Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Deep Learning for Alzheimer's Disease Diagnosis from Brain MRI: Review

Zainab Abdalhussain Kareem[a] , Ahmad Shaker Abdalrada[b]

[a]*College of Computer Science and Information Technology , Wasit university, Iraq*, zaab205@uowasit.edu.com
[b]*College of Computer Science and Information Technology , Wasit university, Iraq*, *aabdalra@uowasit.edu.ia*

ARTICLE INFO

Article history:
Received: 22 /07/2025
Rrevised form: 23 /08/2025
Accepted : 25 /08/2025
Available online: 30/09/2025

Keywords:

Alzheimer's Disease, Deep Learning, Magnetic Resonance Imaging (MRI), Convolutional Neural Network, Preprocessing, Explainable AI, Grad-CAM.

ABSTRACT

Alzheimer's disease (AD) is a progressive neurodegenerative condition that strongly impacts cognition and quality of life. Accurate and early diagnosis is essential for effective management and treatment planning. In recent years, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have been successful methods for automatic AD detection from brain Magnetic Resonance Imaging (MRI). This review synthesizes advances in deep learning using MRI for AD, focusing on preprocessing steps (rescaling, grayscale, normalization) and data augmentation techniques that ensure generalization and account for dataset imbalance. Explainable AI is also promoted for its ability to enhance transparency and clinical confidence. By description of strengths and limitations of existing approaches, this paper aims to guide researchers toward the design of accurate, interpretable, and clinically relevant AI systems for diagnosing Alzheimer's disease.

https://doi.org/10.29304/jqcsm.2025.17.32429

1. Introduction

Brain illnesses encompass a wide variety of disorders affecting physical, cognitive, and emotional health. Traumatic trauma, infections, tumors, and neurodegenerative conditions such as Alzheimer's disease (AD), Parkinson's disease, and Amyotrophic Lateral Sclerosis (ALS) are part of them [1 ,2]. Of these, AD is one of the most common neurodegenerative illnesses where there is abnormally deposited beta-amyloid (Aβ) plaques and tau protein tangles. These pathological changes affect neuronal function, leading to cell death[3]. Their interaction leads to a domino effect of neurodegenerative processes that finally affect memory, cognitive functions, and language [3, 4].

Development of stages of brain in AD is depicted in Figure. 1, where the difference between No Impairment "normal cognition (NC)" as healthy, mild decline "mild cognitive impairment (MCI)" as midle, and severe decline "Alzheimer's disease (AD)" as clinical .

∗Corresponding author: Zainab Abdalhussain Kareem

Email addresses: *zaab205@uowasit.edu.com*

Communicated by 'sub etitor'

AD is the single most frequent cause of dementia in older adults and an important public health problem globally. The primary risk factor is age, and available estimates suggest that the incidence of AD will double in 2050, with roughly half of individuals older than 85 years affected [5]. This anticipated growth will put immense pressure on healthcare systems, families, and caregivers, especially in low- and middle-income countries, highlighting the imperative need for effective healthcare planning and policy interventions [6, 7]. Still, the early diagnosis of AD remains one of the greatest challenges for clinicians considering that the disease typically is diagnosed at late stages after irreversible damage has already occurred [8]. Symptoms will blur with typical aging and other neurologic illnesses and are typically diagnosed by clinical examination and neuroimaging [9]. Earlier correct diagnosis can even enhance treatment greatly to offer interventions that prevent the deterioration of symptoms and enhance quality of life in patients and caregivers [10, 11]. This demand has fueled greater interest in more sophisticated diagnostic tools, namely artificial intelligence (AI) and machine learning (ML), that are able to measure subtle changes in the brain that are linked with AD in its initial phases [12].

Neuroimaging forms a critical part of early detection because progression of brain atrophy in AD can be noted on MRI scans [13]. MRI is capable of precise anatomical brain examination and can detect early structural changes, such as hippocampal atrophy, which are highly related to memory loss and emerge in the early stages of the disease [14]. Although MRI is superior in many ways, it has pragmatic drawbacks such as being very expensive, with time-consuming scanning times, noise, and conflicting with certain medical devices [15]. Furthermore, early AD pathology is still hard to differentiate from aging [16], so AI-based automated image analysis is a good complement to conventional radiological examination [17].

Against this backdrop, ML and deep learning techniques became central drivers of identifying salient patterns from high-dimensional medical imaging data [18]. CNNs, in particular, have proven outstanding capacity to learn discriminative features automatically from MRI scans, bypassing labor-intensive processing, enhancing diagnostic accuracy relative to traditional methods, and facilitating stage-wise classification of AD, ranging from mild cognitive impairment (MCI) to late-stage dementia [19, 20].

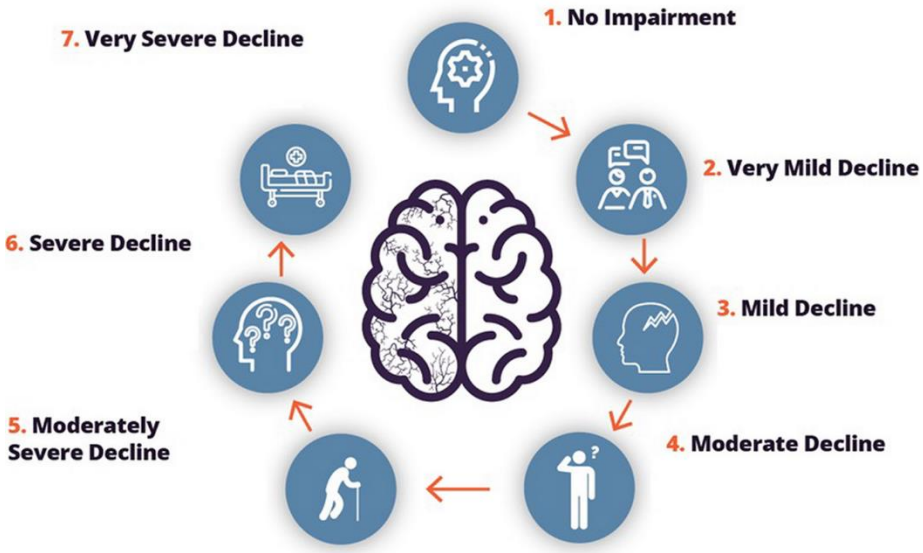

Figure 1. The Progress of brain stages for Alzheimer's disease [21].

This review combines recent advances in MRI-based deep learning for the diagnosis of Alzheimer's, and with a focus on CNN-based feature extraction, preprocessing and augmentation methods, and interpretability issues, the distance is to be bridged between model performance and clinical adoption.

The rest of the paper is organized as follows: Section 2 outlines the literature review. Section 3 outline materials and methods. Section 4 presents preprocessing methods and data augmentation in MRI, and Section 5 is about Interpretability Techniques in Medical AI Models. Section 6 lists Overfitting problem in DL and mitigation methods, then Section 7 on the evaluation measures. Section 8 gives the challenges, and section 9 directions for future work, and Section 10 summarizes the paper.

2. Literature Review

Alzheimer's disease (AD) prediction and diagnosis with brain MRI has emerged as a significantly expanded research area over the last few years, particularly following the application of machine learning (ML) and deep learning (DL) models. In this review article, we provide an overview of ten representative studies, classified by methodological approach (traditional ML, CNN-based models, transfer learning, hybrid/ensemble approaches), and highlighting their respective strengths and weaknesses.

2.1 Traditional Machine Learning Approaches

Kavitha et al. [22] compared classical machine learning models, such as decision trees, random forests, SVM, and XGBoost, on the OASIS and Kaggle datasets for binary classification (non-dementia vs. dementia). Random forest achieved the best performance, with an accuracy of 86.92%. However, the study relied on limited data and focused only on binary classification, which limits the clinical application of the work for diagnosing Alzheimer's disease across multiple stages.

Hala Al-Shamlan et al. [23] employed SVM, Random Forest, and Logistic Regression with feature selection methods (mRMR and MI) on the OASIS-2 dataset (373 scans). mRMR-backed Logistic Regression achieved accuracy of 99.08% for two-class classification. Although promising, the work was based on two classes alone and did not probe intermediate stages of AD.

2.2 CNN-BASED APPROACHES

Fazal Ur Rehman and Kwon [24] constructed a CNN trained on entire-brain MRI to distinguish AD, MCI, and CN subjects. The model attained 96.41% accuracy, surpassing more profound models such as ResNet and VGG. However, dataset size was constrained (489 scans), and there were concerns regarding the generalizability of the model. Further, accuracy in separating AD from MCI (88%) was poorer compared to binary conditions.

Abd El-Latif et al. [25] proposed a lightweight CNN trained on Kaggle Alzheimer's data for four classes: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. The accuracy achieved for binary classification was 99.22%, and for multi-class classification was 95.93%, surpassing ResNet50 and DenseNet201. The model was not yet clinically strong enough for deployment.

De Silva and Kunz [26] applied a CNN to the MIRIAD dataset and reached 89% accuracy for AD vs. HC. Despite respectable performance, the study's binary classification task overlooked the more clinically relevant challenge of multi-class classification.

Sara Esam and Mohammed [27] used Kaggle MRI data with a custom CNN achieving 97% multi-class accuracy. However, resizing images to 150×150 may have reduced spatial detail. Furthermore, training optimization techniques such as early stopping and learning rate scheduling were not employed, and interpretability methods were absent.

2.3 Transfer Learning

Jain et al.[28] employed VGG-16 pre-trained on ImageNet as a feature extractor with entropy-based slice selection for improving AD/MCI/CN classification robustness. It attained 95.73% accuracy despite handling a very small dataset (150 MRIs). Shortcomings include the potential loss of spatial information in 3D-to-2D conversion and disregarding inter-slice relationships.

Duaa AlSaeed and Omar [19] used ResNet-50 as a feature extractor with different classifiers such as Softmax, SVM, and Random Forest, on ADNI and MIRIAD data. They achieved a best performance of 99% (AD vs. NC). While high, the study only considered binary classification and did not include MCI.

2.4 Hybrid and Ensemble Approaches

Diogo et al. [29] proposed combining CNN feature extraction and Random Forest for early AD/MCI classification on ADNI and OASIS. Their HC vs. AD model achieved 90.6% balanced accuracy, whereas multi-class performance (62.1%) was considerably worse, indicating the difficulty in distinguishing MCI from AD.

Tanjim Mahmud et al. [30] presented an explainable AI solution incorporating pre-trained CNNs into ensembles (VGG16+VGG19; DenseNet169+DenseNet201) with Grad-CAM visualizations. The model reached 96% accuracy, which was higher than baseline CNNs. However, despite enhanced interpretability, accuracy was still not sufficient for clinical application.

From the papers reviewed above, several major trends emerged. Convolutional neural network (CNN)-driven approaches dominate the scene since they are able to learn hierarchical features from unprocessed MRI data itself. Performance is very sensitive to the dataset size, preprocessing techniques, and utilization or absence of multi-class classification. Transfer learning has a tendency to improve accuracy but may have domain mismatch between natural images (ImageNet) and medical imaging. Ensemble and hybrid methods yield small improvements at the cost of added complexity, making real-world application difficult. On balance, common disadvantages are reliance on small or imbalanced datasets, better accuracy in binary over multi-class, poor interpretability for all but a few models, and insufficient clinical validation for widespread use. Table 1 provides a comparative summary of studies.

Table 1 Summarize of Related work on AD.

Study, year	Dataset	Technique	Accuracy	Drawback
Kavitha et al. (2022) [22]	OASIS, Kaggle MRI Dataset	Random Forest	Binary 86.92%.	Relatively low accuracy, the limited number of the used images, and the fact that the classification was only binary.
Hala Al-Shamlan et al. (2024) [23]	OASIS	Logistic Regression	Binary 99.08%	However, a limitation of this study is that it classified only two classes (Demented and Non-Demented)
Fazal Ur Rehman and Kwon (2022) [24]	ADNI	CNN	96%	The main limitations of this study is the relatively small dataset size, with only 489 MRIs, which may impact the generalizability of the model
Abd El-Latif et al. (2023) [25]	Alzheimer 4 MRI classes dataset	CNN	95.93%	The achieved accuracy was still insufficient for reliable clinical application.
De Silva and Kunz (2023) [26]	MIRIAD	CNN	Binary 89%	A limitation of this study is that it has a binary classification (AD vs. HC) only.
Sara Esam and Mohammed (2024) [27]	Alzheimer 4 MRI classes dataset	CNN	Multiclass: 97%	Despite the good results, the accuracy still needs to be improved using optimization techniques and to enhance the interpretability of the model.
Jain et al (2019) [28]	ADNI	CNN (VGG-16, Transfer Learning)	95.73%	The sample size was very small (150 MRI images)
Duaa AlSaeed and Omar (2022) [19]	ADNI, MIRIAD	CNN (ResNet50), Softmax, SVM , RF	Binary: Softmax99% SVM 92% RF 85.7%	A limitation of this study is that it classified only two classes (AD and NC),

Diogo et al. (2022) [29]	ADNI, OASIS	CNN + RF	Binary 90.60% Multiclass 62.1%	The accuracy of the triple classification is lower than expected, which means that distinguishing between MCI and AD remains a challenge.
Tanjim Mahmud et al. (2024) [30]	OASIS	EfficientNet + CNN	96%	The achieved accuracy is still critical clinical insufficient for applications.

3. Materials and methods

This review provides a summary of the significant materials and methods reported in Alzheimer's diagnostic research, including medical imaging modalities, publicly available datasets, standard preprocessing, and CNN-based feature extraction approaches.

3.1 Dataset

Brain imaging databases from various modalities are of utmost significance in Alzheimer's disease diagnosis studies. Such datasets serve as the foundation for constructing and validating AI-based diagnostic models of Alzheimer's disease [31], The most widely utilized public databases are:

The Alzheimer's Disease Neuroimaging Initiative (ADNI) - provides extensive information from MRI and positron emission tomography (PET) scans across various stages of the disease, ADNI data is currently divided into four stages, ADNI-GO, ADNI-1, ADNI-2 and ADNI-3 [32].

The Open Access Imaging Study Series (OASIS) - provides MRI scans of healthy individuals and Alzheimer's patients across a broad age spectrum, It provides 3 types of data sets: OASIS-1, OASIS-2, OASIS-3 [31].

(Kaggle) Open Access MRI Datasets - provides expert-selected brain MRI datasets, divided into categories (e.g., no dementia, very mild dementia, mild dementia, moderate dementia) for building and evaluating deep learning models [33].

3.2 Medical Imaging

Medical imaging modalities are crucial in AD diagnosis and monitoring to enable clinicians to visualize structural and functional changes in the brain. Different imaging techniques are utilized in AD clinical practice and research and each offers differential diagnostic information [34].

Positron Emission Tomography (PET) allows visualization of pathologic protein deposits such as β-amyloid and tau, significant Alzheimer's disease pathology biomarkers. PET is highly sensitive in detecting these pathological changes even before the development of clinical symptoms but is hampered by exposure to radiation through radioactive tracers, making it costly to repeat [35].

Computed Tomography (CT) offers rapid imaging with the broad availability of scanners, thus being present in most clinical environments. However, CT has insufficient soft tissue contrast, making it less sensitive for early detection of neurodegeneration, particularly in differentiation of subtle cortical atrophy characteristic of AD [36].

Functional Magnetic Resonance Imaging (fMRI) measures brain function by detecting changes in oxygenation levels of blood and the possibility to map functional deficits in neural networks, such as the Default Mode Network (DMN),

that are compromised in the preclinical stages of AD. Despite being valuable for research purposes, fMRI is less commonly used for everyday clinical diagnosis due to complex acquisition protocols [37].

Of all these techniques, Magnetic Resonance Imaging (MRI) is the most suitable modality for early and non-invasive diagnosis of Alzheimer's disease. Structural MRI (sMRI) provides great images of brain anatomy, well capturing hippocampal atrophy and thinning of the cortex—both being hallmark features of AD development. Moreover, MRI does not use ionizing radiation, making it safer for follow-up. The compatibility of the modality with artificial machine learning and deep learning algorithms, including convolutional neural networks (CNNs), contributes to its value in clinical application and research [38, 39]. Table 2 shows a comparison between imaging methods.

Figure 2 depicts the percentage utilization of various imaging in the diagnosis of Alzheimer's disease.

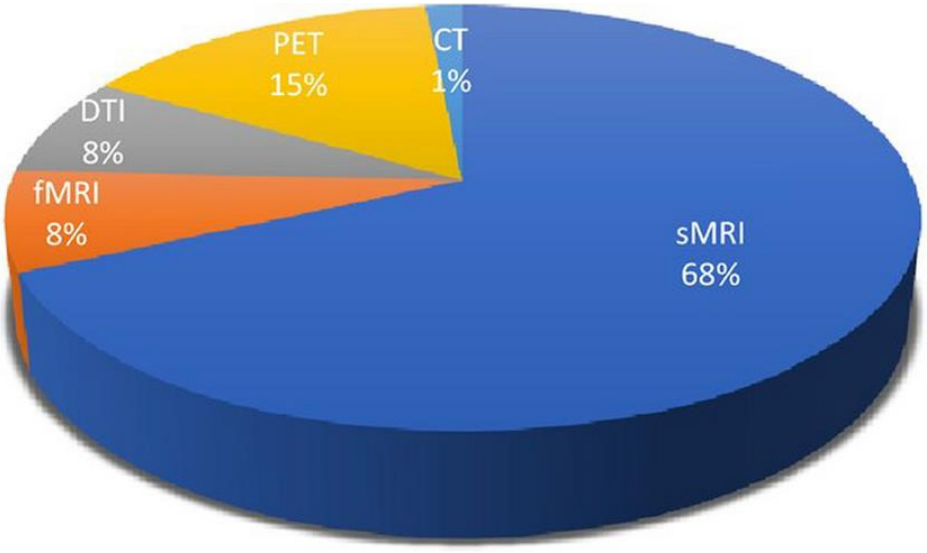

Figure 2. Utilization of each different medical imaging with AD diagnosis [21].

Table 2: Comparison of Medical Imaging methods for AD.

Imaging type	Primary Use	Advantages	Limitations	Suitability for AI/DL Integration
PET	Detects β-amyloid and tau protein deposition	Sensitive to early pathological changes	Radiation exposure, high cost, limited availability	Limited (due to cost and radiation exposure)
CT	General brain structure and atrophy	Fast, widely available	Poor soft tissue contrast, limited sensitivity to early AD	Low (lacks fine structural detail for DL analysis)
fMRI	Measures functional brain activity and connectivity	Maps early functional deficits	Complex acquisition, less available clinically	Moderate (requires careful preprocessing)
MRI	Structural imaging to detect brain atrophy (e.g., hippocampus)	High spatial resolution, non-invasive, no radiation, compatible with AI	Expensive, time-consuming, noisy, contraindicated for some patients	High (best suited for DL models like CNNs)

		models		

3.3 Feature Extraction with CNN

Convolutional Neural Networks (CNNs) have become a cornerstone of Alzheimer's disease (AD) research as they offer an unmatched ability to automatically hierarchically extract features from brain MRI scans from local, low-level textures to global, disease-specific patterns, without manual feature engineering (e.g., intensity thresholds or hand-designed biomarkers) [24]. This allows models to be more accurate, robust, and generalizable to diagnose AD than traditional machine learning methods [40].

Early studies employed simple CNN architectures such as LeNet-5 and AlexNet, which achieved reasonable classification performance in discriminating AD from healthy controls. The models, however, typically underperformed as generalizers due to the unavailability of plentiful training data [41, 42].

The emergence of deeper models—such as ResNet, DenseNet, and EfficientNet—had a significant improvement in AD detection performance. Specifically, ResNet-152 applied to ADNI MRI data had extremely strong discriminative power in multi-stage classification (e.g., AD vs. MCI vs. CN) [19], while a DenseNet-201-transfer learning model attained a 98.24% accuracy in five classes of AD with augmented MRI datasets [43].

all three variants of 3D ResNet with attention and probabilistic fusion provided superior performance for binary and multi-class AD classification [44], While very accurate, these models do come with trade-offs: deeper networks are computationally intensive to run, require large amounts of labeled data, and can overfit unless properly regularized. Hybrid and ensemble models, while powerful, introduce complexity that can make clinical deployment challenging [28].

Hybrid techniques combining CNNs with other techniques, such as RNNs or attention, have further enhanced the identification of mild stages of AD, particularly the Very Mild Demented to Mild Demented [22]. Transfer learning from pre-trained models is also used more and more to circumvent small sizes of available datasets to allow models to utilize features learned on large-scale natural image datasets such as ImageNet [23].

Despite the success, CNN-based approaches are faced with issues like high computation demands, need for large labeled datasets, and tendency for overfitting in data-poor scenarios. Careful regularization, architecture fine-tuning, and data augmentation still remain crucial to transform these models into reliable, real-world clinical tools [45].

4. Preprocessing and Data Augmentation in MRI

Preprocessing and data augmentation are two crucial steps in developing robust MRI-based deep learning models for the diagnosis of Alzheimer's disease. Proper preprocessing ensures that the input images are normalized, noise-free, and model-improving generalization [46].

Before feeding MRI images into a convolutional neural network (CNN), several common preprocessing steps are typically applied, the most prominent of which are:

Resizing: Image resizing is an important preprocessing step in medical image analysis, especially in deep learning models that require images of a uniform size . This process ensures consistency of input data, reduces computational costs, and improves model stability. Input images are typically resized to a relatively small spatial resolution (e.g., 224×224 pixels), and the predefined resolution is applied to both training and testing, which contributes to improving the computational efficiency of the model [47].

Grayscale Conversion: In medical image processing, converting MRI, CT, or X-ray images to grayscale compresses the information into one channel, reducing computational overhead without eliminating essential anatomical details [48]. When diagnosing Alzheimer's disease, grayscale MRI slices allow the model to focus on structural patterns rather than redundant color information, improving the efficiency of classification and reducing input dimensionality without compromising diagnostic accuracy [49].

Intensity Normalization: Intensity normalization is among the most critical preprocessing steps in MRI-based Alzheimer's diagnosis [50]. It converts all pixel intensity values in all images to the scanner setting-induced or light-induced variability-free state so that model learning will not be influenced by these variations. Through conducting operations such as Z-score normalization (scaling around a fixed mean and by a fixed standard deviation) or min–max scaling (scaling intensities to a fixed range), the model achieves faster convergence, more stable training, and invariant input distribution—ultimately leading to improved classification accuracy [51].

Noise Reduction: Noise removal in medical image pre-processing removes unwanted pixel intensity variations caused by sources like scanner noise or patient movement [52]. In MRI-based diagnosis of Alzheimer's disease, methods like Gaussian smoothing or median filtering enhance structural definition without removing critical details. This enhances the signal-to-noise ratio to enable CNN models to focus on critical brain features and enhance classification accuracy[49].

4.1 Role of Data Augmentation

Data augmentation is essential for improving the generalization of deep learning models in MRI-based Alzheimer's diagnosis [53]. It expands the training dataset by creating altered versions of existing images—such as through flipping, rotation, and brightness or contrast changes—introducing realistic variability that prevents overfitting. This helps the model focus on important structural patterns rather than memorizing specific details. In addition, augmentation is particularly valuable for balancing datasets by increasing samples for underrepresented classes, leading to more accurate detection across all stages of the disease [54].

5. Interpretability Techniques in Medical AI Models

Interpretability techniques for medical AI models are required to render predictions transparent, understandable, and trustworthy for both clinicians and patients. In medical uses, where decisions can have significant consequences, it is not enough for a model to simply be accurate—it must also provide insight into why it would make a particular decision. They work to bridge the gap between complex deep learning models and the human intuition required in clinical decision-making [55].

One popular method is saliency-based visualization, such as Grad-CAM, that emphasizes the most relevant regions in medical images that drove the model to its decision. This allows doctors to verify whether the AI is attending to clinically relevant features (e.g., brain areas affected by Alzheimer's disease) and not to irrelevant patterns. Similarly, occlusion sensitivity can be used by sequentially occluding parts of the input image to assess how prediction changes to identify the salient areas for classification [56]. Figure 3 illustrates Grad-CAM applied to MRI images of different stages of AD.

A second important class includes feature attribution methods, e.g., Layer-wise Relevance Propagation (LRP) or Integrated Gradients, which trace back the contribution of each input pixel or feature to the output. These methods are especially valuable for medical imaging, where understanding the mapping from input patterns to diagnostic outcomes can facilitate early detection and reduce bias.

Finally, model-agnostic techniques such as Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are applicable to both image and structured medical data. They approximate the behavior of complex models with more interpretable models locally around individual predictions, enabling clinicians to understand local decision-making logic without needing to fully interpret the deep learning architecture [55].

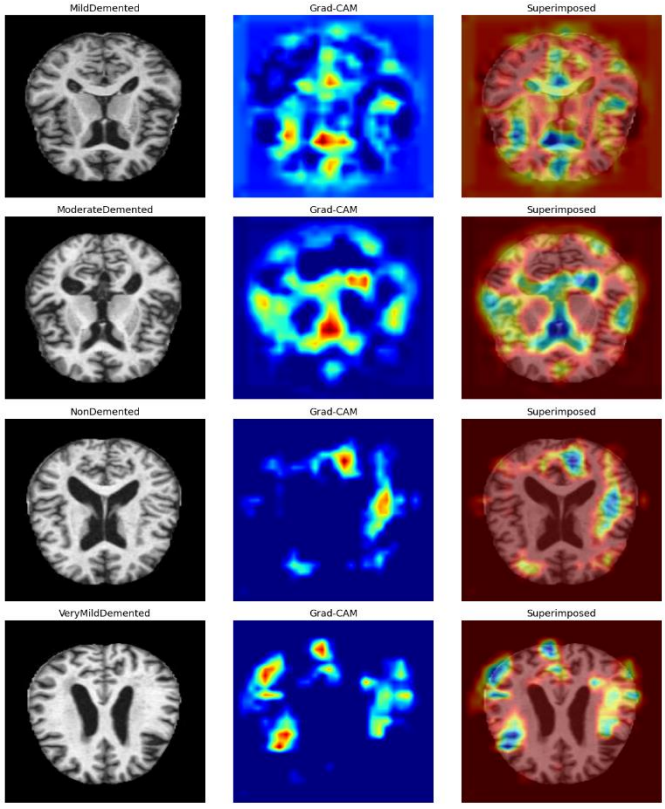

Figure 3 Grad-CAM applied to MRI images

6. Overfitting problem in DL and mitigation methods

Overfitting is a common problem of deep learning (DL) in medical imaging, as the model learns to become overly complex in order to fit the training data and capture noise and irrelevant patterns rather than generalizable features. In the case of Alzheimer's disease (AD) diagnosis from MRI, overfitting is typically induced by small annotated dataset sizes, high-dimensional imaging data, and complex model architectures. This creates poor generalization performance, whereby the model performs well on the training set but significantly less well on novel unseen test data [57].

Some of the mitigation strategies to combat overfitting in DL-based MRI analysis:

Dropout Regularization: Dropout is one of the most effective and widespread regularization techniques in deep learning to combat overfitting. Temporarily disabling a portion of neurons during training forces the model to learn redundant representations, making it more robust and less dependent on specific neurons [58].

Data Augmentation: Applying transformations such as rotation, flip, scale, and brightness modifications increases the diversity of the dataset without capturing new images, thereby improving model generalization [59].

Early Stopping: Tracking the validation loss during training and halting when the performance begins to deteriorate on the validation set prevents overfitting to the training set [60].

Cross-Validation: K-fold cross-validation ensures the model is validated on multiple splits of the data, providing a more accurate estimate of its generalization capability [61].

Batch Normalization: Adding Batch Normalization layers between the hidden layers stabilizes learning and reduces the model's dependence on the specific statistical distribution of training data, thus the better generalization [62].

Reduce the Complexity: Using less complex architectures or fewer layers and neurons can restrict the threat of memorizing noise in the training set. This is particularly handy when the dataset is comparatively small [25].

7. Evaluation metrics used in AD detection

In deep learning detection of Alzheimer's disease (AD), quantifying model performance is of utmost importance to determine reliability, clinical utility, and reproducibility. There should be adequate metrics that provide quantitative evidence of a model's ability to generalize to unknown MRI scans, quantify class-level performance, and address the issue of class imbalance problems common in medical imaging data[63]. The performance metrics are:

7.1 Accuracy

Accuracy is a measure that determines the proportion of instances that were correctly classified out of all the instances. While it is an indicator of overall model performance, it is not representative in imbalanced sets, which are common for AD diagnosis—where one class dominates the distribution, using the following equation 1 [63].

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \qquad (1)$$

7.2 Precision

Precision indicates the proportion of true positive predictions to all positive predictions. High accuracy for AD classification means that the model will rarely misclassify non-AD samples as AD, which is very important in avoiding false diagnoses, using the following equation 2 [64].

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

7.3 Recall (Sensitivity)

Recall quantifies the number of true positive cases that are correctly detected by the model. High levels of recall in the medical setting ensure that most AD patients are detected and few are missed, using the following equation 3 [65, 66].

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

7.4 F1-score

F1-score is the harmonic mean of precision and recall that provides an balanced estimate when false positives and false negatives are both high. It is especially useful in class distribution when it is unbalanced, using the following equation 4 [63].

$$F1_{Score} = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

7.5 MMC

The Matthews Correlation Coefficient (MCC) is a robust evaluation metric that takes into account all elements of the confusion matrix (TP, TN, FP, FN) and is thus particularly suitable for unbalanced classes, using the following equation 5 [67].

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

7.6 Confusion Matrix Analysis

Confusion matrix provides a precise classification of model predictions for each class, offering an effective tool to identify specific misclassification patterns. such as, it highlights the confusion between Very Mild Demented and Mild Demented stages, as illustrated in Figure 4 [68].

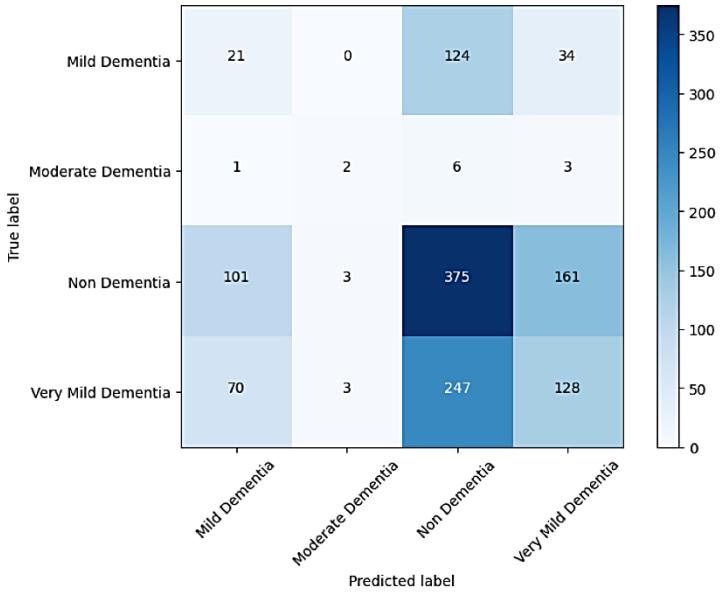

Figure 4 Example of a Confusion Matrix Graph [30].

8. Challenges

While CNN-based models have shown impressive prospects in predicting the stages of Alzheimer's disease from MRI scans, a number of challenges still exist before such models are generally embraced in clinical applications.

Scarcity and heterogeneity of data: Having plenty of and standardized data is rare due to patient privacy restrictions, cost of obtaining them, and heterogeneity in MRI procedures within hospitals. All these factors curtail model robustness and may cause performance decline when applied to new-hospital or new-scan data.

Black-box nature of CNNs: The poor interpretability of CNNs is a challenge to clinical adoption since clinicians require clear and unambiguous reasoning to support the algorithmic outputs. Even with Grad-CAM providing some insight, more advanced tools (e.g., LIME, SHAP) are required in order to achieve increased trust.

Large computational demands: Deep learning model training is computationally demanding, which restricts their application in low-resource clinical environments.

Class imbalance: Minority disease phases (e.g., very mild or moderate dementia) remain underrepresented in data sets, with consequent poor classification for these classes [69].

9. Future Directions

There are several directions for research that can help overcome such challenges and make AI-based diagnosis of Alzheimer's more clinically useful:

Multi-modal integration: Combining MRI with PET scans, genetic information, and clinical tests to provide more accurate diagnoses.

3D volumetric MRI analysis: Moving from 2D slices to complete 3D reconstructions to provide better spatial and structural descriptions.

Lightweight optimized models: Developing efficient architectures to enable real-time deployment on low-power or mobile hardware for low-cost screening.

Federated learning: Applying privacy-preserving collaborative training methods inside institutions without sharing sensitive patient information.

Greater interpretability: Bounding-leading explainable AI methods to enable clinician trust, regulatory sign-off, and safe adoption in healthcare pipelines.

10. Conclusion

This review has examined recent advances, methodologies, and challenges in using deep learning—particularly Convolutional Neural Networks (CNNs)—to diagnose and stage Alzheimer's disease from MRI data. The reviewed literature states the key promise of CNN-based models for high diagnostic performance, early diagnosis, and giving automated, scalable decision support within the clinic. Preprocessing techniques such as grayscale conversion, normalization, removal of noise, and image resizing, as well as data augmentation techniques, have been shown to

enhance model performance and generalizability. Besides, interpretability techniques such as Grad-CAM and other explainable AI methods are crucial in bridging the gap between AI models and clinical uptake through visual and numerical explanations of model decision-making.

In spite of remarkable progress, some challenges remain, including dataset scarcity, imaging protocol heterogeneity, model interpretability limitations, and deployment in low-resource settings. Future research needs to focus on multi-modal data fusion, 3D MRI analysis, real-time deployable light-weight models, and collaborative learning frameworks with privacy preservation. Overcoming these limitations and leveraging novel AI interpretability tools, deep learning models can become more reliable, trustworthy, and widely applicable diagnostic tools against Alzheimer's disease.

References

[1] J. D. Steinmetz *et al.*, "Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021," *The Lancet Neurology*, vol. 23, no. 4, pp. 344-381, 2024, doi: 10.1016/S1474-4422(24)00038-3.

[2] P. Sy, "A Patient's Guide to Brain Disease," *U.S. News & World Report*, 2023 2023. [Online]. Available: https://health.usnews.com/conditions/brain-disease.

[3] A. s. Association, "2019 Alzheimer's disease facts and figures," *Alzheimer's & dementia*, vol. 15, no. 3, pp. 321-387, 2019.

[4] D. J. Selkoe and J. Hardy, "The amyloid hypothesis of Alzheimer's disease at 25 years," *EMBO Molecular Medicine*, vol. 8, no. 6, pp. 595-608, 2016, doi: https://doi.org/10.15252/emmm.201606210.

[5] M. Sudharsan and G. Thailambal, "Alzheimer's disease prediction using machine learning techniques and principal component analysis (PCA)," *Materials Today: Proceedings*, vol. 81, pp. 182-190, 2023.

[6] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri, "The global prevalence of dementia: a systematic review and metaanalysis," *Alzheimer's & dementia*, vol. 9, no. 1, pp. 63-75. e2, 2013.

[7] I. Alzheimer's Disease, "World Alzheimer Report 2015: The Global Impact of Dementia," Alzheimer's Disease International (ADI), London, 2015. [Online]. Available: https://www.alzint.org/u/WorldAlzheimerReport2015.pdf

[8] G. Folego, M. Weiler, R. F. Casseb, R. Pires, and A. Rocha, "Alzheimer's disease detection through whole-brain 3D-CNN MRI," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 534592, 2020.

[9] M. S. Albert *et al.*, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's & dementia*, vol. 7, no. 3, pp. 270-279, 2011.

[10] R. Setyawati, A. Astuti, T. P. Utami, S. Adiwijaya, and D. M. Hasyim, "The importance of early detection in disease management," *Journal of World Future Medicine, Health and Nursing*, vol. 2, no. 1, pp. 51-63, 2024.

[11] G. Livingston *et al.*, "Dementia prevention, intervention, and care," *The Lancet*, vol. 390, no. 10113, pp. 2673-2734, 2017, doi: 10.1016/S0140-6736(17)31363-6.

[12] G. Mirzaei and H. Adeli, "Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia," *Biomedical Signal Processing and Control*, vol. 72, p. 103293, 2022.

[13] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, "Brain imaging in Alzheimer disease," *Cold Spring Harbor perspectives in medicine*, vol. 2, no. 4, p. a006213, 2012.

[14] C. R. Jack Jr *et al.*, "Longitudinal MRI findings from the vitamin E and donepezil treatment study for MCI," *Neurobiology of aging*, vol. 29, no. 9, pp. 1285-1295, 2008.

[15] R. A. a. M. M. A. Ahmed, " MRI Limitations: The Main Aspects and Resolving Techniques," 2020, doi: 10.36106/ijar.

[16] P. Denver and P. L. McClean, "Distinguishing normal brain aging from the development of Alzheimer's disease: inflammation, insulin signaling and cognition," (in eng), *Neural Regen Res*, vol. 13, no. 10, pp. 1719-1730, Oct 2018, doi: 10.4103/1673-5374.238608.

[17] L. Pinto-Coelho, "How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications," (in eng), *Bioengineering (Basel)*, vol. 10, no. 12, Dec 18 2023, doi: 10.3390/bioengineering10121435.

[18] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," (in eng), *Front Public Health*, vol. 11, p. 1273253, 2023, doi: 10.3389/fpubh.2023.1273253.

[19] D. AlSaeed and S. F. Omar, "Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning," *Sensors*, vol. 22, no. 8, p. 2911, 2022.

[20] Z. Pei, Z. Wan, Y. Zhang, M. Wang, C. Leng, and Y.-H. Yang, "Multi-scale attention-based pseudo-3D convolution neural network for Alzheimer's disease diagnosis using structural MRI," *Pattern Recognition*, vol. 131, p. 108825, November 01, 2022 2022, doi: 10.1016/j.patcog.2022.108825.

[21] S. Mohsen, "Alzheimer's disease detection using deep learning and machine learning: a review," *Artificial Intelligence Review*, vol. 58, no. 9, p. 262, 2025/06/04 2025, doi: 10.1007/s10462-025-11258-y.

[22] C. Kavitha, V. Mani, S. Srividhya, O. I. Khalaf, and C. A. Tavera Romero, "Early-stage Alzheimer's disease prediction using machine learning models," *Frontiers in public health*, vol. 10, p. 853294, 2022.

[23] H. Alshamlan, A. Alwassel, A. Banafa, and L. Alsaleem, "Improving Alzheimer's Disease Prediction with Different Machine Learning Approaches and Feature Selection Techniques," *Diagnostics*, vol. 14, no. 19, p. 2237, 2024.

[24] F. U. R. Faisal and G.-R. Kwon, "Automated detection of Alzheimer's disease and mild cognitive impairment using whole brain MRI," *IEEE Access*, vol. 10, pp. 65055-65066, 2022.

[25] A. A. A. El-Latif, S. A. Chelloug, M. Alabdulhafith, and M. Hammad, "Accurate detection of Alzheimer's disease using lightweight deep learning model on MRI data," *Diagnostics*, vol. 13, no. 7, p. 1216, 2023.

[26] K. De Silva and H. Kunz, "Prediction of Alzheimer's disease from magnetic resonance imaging using a convolutional neural network," *Intelligence-Based Medicine*, vol. 7, p. 100091, 2023.

[27] S. Esam and A. Mohammed, "Alzheimer's disease classification for MRI images using Convolutional Neural Networks," in *2024 6th International Conference on Computing and Informatics (ICCI)*, 6-7 March 2024 2024, pp. 01-05, doi: 10.1109/ICCI61671.2024.10485049.

[28] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cognitive Systems Research*, vol. 57, pp. 147-159, 2019.

[29] V. S. Diogo, H. A. Ferreira, D. Prata, and A. s. D. N. Initiative, "Early diagnosis of Alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach," *Alzheimer's Research & Therapy*, vol. 14, no. 1, p. 107, 2022.

[30] T. Mahmud, K. Barua, S. U. Habiba, N. Sharmen, M. S. Hossain, and K. Andersson, "An explainable ai paradigm for alzheimer's diagnosis using deep transfer learning," *Diagnostics*, vol. 14, no. 3, p. 345, 2024.

[31] A. D. Arya *et al.*, "A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease," (in eng), *Brain Inform*, vol. 10, no. 1, p. 17, Jul 14 2023, doi: 10.1186/s40708-023-00195-7.

[32] S. Gao and D. Lima, "A review of the application of deep learning in the detection of Alzheimer's disease," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 1-8, 2022/06/01/ 2022, doi: https://doi.org/10.1016/j.ijcce.2021.12.002.

[33] L. Huang, Y. Jin, Y. Gao, K. H. Thung, and D. Shen, "Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest," (in eng), *Neurobiol Aging*, vol. 46, pp. 180-91, Oct 2016, doi: 10.1016/j.neurobiolaging.2016.07.005.

[34] L. Chouliaras and J. T. O'Brien, "The use of neuroimaging techniques in the early and differential diagnosis of dementia," *Molecular Psychiatry*, vol. 28, no. 10, pp. 4084-4097, 2023/10/01 2023, doi: 10.1038/s41380-023-02215-8.

[35] M. Boccardi *et al.*, "Do beliefs about the pathogenetic role of amyloid affect the interpretation of amyloid PET in the clinic?," *Neurodegenerative Diseases*, vol. 16, no. 1-2, pp. 111-117, 2016.

[36] S. Aramadaka *et al.*, "Neuroimaging in Alzheimer's Disease for Early Diagnosis: A Comprehensive Review," (in eng), *Cureus*, vol. 15, no. 5, p. e38544, May 2023, doi: 10.7759/cureus.38544.

[37] A. Hafkemeijer, J. van der Grond, and S. A. R. B. Rombouts, "Imaging the default mode network in aging and dementia," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, vol. 1822, no. 3, pp. 431-441, 2012/03/01/ 2012, doi: https://doi.org/10.1016/j.bbadis.2011.07.008.

[38] R. S. Desikan *et al.*, "Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease," (in eng), *Brain*, vol. 132, no. Pt 8, pp. 2048-57, Aug 2009, doi: 10.1093/brain/awp123.

[39] S. Liu *et al.*, "Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs," *Scientific Reports*, vol. 12, no. 1, p. 17106, 2022/10/17 2022, doi: 10.1038/s41598-022-20674-x.

[40] F. Sultana, A. Sufian, and P. Dutta, "A review of object detection models based on convolutional neural network," *Intelligent computing: image processing based applications*, pp. 1-16, 2020.

[41] F. Ramzan *et al.*, "A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks," *Journal of Medical Systems*, vol. 44, no. 2, p. 37, 2020, doi: 10.1007/s10916-020-01583-2.

[42] P. Kiran *et al.*, "A New Deep Learning Model based on Neuroimaging for Predicting Alzheimer's Disease," *Neurocomputing*, 2022, doi: 10.1016/j.neucom.2022.08.001.

[43] R. Zia Ur *et al.*, "Classification of Alzheimer disease using DenseNet-201 based on deep transfer learning technique," (in eng), *PLoS One*, vol. 19, no. 9, p. e0304995, 2024, doi: 10.1371/journal.pone.0304995.

[44] X. Yang *et al.*, "An ensemble-based 3D residual network for the classification of Alzheimer's disease," *PLOS ONE*, 2025/06/11 2025, doi: 10.1371/journal.pone.0324520.

[45] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift fuer medizinische Physik*, vol. 29, no. 2, pp. 102-127, 2019.

[46] P. Kragsterman. "The Ultimate Guide to Preprocessing Medical Images: Techniques, Tools, and Best Practices for Enhanced Diagnosis." Collective Minds Radiology. https://about.cmrad.com/articles/the-ultimate-guide-to-preprocessing-medical-images-techniques-tools-and-best-practices-for-enhanced-diagnosis (accessed.

[47] S. Saponara and A. Elhanashi, "Impact of Image Resizing on Deep Learning Detectors for Training Time and Model Performance," in *Applications in Electronics Pervading Industry, Environment and Society*, Cham, S. Saponara and A. De Gloria, Eds., 2022// 2022: Springer International Publishing, pp. 10-17.

[48] Y. Xie and D. Richmond, "Pre-training on grayscale imagenet improves medical image classification," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0-0.

[49] S. Sarraf, D. D. DeSouza, J. Anderson, G. Tofighi, and f. t. A. s. D. N. Initiativ, "DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI," *bioRxiv*, p. 070441, 2017, doi: 10.1101/070441.

[50] A. Carré *et al.*, "Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics," *Scientific Reports*, vol. 10, no. 1, p. 12340, 2020/07/23 2020, doi: 10.1038/s41598-020-69298-z.

[51] S. Schmid, "Image Normalization in Medical Imaging," *Medium*, 2020/05/12 2020. [Online]. Available: https://medium.com/@susanne.schmid/image-normalization-in-medical-imaging-f586c8526bd1.

[52] Y. Li, Z. Zhang, C. Dai, Q. Dong, and S. Badrigilan, "Accuracy of deep learning for automated detection of pneumonia using chest X-Ray images: A systematic review and meta-analysis," *Computers in Biology and Medicine*, vol. 123, p. 103898, 2020/08/01/ 2020, doi: https://doi.org/10.1016/j.compbiomed.2020.103898.

[53] P. Hallaj. "Data Augmentation: Benefits and Disadvantages." https://medium.com/@pouyahallaj/data-augmentation-benefits-and-disadvantages-38d8201aead (accessed.

[54] T. Kumar, R. Brennan, A. Mileo, and M. Bendechache, "Image data augmentation approaches: A comprehensive survey and future directions," *IEEE Access*, 2024.

[55] H. Ying and et al., "Advancing ethical AI in healthcare through interpretability," *Patterns*, vol. 6, no. 6, p. 101290, 2025/06/13 2025, doi: 10.1016/j.patter.2025.101290.

[56] 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," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618-626.

[57] A. T. Tran, T. Zeevi, and S. Payabvash, "Strategies to Improve the Robustness and Generalizability of Deep Learning Segmentation and Classification in Neuroimaging," *BioMedInformatics*, vol. 5, no. 2, p. 20, 2025. [Online]. Available: https://www.mdpi.com/2673-7426/5/2/20.

[58] GeeksforGeeks, "Dropout Regularization in Deep Learning: preventing overfitting by randomly dropping neurons," 2025/06/11 2025. [Online]. Available: https://www.geeksforgeeks.org/dropout-regularization-in-deep-learning/.

[59] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019/07/06 2019, doi: 10.1186/s40537-019-0197-0.

[60] A. Schneppat, "Early Stopping: monitoring validation performance to prevent overfitting," 2025/08/15 2025. [Online]. Available: https://schneppat.com/early-stopping.html?utm_source=chatgpt.com.

[61] GeeksforGeeks, "Cross-Validation in Machine Learning: evaluating model performance through repeated resampling," 2025/08/04 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/cross-validation-machine-learning/.

[62] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, 2015: pmlr, pp. 448-456.

[63] A. Tharwat, "Classification assessment methods," *Applied computing and informatics*, vol. 17, no. 1, pp. 168-192, 2021.

[64] Ž. Vujović, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599-606, 2021.

[65] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Scientific Reports*, vol. 12, no. 1, p. 5979, 2022/04/08 2022, doi: 10.1038/s41598-022-09954-8.

[66] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, no. 1, p. 210, 2022/06/20 2022, doi: 10.1186/s13104-022-06096-y.

[67] H. H. Rashidi, S. Albahra, S. Robertson, N. K. Tran, and B. Hu, "Common statistical concepts in the supervised Machine Learning arena," (in eng), *Front Oncol*, vol. 13, p. 1130229, 2023, doi: 10.3389/fonc.2023.1130229.

[68] GeeksforGeeks, "Understanding the Confusion Matrix in Machine Learning," *GeeksforGeeks*, 2021/05/10 2021. [Online]. Available: https://www.geeksforgeeks.org/confusion-matrix-machine-learning/?utm_source=chatgpt.com.

[69] M. Kale *et al.*, "AI-driven innovations in Alzheimer's disease: Integrating early diagnosis, personalized treatment, and prognostic modelling," (in eng), *Ageing Res Rev*, vol. 101, p. 102497, Nov 2024, doi: 10.1016/j.arr.2024.102497.