Alzheimer's Disease Prediction and Diagnosis Using Machine Learning Techniques

Zainab Abdalhussain Kareem and Ahmad Shaker Abdalrada
College of Computer Science and Information Technology, Wasit University, 52001 Al-Kut, Wasit, Iraq
zaab205@uowasit.edu.iq; aabdalra@uowasit.edu.iq

Keywords: Alzheimer's Disease Prediction, CNN, Deep Learning, MRI.

Abstract: Alzheimer's Disease (AD) is a medical condition that affects the human brain. It is a progressive neurodegenerative disorder that significantly impairs memory, cognitive, and executive functions, particularly among the elderly. Early and accurate diagnosis is essential for effective disease management and slowing its progression. Despite advances in predicting this disease, there is still a need to improve prediction. In this study, we propose a lightweight, custom convolutional neural network (CNN) model for multi-class classification of Alzheimer's disease stages using magnetic resonance images. The model achieved a test accuracy of 98.28%. The F1 scores for the four classes ranged from 0.96 to 0.99, while the average precision and recall were 0.98, demonstrating robust and balanced performance across all diagnostic categories. These results outperform many existing methods while maintaining computational efficiency and complexity-free performance. The proposed model provides an effective and scalable solution for early-stage Alzheimer's disease diagnosis from MRI scans, with potential integration into medical systems that support real-world clinical decisions.

1 INTRODUCTION

Alzheimer's disease (AD) is the most prevalent neurodegenerative disorder, primarily affecting elderly individuals and characterized by the gradual loss of memory, reasoning, and cognitive function [1]. The pathological hallmarks of AD encompass the aggregation of extracellular beta-amyloid (Aβ) plaques and intracellular neurofibrillary tangles composed of hyperphosphorylated tau protein, which collectively disrupt neural communication and result in the cell brain death [2]. Alzheimer's disease (AD) represents a major global public health concern, with its prevalence is expected to double by 2050, approximately 6.9 million people aged 65 and older in the United States have (AD) [3]. And more than 55 million people aged 75 and older with dementia worldwide, an estimated 60% to 70% have AD [4]. Studies indicate that one in two people over the age of 85 will develop the disease [5]. Placing additional burden on medical care systems, especialy in low- and middle-income countries [6]. Therefore, early and accurate diagnosis is critical to limiting disease progression and enhancing patient quality of life [7].

Neuroimaging, particularly magnetic resonance imaging (MRI), has emerged as a powerful tool for early detection of AD [8]. Structural MRI can reveal atrophy in specific brain regions, such as the hippocampus and cortex, even at an early stage of the disease [9]. However, traditional MRI analysis requires expert interpretation and is often limited by high costs and diagnostic variability [10]. Utilizing such a tool for disease detection is considered valuable in the field of Alzheimer's disease diagnosis. Despite advancements in existing studies [11]-[13], there remains a need to further improve the prediction accuracy of Alzheimer's disease [14].

Latest advancements in artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), offer transformative solutions for automated, efficient, and scalable diagnosis [15]-[18]. Deep learning models, like convolutional neural networks (CNNs), have displayed remarkable success in medical imaging tasks by learning complex hierarchical features from unprocessed images, decreasing reliance on manual feature engineering [19], [20]. Also, when diagnosing Alzheimer's disease, CNNs can automatically classify MRI brain scans into various stages of disease severity, including normal, very mild, mild, and moderate dementia with high diagnostic

accuracy [21]. These models not only accelerate diagnosis but also contribute to standardized and reproducible decision-making in clinical practice [22]. This research therefore proposes a customized CNN-based model for early detection and multi-class classification of AD using an augmented MRI dataset.

The research paper design includes the following: Section 2: Review of previous work. Section 3: Presentation of the proposed model. Section 4: Explanation of the materials and methods used in this study. Results are presented in Section 5. Lastly, conclusion is drawn in Section 6.

2 RELATED WORK

Numerous studies have investigated the use of machine learning (ML) and deep learning (DL) techniques for the prediction and classification of Alzheimer's disease (AD). These studies differ in methodology, feature selection strategies, model architectures, and dataset characteristics.

For example, Sudharsan and Thailambal [3] examined the use of principal component analysis (PCA) for feature selection in conjunction with classifiers such as Support Vector Machine (SVM), Regularized Extreme Learning Machine (RELM), and Import Vector Machine (IVM), using sMRI data from the ADNI dataset. The RELM model yielded the highest performance, achieving 81.03% accuracy in binary classification and 62.76% in multi-class classification. However, the relatively low performance in the multi-class setting highlights the difficulty of distinguishing between different disease stages.

Moreover, Jain et al. [13] proposed a classification method based on transfer learning using VGG-16 pretrained on ImageNet. Their approach involved converting 3D MRI scans into 2D slices and applying entropy-based selection to identify the most informative slices. The model achieved 95.73% accuracy on a tri-class classification task (AD, MCI, CN). However, the limited sample size (150 MRIs) and the loss of 3D spatial context due to slice-based processing may have affected the completeness of diagnostic information.

Furthermore, AlSaeed and Omar [23] used ResNet-50 as a feature extractor and evaluated three classifiers: Softmax, SVM, and Random Forest. Their system, applied to datasets such as ADNI and MIRIAD, achieved accuracies ranging from 85.7% to 99%, with the ResNet-50 + Softmax combination performing best (99%) for binary AD vs. NC classification. A key limitation of this study is that only two classes were considered; MCI was excluded, despite its importance for early intervention.

Mahmud et al. [24] introduced an explainable AI framework that combined deep transfer learning with ensemble models. They explored two ensembles – VGG16 + VGG19 and DenseNet169 + DenseNet201 – and demonstrated that an EfficientNet + CNN model with Grad-CAM achieved 96% accuracy. However, this performance was still insufficient for reliable clinical application.

Diogo et al. [25] proposed a generalizable machine learning approach for AD and MCI classification using sMRI data from the ADNI and OASIS datasets. They compared several classifiers and found that a CNN + Random Forest model achieved 90.6% balanced accuracy (BAC) for binary classification (HC vs. AD) and 62.1% BAC for multi-class classification (HC vs. MCI vs. AD), indicating the continued difficulty of accurately distinguishing between AD and MCI.

De Silva and Kunz [26] proposed a convolutional neural network (CNN) model for diagnosing AD using the MIRIAD dataset. Their model achieved 89% accuracy for binary AD vs. HC classification. While effective for initial AD detection, this approach does not address the more clinically relevant and challenging task of multi-class classification, which involves differentiating between intermediate and early stages of dementia.

El-Latif et al. [11] proposed a lightweight CNN model trained on a Kaggle Alzheimer's dataset with four classes: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. The model achieved 99.22% accuracy in binary classification and 95.93% in multi-class classification, outperforming deeper models such as ResNet50 and DenseNet201. Despite the impressive results, the achieved accuracy may still be insufficient for reliable clinical deployment.

Kavitha et al. [27] developed a machine learning framework using classifiers such as Decision Trees, Random Forest, SVM, and XGBoost to differentiate between demented and non-demented individuals. Using data from the OASIS and Kaggle datasets, the best performance was obtained with the Random Forest model, which achieved 86.92% accuracy. Nevertheless, the binary nature of the classification and the limited number of images represent key limitations.

Alshamlan et al. [28] also examined several machine learning algorithms – SVM, Random Forest, and Logistic Regression – combined with feature selection techniques such as minimum Redundancy

Maximum Relevance (mRMR) and Mutual Information (MI). Using the OASIS-2 dataset (373 MRIs), their best-performing configuration (Logistic Regression + mRMR) achieved 99.08% accuracy in binary classification. However, the study did not consider intermediate stages such as MCI.

Despite the promising results reported in the above-mentioned studies [3], [11]–[13], [23]–[28], there remains a critical need to further improve model performance and generalizability [29]. Many existing approaches suffer from limitations such as reduced accuracy in multi-class scenarios and the lack of robust preprocessing strategies. As early and accurate diagnosis of Alzheimer's disease is essential for timely intervention, there is a growing demand for models that can deliver consistently high accuracy across all diagnostic categories. Addressing these challenges requires the development of optimized, well-regularized, and interpretable deep learning frameworks capable of capturing subtle variations in brain structures.

3 THE PROPOSED MODEL

This section describes the proposed model for predicting and classifying Alzheimer's disease. The model was built from scratch, rather than relying on pre-trained models, using a set of effective techniques. Figure 1 shows the steps in building the model proposed in this study.

This model is a convolutional neural network (CNN)-based framework designed to predict Alzheimer's disease (AD) using only image data. This model leverages the capabilities of convolutional neural networks (CNNs) to extract features and analyze MRI data, helping to accurately identify patterns associated with AD. And below explains the steps involved in building the model.

3.1 Data Splitting and Augmentation

In this research, the Alzheimer's MRI image dataset was split into three subsets: training (70%), validation (10%), and testing (20%). Data augmentation was then applied to the training set only, with the number of images in all four categories equal to 5,000 images per category. This offline augmentation included horizontal flipping, slight rotation, brightness enhancement, and contrast adjustment. The data was first segmented and then augmented to avoid data leakage to the testing data.

3.2 Pre-Processing

To ensure that the MRI data utilized for Alzheimer's disease classification was consistent, standardized, and fit for training the convolutional neural network (CNN), a robust preprocessing pipeline was employed after data segmentation. This pipeline was created to address variations in image dimensions, pixel intensity distributions, and spatial orientations, while also introducing controlled transformations that enhance data diversity and decrease the risk of overfitting.

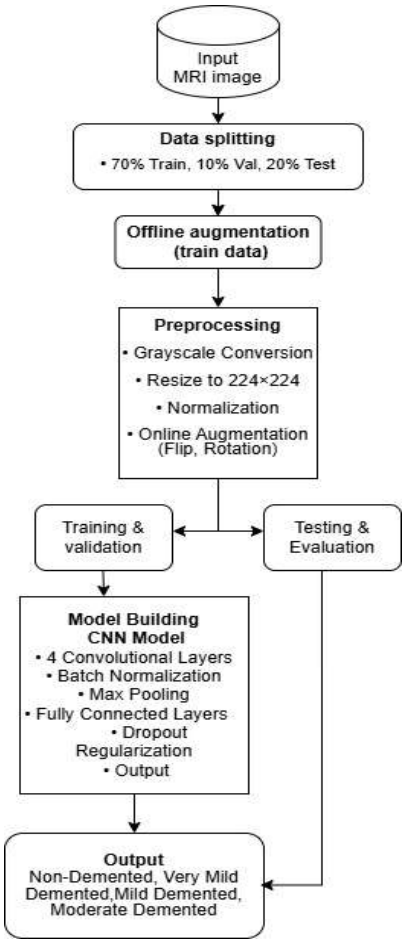

Figure 1: Steps of the proposed model.

The preprocessing first start with resizing all MRI images to 224 × 224 pixels, to make them meet the input requirements of the CNN architecture. This standardization enable the model to process all the images uniformly [30]. Then, all images were converted to grayscale, reducing computational complexity and focusing on brain structure intensity variations that are most pertinent for distinguishing between different stages of Alzheimer's disease [31].

After converting to grayscale , normalization was applied using formula normalization to center the pixel intensity values and scale them inside the range [-1, 1]. This process assists stabilize model training by ensuring that all input data shares consistent statistical properties [32]. In order to enhance the generalization capability of the model, a set of Online augmentation techniques was applied during training Random rotations were presented to simulate variation in head positioning during MRI scan. These subtle angular changes allowed the model to become orientationally stable. In addition, random horizontal flipping was added with 50% probability to introduce anatomical symmetry and further increase the diversity of training input samples. This approach is especially effective in brain imaging analysis, where symmetrical patterns are common [33].

3.3 Components of the Model

This deep learning model for Alzheimer's disease classification it consists of six consecutive layers, combining convolutional operations with regularization and classification techniques. The architecture is carefully designed to extract relevant structural features from brain MRI images and to classify the stage of the disease according to four diagnostic stages, as shown in Figure 2.

The model starts with four convolutional layers which are responsible for extracting spatial and textural features from input MRI images. These layers utilize learnable filters to identify underlying structural patterns in the brain, such as tissue density variations and cortical folds associated with different stages of AD. As the network depth increases, the number of filters also icreases, allowing for the extraction of more abstract and complex features, which is very important for accurately distinguishing between closely related classes .After each conv operation, a Rectified Linear Unit (ReLU) activation function is used. ReLU provides the necessary non-linearity that enables the model to learn complex patterns within the data. ReLU is preferred due to its computational efficiency and ability to mitigate the vanishing gradient problem that commonly occurs in deep networks .Then to enhance training stability and learning efficiency, Batch normalization is applied after each conv layer. This technique normalizes the input distributions within each mini-batch, which helps speed up convergence and allowing for higher learning rates without causing instability in the network .After the feature extraction process is compleate, Max pooling layers are added to reduce the spatial dimensions of the feature maps produced by the convolutional layers. This pooling process focuses on extracting the most important features, while reducing computational complexity and minimizing the risk of overfitting by ignoring less important differences. Then the extracted feature maps are flattened and passed through a fully connected hidden layer. Before this, dropout regularization is employed to randomly deactivate a portion of neurons during training, promoting generalization and preventing the model from becoming overly dependent on specific pathways. The hidden layer integrates spatial features and supports high-level reasoning necessary for final classification. The final component of the model is a fully connected output layer that produces four scalar values corresponding to the diagnostic categories: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. These outputs represent the raw confidence scores (logits) for each class. The model omits an explicit softmax activation in the final layer, as the CrossEntropyLoss function applied during training inherently incorporates the softmax operation internally to compute class probabilities.

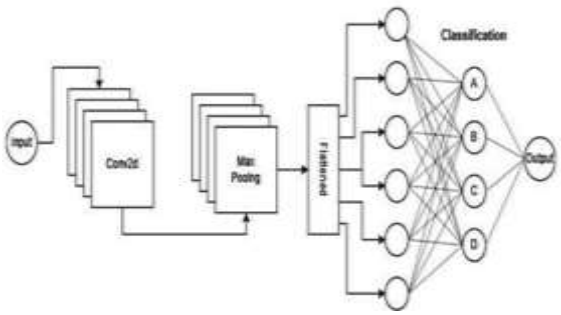

Figure 2: Components of the Model architecture.

3.4 Training the Model with Optimizer (ADAM)

The proposed model is trained using the Adam optimizer with a learning rate of 0.0001. Adam's adaptive learning mechanism adjusts each parameter's learning rate based on gradient moments, ensuring balanced updates across sparse and dense gradients. Additionally, ReduceLROnPlateau was applied to lower the learning rate upon validation loss stagnation. And a 8-epoch early stopping was also implemented to prevent overfitting and select the best performing model. This setup enabled efficient training and superior classification performance across Alzheimer's diagnostic categories.

4 MATERIALS AND METHODS

The assessment and interpretation of the provided predictive model was done. In addition, this part describes the dataset, experiment parameters and the metrics used.

4.1 Dataset

The dataset used in this study was obtained from the publicly available Alzheimer MRI 4 Classes Dataset on Kaggle [34]. This data was originally collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) – a widely recognized and validated source of clinical data. And it consists of approximately 6,400 original 2D brain MRI images, classified into four diagnostic categories: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. This dataset is inherently unbalanced, with a significantly higher number of dementia cases compared to non-dementia dementia and very mild dementia. After splitting the data into training, validation, and test sets, the training set was balanced to contain 5,000 images for each class While the validation and test sets remained in their original unbalanced form to avoid leakage and to simulate real-world evaluation conditions.

4.2 Performance Metrics

In this study, the performance of the proposed CNN model was evaluated using four widely adopted classification metrics: accuracy, precision, recall, and F1-score. Together, these measures provide a comprehensive evaluation of the model's effectiveness in classifying different stages of Alzheimer's disease within a multi-class classification framework.

Accuracy reflects the overall proportion of correctly classified samples across all categories, including both positive and negative predictions. It provides a general indication of the model's ability to make correct decisions regardless of class distribution, with higher values indicating stronger overall classification performance [35].

Precision measures the model's capability to correctly identify positive cases among all cases predicted as positive. This metric is particularly important in medical applications, where false positive predictions may lead to unnecessary anxiety or further clinical procedures. Higher precision values indicate a lower rate of incorrect positive predictions and greater reliability of positive classifications [36].

Recall, also known as sensitivity, evaluates the model's effectiveness in identifying all actual positive cases. It represents the proportion of true positive samples that are correctly detected by the model. High recall is especially critical in medical diagnosis, as failing to identify a true positive case may result in delayed treatment or misdiagnosis [37].

The F1-score provides a balanced performance measure by jointly considering both precision and recall. It is particularly useful in situations involving class imbalance or when it is important to equally account for false positives and false negatives. A high F1-score indicates that the model maintains an effective trade-off between precision and recall, ensuring reliable and consistent classification performance [38].

Collectively, these evaluation metrics offer a multi-dimensional perspective on model performance, enabling a robust assessment that accounts not only for prediction accuracy but also for the balance between different types of classification errors.

4.3 Experiment

These experiments were conducted using Python on a personal laptop. The dataset was pre-organized into four class-specific folders, and each one corresponding to the diagnostic categories: non-dementia, very mild dementia, mild dementia, and moderate dementia. The data was manually split into subsets for training (70%), validation (10%), and testing (20%), utilizing a stratified approach to ensure a balanced distribution of classes across all subsets. This stratification preserved the relative proportion of each class, enabling a fair and consistent evaluation of performance during model training and testing.

5 RESULTS AND DISCUSSIONS

This part presents and review the results gained from the proposed CNN model utilized on the Alzheimer MRI 4 Classes dataset. The model show powerful classification ability in distinguishing between the four stages of Alzheimer's disease: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. The confusion matrix (Fig. 3) indicates that the most of the test samples were correctly categorized across all classes. High classification accuracy was achieved, particularly in the (Moderate Demented) and (Mild Demented) groups, although some confusion happened between (Very Mild

Demented) and (Non-Demented), which can be credited to the overlapping characteristics of early Alzheimer's symptoms. The overall performance reached 98.28% accuracy, and F1 scores for each class ranged from 0.96 to 0.99, demonstrating consistent precision and recall. And the macro-averaged F1-score was 0.98, indicating balanced performance across all classes despite the imbalanced distribution. The weighted average metrics also confirmed the model's effectiveness on the imbalanced test set.

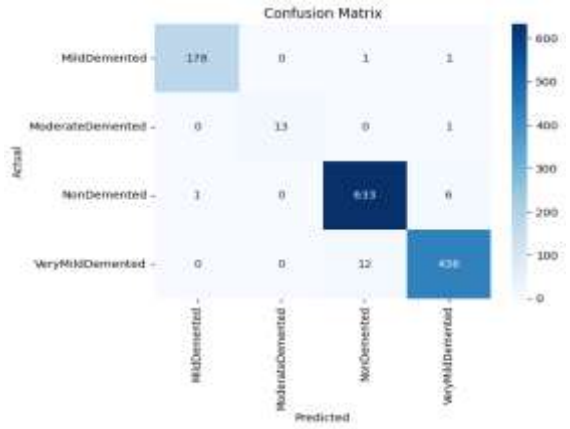

Figure 3: The confusion matrix.

The accuracy curve across training and validation sets Figure 4 illustrates a consistent increase in performance during the training process, with the validation accuracy stabilized around 97.79%, with early stopping triggered after approximately 55 epochs, indicating that the model had converged effectively. Additionally, the loss curves Figure 5 indicate a steady decline, with the training and validation losses remaining closely aligned throughout the process. This balance between the two curves demonstrates the model's ability to generalize well without significantly overfitting. This stability is due to the integration of dropout regulation, batch normalization, and early stopping, all of which contributed to improved training efficiency.

These results demonstrate the effectiveness of the general framework used in this study, including the preprocessing steps, lightweight CNN model architecture, and training strategy. Several factors significantly increased the model's generalization ability and enhanced its accuracy, such as the use of a balanced training set with offline data augmentation, as well as online data augmentation during training. The consistent performance across all classes reflects strong results, even on the imbalanced test set, demonstrating its potential for integration into clinically applicable computer-based diagnostic systems. However, occasional misclassification between similar classes is evident, suggesting opportunities for future improvements, such as incorporating temporal MRI data or clinical metadata, to improve the model's sensitivity in early diagnosis.

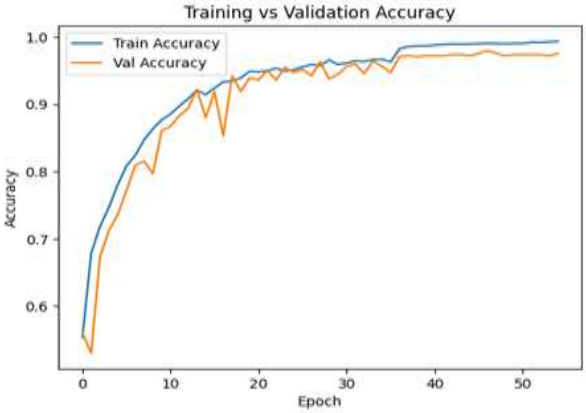

Figure 4: Training and validation accuracy.

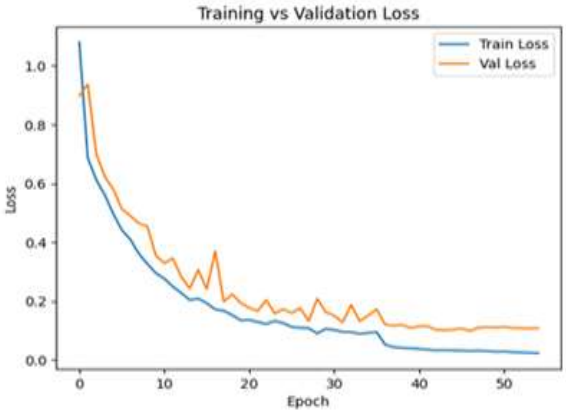

Figure. 5: Training and validation loss.

6 CONCLUSIONS

With the increasing number of patients and the complexity of (AD) diagnosis, manual assessment by physicians has become impractical and error-prone.In this research paper, a customized (CNN) lightweight was developed to classify multiple stages of AD using brain MRI images. The model was trained on a dataset that initially contained 6,400 original and imbalanced images, and this data was divided into four diagnostic categories (Non-Demented, Very Mild Demented, Mild Demented, and Moderate

Demented). After splitting of the data only the training set was balanced to contain 5,000 images per class using offline augmentation techniques. Due to the efficient and accurate training methodology, the model was able to achieve a test accuracy of 98.28%, With good and high results for the rest of the metrics. This strong performance is attributed to the effectiveness of the model architecture, which included batch normalization, dropout, and powerful preprocessing, including converting to grayscale, resizing, normalization, and online data augmentation techniques. And cause of using these components, enables the model to be generalizable, effectively detecting subtle structural variations in the brain.

Despite these good results, the research has some drawbacks. Generalizability may be impacted by the dataset's incomplete representation of actual clinical situations. To assess the model's applicability, further research should validate it on longer-term, multi-center datasets. Additionally, using 3D MRI data could increase classification accuracy and better capture spatial features. Clinical trust may be improved and model judgments clarified with the use of explainable AI (XAI) tools like Grad-CAM. Diagnostic accuracy may be further increased by combining MRI data with information from other modalities, such as genetic or cognitive data. Lastly, the model's use in clinical settings would be facilitated by real-time optimization.

REFERENCES

[1] D. Selkoe, E. Mandelkow, and D. Holtzman, "Deciphering Alzheimer disease," Cold Spring Harbor Perspectives in Medicine, vol. 2, no. 1, p. a011460, 2012.

[2] A. S. Association, "2019 Alzheimer's disease facts and figures," Alzheimer's & Dementia, vol. 15, no. 3, pp. 321–387, 2019.

[3] 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.

[4] Mayo Clinic Staff, "Alzheimer's disease: Symptoms and causes," Mayo Clinic. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447.

[5] E. Nichols et al., "Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019," The Lancet Public Health, vol. 7, no. 2, pp. e105–e125, 2022.

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

[7] 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.

[8] 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.

[9] G. B. Frisoni, N. C. Fox, C. R. Jack Jr., P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," Nature Reviews Neurology, vol. 6, no. 2, pp. 67–77, 2010.

[10] E. J. van Beek et al., "Value of MRI in medicine: More than just another test?" Journal of Magnetic Resonance Imaging, vol. 49, no. 7, pp. e14–e25, 2019.

[11] 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.

[12] 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.

[13] 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.

[14] Z. Yuan, X. Li, Z. Hao, Z. Tang, X. Yao, and T. Wu, "Intelligent prediction of Alzheimer's disease via improved multifeature squeeze-and-excitation-dilated residual network," Scientific Reports, vol. 14, no. 1, p. 11994, 2024.

[15] P. Rajpurkar et al., "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," PLoS Medicine, vol. 15, no. 11, p. e1002686, 2018.

[16] A. S. Abdalrada, J. Abawajy, M. Chowdhury, S. Rajasegarar, T. Al-Quraishi, and H. F. Jelinek, "Meta learning ensemble technique for diagnosis of cardiac autonomic neuropathy based on heart rate variability features," in Proc. 30th Int. Conf. Computer Applications in Industry and Engineering (CAINE), 2017, pp. 169–175.

[17] A. Abdalrada, A. F. Neamah, and H. Murad, "Predicting diabetes disease occurrence using logistic regression: An early detection approach," Iraqi Journal for Computer Science and Mathematics, vol. 5, no. 1, pp. 160–167, 2024.

[18] A. S. Abdalrada, J. Abawajy, T. Al-Quraishi, and S. M. S. Islam, "Machine learning models for prediction of co-occurrence of diabetes and cardiovascular diseases: A retrospective cohort study," Journal of Diabetes & Metabolic Disorders, vol. 21, no. 1, pp. 251–261, 2022.

[19] S. Hussain, B. Shah, A. Khan, and S. Tanvir, "A novel prediction model for Alzheimer classification using deep learning," in Proc. Int. Conf. Asia Pacific Advanced Network, 2024, pp. 240–253.

[20] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," Frontiers in Public Health, vol. 11, p. 1273253, 2023.

[21] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," Brain Informatics, vol. 5, pp. 1–14, 2018.

[22] S. Koga, A. Ikeda, and D. W. Dickson, "Deep learning-based model for diagnosing Alzheimer's disease and tauopathies," Neuropathology and Applied Neurobiology, vol. 48, no. 1, p. e12759, 2022.

[23] 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.

[24] 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.

[25] V. S. Diogo, H. A. Ferreira, D. Prata, and Alzheimer's Disease Neuroimaging 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.

[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] 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.

[28] 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.

[29] 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.

[30] S. Alsubai, H. U. Khan, A. Alqahtani, M. Sha, S. Abbas, and U. G. Mohammad, "Ensemble deep learning for brain tumor detection," Frontiers in Computational Neuroscience, vol. 16, p. 1005617, 2022.

[31] GrayscaleImage.com, "Why use grayscale conversion when processing images?" [Online]. Available: https://www.grayscaleimage.com/why-use-grayscale-conversion-when-processing-images/.

[32] S. Schmid, "Image intensity normalization in medical imaging," Medium. [Online]. Available: https://medium.com/@susanne.schmid/image-normalization-in-medical-imaging-f586c8526bd1.

[33] IMAIOS, "Data augmentation for medical image analysis in deep learning." [Online]. Available: https://www.imaios.com/en/resources/blog/ai-for-medical-imaging-data-augmentation.

[34] "Alzheimer MRI 4 Classes Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/marcopinamonti/alzheimer-mri-4-classes-dataset.

[35] L. Zeng et al., "MAMC-Net: An effective deep learning framework for whole-slide image tumor segmentation," Multimedia Tools and Applications, vol. 82, no. 25, pp. 39349–39369, 2023.

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

[37] S. A. Hicks et al., "On evaluation metrics for medical applications of artificial intelligence," Scientific Reports, vol. 12, no. 1, p. 5979, 2022.

[38] A. Tharwat, "Classification assessment methods," Applied Computing and Informatics, vol. 17, no. 1, pp. 168–192, 2021.