Automated Alzheimer's Disease Diagnosis using Convolutional Neural Networks and Magnetic Resonance Imaging

ASMAA NASR MOHAMMED[1], ABDULGANI ALBAGUL[2], MOAMER MUSBAH AHMAD[3]
[1]College of Electronic Technology Baniwalid
[2]Libyan Center for Engineering Research and Information Technology
[3]University of Baniwalid
Baniwalid
LIBYA

Abstract: - Alzheimer's disease is a debilitating neuro-logical condition affecting millions globally; therefore, correct diagnosis plays a significant role in treating or managing it effectively. Convolutional neural networks (CNNs), which are popular deep learning algorithms are applied to image processing tasks, offer a good technique to study and investigate images processing. In this study, a CNN model for classifying Alzheimer's patients is proposed. The research yielded impressive results: recall and precision scores as high as 0.9958 which indicate trustworthy identification of true positives while maintaining few false positives; test accuracy exceeding 99% confirming desirable generalization capabilities from the training dataset to live scenarios; ROC AUC score at an astronomical height of 0.9999 signifying great potential in distinguishing between afflicted individuals from their non-affected counterparts accurately. The proposed network achieved a classification accuracy of 99.94% on LMCI vs EMCI, 99.87% on LMCI vs MCI, 99.95% on LMCI vs AD, 99.94% on LMCI vs CN, 99.99% on CN vs AD, 99.99% on CN vs EMCI, 99.99% on CN vs MCI, 99.99% on AD vs EMCI, 99.98% on AD vs MCI, and 99.96% on MCI vs EMCI. The proposed CNNs model is compared with two ultramodern models such as VGG19 and ResNet50. The results show that the proposed model achieved a superior performance in diagnostic precision and effectiveness of Alzheimer's disease, leading to early detection, enhanced treatment plans, and enriching the quality of life for those affected.

Key-words: - Leave Alzheimer's disease, MRI, CNN, Transfer Learning.

Received: June 16, 2022. Revised: August 19, 2023. Accepted: September 21, 2023. Published: October 4, 2023.

1 Introduction

Alzheimer's disease is a lasting condition that affects countless individuals, across the globe. It stands as the contributor to dementia accounting for 60-80% of diagnosed cases [1]. The illness is defined by the buildup of proteins in the brain such as beta amyloid plaques and tau tangles. These contribute to the depletion of neurons and synapses resulting in a decline in abilities, over time [2]. Alzheimer's disease usually starts with memory loss and confusion which then worsens into significant cognitive difficulties. These challenges can include issues with language, abilities, and executive function. Sadly, there is no cure for Alzheimer's disease now. The treatments available only aim to alleviate symptoms and slow down the progression of the illness. Ongoing research is being conducted to explore the factors behind Alzheimer's disease and find treatments. The main aim is to understand the mechanisms that drive this disease and develop therapies to address it. Promising areas of study involve the creation of medications that specifically target beta amyloid and tau proteins alongside pharmacological approaches, like making lifestyle adjustments and engaging in cognitive training [3]. Studies conducted recently have also showed a connection between Alzheimer's disease and other health conditions like diabetes, high blood pressure and depression. These findings underscore the significance of adopting an approach towards treating and preventing Alzheimer's disease [4, 5, 6]. In this research we aim to explore the effectiveness of a CNN model in diagnosing AD using MRI scans. Currently AD diagnosis relies on behavioural tests which may have exabit margin for error. One of the changes in the brains of AD patients is the atrophy of the hippocampus and cortex well as other structural alterations visible in MRI scans. CNN models have shown results in using MRI images to find signs of AD. These models could understand patterns within pictures. The primary aim of our project is to develop and evaluate a CNN model specifically designed for diagnosing AD based on MRI scans. We will use a dataset consisting of MRI scans from both AD patients and healthy individuals to train and evaluate our model's accuracy, sensitivity, and specificity. We will also compare its performance with existing techniques used for AD diagnosis. This research has the potential to contribute towards creating an effective technique for diagnosing AD. Timely and correct diagnosis can lead to interventions and improved outcomes for individuals affected by this condition. Integrating CNN technology with MRI scans in the diagnosis process may offer a systematic approach towards finding cases of Alzheimer's disease.

2 Methodology

Study design: We conducted a retrospective study using MRI data from patients with clinically diagnosed AD and age-matched healthy controls. We collected T1-weighted axial MRI scans from the ADNI dataset and OASIS3 dataset. The MRI data were pre-processed using the following steps.

2.1 Skull Stripping

Removing the skull, known as skull stripping, is an initial step in neuroimaging. It involves separating the brain tissue from the surrounding brain tissue and skull. This process is commonly performed to enhance the precision of image analysis techniques, such, as segmentation and registration [7, 8]. There are methods for skull stripping, such as thresholding, region growing and utilizing machine learning based approaches [9, 10]. Properly removing the skull is extremely important in neuroimaging applications such as functional MRI, diffusion tensor imaging and positron emission tomography. Skull stripping was done by examining the image, cropping it to eliminate any surrounding light box. Producing an image by applying a threshold. It then gets rid of specks of noise from the image and seals off the bottom part of the image. To create a version the binary image undergoes erosion before being used to mask the initial grayscale image effectively eliminating any gaps in the binary image as shown in Fig. 1 and Fig. 2.

2.2 Image Cropping

With precision in mind, we perform a series of steps aimed at obtaining an optimally sized and shaped image. Following calculations designed for scaling purposes, our discrepancy metric reduces to a diminutive proportion at just -10 (-2% of original size). Through careful manipulation of crop parameters based on user-defined bounds via bbox, we carve out only those areas necessary for proper viewing. The result is an accurately sized rectangle with dimensions exactly totaling up to no more than 20 total pixels (measuring both width and height) resulting in an image as shown in figure 3.

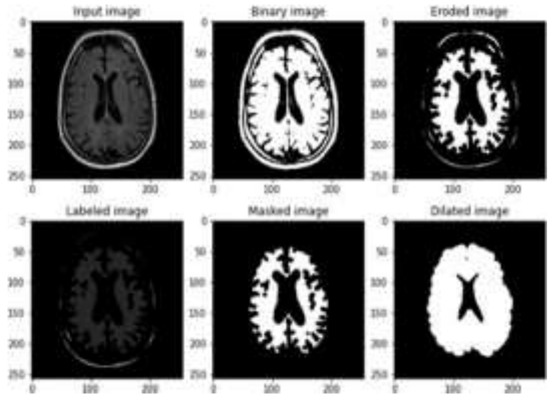

Fig. 1: Skull stripping stages

Fig. 2: Skull stripped image.

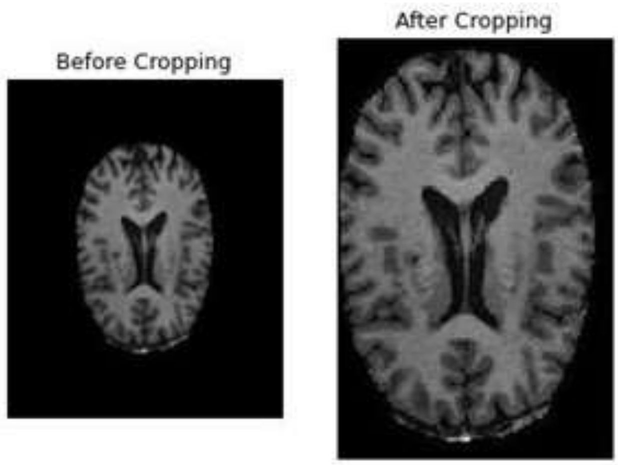

Fig. 3: Image cropping

2.3 Dataset Size

The MRI dataset for Alzheimer's disease classification consists of five classes: AD, CN, EMCI, LMCI, and MCI. The AD class includes individuals with Alzheimer's disease, while the CN class includes cognitively normal individuals. The EMCI, LMCI, and MCI classes include individuals with mild cognitive impairment in early, late, and progressive stages, respectively. The dataset is quite large, with a total of 40,077 MRI images, and it provides valuable insights into the brain changes that occur in individuals with Alzheimer's disease and related conditions. Each class in the dataset has a substantial number of MRI brain images, with 8000 pictures per class.

2.4 Convolutional Neural Network

Convolutional neural networks (CNNs) are a kind of deep learning technique employed to analyse images and videos. They consist of layers that execute convolutions, pooling, and nonl inear activations. These networks are trained using backpropagation to enhance a loss function [11, 12]. Convolutional Neural Networks (CNNs) have proven performance in computer vision tasks, such as accurately classifying images detecting objects and segmenting visual data [13, 14, 15]. They have also been used in fields, including natural language processing and the recognition of speech [16, 17]. Recent developments in CNNs include the creation of deeper and more intricate structures, such attention mechanisms, and residual networks, as well as the incorporation of CNNs with other deep learning methods, like generative adversarial networks and reinforcement learning [18, 19]. The CNN architecture primarily consists of five layers; Input Layer, Convolutional Layer, Pooling Layer, Fully Connected Layer and Output Layer [20]. These 5 CNN neural network layers as shown in fig 4.

Fig 4 Convolutional Neural Network [21]

2.5 The Proposed Architecture

Convolutional neural networks (CNNs), a type of deep learning technology [21], have recently demonstrated potential in the detection of Alzheimer's disease using MRI images. In this paper, we suggest a CNN architecture for MRI scan-based Alzheimer's disease identification. A few convolutional, pooling, and fully connected layers make up the proposed model, which is intended to extract and learn characteristics from the MRI images. To avoid overfitting, the model additionally uses regularization techniques like Dropout and L2 regularization. We think that our suggested model will increase the reliability of MRI scans used to identify Alzheimer's disease and assist physicians in making an early diagnosis when therapy is most successful. The proposed CNN model has 29 layers, including: 2 Conv2D layers, 2 MaxPooling2D layers, 6 SeparableConv2D layers, 6 Batch- Normalization layers, 3 MaxPool2D layers, 5 Dropout layers, 1 Flatten layer, 4 Dense layers. The proposed architecture is shown in Fig 5, and the proposed structure layout is shown in Fig. 6.

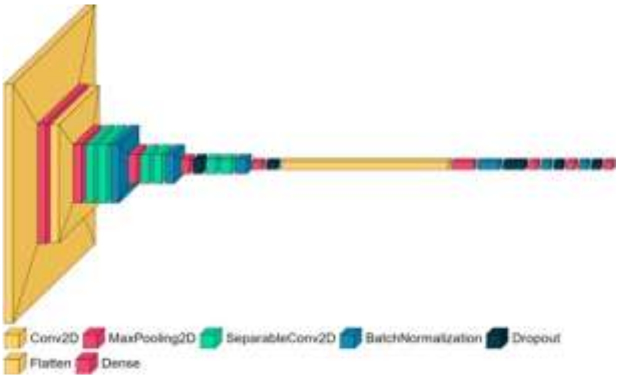

Fig. 5: The proposed architecture

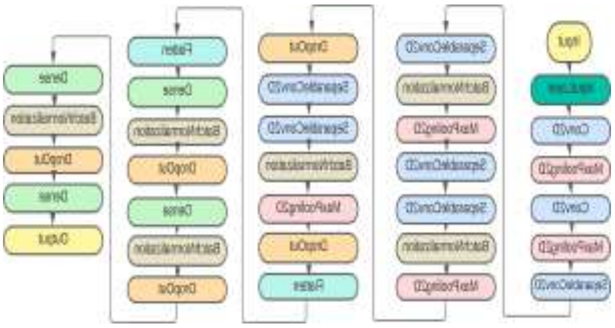

Fig. 6: The proposed network structural layout

2.6 Transfer Learning

In recent times people have been taking advantage of a technique called" transfer learning" when conducting image classification tasks due to its incredible utility potential. Pre trained models are key players in this technique since they allow features extraction from images with ease. Conventionally these models include deep neural networks trained on vast datasets such as ImageNet; their primary goal being able in finding varying objects and patterns within images. Essentially transfer learning exploits this by using the preexisting model to form a foundation for developing recognition capabilities for novel object categories [22]. As part of our comparison exercise, we will be using two of the most successful pre-trained models - VGG19 and ResNet50 -for image classification. Their excellent history in performing various image classification tasks has made them a preferred choice for transfer learning applications.

2.7 Model Training

The CNN training pipeline for Alzheimer's detection is shown in Fig. 7.

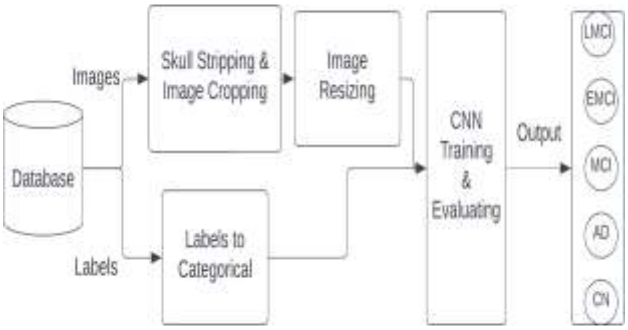

Fig. 7: The CNN Training Pipeline Diagram

The training procedure of the CNN model to classify images is carried out using Python and well-known packages such as Tensor-Flow and Keras. We also constructed our image processing in Python, utilizing tools like OpenCV and NumPy for operations like resizing, cropping, and normalization. To start we resized all the images in the dataset to a size of 150x150 pixels. This helped us ensure that the input size for the model was standardized. Additionally, we transformed the labels from their format into a format enabling us to employ a multi class classification model. Afterwards we divided the dataset into two parts. One, for training and the other for testing. We made sure to allocate 80% of the data for training purposes while keeping 20% aside to assess how well the model performed on unseen data. In our training process we utilized the entropy loss function, a widely employed method, for solving multi class classification problems. Additionally, we employed the Adam optimizer, which's an adaptive learning rate optimization algorithm that proves to be highly effective for learning models. To avoid overfitting, we implemented a stopping mechanism, with a patience of 10 epochs. Essentially this means that if the validation loss didn't show any improvement, for 10 epochs the training process would be halted prematurely to prevent the model from fitting to the training data. To maintain the model's convergence, towards a solution, without any overshooting or divergence we decided to decrease the learning rate by a factor of 10 after each epoch. This adjustment significantly enhanced the stability and performance of the model during training. In the end we made sure to train the model for 50 epochs allowing time for it to grasp the underlying patterns, in the data and attain performance. Through these methods we successfully trained a learning model that excels at classifying images, across categories.

2.7 Evaluation Metrics

Once the training of the network (CNN) model was complete, a range of metrics was tried to assess its effectiveness. These metrics encompassed accuracy, sensitivity, specificity positive predictive value (PPV) negative predictive value (NPV) and the area under the operating curve (AUC ROC). Accuracy is a used measure to evaluate how well a model performs. It quantifies the percentage of classified samples, among the number of samples. Sensitivity and specificity are metrics used to assess how well a model can accurately identify negative samples. Sensitivity gauges the ratio of identified samples to all positive samples whereas specificity calculates the ratio of correctly identified negative samples, to all negative samples [23]. PPV and NPV are metrics used to evaluate the negative values of a model. PPV calculates the ratio of predictions out of all samples predicted as positive while NPV calculates the ratio of correct negative predictions, out of all samples predicted as negative [24]. The AUC ROC is a metric that evaluates how well a model can differentiate between negative samples. It is calculated by plotting the sensitivity (rate) against the specificity (1. False positive rate), at different classification thresholds. To evaluate how well the model performed on the testing data we calculated the confusion matrix and the classification report. The confusion matrix gives us an overview of the model's predictions in terms of positives true negatives, false positives, and false negatives. The classification report provides a summary of the precision, recall and F1 score for each class predicted by the model. We utilized these measurements to assess how the CNN model performed on the test data and acquire an understanding of its merits and drawbacks. This enabled us to pinpoint areas that needed enhancement and implement any required modifications to the model.

3 Results and Discussion

CNN models are commonly used for image classification tasks, and their performance is typically evaluated using metrics such as accuracy and loss. Accuracy measures the proportion of correctly classified samples out of the total number of samples in the dataset, while loss measures the difference between the predicted and actual values for each sample, with lower values indicating better performance. During the training process, the model's accuracy and loss are constantly updated based on its performance on the validation dataset. As the model learns to recognize patterns in the images, its accuracy improves, and its loss decreases as shown in Fig. 8 and Fig. 9.

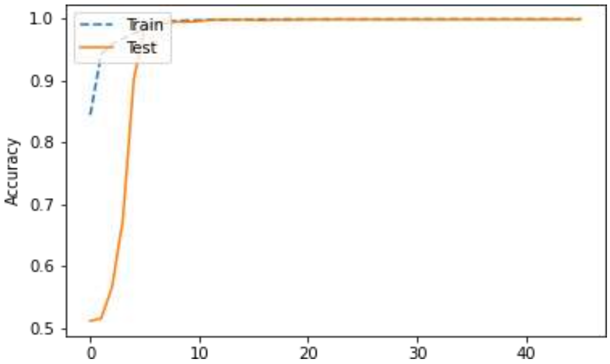

Fig. 8: Model accuracy

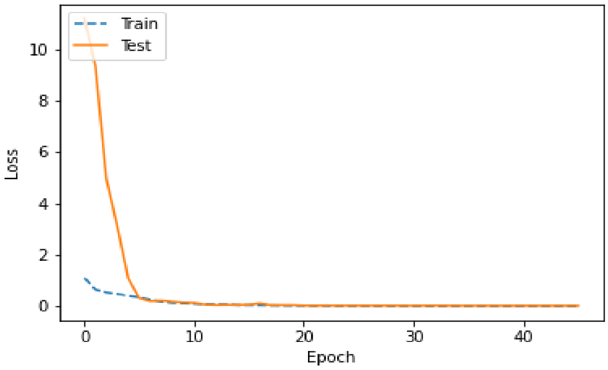

Fig. 9: Model loss

The proposed CNN's test accuracy of 99.92% is also very high, indicating that the model is performing well on the test set. This high accuracy suggests that the model is generalizing well to new data and is not overfitting to the training set. Analysis of Fig. 10 confusion matrix suggests that the proposed CNN model is doing an exceptional job with its designated task. The proposed CNN model has high precision and recall for each class and f1-score of the accuracy of 100%, additionally the proposed CNN's recall or sensitivity and precision scores of 99.58% and specificity of 99.89% as shown in table 1 indicate that it is performing very well in correctly identifying positive cases while minimizing false positives. Furthermore, as shown in fig 11 the model's ROC AUC score of 0.9999 indicates that it can effectively distinguish between positive and negative cases, with very few misclassifications.

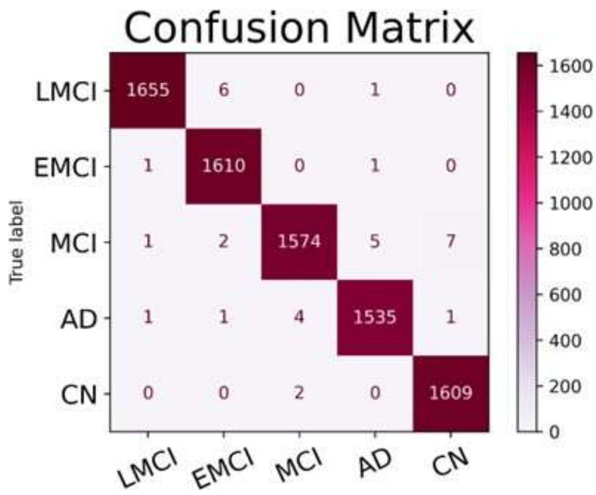

Fig. 10. The proposed model's Confusion

Table 1: The proposed model's classification report

	Precision	Recall	f1-score	Support
LMCI (0)	1.00	1.00	1.00	1662
EMCI (1)	0.99	1.00	1.00	1612
MCI (2)	1.00	0.99	0.99	1589
AD (3)	1.00	1.00	1.00	1542
CN (4)	1.00	1.00	1.00	1611
Accuracy			1.00	8016
Macro avg	1.00	1.00	1.00	8061
Weighted avg	1.00	1.00	1.00	8061

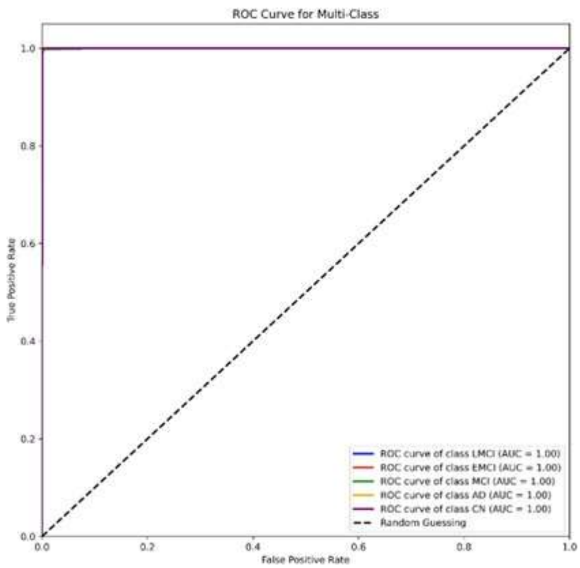

Fig. 11: The proposed model's ROC curve

This is a very impressive score and suggests that the model is highly discriminative. The PPV and NPV for all classes are 1.00, indicating that the classification model has a very high accuracy for both positive and negative predictions. Our network achieved a classification accuracy of 99.94% on

LMCI vs EMCI, 99.87% on LMCI vs MCI, 99.95% on LMCI vs AD, 99.94% on LMCI vs CN, 99.99% on CN vs AD, 99.99% on CN vs EMCI, 99.99% on CN vs MCI, 99.99% on AD vs EMCI, 99.98% on AD vs MCI, and 99.96% on MCI vs EMCI. The high accuracy values indicate that the proposed CNN can effectively differentiate between different stages of the disease. The CNN proposed model, which underwent comprehensive training on a dataset pertaining to Alzheimer's ailment comprising five distinct classes. Upon assessment, it demonstrated an excellent F1 score accuracy of 1.00 as explained in table 1 and a remarkable ROC AUC score of 0.9999 as shown in Fig. 6. Furthermore, it also displayed admirable precision values of 0.9989 along with substantial recall and sensitivity scores of 0.9958. Finally, its resulting test accuracy amounted to an impressive 99.92% of the system as shown in Fig. 8. According to the Fig. 12, Fig. 13, and the confusion matrix in Fig. 14.

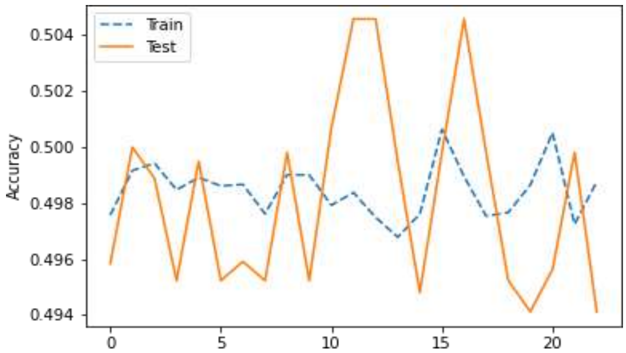

Fig. 12. VGG19 accuracy

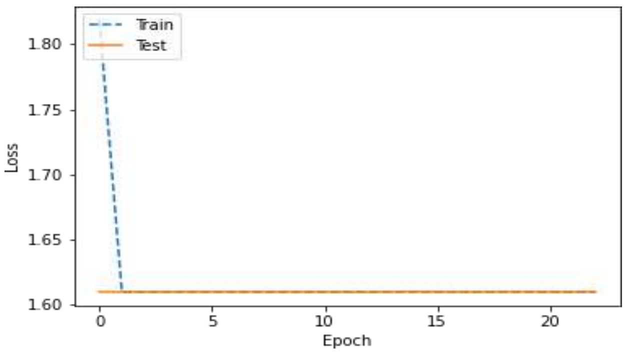

Fig. 13. VGG19 loss

It became rather obvious that the VGG19 model is failing in producing any accurate predictions across any available classes. To be precise, all samples from every class have been incorrectly predicted as belonging to an entirely separate class (class 0). Ultimately this led to a diagnostic diagram featuring values placed only along its main diagonal and corresponding strictly to their respective total sample count within different classes.

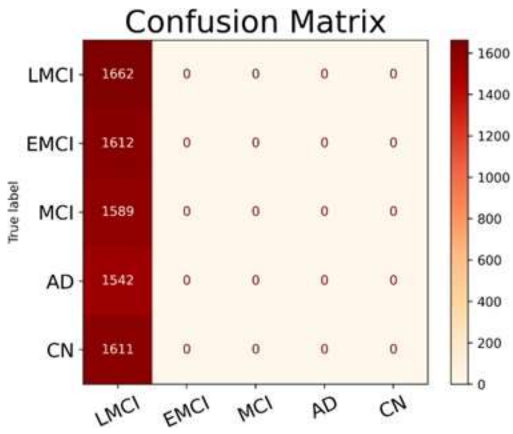

Fig. 14. VGG19 Confusion Matrix

The (PPV) and (NPV) cannot be calculated because all the predicted values for each class are 0, except for the true positives on the diagonal as shown in Fig. 15.

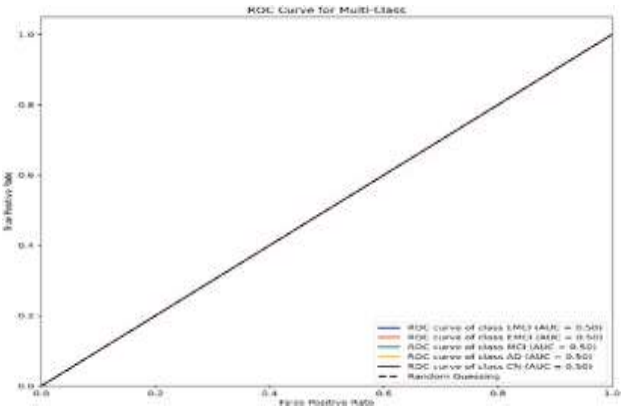

Fig. 15. ROC AUC of VGG19

Based on the evaluation of table 2 classification report surrounding VGG19's capabilities regarding task at hand showcased noticeable deficiencies in performance quality standards specifically relating to several shortcomings during precision-recall assessment tests for all assigned categories incorporating considerably lower than expected f1 scores concerning prediction effectiveness hence underscoring these models' inadequacies furthermore while providing additional insight into its low levels of accuracy observed at a paltry rate of only 21% able to provide accurate forecasts reflecting insufficient predictive capacity resultant from inadequate results witnessed on both macro and weighted average performance evaluation metrics reflecting the model's subpar level of effectiveness across assigned categories with an overarching macro-average f1-score of just 0.07

attesting to this observation.

Table 2: Vgg19 classification report

	Precision	Recall	f1-score	Support
LMCI (0)	0.21	1.00	0.34	1662
EMCI (1)	0.00	0.00	0.00	1612
MCI (2)	0.00	0.00	0.00	1589
AD (3)	0.00	0.00	0.00	1542
CN (4)	0.00	0.00	0.00	1611
Accuracy			0.21	8016
Macro avg	0.04	0.20	0.07	8061
Weighted avg	0.04	0.21	0.07	8061

According to Fig. 16 and Fig. 17, the model learns to recognize patterns in the images, its accuracy improves and its loss decreases.

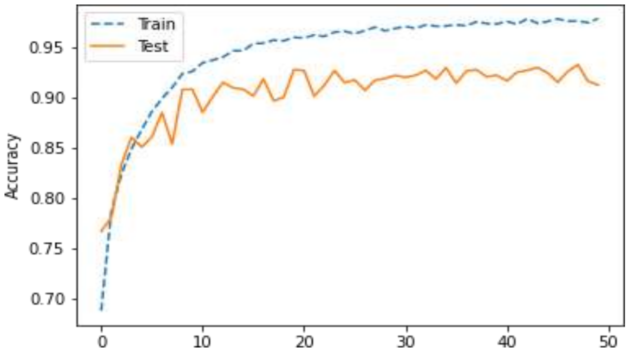

Fig. 16: Model accuracy

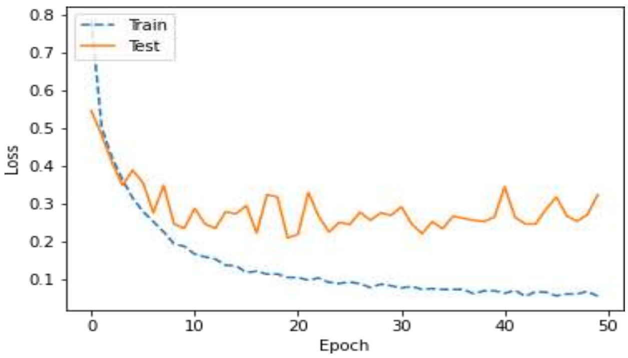

Fig. 17. Model loss

The ResNet50's test accuracy recorded stood at an excellent figure at 91.31% is also extremely high, showing that the model is performing well on the test set with a decreased loss recorded at around 0.323. From Fig, 18, The overall PPV of 0.923 means that out of all the samples that were predicted to be positive, 92.3% of them were positive and correctly classified. On the other hand, the overall NPV of 0.98 means that out of all the samples that were predicted to be negative, 98% of them were negative and correctly classified. These values show that the model has a high degree of accuracy in predicting both positive and negative classes.

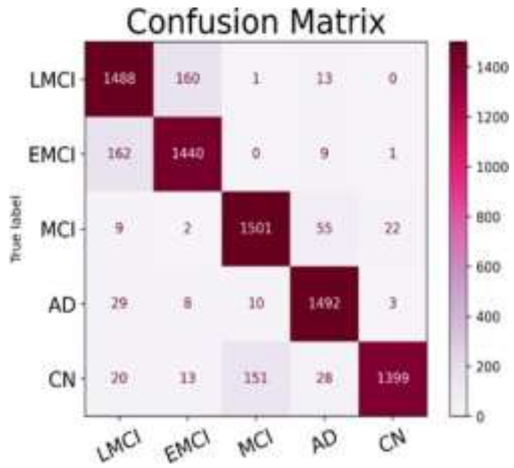

Fig. 18. ResNet50 Confusion Matrix

According to Table 3's classification report, it appears that the ResNet50 outperformed the VGG19 on our current task at hand by a significant margin. Across all categories, values for precision, recall, and f1-score were significantly superior within the ResNet50 model - suggesting its superior accuracy when making forecasts. Moreover, with an accuracy of 91%, it is correct with almost every nine out of ten predictions made by this model implemented correctly! Furthermore, measuring against macro-average and weighted-average metrics across all categories also indicates progressively better results for ResNet50 than for its counterpart - VGG19. Lastly mentioned is a strong indication towards an excellent overall summation characterizing performance results by presenting a Macro- Average F1-Score of 0.91.

Table 3**:** ResNet50 Classification report

	Precision	Recall	f1-score	Support
LMCI (0)	0.87	0.90	0.88	1662
EMCI (1)	0.89	0.89	0.89	1612
MCI (2)	0.90	0.94	0.92	1589
AD (3)	0.93	0.97	0.95	1542
CN (4)	0.98	0.87	0.92	1611
Accuracy			0.91	8016
Macro avg	0.92	0.91	0.91	8061
Weighted avg	0.91	0.91	0.91	8061

According to our findings, while analyzing three models namely, VGG19, ResNet50 and our newly introduced CNN architecture, we discovered that VGG19 had a higher loss value compared to its counterparts indicating lower accuracy in predictions (with an overall loss value of 1.6093) as shown in Fig

19.

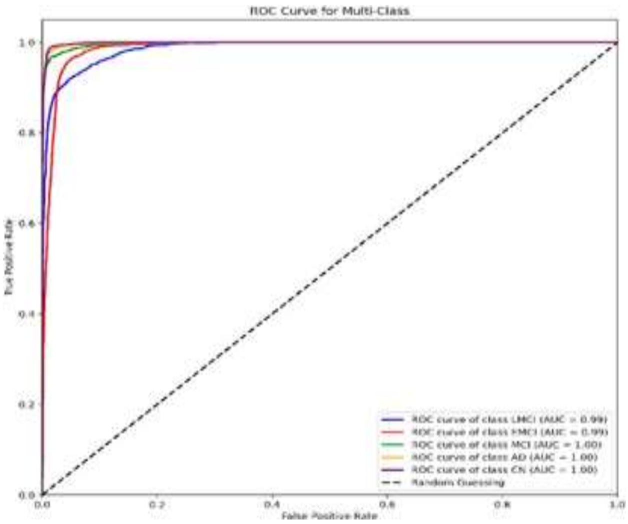

Fig. 19: ResNet50 ROC curve

One possible explanation behind this discrepancy previously observed by computer vision researchers ould stem from the depth and complexity featured within VGG19's design as it has many parameters and hence may be prone to over-fit. On another note, having described fewer complex techniques applied in reducing data variances for regression problems i.e., applying regularization techniques such as embedding dropout layers within modeling architectures amongst others; The use residual connections present in ResNet50 may account for why ResNet50 displayed better forecasting ability with a decreased loss recorded at around 0.323 as shown in Fig. 17. The proposed CNN architecture yielded even better prediction ability with a low-loss value at about 0.006862 as shown in Fig. 9, setting an impressive standard for true precision, hence championing it as a preferred choice for the task at hand. The exceptional performance of the CNN model that has been suggested on the Alzheimer's dataset with five classes can be traced back to its specificity to this dataset. As it has been trained specifically using this dataset, it may have learned unique features that are specific to this task. These acquired features have resulted in its outstanding F1 score accuracy, high ROC AUC score, recall, sensitivity, precision, specificity, and test accuracy. The less-than-ideal results generated by VGG19's classification attempts on Alzheimer's dataset are seemingly rooted in its pre-training on ImageNet dataset. The dissonance regarding class and feature factors between these two datasets meant VGG19 lacked adequate requisite-feature knowledge needed for precise classifications under current five-class diagnosis models. While ResNet50 excelled beyond VGG19, its performance pales when evaluated against that of the proposed CNN counterpart precisely tailored for accommodating a forementioned specifications specific to Alzheimer's dataset with five classes. To sum up, the results from the Alzheimer's dataset with five classes indicate that the CNN model surpasses pre-trained models like VGG19 and ResNet50 in terms of F1 score accuracy, ROC AUC score, recall, sensitivity, precision, specificity, and test accuracy.

4 Conclusion

The proposed CNN model demonstrated exceptional accuracy and sensitivity in detecting AD and exhibited an impeccable test accuracy score of 99.92%. The model's ROC AUC score of 0.9999 indicates that it can effectively distinguish between positive and negative cases of AD. PPV and NPV for all classes are 1.00 which indicate a high accuracy for both positive and negative predictions. The proposed model outperformed pre-trained models VGG19 and ResNet50 in all measures, with a low-loss value and exceptional F1 score accuracy, ROC AUC score, recall rate, precision, specificity, and test accuracy. In contrast, VGG19 showed poor performance with F1 grade accuracy of 0.21, ROC AUC score of 0.5, recall of 0.2, and low accuracy, resulting in a test accuracy of only 20.73%. ResNet50 performed better than VGG19, but the accuracy of the F1 score of 0.91 and the ROC AUC score of 0.99374 were lower than that of the proposed CNN model. The proposed CNN model could recognise the specific brain regions such as the corpus callosum and thalamus which play a significant role in identifying Alzheimer's disease images and acknowledges their importance as feature maps. In general, it can be notice that MRI and CNN models hold promise in improving both the accuracy and efficiency of AD diagnosis. Further exploration in this field may result in the development of a standardized approach to diagnosing AD ultimately benefiting individuals who suffer from it.

References:

[1] Alzheimer's Disease and Dementia. What is Alzheimer's?[Online].Available:https://www.alz.org/alzheimers-dementia/what-is-alzheimers. [Accessed: Dec. 21, 2022].

[2] J. Hardy and D. J. Selkoe," The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics," Science, vol. 297, pp. 353-356, 2002. doi: 10.1126/science.1072994.

[3] J. Cummings, G. Lee, A. Ritter, M. Sabbagh, and K. Zhong," Alzheimer's disease drug

development pipeline: 2019," Alzheimer's Dementia: Translational Research Clinical Interventions, vol. 5, pp. 272-293, 2019. doi: 10.1016/j.trci.2019.05.008.

[4] R. F. Gottesman, A. L. C. Schneider, Y. Zhou, J. Coresh, E. Green, N. Gupta, D. S. Knopman, A. Mintz, A. Rahmim, A. R. Sharrett, L. E. Wagenknecht, D. F. Wong, and T. H. Mosley," Association Between Midlife Vascular Risk Factors and Estimated Brain Amyloid Deposition," JAMA, vol. 317, no. 14, pp. 1443-1450, 2017. doi: 10.1001/jama.2017.3090.

[5] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, C. Ballard, S. Banerjee, A. Burns, J. Cohen-Mansfield, C. Cooper, N. Fox, L. N. Gitlin, R. Howard, H. C. Kales, E. B. Larson, K. Ritchie, K. Rockwood, E. L. Sampson, Q. Samus, L. S. Schneider, G. Selbæk, L. Teri, and N. Mukadam, "Dementia prevention, intervention, and care," The Lancet, vol. 390, no. 10113, pp. 2673-2734, 2017.doi:10.1016/S0140-6736(17)31363-6.

[6] P. B. Rosenberg, M. M. Mielke, B. S. Appleby, E. S. Oh, Y. E. Geda, and C. G. Lyketsos," The association of neuropsychiatric symptoms in MCI with incident dementia and Alzheimer disease," The American Journal of Geriatric Psychiatry, vol. 26, no. 10, pp. 1015-1023, 2018. doi: 10.1016/j.jagp.2018.05.002.

[7] D. W. Shattuck and R. M. Leahy," Automated graph-based analysis and correction of cortical volume topology," IEEE Transactions on Medical Imaging, vol. 20, no. 11, pp. 1167-1177, 2001. doi: 10.1109/42.963819.

[8] S. M. Smith," Fast robust automated brain extraction," Human Brain Mapping, vol. 17, no. 3, pp. 143-155, 2002. doi: 10.1002/hbm.10062.

[9] C. Fennema-Notestine, I. B. Ozyurt, K. A. Clark, S. Morris, A. Bischoff-Grethe, M. W. Bondi, T. L. Jernigan, B. Fischl, F. Segonne, D. W. Shattuck, R. M. Leahy, D. E. Rex, A. W. Toga, K. H. Zou, and G. G. Brown, "Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: Effects of diagnosis, bias correction, and slice location," Human Brain Mapping, vol. 27, no. 2, pp. 99-113, 2006. doi: 10.1002/hbm.20161.

[10] R. Li, W. Zhang, H. I. Suk, L. Wang, and D. Shen," Deep learn- ing based imaging data completion for improved brain disease di- agnosis," Medical Image Analysis, vol. 63, p. 101691, 2020. doi: 10.1016/j.media.2020.101691.

[11] Y. LeCun, Y. Bengio, and G. Hinton," Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015. doi: 10.1038/nature14539.

[12] I. Goodfellow, Y. Bengio, and A. Courville," Deep Learning," MIT Press, 2016.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton," ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems, vol. 25, pp. 1097-1105, 2012.

[14] K. He, X. Zhang, S. Ren, and J. Sun," Deep Residual Learning for Image Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016. doi: 10.1109/CVPR.2016.90.

[15] J. Long, E. Shelhamer, and T. Darrell," Fully Convolutional Networks for Semantic Segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440, 2015. doi: 10.1109/CVPR.2015.7298965.

[16] Y. Kim," Convolutional neural networks for sentence classification," in Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 1746-1751, 2014. doi: 10.3115/v1/D14-1181.

[17] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao," Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, 2017. doi: 10.1109/LSP.2016.2640551.

[18] Y. Zhang, H. Chen, L. Chen, and T. Huang," Joint training of cascaded CNN for face detection," Pattern Recognition, vol. 74, pp. 308-316, 2018. doi: 10.1016/j.patcog.2017.10.005.

[19] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis," Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529-533, 2015. doi: 10.1038/nature14236.

[20] D.R. Sarvamangala and R.V. Kulkarni," Convolutional neural networks in medical image understanding: a survey," Evolutionary Intelligence, vol. 15, pp. 1-22, 2022. DOI: 10.1007/s12065-020-00540-3.

[21] A. Singh," Introduction to neural network: Convolutional Neural Network," Analytics Vidhya, 2020. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/02/mathematics-behind- convolutional-neural-

network/?utmsource = learn − image − classification − cnn − convolutional − neural − networks − 5 − datasets. [Accessed: Dec.21, 2022].

[22] J. Brownlee," Transfer learning in Keras with computer vision models," Machine Learning Mastery, 2020. [Online]. Available: https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models [Accessed: May 25, 2023].

[23] A. Rojatkar," Precision, Recall, Sensitivity and Specificity," OpenGenus IQ: Computing Expertise Legacy, Nov. 4, 2021. [Online]. Available: https://iq.opengenus.org/ precision-recall-sensitivity-specificity/

[24] S. R. Steadman, M. D. Robertson, and M. J. Morrell," Performance Evaluation Metrics for Convolutional Neural Networks: Overview and Systematic Eval- uation," IEEE Access, vol. 7, pp. 122,190-122,205, 2019. doi: 10.1109/AC-CESS.2019.2931563.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.