Multi-Stages of Alzheimer's Disease Classification Using Deep Learning

Firzana Eiwany Mashi[1], Marshima Mohd Rosli[1*,2], Haryani Haron[1], Waleed A Hammood[3] and Salah A. Aliesawi[3]

[1] College of Computing, Informatics & Mathematics, Universiti Teknologi MARA, 40450, Selangor, Malaysia
[2] Institute for Pathology, Laboratory and Forensic Medicine, University Teknologi MARA, 47000 UiTM, Sungai Buloh, Selangor, Malaysia
[3] Faculty of Computer Science and Information Technology, University of Anbar, Iraq
marshima@uitm.edu.my

Abstract.

Alzheimer's disease (AD) is a genetic disorder that is characterized by a gradual deterioration in cognitive function and is the most prevalent cause of dementia. Early and precise identification of AD stages is essential for efficient therapy and control. The existing diagnostic techniques, which mainly depend on Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans, are inefficient and susceptible to mistakes made by humans, resulting in delays in diagnosing and treating patients. This study aims to propose a deep learning model for classifying the different stages of AD using scans. This study evaluates the performance of three deep learning models, including Convolutional Neural Networks (CNN), ResNet50 and VGG16, on the MRI dataset collected from Kaggle, consists of images classified into four categories: no dementia, very mild dementia, mild dementia, and moderate dementia. Our results show that CNN outperformed the other models with an accuracy of 88.4%, demonstrating high sensitivity and specificity across different stages. The system achieved an accuracy of 97% in classifying AD into multiple categories. CNN exhibited strong performance in accurately differentiating between various stages of AD, as indicated by the results of the receiver operating characteristic (ROC) curve. Future research will prioritize the integration of supplementary clinical datasets to enhance the resilience of the model and broaden its diagnostic capabilities to encompass other neurodegenerative disorders.

Keywords: Alzheimer’s disease, DenseNet169, ResNet50, CNN

1 Introduction

Alzheimer's disease (AD) is a severe disorder characterised by the buildup of amyloid plaques and neurofibrillary tangles, resulting in the death of neurons [1], [2]. It accounts for 60% to 80% of dementia cases worldwide, predominantly affects individuals aged

N. A. S. Abdullah et al. (eds.), *Proceedings of the International Conference on Innovation & Entrepreneurship in Computing, Engineering & Science Education (InvENT 2024)*, Advances in Computer Science Research 117,
https://doi.org/10.2991/978-94-6463-589-8_42

accounts for 60% to 80% of dementia cases worldwide, predominantly affects individuals aged 65 and above. Early detection and accurate classification of AD are essential for efficient therapy and management of illness. However, existing diagnostic methods, which mostly rely on Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans, are time consuming and prone to human error, leading to unreliable diagnosis and delayed treatment.

Prior research has mostly focused on the binary categorization of AD versus healthy brains, but these studies have not thoroughly explored the requirement for multi-stage classification. The multi-stage classification is crucial for developing detailed treatment strategies tailored to the specific stages of AD: mild, moderate and severe [3], [4]. MRI is widely used to visualise structural changes in the brain, such as the reduction in size of cerebral and hippocampal areas [5] Several Computer-Aided Diagnosis Systems (CADS) have been created to improve the precision of MRI-based diagnoses by identifying and categorising characteristics linked to AD.

Artificial Intelligence (AI) has transformed numerous domains, such as healthcare, by enabling the processing of large datasets swiftly and accurately [7]. Machine Learning (ML) and Deep Learning (DL) are highly efficient techniques within the field of AI. DL, particularly Convolutional Neural Networks (CNNs), have demonstrated significant potential in medical imaging because of their exceptional accuracy in classifying images [8]. Several AI-based models have been suggested for AD diagnosis. Klöppel et al. [9] employed a linear support vector machine to analyse T1-weighted MRI data, whereas Grey, Aljabar et al. [10] utilised a multimodal classification approach employing PET and MRI data along with a random forest classifier. Morra, Tu et al. [11] conducted a comparison of different models, such as hierarchical AdaBoost and SVM, to identify AD in MRI data.

This study aims to evaluate the performance of three deep learning models using CNN architecture for classifying the stages of Alzheimer's Disease from MRI data. The suggested method seeks to streamline the classification process through automation, with the aim of reducing the diagnostic workload on radiologists, enhancing diagnostic accuracy, and delivering timely information on patient situations. The proposed method aims to enhance the efficiency of the categorization process by utilising a web-based interface to optimise the workflow for radiologists. The web-based interface will streamline the process of entering and organising data, as well as managing patient records, thereby improving the efficiency and precision of AD diagnosis.

2 Related work

Researchers have utilised a range of AI methods, including ML such as Support Vector Machine, as well as advanced DL techniques, to train and test data related to AD. The accuracy of AD classification could significantly improve patient care. Neural networks like Residual Network (ResNet), Visual Geometry Group (VGG), and CNN have proven to be highly efficient at extracting features from intricate data [6]. AD is a serious worldwide health issue that is characterized by memory loss and a slow decline in cognitive function that leads to dementia. Since AD is the main cause of dementia, controlling and putting into practice effective treatment options requires an accurate and timely diagnosis of the illness at an early stage. Unfortunately, the current diagnostic techniques which mostly depend on PET and MRI scans come with inefficiencies and hazards related to human error. The diagnosis and initiation of treatment for patients are thus postponed.

CNN is a DL technique commonly used for image categorization and recognition. It has convolution, pooling, fully connected that work together to classify and predict input [12]. Helaly et al. [13] emphasised the capability of CNNs to acquire hierarchical data representations, encompassing both low-level and high-level characteristics. Majdah et al. [22] developed a system based on CNN and utilised an Adam optimizer. The system achieved an accuracy of 97% in classifying AD into multiple categories. Shiny Pershiya et.al [14] introduced the LeNet5 model for classifying AD into two categories, obtaining an accuracy of 98.64%.

ResNet is a CNN with deep architecture, developed by Microsoft Research in 2015 [15]. It includes several variations, namely ResNet-50, ResNet-34, ResNet-101, and ResNet-152, where the numerical value represents the number of layers. ResNet efficiently tackles the problems of degradation and vanishing gradient by employing skip connections, which allow the network to learn residual functions [6]. Roy et al. [16] found that ResNet-152 accurately categorises early AD with 98.79% multi-classification accuracy. The hyperparameters were optimised using gradient-based optimisation, and a dropout layer was incorporated to mitigate overfitting. Dwivedi et al. [6] employed ResNet-50 and achieved accuracy of 97%, surpassing other pre-trained architectures. Li et al. [14] conducted a comparison between VGG19, AlexNet, and ResNet-50 to categorise AD stages and results showed that ResNet-50 achieved an accuracy of 78.04%.

VGG Net is recognised for its efficiency and simplicity [17]. Variations such as VGG16 and VGG19 use several convolutional layers, max-pooling layers, and fully linked layers for classification. The employment of 3×3 kernel-sized filters in VGG Net enables it to acquire intricate features while minimising the number of parameters [18]. Abed et al. [19] utilised the VGG19 with transfer learning to classify different phases of Alzheimer's disease, attaining a validation accuracy of 93%. Helaly et al. [13] employed an enhanced VGG19 to analyse 2D and 3D MRI images, resulting in a 97% accuracy. Sharma et al. [20] found that by utilising the VGG16 with an ensemble classifier and data augmentation, they obtained an accuracy of 96.8% in classifying brain tumours into multiple categories [24].

AI, especially ML and DL, is being used to automate and improve AD diagnosis. ResNet accurately classifies AD stages despite deterioration and vanishing gradients. CNNs' hierarchical data representation and VGG networks' simplicity make them accurate cognitive condition discriminators. These advances drive research in AI-based diagnostic tools, aiming to detect AD early and accurately, increasing patient outcomes and neurodegenerative disease diagnosis [25].

3 Methodology

This study has constructed a workflow model for classifying Alzheimer disease stages as illustrated in Figure 1. The workflow model has three phases: 1) Data preprocessing. 2) Alzheimer Disease Classification. 3) Performance measurement. In Data Preprocessing, the MRI dataset is processed through three methods which are cropping and median filtering, contrast limited adaptive histogram equalisation (CLAHE) and resize, and data augmentation. In Alzheimer Disease Classification, the dataset is used for classifying four Alzheimer disease stages using three deep learning models known as CNN, ResNet, and VGG. In the Performance Measurement, the effectiveness of each model is assessed using important metrics such as accuracy, sensitivity, and specificity.

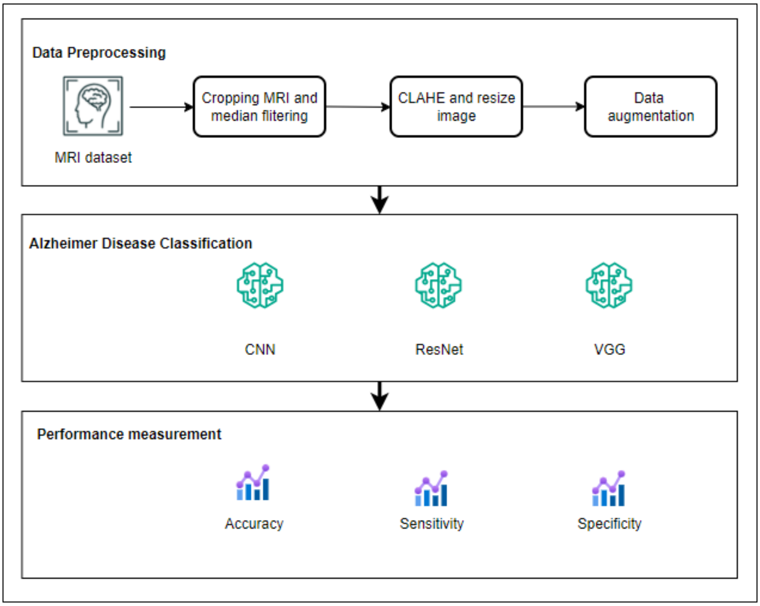

Fig. 1. Workflow model

3.1 Data Preprocessing

The dataset was collected from the Kaggle source and comprises 5888 MRI pictures. The dataset is divided into four distinct categories: Non-Dementia, Very Mild-Dementia, Mild Dementia, and Moderate Dementia. Every image was uniquely tagged for the purpose of analysis. Figure 2 depicts the four stages of Alzheimer's disease.

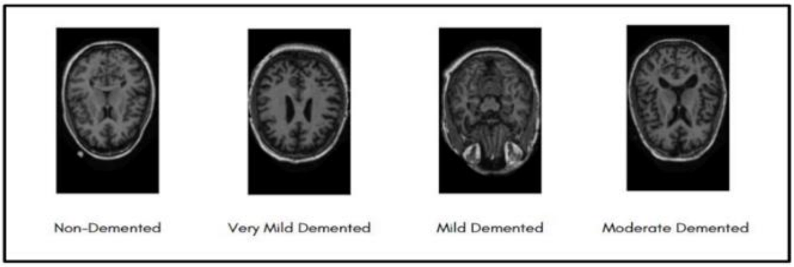

Fig. 2. Four stages of Alzheimer's disease

The preprocessing encompassed the cropping of MRI images, conversion to grayscale, enhancement using CLAHE, and the application of data augmentation techniques. These methods enhance the quality of the dataset and increase the performance of the model. In the cropping and grayscale conversion stage, the MRI images were cropped to eliminate extraneous background and noise, and were subsequently converted to grayscale to decrease computing complexity. Figure 3 shows the results of cropping MRI and median blur.

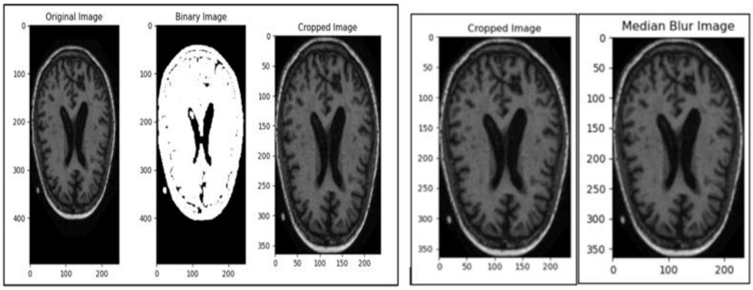

Fig. 3 Result of cropping MRI and median blur

In the enhancement stage, the application of contrast limited adaptive histogram equalisation (CLAHE) improved image contrast, hence enhancing the visibility of significant features. The CLAHE function is utilised with clip limit and tile grid size. The clip limit determine the maximum slope of the cumulative distribution function during the contrast enhancement and the tile grid size parameter defines the size of regions used of adaptive histogram equalization. Figure 4 shows the results of CLAHE.

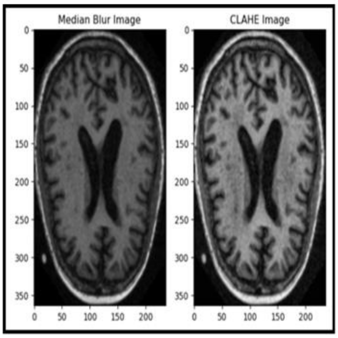

Fig. 4. Results of CLAHE

In the resize and data augmentation stage, we resized the images and applied data augmentation including rotation, zoom, and horizontal flipping to expand the size of the training images and introduce more variety to the dataset. The augmentation process is performed using the Keras ImageDataGenerator library.

A rotation range of 30 degrees is applied to introduce multiple image orientation, enhancing the model's ability to generalize across various angles. Additionally, a horizontal lip with a probability of 50% creates mirror images, diversifying the dataset. A random width shift of 10% is introduces lateral shifts to enhance the model's adaptability to changes in object positions. Figure 5 shows the augmented images for moderate demented.

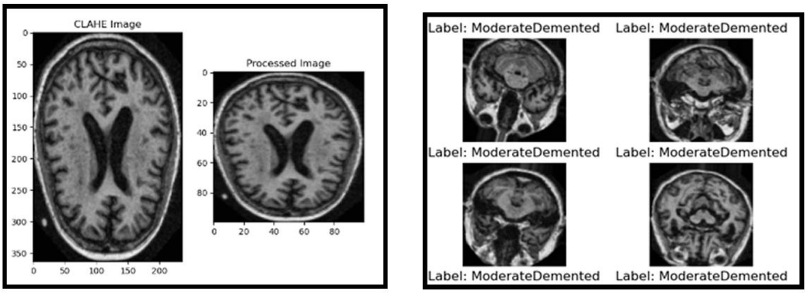

Fig. 5. Results of resize and data augmentation

3.2 Alzheimer Disease Classification

This study developed predictive models using Python, utilising libraries such as Scikit-learn, Tensorflow, Keras for feature extraction, model training, testing and validation. The clean dataset is divided into training 80%, testing 10% and validation 10% which belonged to four classes. Three deep learning models based on CNN architecture were employed for the classification of Alzheimer's disease: CNN, ResNet50 and VGG16.

A CNN for feature classification consists of an input layer that defines images with dimensions of 240x240 pixels and 3 colour channels. The convolutional layers contain important parameters and feature maps, with padding used to align the input and output dimensions. Batch normalisation is typically applied before ReLU activation functions. Max-pooling layers decrease the spatial dimensions of feature maps, getting rid of unnecessary data and decreasing the computing requirements. Fully connected layers handle the final classification and the Softmax layer normalizes outputs and generates positive integers for categorization. The final classification layer utilises the Softmax function to classify inputs, calculating probability and loss for each image.

ResNet50 is a CNN consisting of 50 layers, incorporating residual blocks and shortcut connections to improve gradients flow and prevent vanishing gradients. The process begins with a 7x7 initial convolution layer, followed by successive layers with 3x3 filters. The model also incorporates max-pooling and global average pooling layers. Residual learning in ResNet50 enables more efficient training of deep networks, provides exceptional precision and optimized gradient flow; however, it is intricate and necessitates meticulous calibration.

VGG16 is derived from CNN that consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. The model utilizes max-pooling layers to decrease the size of spatial dimensions. VGG16 does not employ residual learning, unlike ResNet50. The model attains a high level of accuracy, but it requires substantial processing resources and may face gradient problems as a result of its depth and parameter count.

3.3 Performance Measurement

This study assessed the effectiveness of deep learning models by measuring their accuracy, specificity, and sensitivity. Accuracy refers to the degree of correctness in making predictions. Sensitivity quantifies the model's capability to accurately identify cases of AD, whereas specificity evaluates its capability to accurately identify situations that are not AD. The evaluation metrics used four fundamental qualities obtained from the confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The accuracy values were computed using Equation (1).

$$\text{Accuracy} = TP + TN/(TP + FP + TN + FN) \quad (1)$$

We utilised sensitivity and specificity as measures to validate the accuracy values. The procedure for calculating these measures is explained in Eq. (2) and Eq. (3).

$$\text{Sensitivity} = TP/(TP + FN) \quad (2)$$

$$\text{Specificity} = TN/(TN + FP) \quad (3)$$

4 Results and Discussion

This study examines the performance accuracy of three deep learning models (CNN, ResNet50, and VGG16) by utilising the ReLU layers to prevent the exponential growth of neural networks by changing any negative input values of neuron to zero and only activates positive inputs which can reduced the computational load. The classification results obtained by the three models were evaluated using the three different performance metrics: accuracy, sensitivity, and specificity. Table 1 shows the statistical results of the training and testing after the successful execution of the deep learning models. CNN outperformed ResNet50 and VGG16 in both training and testing. Based on these results, this study decided to use the CNN model for further processing.

Table 1. Deep Learning Model Performance Results

Deep Learning Model	Training	Testing
CNN	88.4%	77.4%
ResNet50	62.2%	61.0%

VGG16	70.1%	65.2%

Figure 6 shows the training and validation accuracy of the CNN with data augmentation and noise removal on the MRI dataset over 50 epochs. Initially (epochs 1-10), training accuracy was 28-36%. In the middle epochs (11-30), training accuracy increased to 78-82%, with validation accuracy at 69-70%. In the final epochs (31-50), training accuracy reached 87-88%, and validation accuracy was 77-78%. Overall, the CNN model achieved a training accuracy of 88.4%.

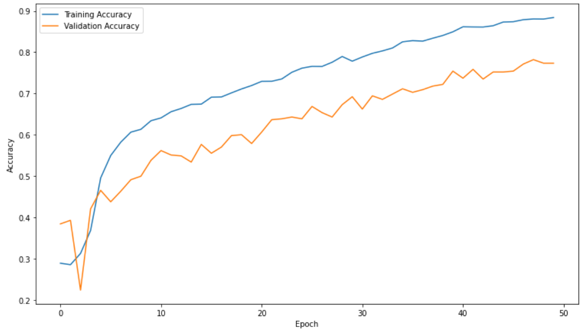

Fig. 6. CNN model training and validation accuracy

Figure 7 shows an orange line for validation loss and a blue line for training loss. In CNN, both the training and validation loss decrease in the initial epoch (1-10), indicate that the CNN is learning and improving. In the epoch (31-50), the validation loss indicates a significant decrease that indicates improve generalization.

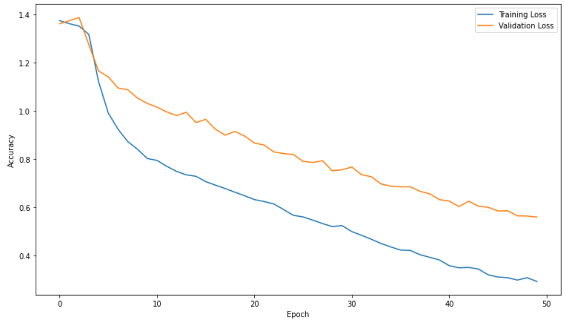

Fig. 7 CNN model training and validation loss

This study used sensitivity and specificity to evaluate CNN's performance in classifying AD stages. High sensitivity indicates the model effectively identifies patients with AD at various stages (Very Mild Demented, Mild Demented, and Moderate Demented). High specificity shows the model has the ability to accurately identify NonDemented patients. Table 2 shows the sensitivity and specificity results for the CNN model accross the four stages of AD. The high values for both metrics demonstrate the model successfully reduces false positives and false negatives.

Table 2. Measurement of CNN

Stages	Sensitivity	Specificity

NonDemented	77.2%	90.5%
VeryMildDemented	82.2%	83.3%
MildDemented	84.3%	95.6%
ModerateDemented	54.5%	99.6%

In this study, Receiver Operating Characteristic (ROC) curve is utilised to illustrate the performance of the CNN at various classification thresholds. Figure 9 shows the result of the ROC curve for the CNN, which plots different classification thresholds to illustrate the difference between the True Positive Rate and False Positive Rate. The ROC scores for the four classes (NonDemented, Very Mild Demented, Mild Demented, and Moderate Demented) are between 92% and 98%. These results show that the model is performing well in differentiating between the different stages of Alzheimer's disease. The ROC scores are consistently high, indicate robust performance in distinguishing each class from the others, thus suggesting that the CNN model can accurately identify patients across the various stages of disease.

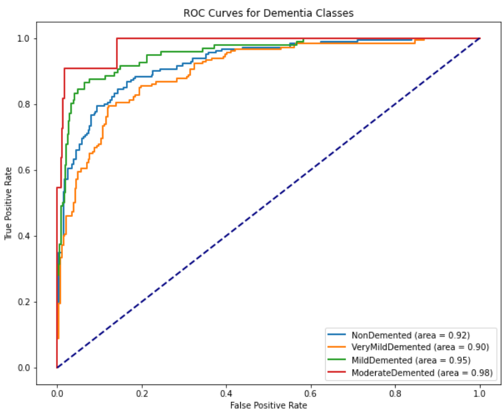

Fig 8. ROC Curve of CNN

This study developed a prototype system to simulate the classification of Alzheimer’s stages using the proposed CNN model. Figure 9 shows the user interface for the system that assists radiologists to interact with the system. The interface includes several functionalities: the radiologist can create an account, log into the system, register patient information, predict the stage of Alzheimer’s disease from patient MRI scan image, and track the MRI records for each patient. The dashboard provides an overview of the classification counts for each stage of Alzheimer’s disease.

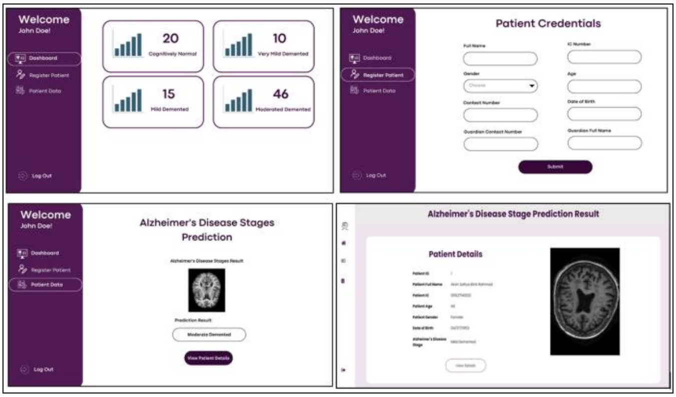

Fig. 9. A prototype of Alzheimer's stages classification system.

5 Conclusion

Alzheimer disease (AD) is a major global health concern that is typified by a gradual deterioration in cognitive abilities and memory loss, which eventually results in dementia. Since AD is the primary cause of dementia, early and precise diagnosis of the disease's stages is essential for managing and implementing successful treatment approaches. Unfortunately, there are inefficiencies and human error risks associated with the present diagnostic methods, which mostly rely on PET and MRI scans. As a result, patients' diagnosis and start of treatment are delayed.

This research addressed this issue by presenting a novel method for classifying distinct phases of AD using neuroimaging scans and deep learning models. This study assesses the performance of the ResNet50, VGG16, and CNN architectures using an MRI dataset that was obtained from Kaggle. Images from four classes: no dementia, very mild dementia, moderate dementia, and mild dementia are included in this dataset. This study evaluated the accuracy of AD stage classification using deep learning models and compared CNN, ResNet50, and VGG16 performance in this regard. After a thorough study, the findings showed that CNN performs better and has higher sensitivity and specificity at different stages of AD. Positive results on ROC curve analysis indicate that CNN is particularly effective in differentiating between AD stages.

Acknowledgments. The authors would like to thank the College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Selangor, Malaysia for their financial support to this research project.

References

1. R. Sivakani and G. A. Ansari, "Machine Learning Framework for Implementing Alzheimer's Disease," in *Proceedings of the 2020 IEEE International Conference on Communication and Signal Processing, ICCSP 2020*, 2020.

2. E. K. Lee, F. Yuan, B. J. Man, and B. Egan, "A General-Purpose Multi-stage Multi-group Machine Learning Framework for Knowledge Discovery and Decision Support," in *Communications in Computer and Information Science*, 2023.
3. B. D. W. Scharre, "Preclinical, Prodromal, and Dementia Stages of Alzheimer's Disease," *Practical Neurology Magazine*, no. June, 2019.
4. F. Ahmad, M. Javed, M. Athar, and S. Shahzadi, "Determination of affected brain regions at various stages of Alzheimer's disease," *Neurosci Res*, vol. 192, 2023.
5. A. W. Salehi, P. Baglat, B. B. Sharma, G. Gupta, and A. Upadhya, "A CNN Model: Earlier Diagnosis and Classification of Alzheimer Disease using MRI," in *Proceedings - International Conference on Smart Electronics and Communication, ICOSEC 2020*, 2020.
6. S. Dwivedi, T. Goel, M. Tanveer, R. Murugan, and R. Sharma, "Multimodal Fusion-Based Deep Learning Network for Effective Diagnosis of Alzheimer's Disease," *IEEE Multimedia*, vol. 29, no. 2, 2022.
7. K. Chhaya, A. Khanzode, and R. D. Sarode, "ADVANTAGES AND DISADVANTAGES OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING: A LITERATURE REVIEW," *International Journal of Library & Information Science*, 2020.
8. A. Mehmood, M. Maqsood, M. Bashir, and Y. Shuyuan, "A deep siamese convolution neural network for multi-class classification of alzheimer disease," *Brain Sci*, vol. 10, no. 2, 2020.
9. S. Klöppel *et al.*, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, 2008.
10. K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *Neuroimage*, vol. 65, 2013.
11. J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson, "Comparison of adaboost and support vector machines for detecting alzheimer's disease through automated hippocampal segmentation," *IEEE Trans Med Imaging*, vol. 29, no. 1, 2010.
12. M. Shafiq and Z. Gu, "Deep Residual Learning for Image Recognition: A Survey," *Applied Sciences (Switzerland)*, vol. 12, no. 18. 2022.
13. P. Roy, M. M. Oddin Chisty, and H. M. Abdul Fattah, "Alzheimer's Disease Diagnosis from MRI images using ResNet-152 Neural Network Architecture," in *2021 5th International Conference on Electrical Information and Communication Technology, EICT 2021*, 2021.
14. Q. Li, W. He, X. Zhang, Y. Yang, and T. T. Toe, "Comparative study of Transfer Learning and VGGNet based classification of Alzheimer's disease," in *Proceedings - 2022 12th International Conference on Information Technology in Medicine and Education, ITME 2022*, 2022.
15. G. Boesch, "VGG Very Deep Convolutional Networks (VGGNet) - What You Need to Know," *viso.ai*. 2021.
16. A. Namachivayam and N. Puviarasan, "Computerized Brain Disease Classification Using Transfer Learning," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 7s, 2023.
17. H. A. Helaly, M. Badawy, and A. Y. Haikal, "Deep Learning Approach for Early Detection of Alzheimer's Disease," *Cognit Comput*, vol. 14, no. 5, 2022.
18. S. Sharma, K. Guleria, S. Tiwari, and S. Kumar, "A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans," *Measurement: Sensors*, vol. 24, 2022.
19. A. Priyatama, Z. Sari, and Y. Azhar, "Deep Learning Implementation using Convolutional Neural Network for Alzheimer's Classification," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, 2023.
20. B. Y. Lim *et al.*, "Deep Learning Model for Prediction of Progressive Mild Cognitive Impairment to Alzheimer's Disease Using Structural MRI," *Front Aging Neurosci*, vol. 14, 2022.
21. R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4. 2018.
22. M. Alshammari and M. Mezher, "A Modified Convolutional Neural Networks for MRI-based Images for Detection and Stage Classification of Alzheimer Disease," in *Proceedings - 2021 IEEE 4th National Computing Colleges Conference, NCCC 2021*, 2021.
23. W. A., et al. "Conceptual model of internet banking adoption with perceived risk and trust factors." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 21.5 (2023): 1013-1019.

24. Cucus, Ahmad, et al. "Systematic Review of IT Risk Management Using a Scientometric Approach." *2023 6th International Conference on Information and Communications Technology (ICOIACT)*. IEEE, 2023.
25. A. Shiny Pershiya, S. Suganyadevi, A. S. Rajasekaran, P. Satheesh N, R. Suganthi, and R. Naveenkumar, "Alzheimer's Disease Diagnosis using Deep Learning Approach," in *Proceedings of the 2023 2nd International Conference on Electronics and Renewable Systems, ICEARS 2023*, 2023.