Computing*, vol. 120, p. 108660, May 2022,doi: 10.1016/j.asoc.2022.108660.
- [25] L.Wangetal.,“Dementiaanalysisfromfunctionalconnectivitynetworkwith graph neural networks,” *Information Processing & Management*, vol. 59, no. 3, p. 102901, May 2022, doi: 10.1016/j.ipm.2022.102901.
- [26] F. Zhang et al., “A Single Model Deep Learning Approach forAlzheimer’s Disease Diagnosis,” *Neuroscience*, vol. 491, pp. 200–214,May 2022, doi: 10.1016/j.neuroscience.2022.03.026.
- [27] L. Zeng, H. Li, T. Xiao, F. Shen, and Z. Zhong, “Graph convolutionalnetwork with sample and feature weights for Alzheimer’s diseasediagnosis,” *Information Processing & Management*, vol. 59, no. 4, p.102952, Jul. 2022, doi: 10.1016/j.ipm.2022.102952.
- [28] M. Khojaste-Sarakhsi, S. S. Haghighi, S. M. T. F. Ghomi, and E.Marchiori, “DeeplearningforAlzheimer’sdiseasediagnosis:Asurvey,”*Artificial Intelligence in Medicine*, vol. 130, p. 102332, Aug. 2022, doi:10.1016/j.artmed.2022.102332.
- [29] Z. Yılmaz Acar, F. Bas,ç.iftç,i, and A. H. Ekmekci, “Future activityprediction of multiple sclerosis with 3D MRI using 3D discrete wavelettransform,” *Biomedical Signal Processing and Control*, vol. 78, p.103940, Sep. 2022, doi: 10.1016/j.bspc.2022.103940.
- [30] L.Yu,W.Xiang,J.Fang,Y.-P.PhoebChen,andR.Zhu,“AnovelexplainableneuralnetworkforAlzheimer’sdiseasediagno-sis,” *Pattern Recognition*, vol. 131, p. 108876, Nov. 2022, doi:10.1016/j.patcog.2022.108876.
- [31] M. Menagadevi, S. Mangai, N. Madian, and D. Thiyagarajan, “Auto-matedpredictionsystemforAlzheimerdetectionbasedondeepresidualautoencoder and support vector machine,” *Optik*, vol. 272, p. 170212, Feb. 2023, doi: 10.1016/j.ijleo.2022.170212.
- [32] H.-D. Nguyen, M. Cle’ment, B. Mansencal, and P. Coupe’, “TowardsbetterInterpretableandGeneralizableADdetectionusingCollectiveAr-tificialIntelligence,”*ComputerizedMedicalImagingandGraphics*, vol. 104, p. 102171, Mar. 2023, doi: 10.1016/j.compmedimag.2022.102171.
- [33] D. Wang et al., “Deep neural network heatmaps capture Alzheimer’sdiseasepatternsreportedinalargemeta-analysisofneuroimag-ing studies,” *NeuroImage*, vol. 269, p. 119929, Apr. 2023, doi:10.1016/j.neuroimage.2023.119929.
- [34] F.Liuetal.,“MPS-FFA:Amultiplaneandmultiscalefeaturefu-sion attention network for Alzheimer’s disease prediction with struc-tural MRI,” *Comput Biol Med*, vol. 157, p. 106790, May 2023, doi:10.1016/j.compbiomed.2023.106790.
- [35] S. Shojaei, M. Saniee Abadeh, and Z. Momeni, “An evolutionaryexplainabledeeplearningapproachforAlzheimer’sMRIclassification,”*Expert Systems with Applications*, vol. 220, p. 119709, Jun. 2023, doi:10.1016/j.eswa.2023.119709.

- [36] Z. Yao et al., "Artificial intelligence-based diagnosis of Alzheimer's disease with brain MRI images," *Eur J Radiol*, vol. 165, p. 110934, Aug. 2023, doi: 10.1016/j.ejrad.2023.110934.
- [37] G. Ekuma, D. B. Hier, and T. Obafemi-Ajayi, "An Explainable Deep Learning Model for Prediction of Severity of Alzheimer's Disease," in *2023 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Eindhoven, Netherlands: IEEE, Aug. 2023, pp. 1–8, doi: 10.1109/CIBCB56990.2023.10264880.
- [38] I. A. Fouad and F. El-Zahraa M. Labib, "Identification of Alzheimer's disease from central lobe EEG signals utilizing machine learning and residual neural network," *Biomedical Signal Processing and Control*, vol. 86, p. 105266, Sep. 2023, doi: 10.1016/j.bspc.2023.105266.
- [39] N. Y. Murad, M. H. Hasan, M. H. Azam, N. Yousof, and J. S. Yalli, "Unraveling the Black Box: A Review of Explainable Deep Learning Healthcare Techniques," *IEEE Access*, vol. 12, pp. 66556–66568, 2024, doi: 10.1109/ACCESS.2024.3398203.
- [40] T. Mahmud, K. Barua, S. U. Habiba, N. Sharmen, M. S. Hossain, and
- [41] K. Andersson, "An Explainable AI Paradigm for Alzheimer's Diagnosis Using Deep Transfer Learning," *Diagnostics*, vol. 14, no. 3, Art. no. 3, Jan. 2024, doi: 10.3390/diagnostics14030345.
- [42] D. A. Arafa, H. E. -D. Moustafa, H. A. Ali, A. M. T. Ali-Eldin, and S. F. Saraya, "A deep learning framework for early diagnosis of Alzheimer's disease on MRI images," *Multimed Tools Appl*, vol. 83, no. 2, pp. 3767–3799, Jan. 2024, doi: 10.1007/s11042-023-15738-7.
- [43] A. S. Alatrany, W. Khan, A. Hussain, H. Kolivand, and D. Al-Jumeily, "An explainable machine learning approach for Alzheimer's disease classification," *Sci Rep*, vol. 14, no. 1, p. 2637, Feb. 2024, doi: 10.1038/s41598-024-51985-w.
- [44] M. S. K. Inan et al., "A slice selection guided deep integrated pipeline for Alzheimer's prediction from Structural Brain MRI," *Biomedical Signal Processing and Control*, vol. 89, p. 105773, Mar. 2024, doi: 10.1016/j.bspc.2023.105773.
- [45] Q. K. Nguyen, T. T. H. Nguyen, V. T. K. Nguyen, V. B. Truong,
- [46] T. Phan, and H. Cao, "Efficient and Concise Explanations for Object Detection with Gaussian-Class Activation Mapping Explainer," Apr. 20, 2024, arXiv:2404.13417. doi: 10.48550/arXiv.2404.13417.
- [47] W. Zhu et al., "Predicting Risk of Alzheimer's Diseases and Related Dementias with AI Foundation Model on Electronic Health Records," medRxiv, p. 2024.04.26.24306180, Apr. 2024, doi: 10.1101/2024.04.26.24306180.
- [48] V. Dhore, A. Bhat, V. Nerlekar, K. Chavhan, and A. Umare, "Enhancing Explainable AI: A Hybrid Approach Combining GradCAM and LRP for CNN Interpretability," May 20, 2024, arXiv:2405.12175. doi: 10.48550/arXiv.2405.12175.
- [49] J. Xu, C. Yuan, X. Ma, H. Shang, X. Shi, and X. Zhu, "Inter-pretable medical deep framework by logits-constrained attention-guided graph-based multi-scale learning for Alzheimer's disease analysis," *Pattern Recognition*, vol. 152, p. 110450, Aug. 2024, doi: 10.1016/j.patcog.2024.110450.
- [50] G. Lozupone, A. Bria, F. Fontanella, F. J. A. Meijer, and C. D. Stefano, "AXIAL: Attention-based eXplainability for Interpretable Alzheimer's Localized Diagnosis using 2D CNNs on 3D MRI brain scans," Oct. 01, 2024, arXiv:2407.02418. doi: 10.48550/arXiv.2407.02418.
- [51] A. R. Monteiro, D. J. Barbosa, F. Remião, and R. Silva, "Alzheimer's disease: Insights and new prospects in disease pathophysiology, biomarkers and disease-modifying drugs," *Biochemical Pharmacology*, vol. 211, p. 115522, 2023, doi: 10.1016/j.bcp.2023.115522.
- [52] M. V. F. Silva, C. d. M. G. Loures, L. C. V. Alves, et al., "Alzheimer's disease: risk factors and potentially protective measures," *Journal of Biomedical Science*, vol. 26, no. 33, 2019, doi: 10.1186/s12929-019-0524-y.
- [53] Z. Sadeghi, R. Alizadehsani, M. A. CIFCI, S. Kausar, R. Rehman, P. Mahanta, P. K. Bora, A. Almasri, R. S. Alkhalwaldeh, S. Hussain, B. Alatas, A. Shoeibi, H. Moosaei, M. Hlad'ik, S. Nahavandi, and P. M. Pardalos, "A review of Explainable Artificial Intelligence in healthcare," *Computers and Electrical Engineering*, vol. 118, Part A, 2024, Art. no. 109370, doi: 10.1016/j.compeleceng.2024.109370.
- [54] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, "Application of explainable artificial intelligence for health-care: A systematic review of the last decade (2011–2022)," *Computer Methods and Programs in Biomedicine*, vol. 226, 2022, Art. no. 107161, doi: 10.1016/j.cmpb.2022.107161.
- [55] A. B. Arrieta, N. D'iaz Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," *CoRR*, vol. abs/1910.10045, 2019. [Online]. Available: <http://arxiv.org/abs/1910.10045>.
- [56] A. Shahroudjedjad, "A Survey on Understanding, Visualizations, and Explanation of Deep Neural Networks," *CoRR*, vol. abs/2102.01792, 2021. [Online]. Available: <https://arxiv.org/abs/2102.01792>.
- [57] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," *CoRR*, vol. abs/1703.01365, 2017. [Online]. Available: <http://arxiv.org/abs/1703.01365>.
- [58] M. R. Karim, T. Islam, O. Beyan, C. Lange, M. Cochez, D. Rebolz-Schuhmann, and S. Decker, "Explainable AI for Bioinformatics: Methods, Tools, and Applications," arXiv preprint, arXiv:2212.13261, 2023. [Online]. Available: <https://arxiv.org/abs/2212.13261>.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)