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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.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 neu-rodegenerative 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 learn-ing 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 advancesinvisualinterpretabilityyetpointingtowardspersisting issues like a deficiency of standard metrics in evaluation, sparse clinical validation, and limited generalizability. Several studies lackrule-basedexplanationsorfailtoquantifythereliabil-ityofmodeldecisionsandthereisanecessitytointegrate more longitudinal, multimodal, and demographically diversified datainADresearch.Thisreviewhighlightstheimportanceof standardised, clinically validated explainability frameworksto connect research with real-world usage. It seeks to direct future efforts in the creation of responsible, interpretable AItoolsforhealthcare,providingbeneficialinsightsforresearchers, 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 press-ing public health challenges that is characterised by early but subtle alterations in the brain much before to the onset of noticeablememoryimpairment.Alzheimer'sDisease(AD)by natureisagradualandthepredominantsourceofdementia. 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 analy-ses, somatic plus neural scrutinies, diagnostic assays, MRI or PET scans. These investigations have found that pathological changescommencefarpriortosymptomsbecomingnoticeable 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.Contemporaryadvancementsinartificialintelli-gence (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 scrutinizedand assimilated in healthcare applications, presenting visual perceptions into model decisions.

ThereisanincreasingdemandforExplainableAI(XAI) [53] frameworks capable of delivering clinically interpretable insights with high diagnostic accuracy to close the gap be-tweenperformanceandtrust.SHAP(ShapleyAdditiveExpla-nations) [4], LIME (Local Interpretable Model-Agnostic Ex-planations) [3], Grad-CAM (Gradient-weighted Class Activa-tionMapping)[9],IntegratedGradients(IG)[55],DeepLIFT [7] and Layer-wise Relevance Propagation (LRP) [1] not only confirmmodelchoicesbutalsoenableregulatoryapprovaland ethical AI deployment in healthcare.

ObjectivesandPurposeoftheReview

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 withtheincreasedapplicationofAIinhealthcare.Thisreview 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 stud-ies involving imaging and clinical data for their clinical utility.
- Mapemergingtrendsandlimitationsinexistingliterature.

II. LITERATURE REVIEW

The reviewed studies have been divided into four distinct clustersaccordingtothemethodologyused:(i)reviewstudies relatedtoADresearch,(ii)studiesthatusemachinelearningor deeplearningmodelswithoutincorporatingexplainability,(iii) studiesthatintegrateXAImethodswithinML/DLframeworks, and (iv) studies focusing on the novel explainable AI (XAI) techniques.

A. ReviewstudiesinADresearch

Numerousstudieshaveexaminedtheimplementationof DL techniques for AD diagnosis, emphasising both benefits and limitations.WhileYangetal.[13]investigateapproachessuch 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 compar-ative 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 offerarestrictedevaluation ofspecificXAImethodologies.

[36]emphasisesMRI-basedneuroimagingandhighlightsLRP and counterfactuals for trust; nonetheless, issues pertaining to clinicalvalidationandstandardisationpersist.Muradetal.[39] group XAI methodologies into intrinsic, visualization-based, and distillation categories, and suggest the use of fuzzy logic toaddressambiguityinclinicaldata.Thesestudiescollectively emphasise the necessity for enhanced integration of XAI and real-world validation in Alzheimer’s disease diagnosis.

B. StudiesUsingMLorDLModelsWithoutXAIIntegration

CNNs, ensemble methods, and lightweight architectures have demonstrated significant improvement in diagnostic ac-curacy but they fall short in offering interpretable decisions, limiting their transparency and clinical adoption. Ensemble techniques have been used by researchers to boost model performanceinADdiagnosis.IslamandZhangetal.[5] used an ensemble of deep CNNs on the OASIS dataset, reportingimprovedaccuracyovertraditionalmodels.Loddoet 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 andclinicaldatathroughanensembleofCNNs,demonstrating improved diagnosis accuracy. Yet in spite of their efficiency, these models lack transparency, as they fail to provide inter-pretationfortheirpredictions,hencerestrictingtheirreliability 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 gen-eralisation to external datasets such as ADNI and lacked interpretability.Shanmugametal.[18]usedT1-weightedMRI 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 automatedtechniquethatintegrateswhitemattersegmentation andmachinelearningforthediagnosisofAlzheimer’sdisease. Improved image processing and feature extraction result in highly accurate performance. Arafa et al. [41] assessed two methodologies:acustomCNNdevelopedfromscratchand a transfer learning model employing VGG16 for MRI-based categorisation. Fouad et al. [38] developed a fully automated systemutilisingEEGsignals,employingResNet-50toidentify significant patterns in cerebral activity. Although the models demonstrate accuracy, they lack insight into feature contri-butions, hence limiting their transparency and potential for clinical use. Another fully automated deep learning pipelineby 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.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

C. StudiesIntegratingExplainability

IntegratingXAIintodeeplearningframeworksinhealthcare 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 a 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 attentionmapsfromCNNmodelsandidentifydisease-relevant areas. It was observed that the method had limitations in dif-ferentiating stable from progressive MCI. Other studies—like thosebyShojaeietal.[35],Z.Yaoetal.[37],Nguyenet al. [32], and D. Wang et al. [33]—examined SHAP, voxel grading,andquantitativeheatmapvalidationtoenhancemodel transparency. Most common limitations encountered were small datasets or lack of resolution. More advanced models, suchasMAXNet[30],MPS-FFA[34],andLA-GMF[47], utilizedattentionmechanismsandgraph-basedfusiontobetter 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 insightsintoCNNpredictionsbasedonEEGsignalstoidentify keybrainregions.Vitonetal.[11]madeuseofheatmapsin 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. Abuhmedet 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,themodelpredictionswerebasedonlifestyleand cognitivefeatures,revealingnovelinsightsintoADrisk.Rule-based methods were used with SHAP and LIME by Alatrany et al. [42] to make predictions more transparent, while visual explanationswiththehelpofsaliencymapsinensembleCNNs 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 be-tween different brain regions. L. Wang et al. [25] formulateda GNN model that uses self-attention to demonstrate the interactionsbetweendifferentpartsofthebrain.Zengetal.

[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 in-troduced by Lozupone et al. [48] is notable because of its clinically validated andinterpretable outcomes. The model utilises 2D CNNs for processing 3D MRI images and uses a soft attention mechanism to produce voxel-level maps for lo-calising disease-relevant regions. It employs double transfer learning from AD vs CN classification to the more intricate pMCI versus sMCI differentiation. The model reveals 17brainareas,includingthehippocampus,parahippocampus,and amygdala,thatstronglyalignwithradiologistverdicts,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 frameworksstilllackclinicalvalidation,rule-basedreasoning, or generalization across populations. These remain important areas for future research.

D. XAITechniques

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 researcherstogetinsightsaboutthemostinfluentialfeaturesin the context of model's prediction, thereby making the models more interpretable and trustworthy in clinical settings. Some studiesalsoemployattentionmechanisms.Recentresearchhasalsoincludedhybridapproachesthatcombinevisualheatmapswithquantitative attr ibutionsandgraph-basedembeddings 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

Consideringthegrowinguseofdeeplearningandma-chine learning frameworks for Alzheimer's diagnosis, certain challenges remain. Studies make improving predictive per-formance the priority but overlook the explainability aspect, which is essential for clinical credibility.

Despite the use of visualexplanations,anacceptedstandardforassessingtheir reliability or effectiveness remains absent. Heatmaps, though common, seldom offer explicit, rule-based justification for predicted outcomes. Furthermore, a majority of approachesare evaluated on a singular dataset, such as ADNI, which restrains their applicability across diverse settings or patient demographics. Clinical validation for the predictions and ex-planations produced by these models is available for only a few studies conducted. As early identification is crucial, afew 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 tothe 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. *DeepLearninginADResearch*

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 pre-dictionperformanceespeciallywhentheavailablityofdata is limited. Transformers and attention-based models have shown promise in predicting the AD risk and are applied increasingly to handle sequential data suc 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 improtant features, GANs for generating synthetic data and hybrid models using different methods and multimodal data.

B. *XAITechniquesforADDiagnosis*

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

- Visualization Techniques: The techniques provide visual perceptionsformodeldecisions.GradientGradient-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 advantageofattention-basedexplainabilitytoenableinterpretability in localized diagnosis. [30] uses Dual Attention Module andHigh-resolutionActivationMapping(HAM)foraddi-tional visual interpretability. Occlusion-based techniques such as the Occlusion Test and Swap Test [10] measure featureimportancebyoccludingportionsoftheinput in a systematic manner and measuring it’s effect on the model’s prediction.
- Feature-based Methods: LIME, SHAP provide impor-tance 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 employedtodeterminethecausalrelevanceoffeatures inCNNpredictions.Rule-basedtechniquessuchasClass Association Rule Mining (CAR) and SIRUS [42] derive human-readable rules from the model predictions, giving globalinterpretability.Anchorsalsoofferexplanationsas if-then rules.

TABLEI
XAI METHODS REVIEWED

Ref	XAIMethod	KeyFindings
[2]	ClassActivationMap(CAM)	ClassActivationMapsaregeneratedusingglobalaveragepoolingandlinearclassification.CAM allowsforthevisualizationofclass-specificunitsinCNNs,providinginsightsintowhichfeaturesand image regions are most discriminative for different classes.
[1]	Layer-wiseRelevancePropa-gation	Providespixel-levelattributionbyredistributingpredictionscoresbackwardthroughthe network and can provide meaningful explanations for the decisions of non-linear image classifiers. Thegenerated heatmaps highlight image regions that are relevant for a particular classificationdecision, even for complex models like deep neural networks. While visual assessment isimportant, there is a need for more robust, quantitative methods to evaluate the quality of pixel-wise explanations.

[3]	LIME	Perturbsinputfeaturesandtrainslocalsurrogatemodels.SP-LIME,byselectingadiverse and representative set of explanations, provides a valuable global understanding of the model'sbehaviour and helps users identify potential issues and limitations.
[4]	SHAP	SHAPprovidesauniquesolutionamongadditivefeatureattributionmethodsthatguaranteeslocal accuracy,missingness,andconsistency.KernelSHAPismoresample-efficientandaccuratethanShapley sampling and LIME.
[7]	DeepLIFT	Tracksdifferencesfromareferenceinputacrosslayers.ThechoiceoftheDeepLIFTrule(Linear, Rescale,RevealCancel)canimpactperformance,andtheoptimalchoicemaydependonthenetwork architecture and the task.
[9]	Grad-CAM	Grad-CAMgeneralizestheCAMtechnique,makinggradient-basedlocalizationapplicabletoa much wider range of CNN architectures. It produces class-discriminative visual explanationsusing gradient information. It can be effectively used to diagnose failure modes in CNNs,understand vulnerabilities to adversarial attacks, and identify biases in training datasets. Com-bining Grad-CAM with high-resolution gradient-based methods like Guided Backpropagationyields Guided Grad-CAM, which provides both class-discriminative and fine-grained visualexplanations.
[46]	AHybridApproachCombin-ingGradCAMandLRPforCNN Interpretability	Aninnovativetechniquetoimprovetheinterpretabilityof CNNsbyproducingclearerand more precise visual explanations. The combined method offers superior interpretability andcomprehensibleresultscomparedtostandaloneGradCAM,LRP,andcombinationsofGradCAMwith other visualization techniques.
[44]	Gaussian-Class Activation MappingExplainer(G-CAM)	Gaussian-weightedClassActivationMapsforobjectlocalization.G-CAMEachievesstate-of-the-artperformanceinspeedandaccuracyforobjectdetectionexplainability.Especiallyeffectivefortinyobjectdetectionwhereothermethodsstruggle.Reducessaliencynoise,improvesexplanationtrust, and offers real-time applicability.

- Example-based Methods: These methods explain a pre-diction by relating it to similar or contrasting examples. Methodsinvolvecounterfactualexplanations,Contrastive Explanation Method (CEM)[12], and employing proto-types. MAXNet's [30] Prediction-basis Creation and Re-trieval (PCR) module employs a reference set of relevant labeledsamplestosupplyevidenceformodelpredictions.

IV. CONCLUSION AND FUTURE WORK

Thisreviewhighlightstheincreasingoverlapbetween 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.Sincemanyheatmap-basedoutputsdostill fail to provide actionable rule-based perceptions, the fieldnowlacksstandardizedbenchmarksforanyevaluatingof explainabilitymethods.Smallorimbalanceddatasetsgenerate concern, plus underrepresentation of diverse patient groups and limited external validation generate concern.

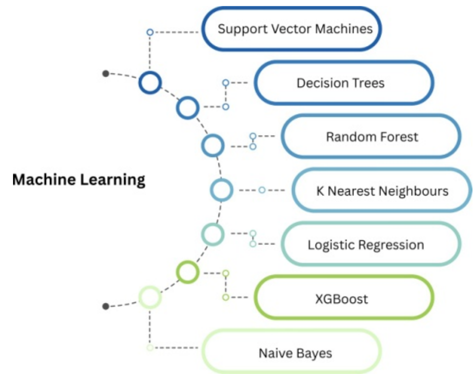

Fig.1.MachineLearningTechniquesUsedInADReasearch

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 withclinicalstakeholdersforexplainableAIinADdiagnosis.

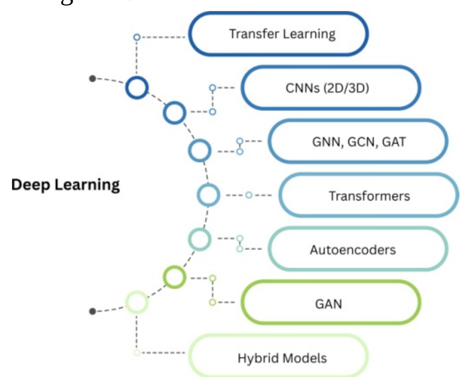

Fig.2.DeepLearningTechniquesUsedInADReasearch

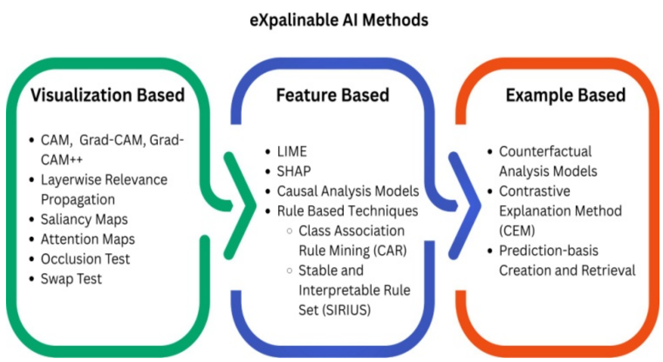

Fig.3.XAITechniquesUsedInADReasearch

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.013014010.1371/journal.pone.0130140.

[2] B.Zhou,A.Khosla,A.Lapedriza,A.Oliva,and a. Torralba, “Learning Deep Features for DiscriminativeLocalization,” arXiv preprint arXiv:1512.04150, Dec. 2015, doi:https://doi.org/10.48550/arXiv.1512.0415010.48550/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.1602.0493810.48550/arXiv.1602.04938.

[4] S.LundbergandS.-I.Lee,“AUnifiedApproachtoInterpretingModel Predictions,” arXiv preprint arXiv:1705.07874, Nov. 2017, doi:https://doi.org/10.48550/arXiv.1705.0787410.48550/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/s40708-018-0080-310.1186/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.2018.12.00810.1016/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.1704.0268510.48550/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.knosys.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,”in2021IEEEAppliedImageryPatternRecognitionWorkshop(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.knosys.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.knosys.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¸c¸iftc¸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.PhoebeChen,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’sdisease 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 DeepLearning Model for Prediction of Severity of Alzheimer’s Disease,” in2023IEEEConferenceonComputationalIntelligenceinBioinformaticsand 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’sdisease from central lobe EEG signals utilizing machine learning andresidual 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. Yousuf, and J. S. Yalli,“Unraveling the Black Box: A Review of Explainable Deep LearningHealthcare 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,“AnExplainableAIParadigmforAlzheimer’sDiagnosisUsing 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,andS.F.Saraya, “A deep learning framework for early diagnosis of Alzheimer’sdiseaseonMRIimages,”MultimedToolsAppl,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 diseaseclassification,” 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 pipelinefor Alzheimer’s prediction from Structural Brain MRI,” BiomedicalSignal 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 ObjectDetectionwithGaussian-ClassActivationMappingExplainer,”Apr.20,2024, arXiv:2404.13417. doi: 10.48550/arXiv.2404.13417.

[47] W.Zhuetal.,“PredictingRiskofAlzheimer’sDiseasesandRelatedDementiaswithAIFoundationModelonElectronicHealth 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, “Enhanc-ingExplainableAI:AHybridApproachCombiningGradCAMand 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,andX.Zhu,“Inter-pretablemedicaldeepframeworkbylogits-constraintattentionguid-inggraph-basedmulti-scalefusionforAlzheimer’sdiseaseanaly-sis,” 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’sLocalized 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. Remia˜o, and R. Silva, “Alzheimer’sdisease: 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’sdisease: risk factors and potentially protective measures,” Journal ofBiomedical 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. Alkhawaldeh, S. Hussain, B.Alatas, A. Shoeibi, H. Moosaei, M. Hlad´ık, S. Nahavandi, and P. M.Pardalos,“AreviewofExplainableArtificialIntelligenceinhealthcare,”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),” ComputerMethodsandProgramsinBiomedicine,vol.226,2022,Art.no.107161,doi: 10.1016/j.cmpb.2022.107161.

[55] A.B.Arrieta,N.D´ıazRodr´ıguez,J.DelSer,A.Bennetot,S.Tabik,A.Barbado,S.Garc´ıa,S.Gil-Lo´pez,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. Shahroudnejad, “A Survey on Understanding, Visualizations, andExplanation of Deep Neural Networks,” CoRR, vol. abs/2102.01792,2021. [Online]. Available: https://arxiv.org/abs/2102.01792.

[57] M.Sundararajan,A.Taly,andQ.Yan,“AxiomaticAttributionforDeepNetworks,”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. Rebholz-Schuhmann, and S. Decker, “Explainable AI for Bioinformatics: Meth-ods, Tools, and Applications,” arXiv preprint, arXiv:2212.13261, 2023.[Online]. Available: https://arxiv.org/abs/2212.13261.