DOI: https://doi.org/1036602/jsba.2025.20.60

Deep Learning-Based Detection and Diagnosis of Alzheimer's Disease from MRI Images: A Comparative Approach

Abdelkader Alrabai
Department of Physics, Wadi Alshatti University, Libya
a.alrabai@wau.edu.ly

Submission date: 25.07.2025 Acceptance date:21.10.2025 Electronic publishing date:02.11.2025

Abstract: Alzheimer's disease gradually erodes brain function, stringently disrupting memory and reasoning, expressly among older adults. Identifying the condition in its preliminary stages is decisive for timely support and potentially more operative care.This study investigated the application of deep learning models for the automated detection of AD from MRI images. Three Convolutional Neural Network (CNN) architectures are utilized specifically—VGG16, Xception, and ResNet50. The models are evaluated in both binary classification and multi-class classification. Standard evaluation metrics are used to assess model performance. For binary classification, ResNet50 had the highest accuracy (97.96%), followed by VGG16 (97.10%) and Xception (95.93%). In multi-class classification, ResNet50 additionally led (95.39%), slightly ahead of VGG16 (94.92%) and Xception (94.93%).These results underscore the strong potential of ResNet50, in particular, for clinical application, demonstrating reliable generalization to previously unseen MRI images. The study highlights the potential of deep learning models to enhance early detection of Alzheimer's disease by supporting clinical diagnosis, improving accuracy, and enabling timely interventions. Automated MRI analysis may also reduce costs and expand access to quality screening, especially in resource-limited settings—reinforcing the growing case for integrating AI into medical imaging workflows.
Keywords: Alzheimer, CNN, Demented, MRI

Introduction

Alzheimer's disease (AD) is the predominant driver of dementia, originally demonstrating as short-term memory breaks and progressively influencing thinking, reasoning, and behavior. Currently, around 50 million individuals live with dementia globally, and this number is expected to triple by 2050 due to an aging population. This trend poses serious challenges, including increased disability, healthcare costs, and disease load [1].AD follows a progressive course, beginning with subtle cognitive changes that may not be measurable through standard testing. Over time, these changes become more evident, affecting specific mental occupations. As the condition worsens, it can lead to dementia, where cognitive decline is severe enough to interfere with independence and daily life. Dementia is a hallmark of the later stages of AD, often characterized by significant memory loss and declined functional capacity [2].AD is generally delineated into three progressive phases—early, middle, and advanced—corresponding to the intensifying nature of cognitive and functional decline. In mild AD, individuals may have difficulty remembering daily activities but can still function independently, often without medical assistance. Early treatment at this stage is crucial to slowing the progression of the disease. In moderate AD, symptoms worsen due to increased brain damage, leading to increasing dependence on others for daily activities. In severe AD, extensive brain cell loss caused by extensive plaques and tangles causes significant brain shrinkage. Patients in this stage often lose mobility and the ability to communicate. Early detection is crucial to prevent the disease from progressing to this devastating stage [3]. Recognizing AD in its initial stages is decisive, as it enables individuals to make informed decisions about their health and adopt protective measures. It also assists medical professionals in assessing how likely the condition is to advance. For those affected, understanding the potential severity can be a powerful motivator to pursue lifestyle adjustments or begin treatment to slow its progression [4].Conventional procedures for diagnosing AD often rely on cognitive assessments conducted after noticeable symptoms emerge, resulting in many cases being recognized only in the later stages, when treatment options have limited impact. MRI has proven especially useful for examining structural brain changes, offering detailed differentiation between grey and white matter, and revealing alterations in regions such as the hippocampus and amygdala, which show early signs of atrophy linked to AD. However, interpretation of MRI scans can be subjective, highlighting the need for AIbased techniques to improve diagnostic accuracy and consistency [5].Deep learning techniques are particularly effective at identifying complex and informative features within extensive biomedical imaging datasets, including modalities like MRI and PET. These models, built from hierarchically structured neural layers, progressively enhance their predictive performance as they are exposed to larger volumes of training data. In AD prediction, leveraging extensive imaging datasets with deep learning offers promising potential to detect early signs and predict disease onset by identifying critical features that might be missed by traditional methods [6].

Numerous prior studies have explored the application of AI-based techniques for the classification and detection of AD. These efforts have employed a range of deep learning models and diverse neuroimaging datasets to enhance diagnostic accuracy. Several studies—such as those referenced in [7–11]—focused specifically on binary classification, using MRI images to distinguish between demented and non-demented individuals. In addition, works cited in [12] and [13] addressed both binary and multi-class classification, categorizing MRI scans into four clinical stages. Moreover, studies [14–18] have concentrated exclusively on multi-class classification using these same four AD progression categories.

The variety in model choices and dataset compositions across these studies highlights the ongoing efforts to optimize automated detection methods for AD. However, despite promising results, further research is still needed to address challenges such as model generalization across diverse populations, integration of multi-modal data, and validation on larger, more representative datasets. Advancements in these areas will be crucial for developing robust, clinically applicable tools for early and accurate AD diagnosis.

This study aims to automate and compare AD detection by CNN architectures, including VGG16, Xception, and ResNet50, using MRI images. This work provides a comprehensive comparative analysis, emphasizing model performance, generalization, and practical applicability. The research includes identifying the most effective CNN architecture for this classification task and offering insights into model behaviour.At its core, this study aims to contribute to the construction of dependable, AI-driven diagnostic systems that can aid medical specialists in identifying AD at its earliest phases with greater precision. By enhancing diagnostic accuracy and enabling earlier recognition, such techniques have the potential to improve clinical decision-making, facilitate prompt care approaches, and positively encouragement long-term patient trajectories.

Methodology

This study investigates AD detection and diagnosis from MRI images using three CNN architectures—VGG16 [19], Xception [20], and ResNet50 [21]. The models classify individuals into demented or non-demented categories for binary classification and into four classes—mild demented, moderate demented, very mild demented, and non-demented—for multi-class classification. The models will be evaluated on test datasets, and their results will be compared to identify the most effective architecture for each classification task.An overview of the method's sequential steps is illustrated in Figure 1.

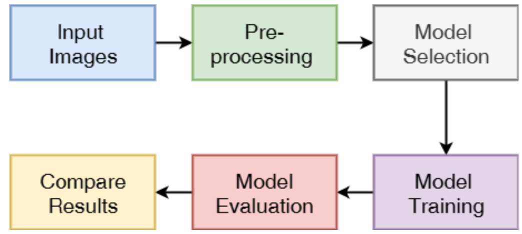

Fig. 1. Method key phases

The study utilized a Kaggle dataset[22]consisting of 6400 MRI images categorized into four classes: milddemented 896 images, moderatedemented 64 images, non-demented 3200 images, and verymilddemented 2240 images. For binary classification, the three demented categoriesare merged into a single demented class, resulting in a balanced dataset with 3,200 images for both demented and non-demented classes. Sample images from the dataset are shown in Figure 2.

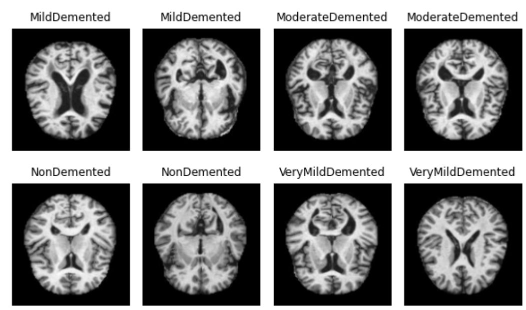

Fig. 2.Visual samples

Images are standardized to 128×128 pixels and normalized. All models, except a custom CNN, are initialized with ImageNet[23] weights, and their final layers are adapted for classification tasks. For binary classification, the output layer uses a sigmoid activation, while for multi-class classification (four classes), a softmax output layer is used. In both cases, the architecture includes a global average pooling layer followed by a dense layer with 512 units and ReLU activation. Training is performed using the Adam optimizer with a learning rate of 0.0001, a batch size of 32, and binary or categorical cross-entropy loss, depending on the task (binary or multi classification). Each model is trained for up to 30 epochs.For multiclass classification, the dataset is imbalanced; therefore, class weights are used to address the imbalance. Model effectiveness is measured using standard evaluation metrics. All experiments share the same settings and data partitions to ensure fairness, and results are validated on an isolated test set to ensure unbiased performance comparisons.

Results and Discussion

The models utilized in this study are evaluated using unseen MRI images from the test subset, which constitutes 20% of the original dataset and was separated before training. The classification accuracy results are summarized in Table 1, offering a comprehensive assessment of each model's effectiveness.Performance assessment mainly centres on four key outcome categories: true positives, true negatives, false positives, and false negatives. A true positive is recorded when the system accurately detects the presence of AD in a patient, whereas a true negative denotes the correct exclusion of the disease in a healthy individual. Conversely, false positives occur when the algorithm erroneously flags a non-affected person as having AD, while false negatives arise when the model overlooks the disease in an afflicted individual. These distinctions are crucial for quantifying the effectiveness of diagnostic models, with correct classifications indicating precision and misclassifications exposing vulnerabilities—particularly false negatives, whose reduction is vital to avoid postponement ininitiating essential care and intervention.

Table 1.Accuraciesresults

Classification	VGG16	Xception	ResNet50
Binary	0.9710	0.9593	0.9796
Multi-class	0.9492	0.9493	0.9539

The results presented in Table 1 provide a clear comparison of the classification accuracies achieved by three models utilized —across both binary and multi-class tasks in the context of dementia detection.In the binary classification task, ResNet50 outperforms the other models with an accuracy of 0.9796. VGG16 follows closely with an accuracy of 0.9710, demonstrating strong performance. Xception, lags slightly behind with an accuracy of 0.9593.

In the more complex multi-class classification task, which involves distinguishing between four different stages or types of cognitive conditions, the accuracy scores are slightly lower across all models. However, ResNet50 again leads with an accuracy of 0.9539, indicating that it maintains strong performance even as the classification task becomes more challenging. Interestingly, VGG16 and Xception show nearly identical accuracies—0.9492 and 0.9493 respectively—in this scenario.Overall, these results highlight ResNet50's consistent superiority across both binary and multi-class classification, making it the most effective and reliable model among the three for this application. VGG16 remains a strong contender, especially in binary classification, offering near-matching performance with lower computational demands. Xception, while effective in multi-class classification, underperforms in the binary case and may require further tuning, augmentation strategies, or architectural adjustments to fully leverage its capabilities. These findings are critical when selecting a model for medical imaging tasks where both accuracy and efficiency are essential.

The evaluation overview in Table 2 highlights a clear distinction in performance among the three architecturesutilized—when applied to the binary classification task distinguishing between demented and non-demented classes.

Table 2.Evaluation overview for binary classification

Metrics	Class	VGG16	Xception	ResNet50	Support
Precision	Demented	0.9718	0.9481	0.9767	640
	ND	0.9703	0.9711	0.9827	640
Recall	Demented	0.9703	0.9718	0.9828	640
	ND	0.9718	0.9468	0.9765	640
F1-score	Demented	0.9710	0.9598	0.9797	640
	ND	0.9711	0.9588	0.9796	640

ResNet50 clearly outperforms VGG16 and Xception across all key metrics, demonstrating its strong ability to accurately identify true positives while minimizing false predictions. It achieves the highest precision scores—0.9767 for demented and 0.9827 for non-demented (ND)—indicating fewer false positives. ResNet50 also leads in recall, with values of 0.9828 for demented and 0.9765 for ND, meaning it effectively detects most true cases and reduces missed diagnoses. Its F1-scores, 0.9797 for demented and 0.9796 for ND, highlight a well-balanced trade-off between precision and recall.VGG16 follows closely, with precision scores of 0.9718 demented and 0.9703 ND and recall values of 0.9703 demented and 0.9718 ND, showing reliable detection capabilities. Its F1-scores 0.9710 for demented and 0.9711 for ND confirm this balance, underscoring VGG16's robustness. This makes VGG16 a solid choice when computational resources or training time are limited.Xception shows comparatively lower performance, with a precision of 0.9481 for demented, suggesting more false positives. Although its recall for demented is fairly high at 0.9718, it drops to 0.9468 for ND, indicating more missed true negatives. This imbalance could risk overlooking healthy individuals in clinical settings. Its F1-scores 0.9598 for demented and 0.9588 for ND reflect a weaker balance between precision and recall compared to the other models.Since each model is evaluated on equal support of 640 samples per class, these results fairly reflect true performance differences. Overall, ResNet50 is the best-performing model for this binary dementia classification, critical in medical contexts where both false positives and negatives have serious consequences. VGG16 remains a reliable alternative, especially in resource-limited scenarios, while Xception needs more optimization to be competitive. This analysis guides model selection for dementia detection using MRI images, emphasizing accuracy and reliability.

Table 3. Evaluation overview for multi-class

Metric	Class	VGG16	Xception	ResNet50	Support
Precision	0	0.9548	0.9418	0.9719	180
	1	0.9166	0.8750	1.0000	14
	2	0.9625	0.9608	0.9509	640
	3	0.9292	0.9384	0.9496	448
Recall	0	0.9388	0.9000	0.9611	180
	1	0.7857	1.0000	1.0000	14
	2	0.9640	0.9593	0.9703	640
	3	0.9375	0.9531	0.9263	448
F1-score	0	0.9467	0.9204	0.9664	180
	1	0.8461	0.9333	1.0000	14
	2	0.9633	0.9601	0.9605	640
	3	0.9333	0.9457	0.9378	448

Mild demented = 0, moderatedemented = 1, non-demented = 2, very mild demented = 3

The multi-class evaluation metrics provided in Table 3 offer a comprehensive view of how the models classifying MRI images into four cognitive categories. The analysis of precision, recall, and F1-score across these classes reveals both the strengths and limitations of each model.ResNet50 model shows the most consistent and highest performance across nearly all metrics, particularly in the moderate demented class, where it achieves a perfect precision, recall, and F1-score of 1.0000, despite the small support size (14 samples). This indicates that ResNet50 was able to classify all samples of this class correctly, which is particularly impressive given the class imbalance. It also demonstrates high precision and recall for the mild demented class (0.9719 and 0.9611, respectively), leading to an F1-score of 0.9664. For non-demented, it achieves strong recall (0.9703) and solid precision (0.9509), suggesting balanced performance. Even in the very mild demented class, ResNet50 maintains good precision (0.9496) and a slightly lower recall (0.9263), indicating it missed a few true positives but made few false positives. These metrics confirm ResNet50's robust generalization and reliability, even in the presence of class imbalance.VGG16, while slightly behind ResNet50 in several areas, still delivers strong and stable performance, especially in the non-demented and very mild demented categories. It achieves precision scores of 0.9625 and 0.9292, respectively, with corresponding recall values of 0.9640 and 0.9375. The resulting F1-scores are also high (0.9633 and 0.9333), which underscores its reliability for these more prevalent classes. For mild demented, VGG16 performs well with an F1-score of 0.9467, though slightly lower than ResNet50. However, its performance drops in the moderate demented class, with a recall of 0.7857 and an F1-score of 0.8461. This suggests that VGG16 struggles more than ResNet50 to accurately detect rare or underrepresented cases, likely due to its shallower architecture and limited capacity to capture more subtle patterns.

In contrast, Xceptionshows more variable results across classes. It performs competitively in the very mild demented class, achieving the highest recall of 0.9531 and a strong F1-score of 0.9457. For moderate demented, Xception reaches a perfect recall of 1.0000 and a high F1-score of 0.9333, though its precision is lower at 0.8750—indicating more false positives. In contrast, it underperforms in the mild demented class with lower precision (0.9418), recall (0.9000), and F1-score (0.9204) compared to ResNet50 and VGG16. Similarly, while its performance on the non-demented class is quite strong, it is marginally behind the others in F1-score (0.9601). Overall, Xception's performance is competitive but slightly less stable, possibly due to its architectural sensitivity or training dynamics, especially in imbalanced scenarios.

Ultimately, ResNet50 consistently delivers the best results across all four classes, showing exceptional accuracy and generalization, particularly in detecting the rare moderate demented cases. VGG16 performs nearly as well in common classes and remains a strong option when computational resources are limited. Xception, while capable, exhibits more variability across classes and may require further tuning or augmentation to match the robustness of the other two models. The results validate ResNet50 as the top performer for multi-class dementia classification in MRI analysis, especially when accurate detection of all cognitive stages is critical.

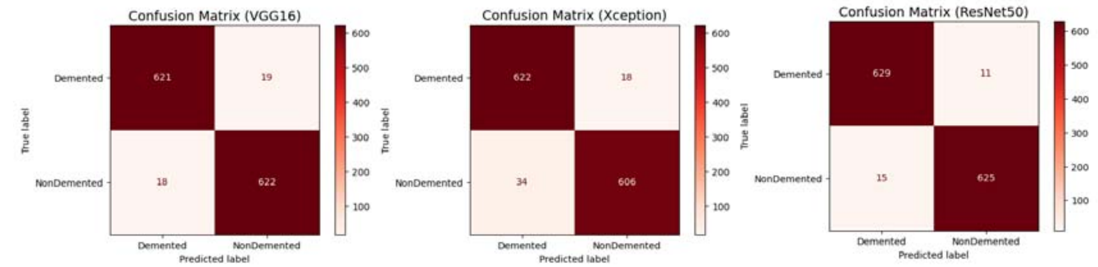

Fig. 3. Performance matrices for binary classification

Figure 3 displays matrices of the models utilized. All three models show a high degree of accuracy in correctly classifying both classes.Comparing the models, ResNet50 appears to have the highest overall performance. It correctly identifies 629 demented and 625 non-demented subjects, with the lowest number of false positives 11 cases and false negatives 15 cases. VGG16 and Xception show very similar performance. VGG16 correctly identifies 621 demented and 622 non-demented subjects, with 19 false positives and 18 false negatives. Xception's results are 622 correct demented and 606 correct non-demented, with 18 false positives and 34 false negatives. In summary, all three models are effective classifiers for this task. However, ResNet50 slightly outperforms the other two by achieving the highest number of correct classifications and the lowest number of misclassifications, making it the most robust model among the three for this specific dataset.

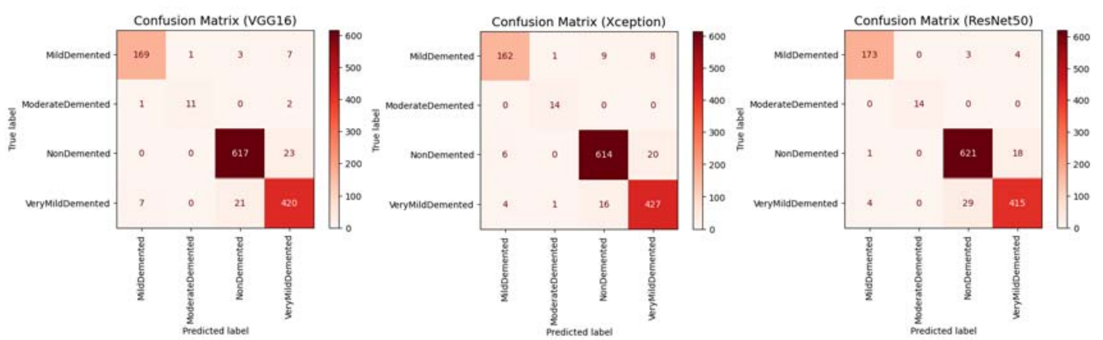

Fig. 4. Performance matrices for multi-class

Figure 4 demonstrates the matrices, representing the performance of the used models on multi-class classification task. A general analysis of all three models reveals that they are highly effective at classifying non-demented and very mildly demented cases, as indicated by the large numbers on the diagonal for these classes. However, they struggle significantly with the mildly demented and moderately demented categories. This is evidenced by the substantial misclassification of mildly demented cases as non-demented or very mildly demented, and the almost complete misclassification of moderately demented cases, with the models often predicting them as non-demented.Comparing the performance of the three models, ResNet50 appears to be the most accurate overall. It correctly identifies 173 mildly demented cases, which is slightly better than VGG16 (169) and Xception (162). ResNet50 also performs best in classifying non-demented cases with 621 correct predictions, compared to VGG16's 617 and Xception's 614. All three models have the most difficulty with the moderately demented class, and they all achieve the same low number of correct predictions (14). However, VGG16 and ResNet50 are better at correctly classifying very mildly demented cases, with 420 and 415 respectively, compared to Xception's 427. It's worth noting that VGG16 misclassifies 21 very mildly demented cases as non-demented, while Xception and ResNet50 only misclassify 16 and 29 respectively.

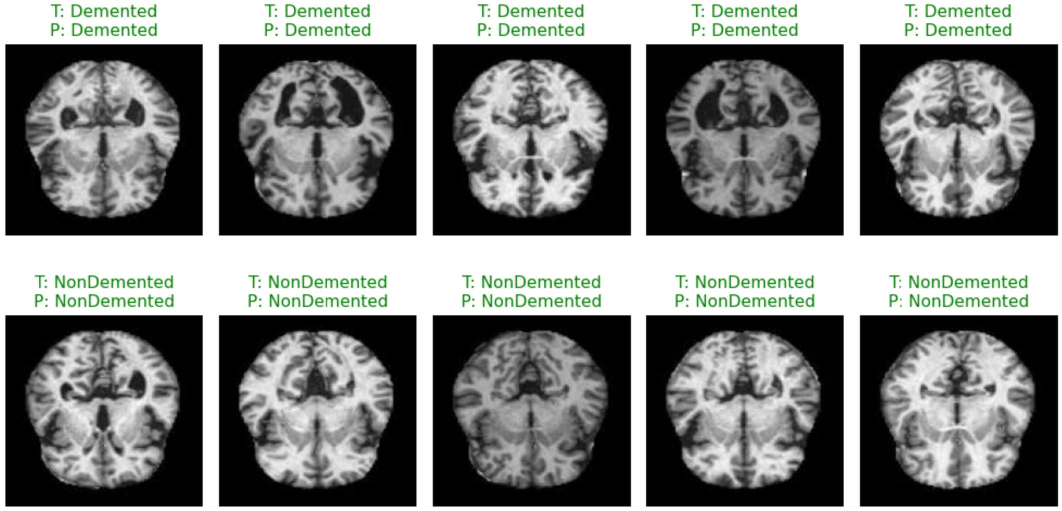

Fig. 5.Binary ResNet50 predictions

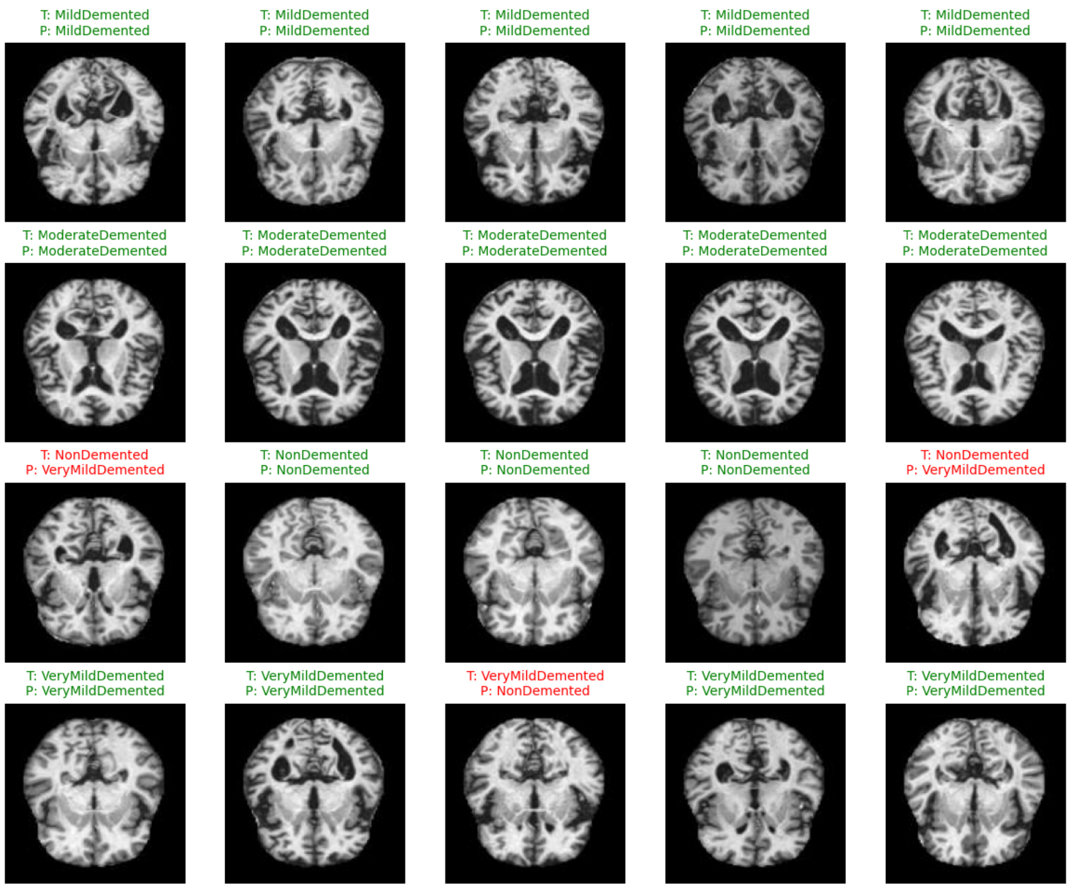

Fig. 6.Muti-class ResNet50 predictions

All three models exhibit a similar pattern of strengths and weaknesses. They are all highly proficient at identifying non-demented and very mildly demented individuals but struggle to distinguish between the various stages of dementia, particularly the moderately demented class. While the differences are subtle, ResNet50 demonstrates a slight edge in overall classification accuracy, particularly for the mildly demented and non-demented classes, making it the best-performing model among the three for this specific task. However, all three models would benefit from further improvements to address the significant misclassification of the moderately demented and mildly demented categories. Figure 5 demonstrates the ResNet50 prediction for binary classification. The top row displays five samples where the true label (T) is demented, and the ResNet50 model has correctly predicted (P) them as demented. This demonstrates the model's proficiency in identifying the characteristic features of dementia from the brain images. Similarly, the bottom row exhibits five samples where the true label is non-demented, and the model has accurately predicted them as non-demented. This highlights the model's ability to discern healthy brain structures from those indicative of dementia. The perfect concordance between the true labels and the predicted labels across all ten samples in Figure 5 serves as a visual testament to the high accuracy and effectiveness of the ResNet50 model in distinguishing between demented and non-demented brain scans, reinforcing the quantitative performance observed in its confusion matrix. Figure 6 shows sample predictions from the ResNet50 model for multi-class classification. The model performs well in identifying non-demented and very mildly demented cases, correctly classifying 4 out of 5 images in each category. It also correctly classifies all five mild demented samples shown, though the small sample size limits conclusions. While 4 out of 5 moderate demented predictions are correct, this likely reflects selective sampling, as the confusion matrix previously indicated low accuracy for this class. A key misclassification is a non-demented case labelled as very mildly demented, highlighting the model's difficulty in distinguishing between these similar categories.

Conclusion

This study presented a comprehensive evaluation of three deep learning-based CNN architectures for the automated detection and diagnosis of AD using MRI data. By comparing the models on a carefully curated dataset, the study highlighted the strengths of each architecture through common metrics, and generalization capability. Among the models evaluated, ResNet50 demonstrated the highest performance for both binary and muli-class classifications, suggesting its potential suitability for clinical decision support in AD. The findings reinforce the promise of deep learning as a tool for early and accurate detection of AD, which is critical for timely intervention and improved patient outcomes. In addition, this study contributes to existing literature by offering a comparative framework that can guide future model selection and development in the medical imaging domain. Future investigations should focus on enhancing the consistency of model performance across varied patient groups, incorporating relevant clinical data, and evaluating the systems in practical, clinical settings. Advancing in these areas is key to turning algorithm-driven diagnostics into dependable and widely usable tools in medical practice.

References:

[1] Zhang, X. X., Tian, Y., Wang, Z. T., Ma, Y. H., Tan, L., & Yu, J. T. (2021). The epidemiology of Alzheimer's disease modifiable risk factors and prevention. The journal of prevention of Alzheimer's disease, 8(3), 313-321.

[2] Knopman, D. S., Amieva, H., Petersen, R. C., Chételat, G., Holtzman, D. M., Hyman, B. T., ... & Jones, D. T. (2021). Alzheimer disease. Nature reviews Disease primers, 7(1), 33.

[3] Illakiya, T., & Karthik, R. (2023). Automatic detection of Alzheimer's disease using deep learning models and neuro-imaging: current trends and future perspectives. Neuroinformatics, 21(2), 339-364.

[4] Arafa, D. A., Moustafa, H. E. D., Ali-Eldin, A. M., & Ali, H. A. (2022). Early detection of Alzheimer’s disease based on the state-of-the-art deep learning approach: a comprehensive survey. Multimedia Tools and Applications, 81(17), 23735-23776.

[5] Yao, Z., Wang, H., Yan, W., Wang, Z., Zhang, W., Wang, Z., & Zhang, G. (2023). Artificial intelligence-based diagnosis of Alzheimer's disease with brain MRI images. European Journal of Radiology, 165, 110934.

[6] Vrahatis, A. G., Skolariki, K., Krokidis, M. G., Lazaros, K., Exarchos, T. P., &Vlamos, P. (2023). Revolutionizing the early detection of Alzheimer’s disease through non-invasive biomarkers: the role of artificial intelligence and deep learning. Sensors, 23(9), 4184.

[7] Sorour, S. E., Abd El-Mageed, A. A., Albarrak, K. M., Alnaim, A. K., Wafa, A. A., & El-Shafeiy, E. (2024). Classification of Alzheimer’s disease using MRI data based on Deep Learning Techniques. Journal of King Saud University-Computer and Information Sciences, 36(2), 101940.

[8] Tekin, R., &Onur, T. Ö. (2025). Binary Classification of Alzheimer's Disease Using Siamese Neural Network for Early Stage

Diagnosis. Celal Bayar University Journal of Science, 21(2), 152-158.

[9] Mohammed, R. K., Jawad, M. Q., & Jasim, O. M. (2025). Alzheimer Detection Using Deep Learning Methods. Journal of Intelligent Systems & Internet of Things, 16(2).

[10] Prajapati, R., & Kwon, G. R. (2022). A binary classifier using fully connected neural network for Alzheimer's disease classification. Journal of Multimedia Information System, 9(1), 21-32.

[11] Mamun, M., Shawkat, S. B., Ahammed, M. S., Uddin, M. M., Mahmud, M. I., & Islam, A. M. (2022, October). Deep learning based model for alzheimer's disease detection using brain mri images. In 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 0510-0516). IEEE.

[12] El-Latif, A. A. A., Chelloug, S. A., Alabdulhafith, M., & Hammad, M. (2023). Accurate detection of Alzheimer’s disease using lightweight deep learning model on MRI data. Diagnostics, 13(7), 1216.

[13] Pradhan, N., Sagar, S., & Singh, A. S. (2024). Analysis of MRI image data for Alzheimer disease detection using deep learning techniques. Multimedia Tools and Applications, 83(6), 17729-17752.

[14] Marwa, E. G., Moustafa, H. E. D., Khalifa, F., Khater, H., &AbdElhalim, E. (2023). An MRI-based deep learning approach for accurate detection of Alzheimer’s disease. Alexandria Engineering Journal, 63, 211-221.

[15] Sharma, S., Guleria, K., Tiwari, S., & Kumar, S. (2022). A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans. Measurement: Sensors, 24, 100506.

[16] Fareed, M. M. S., Zikria, S., Ahmed, G., Mahmood, S., Aslam, M., Jillani, S. F., ... &Asad, M. (2022). ADD-Net: an effective deep learning model for early detection of Alzheimer disease in MRI scans. IEEE Access, 10, 96930-96951.

[17] Murugan, S., Venkatesan, C., Sumithra, M. G., Gao, X. Z., Elakkiya, B., Akila, M., & Manoharan, S. (2021). DEMNET: a deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images. Ieee Access, 9, 90319-90329.

[18] Helaly, H. A., Badawy, M., &Haikal, A. Y. (2022). Deep learning approach for early detection of Alzheimer’s disease. Cognitive computation, 14(5), 1711-1727.

[19] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[20] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[21] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[22] Alzheimer MRI 4 classes dataset, https://www.kaggle.com/datasets/marcopinamonti/alzheimer-mri-4-classes-dataset

[23] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.