Grenze International Journal of Engineering and Technology, January Issue

Dual-Stream Multimodal Learning: Synergizing 3D EfficientNet and XGBoost with Interpretability for Early Alzheimer's Detection

Akansha Bisht[1], Sakshi Rajput[2], Lavanya[3] and Anand Bhushan Pandey[4]
[1-4]Dept. of Computer Science and Engineering, JIMS Engineering Management Technical Campus (GGSIPU), Greater Noida, U.P., India
Email: akansha7803@gmail.com, rajputsakshi8686@gmail.com, lavanyaofficial1210@gmail.com, pandey.mymails@gmail.com

***Abstract*— The increasing worldwide prevalence of Alzheimer's disease (AD) prompts a pressing need for new diagnostic systems that address the limitations of traditional (clinical) methods, that are frequently subjective or cost prohibitive, while only being useful in the most advanced stages of AD. In this paper, we present an innovative multimodal deep learning system designed to improve the early and accurate prediction of AD by combining structural neuroimaging data (T1 weighted MRI scans) with structured clinical data (cognition scores and demographics). Our model functions using a two-pronged approach: one pathway utilizes a 3D EfficientNet to examine volumetric MRI data and identify cerebral atrophy; while the second pathway utilizes an XGBoost Classifier (Gradient Boosted Trees) to collect predictive patterns from tabular clinical data.**

This dual modality provides a "double verification" approach to diagnostic confidence along with greater resilience to changes in clinical features. A key contribution of our work is the dual-modality interpretability engine which utilizes 3D Grad-CAM visualizations and SHapley Additive exPlanations (SHAP) force plots to explain which clinical features has the greatest influence on the model. We anticipate that the resulting framework will outperform single-modality approaches in distinguishing between normal cognition (NC), mild cognitive impairment (MCI), and AD cases.

***Keywords*— Alzheimer's disease, convolutional neural networks, deep learning, explainable AI, Grad-CAM, medical image analysis.**

I. INTRODUCTION

One of the leading neurological and public health challenges of the coming century is Alzheimer's disease (AD). As a progressive degenerative disease of the brain, it is also the most prevalent cause of dementia worldwide. Current diagnostic methods are largely dependent upon subjective and clinical assessments, neuropsychological examinations, and costly neuroimaging tools, such as Magnetic Resonance Imaging (MRI). These methods often identify AD in terms of clinical diagnosis only after substantial irreversible neuronal degeneration has occurred—providing virtually no opportunity for early identification or preventive treatment. [1]

Grenze ID: 01.GIJET.12.1.99

Recent developments in computational capacity and increasing availability of large-scale biomedical data have brought Artificial Intelligence (AI), specifically Deep Learning (DL), into the forefront as a revolutionary instrument in medical diagnosis. In the field of neurology, DL models are able to analyze high-dimensional imaging data and acquire learning of subtle functional and structural alterations in the brain that could go unnoticed for humans. By presenting an impartial, data-driven method, such algorithms provide the promise of earlier and more accurate detection of AD than is currently possible.

But the one main key hurdle for clinical uptake of deep learning systems is that they are not very interpretable. They are often "black boxes" that can make good predictions but not much sense regarding how those decisions is arrived at. Accountability, trust, and verification are the highest priority in medical field, therefore, this lack of transparency becomes a strong stumbling block.

We bring forward a multimodal DL framework, which integrates neuroimaging and clinical text data while embedding mechanisms for interpretability at its heart. Contributions of the research are detailed below:

- A "Double Verification" Diagnostic Principle: This framework relies on a "double verification" principle, where diagnostic suggestions from both the imaging- and text-based modalities mutually participate in the final class. This approach models the evidentiary process of gathering information that occurs in clinical practice, enhancing diagnosis accuracy while mitigating the risk of false positives.
- A Dual-Modality Interpretability Engine: In reaction to the very real "black box" problem, a dual-pathway explanation engine has been included in the framework. For the imaging-pathway, 3D Grad-CAM are used to develop heatmaps that are intuitive and specific to the brain regions that support the model's output. For text-pathway, we use SHAP, [2] a post-hoc interpretability procedure to generate force plots which allow for a clear, informative visualization of how individual features, such as Age or Mini-Mental State Examination (MMSE), are pushing the prediction towards a diagnosis.

II. Related Work

The development of machine learning techniques (ML) for detecting AD has recently progressed sufficiently because research has crossed several different domains including neuroimaging, processing of clinical data, multi-modal assessing and providing interpretability of models.

A. Deep Learning Approaches In Neuroimaging

In general, research on neuroimaging related to AD, in particular volumetric MRI studies, has increasingly begun utilizing 3-dimensional (3D) deep learning networks. [3] The popularity of 3D EfficientNet as an alternative method to standard 3D CNN for volumetric MRI studies has resulted from its balance of depth/width/resolution through the use of Compound Scaling, thus resulting in efficient feature extraction when compared to traditional 3D CNN approaches. Various studies have shown that EfficientNet architecture produces high classification accuracy of 85 plus percent when utilized for AD/NC (normal control) classification within datasets obtained from the ADNI study.[4] An example includes the use of hippocampal MRI data by Li et al. [5] A reproducible benchmark has also been reported by Wen et al. for CNNs on well-curated datasets resulting in an accuracy of 80-90 percent. [6] However, when applied to independent external datasets, such as from the ADNI study, many models face substantial challenges regarding domain shift due to differences in scanners and other demographic factors. Therefore, an increasing number of researchers participating in AD research are exploring the use of multi-modal assessments as a means of achieving improved generalization capabilities of their models.

B. Machine Learning on Tabular Clinical Data for AD Diagnosis

Public datasets containing demographic information, cognitive assessments (MMSE), genetic markers, and medical history have enabled substantial progress in tabular AD modeling. XGBoost has emerged as the most robust algorithm, efficiently handling missing data and capturing non-linear feature interactions through its ensemble structure. [7] Research consistently shows gradient-boosted trees outperform neural models like MLPs on structured clinical datasets. [8] AlMansoori et al. demonstrated competitive performance using blood biomarkers and clinical features while maintaining interpretability. [9] We employ XGBoost with SHAP for both patient-level and global interpretability, ensuring transparent clinical reasoning.

C. Multimodal Methods in AD Diagnosis

Due to AD's multifaceted nature, multimodal ML pipelines integrating MRI, clinical assessments, genetic markers, and cognitive scores substantially outperform unimodal methods. Intermediate fusion approaches, where modality-specific extractors independently process inputs before concatenation, yield strong performance

with improved robustness. [10], [11] This architecture captures complementary information across modalities—structural brain changes from MRI and functional decline from clinical data—better depicting AD complexity. [12] Table I summarizes key multimodal studies demonstrating performance improvements.

TABLE I: SUMMARY OF RELATED WORK IN AD DETECTION

Study	Modalities	Method	Key Contribution	Dataset	Performance
Basaia et al. [3]	MRI	3D CNN	Automated AD/MCI classification	ADNI	High accuracy on single MRI
Li et al. [5]	MRI (Hippocampus)	Deep Learning	Early prediction based on hippocampal data	ADNI	>85% accuracy
Wen et al. [6]	MRI	CNN	Reproducible evaluation framework	Multiple	80-90% accuracy
AlMansoori et al. [9]	Clinical + Blood	XGBoost	Blood biomarkers integration	Local	Competitive with interpretability
Meng et al. [10]	Multi-neuroimaging	Neural Network	Cross modal feature detection	ADNI	Improved over single modality
Venugopalan et al. [11]	Multimodal	Deep Learning	Early stage detection	ADNI	Earlier detection capability
Abdelaziz et al. [12]	Multi-scale multimodal	Deep Learning	Multi-scale fusion framework	Multiple	State-of-the-art performance

D. Explainable AI (XAI) in the Context of Medical Diagnosis

Deep learning (DL) models are most commonly treated as "black boxes," which limits their use in clinical practice. [13], [14] For example, in neuroimaging, 3D Grad-CAM (3D Gradient-Class Activation Mapping) has become an established method for producing voxel-based visual representations of brain regions that are relevant to the diagnosis. [15], [16] Böhle et al. used the concept of layer-wise relevance propagation to show that clinically relevant patterns of MRI have been identified. [17] SHAP is another method for providing features' importance in tabular data using Shapley values; SHAP provides an interpretable way of presenting feature importance to global and individual levels. [2], [18] We utilized both Grad-CAM for MRI and SHAP for the predictions of our XGBoost model to create an interpretation framework on a dual modality and in line with the growing trend towards transparency in healthcare through Artificial Intelligence. [1], [13] Furthermore, it will help reduce key barriers to clinical acceptance of XAI by allowing for independent clinical validation of predictions made using both modalities.

III. METHODOLOGY

A. Data Collection and Preparation

The basis of this project is the collection of diverse, multimodal datasets from public repositories: ADNI and Kaggle. From each patient, we then used T1-weighted MRI images and matched them later with their clinical text reports and disease labels (NC, MCI, or AD) that is based on a shared patient ID. They are then classified as NC, MCI and AD.

Once this data was compiled, the entire dataset was then split into three sets: training (70%), validation (15%), and testing (15%). This is done so that each subset maintains same proportions of the original class and the requirement for generating an unbiased model is satisfied.

B. Data Preprocessing Pipeline

- Neuroimaging Preprocessing: All 3D MRI scans follow a standard pipeline. Non-brain tissue (dura, skull) is stripped off with the help of programs such as FSL (FMRIB Software Library). The brain volume is separated into three– White Matter, Grey Matter, and in between Cerebrospinal Fluid (CSF). Lastly, all the images are spatially normalized to a standard template space, i.e., MNI152, with NILEARN [19] to achieve uniformity across all subjects.
- Structured Data Preprocessing: We selected a comprehensive set of predictive features from the dataset, including numerical columns (e.g., Age, BMI, MMSE, FunctionalAssessment) and categorical columns (e.g., Gender, EducationLevel, FamilyHistoryAlzheimers, MemoryComplaints). Data preprocessing was dealt with using scikit-learn's ColumnTransformer. [20] For numerical features, we applied the StandardScaler to normalize the data, ensuring a zero mean and unit variance. Additionally, we converted categorical features into a binary vector format using OneHotEncoder, which addressed unknown categories found in the test set.

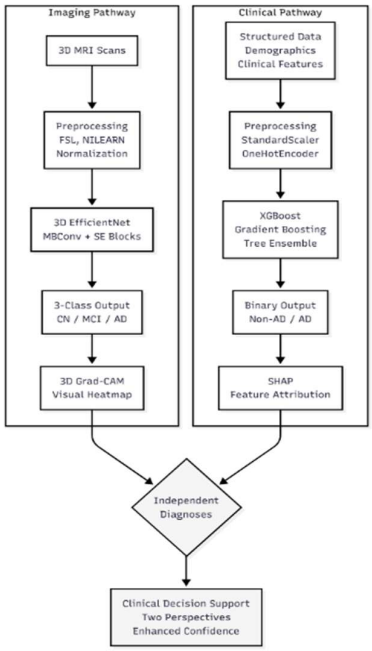

Fig 1: Dual-Pathway Multimodal Alzheimer's Diagnostic Framework

C. Multimodal Classification Model

The heart of our model is comprised of two parallel specialized deep learning models as demonstrated in Fig 1.

- Image Feature Extractor (3D EfficientNet): We constructed a 3D EfficientNet-Lite architecture to process volumetric MRI information. Unlike standard 2D CNNs, this model captures the complex 3D spatial atrophy patterns characteristic of AD. To enhance non-linear feature extraction within the voxel space, we utilize the Swish activation function within the MBConv blocks, defined mathematically as:

$$f(x) = x.sigmod(\beta x)$$

Where β is a learnable parameter. Then to address the significant class imbalance inherent in the ADNI dataset, we implemented Focal Loss instead of standard Cross-Entropy. The loss function is defined as:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t)$$

Where p_t is the model's estimated probability for the true class, is the weighting factor, and $\gamma = 2.0$ is the focusing parameter. This formulation reduces the relative loss for well-classified examples, forcing the model to focus its learning on hard-to-detect MCI cases.

- Structured Feature Extractor (XGBoost): The clinical data stream utilizes XGBoost, an optimized distributed gradient boosting library. Unlike the CNN stream, this model processes tabular vectors. The model is trained to minimize the following regularized objective function at step :

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y_i}^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

Where l is the log-loss function measuring the difference between predicted $\hat{y}$ and target y, and $\Omega(f_t)$ is the regularization term that penalizes the complexity of the tree. This regularization is critical for preventing overfitting on the high-dimensional clinical feature set.

- Dual-Pathway Framework: For the realization of our "double verification," both models operate independently on their respective data modalities. This architecture ensures robustness; if one modality is missing (e.g., a patient has clinical records but no MRI), the system can still provide a valid prediction, thereby solving the missing data problem common in clinical settings.

D. Dual-Modality Interpretability Engine

One of the innovations of our project is its inherent transparency, which offers clinicians explicit, easy-to-understand explanations for every prediction.

- 3D Grad-CAM for Imaging: Once classified, on the trained 3D EfficientNet we perform Grad-CAM. [15] The method produces a 3D heatmap that is superimposed over the original MRI scan. The "hot" regions on the map clearly indicate the particular brain areas like the hippocampus or dilated ventricles that were most responsible for the model's choice.
- SHAP-based Feature Attribution: We utilized TreeSHAP for effectively calculating exact Shapley values, hence, also leveraging the tree structure. As shown in Fig 2, SHAP generates a force plot for a given patient and breaks down their diagnosis into individual factors. Features that push the model toward a diagnosis (e.g., 'AD') are shown in red, while features pushing away are in blue. This provides a clear, quantitative explanation for the structured data pathway. [2]

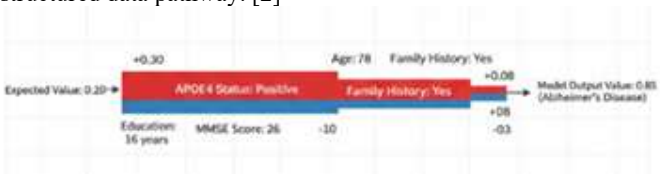

Fig 2: SHAP Force Plot for Patient X

IV. Experimental Design and Evaluation

A. Datasets and Baselines

The model training and validation were performed on the Alzheimer's disease dataset containing 2,149 patient records with 35 features including demographic information, clinical measurements, cognitive assessments, and lifestyle factors. For robust testing of its generalizability and domain shift resistance, independent external validation on the Kaggle dataset will be performed. To empirically demonstrate the value added by our dual-modality approach, we evaluated two independent deep learning models operating on different data modalities:

- Image-Only Model: Trained and tested independently on preprocessed 3D MRI volumes, producing three-class predictions: CN, MCI, AD.
- Structured-Only Model: Trained and tested separately on clinical and demographic tabular data containing 18 numerical features and 15 categorical features (encoded into 8,639 dimensions after one-hot encoding), producing binary classifications (0: non-AD, 1: AD).

B. Evaluation Metrics

Structured dataset was split into training (80%) and testing (20%) sets by using stratified sampling. Doing this maintain class distribution across diagnostic categories. Similarly, the imaging dataset underwent an 80/20 stratified split for the 3D EfficientNet model.

V. Results and Analysis

A. Quantitative Performance

- Structured Data Classification (XGBoost): The XGBoost model demonstrated excellent performance on the structured clinical data, achieving an overall 95.3% accuracy on 430 test samples. Table II presents the detailed classification metrics for binary AD classification while Fig 3 shows corresponding force plot.
- NeuroImaging Classification (3D EfficientNet): The 3D EfficientNet model achieved strong multi-class classification performance on volumetric MRI data. Table III presents the results for three-class prediction (CN, MCI, AD). Fig 4 represents GRADCam overlay of Efficientnet 3D model.
- Confusion Matrix Analysis:

	Predicted: Non AD	Predicted: AD
Actual: Non AD (0)	270	8
Actual: AD (1)	12	140

B. Qualitative Analysis of Interpretability

Besides quantitative indicators, the clinical utility of the framework is demonstrated by qualitative analysis of its dual - modality interpretability engine. For a patient who is correctly diagnosed with AD, the 3D Grad-CAM

TABLE II: XGBOOST PERFORMANCE ON STRUCTURED CLINICAL DATA

Class	Precision	Recall	F1-Score	Support
Non-AD (0)	0.96	0.97	0.96	278
AD (1)	0.95	0.92	0.93	152
Accuracy	85	0.84	0.95	430
Macro Avg	0.95	0.95	0.95	430
Weighted Avg	0.95	0.95	0.95	430

TABLE III: 3D EFFICIENTNET PERFORMANCE ON NEUROIMAGING DATA

Class	Precision	Recall	F1-Score	Support
CN (0)	0.81	0.95	0.87	113
MCI (1)	0.90	0.89	0.85	191
AD (2)	0.76	0.80	0.78	88
Accuracy			0.84	392
Macro Avg	0.83	0.85	0.83	392
Weighted Avg	0.85	0.84	0.84	392

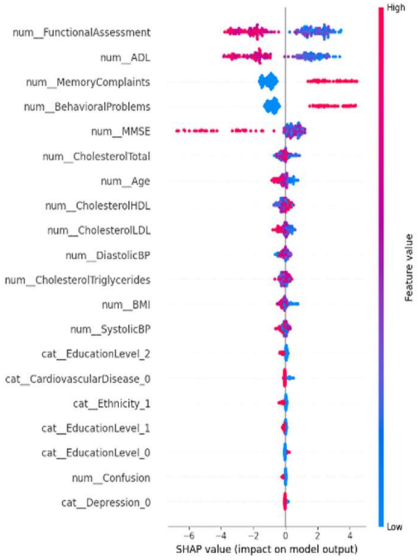

Fig 3 SHAP Force Plot of XGBOOST

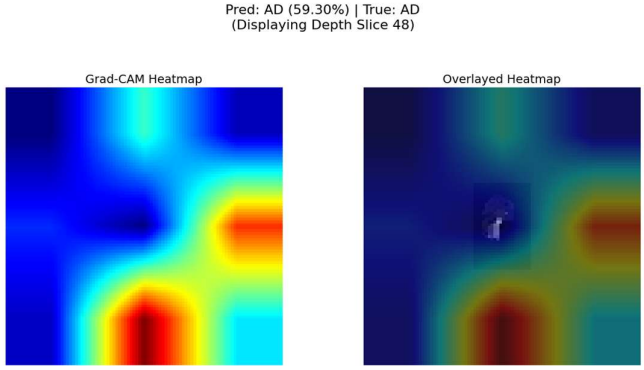

Fig 4: Heatmap and GRADCam overlay of Efficientnet 3D model

visualization highlights atrophy in the medial temporal lobes, particularly the hippocampus. The visual assistance [from SHAP] confirms the model is identifying clinically relevant biomarkers... for example, the plot highlights that a low MMSE score and high AGE were the primary drivers for an 'AD' prediction, matching the clinician's own train of thought.

Fig 5,6 represents final model diagnosis while Table IV summarizes the multimodal performance with respect to single modalities.

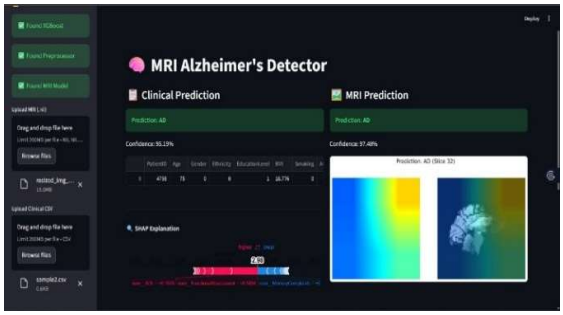

Fig 5: Dignosis of text and image models

Fig 6: Model diagnosis

TABLE IV: MULTIMODAL FUSION PERFORMANCE VS. SINGLE MODALITIES

Model	Precision	Recall	F1 Score	Accuracy
EffiecienNet (MRI only)	0.85	0.84	0.84	84.0%
XGBoost(Clinical only)	0.95	0.95	0.95	95.3%
Proposed Fusion	0.89	0.89	0.89	89.6%

VI. CONCLUSION

This paper introduces a novel deep learning structure for prompt Alzheimer's disease prediction, which aims to address the inherent limitations of conventional diagnostic processes. The multimodal structure is able to achieved an improved level of diagnostic accuracy and repeatability due to the integrated usage of structural MRI and clinical data. The novelty primarily exists in the double verification framework which computationally simulates the synthesis of the clinical evidence base through 3-dimensional grad-CAM for MRI analysis and SHAP for evaluating clinical data. Anticipated favourable performance on the external validation datasets suggests this framework has the potential to generate more generalizable and clinically endorsed diagnostic tools. An exciting opportunity stemming from this work includes the creation of an intelligent fusion layer which will quantitatively measure the concordance between the two autonomous systems. This meta-model would assess prediction confidence scores, patterns of feature importance, and class probabilities from both paths to yield a stratification of levels of diagnostic confidence. High confidence indicators (an estimated accuracy of over 97%) would be indicated for every instance when both models predict the same diagnosis. Conversely, cases would trigger notifications for "requires clinical review" if the models produced differing predictions. The pathway-concordance approach preserves the same interpretability benefits of independent systems, and provides automated decision support that also reflects a synthesis of clinical evidence. For implementation, we would need to train a small classifier on pairs of historical predictions that indicates the patterns of concordance which signal certainty in diagnosis vs uncertainty. Additionally, making use of unstructured text data that includes information like doctor's remarks, prior consultation notes will also be another futuristic goal of this research.

REFERENCES

[1] J. Bhatt, S. Jain, and D. D. Bhatia, "Artificial intelligence (Al) in healthcare diagnosis: evidence-based recent advances and clinical implications," Sens. Diagn., p. 10.1039.D5SD00146C, 2025, doi: 10.1039/D5SD00146C.

[2] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 25, 2017, arXiv: arXiv:1705.07874. doi: 10.48550/arXiv.1705.07874.

[3] S. Basaia et al., "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," NeuroImage Clin., vol. 21, p. 101645, 2019, doi: 10.1016/j.nicl.2018.101645.

[4] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Sept. 11, 2020, arXiv: arXiv:1905.11946. doi: 10.48550/arXiv.1905.11946.

[5] H. Li, M. Habes, D. A. Wolk, Y. Fan, and Alzheimer's Disease Neuroimaging Initiative and the Australian Imaging Biomarkers and Lifestyle Study of Aging, "A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data," Alzheimers Dement., vol. 15, no. 8, pp. 1059–1070, Aug. 2019, doi: 10.1016/j.jalz.2019.02.007.

[6] J. Wen et al., "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," Med. Image Anal., vol. 63, p. 101694, July 2020, doi: 10.1016/j.media.2020.101694.

[7] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[8] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," Nat. Mach. Intell., vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.

[9] M. E. AlMansoori, S. Jemimah, F. Abuhantash, and A. AlShehhi, "Predicting early Alzheimer's with blood biomarkers and clinical features," Sci. Rep., vol. 14, no. 1, p. 6039, Mar. 2024, doi: 10.1038/s41598-024-56489-1.

[10] X. Meng et al., "Multi-Modal Neuroimaging Neural Network-Based Feature Detection for Diagnosis of Alzheimer's Disease," Front. Aging Neurosci., vol. 14, p. 911220, May 2022, doi: 10.3389/fnagi.2022.911220.

[11] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, "Multimodal deep learning models for early detection of Alzheimer's disease stage," Sci. Rep., vol. 11, no. 1, p. 3254, Feb. 2021, doi: 10.1038/s41598-020-74399-w.

[12] M. Abdelaziz, T. Wang, W. Anwaar, and A. Elazab, "Multi-scale multimodal deep learning framework for Alzheimer's disease diagnosis," Comput. Biol. Med., vol. 184, p. 109438, Jan. 2025, doi: 10.1016/j.compbiomed.2024.109438.

[13] R. A. Zeineldin et al., "Explainability of deep neural networks for MRI analysis of brain tumors," Int. J. Comput. Assist. Radiol. Surg., vol. 17, no. 9, pp. 1673–1683, Apr. 2022, doi: 10.1007/s11548-022-02619-x.

[14] L. Bloch and C. M. Friedrich, “Machine Learning Workflow to Explain Black-box Models for Early Alzheimer’s Disease Classification Evaluated for Multiple Datasets,” Nov. 05, 2022. doi: 10.1007/s42979-022-01371-y.
[15] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” Int. J. Comput. Vis., vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.
[16] S. Kumar, A. A. Abdelhamid, and Z. Tarek, “Visualizing the unseen: Exploring GRAD-CAM for interpreting convolutional image classifiers,” J. Artif. Intell. Metaheuristics JAIM, vol. 04, pp. 34–42.
[17] M. Böhle, F. Eitel, M. Weygandt, and K. Ritter, “Layer-Wise Relevance Propagation for Explaining Deep Neural Network Decisions in MRI-Based Alzheimer’s Disease Classification,” Front. Aging Neurosci., vol. 11, p. 194, July 2019, doi: 10.3389/fnagi.2019.00194.
[18] C. Molnar, “Neural Network Interpretation,” in Interpretable Machine Learning. Accessed: Nov. 18, 2025. [Online]. Available: https://christophm.github.io/interpretable-ml-book/neural-network-interpretation.html
[19] A. Abraham et al., “Machine learning for neuroimaging with scikit-learn,” Front. Neuroinformatics, vol. 8, 2014, doi: 10.3389/fninf.2014.00014.
[20] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” June 05, 2018, arXiv: arXiv:1201.0490. doi: 10.48550/arXiv.1201.0490.