Automated Alzheimer's disease detection and classification based on optimized deep learning models using MRI

Rashmi Saini[1], Suraj Singh[1], Prabhakar Semwal[2]
[1]Govind Ballabh Pant Institute of Engineering and Technology, Pauri Garhwal, India
[2]Tehri Hydro Development Corporation Institute of Hydropower Engineering and Technology, Tehri Garhwal, India

Article Info

Article history:

Received Jan 10, 2024
Revised Mar 3, 2024
Accepted Apr 13, 2024

Keywords:

Alzheimer's disease detection
Brain MRI
Deep learning
Image classification
InceptionV3
ResNet50
VGG16

ABSTRACT

Alzheimer's disease (AD) is a devastating neurologic condition characterized by brain atrophy and neuronal loss, posing a significant global health challenge. Early detection is paramount to impede its progression. This study aims to construct an optimized deep learning (DL) framework for early AD detection and classification using magnetic resonance images (MRI) scans. The classification task involves distinguishing between four AD stages: mild demented (MD), very mild demented (VmD), moderate demented (MoD), and non-demented (ND). To achieve effective classification, three DL models (VGG16, InceptionV3, and ResNet50) are implemented and fine-tuned. A systematic evaluation is conducted to optimize hyper-parameters, with extensive experimentation. The results demonstrate superior classification performance of the customized DL models compared to state-of-the-art methods. Specifically, visual geometry group 16 (VGG16) achieves the highest accuracy of 95.85%, followed by ResNet50 with 89.38%, while InceptionV3 yields the lowest accuracy of 87.23%. This study highlights the critical role of selecting appropriate DL models and customizing them for accurate AD detection and classification across various stages, offering significant insights for advancing clinical diagnosis and treatment strategies.

Corresponding Author:

Rashmi Saini
Govind Ballabh Pant Institute of Engineering and Technology
Pauri Garhwal, India
Email: 2rashmisaini@gmail.com

1. INTRODUCTION

Alzheimer's disease (AD), a relentless neurodegenerative condition causing gradual cognitive decline, stands as a formidable global health issue, ranking fourth among leading causes of death worldwide [1]. AD is responsible for 60–80% of dementia cases and exerts a significant economic burden on developed nations. As it affects 4–8% of people at the age 65 and poses a 35% risk after 85, AD emphasizes the urgency for focused research and targeted interventions to address its widespread public health implications [1], [2]. The global prevalence of dementia, projected to reach 82 million by 2030 and 152 million by 2050, is increasing, particularly in developing countries. AD, characterized by extensive brain tissue loss and cognitive decline, significantly affects daily functioning, and interpersonal recognition [3].

Researchers suggests that AD may begin up to 20 years before symptoms appear, with subtle brain changes leading to neuron damage. Over time, symptoms like memory loss and language difficulties become noticeable. AD patients typically endure worsening symptoms over years, hindering daily tasks. With no cure, current therapies aim to slow disease progression. To improve patients' quality of life and manage their declining decision-making abilities, more effective interventions are needed [1]-[4].

Valsala and Kariputtaiah [5] employ magnetic resonance images (MRI) scans to detect AD and fuses them with PET scans for enhanced accuracy. The study computes brain matter volumes and ratios, yielding promising outcomes, including a peak signal-to-noise ratio of 60.6 DB and a structural similarity index of 0.8. The disease advances through early, moderate, and late stages, with aggressive behavior and respiratory failure in advanced cases, ultimately leading to mortality [6]. Structural MRI (sMRI) plays a pivotal role in identifying brain damage, aiding in distinguishing AD from other factors for accurate diagnostic assessments [7]. The utilization of convolutional neural networks (CNNs) has expanded in medical imaging, particularly for direct inputs of 2D or 3D images [7], [8].

This study presents an optimized deep learning (DL) framework for early AD detection and classification using MRI, categorizing AD into four stages: mild demented (MD), very mild demented (VmD), moderate demented (MoD), and non-demented (ND). Three DL models-VGG16, InceptionV3, and ResNet50 are implemented and optimized for AD classification. A systematic evaluation is conducted to enhance the models' effectiveness in AD detection and classification. The study aims to build an end-to-end DL model for accurate AD detection and categorization using MRI, optimizing selected DL models (VGG16, ResNet50, and InceptionV3) to improve classification accuracy across MD, VmD, MoD, and ND classes.

In this research endeavor, our primary objective is to address the obstacles associated with AD by leveraging MRI data through the introduction of an optimized DL methodology for early detection of AD. We suggest employing various transfer learning approaches as DL models, such as InceptionV3, ResNet50, and VGG16 with augmentation. The efficacy of these DL models is assessed using a range of metrics, including accuracy, specificity, precision, recall, F1-score, and processing time.

The methodology of our investigation encompasses four fundamental stages. Initially, we acquire a suitable dataset for AD. Subsequently, we pre-process the data, transforming it from unstructured to structured form conducive to classification. Following this, DL model engage in feature extraction and classification of the pre-defined classes. Lastly, we evaluated the performance of these DL models based on pre-defined metrics. Through this evaluation process, we have found the DL model exhibiting superior efficacy in early AD detection.

The research introduces a novel method for early AD detection utilizing DL models. The key contributions highlighting the uniqueness of this approach are elucidated as follows:

- An optimized DL framework tailored for early detection and classification of AD using MRI, addressing the critical need for timely intervention in AD progression.
- Categorization of AD into four stages-MD, VMD, MoD, and ND-enabling finer diagnostic granularity and personalized treatment strategies.
- Implementation and optimization of three DL models-VGG16, InceptionV3, and ResNet50-for AD classification, showcasing advancements in medical imaging analysis.
- Systematic evaluation and customization of DL models through rigorous experimentation, emphasizing the importance of hyper-parameter optimization for improved AD detection.
- Validation of the superiority of the customized DL models over existing methods, with VGG16 demonstrating the highest classification accuracy of 95.85%, followed by ResNet50 at 89.38%, and InceptionV3 at 87.23%.
- Insights into the significance of appropriate model selection and customization for accurate detection and classification of AD across diverse stages, paving the way for improved diagnostic precision and patient care.

2. LITERATURE REVIEW

This section offers a comprehensive literature review, emphasizing the pivotal role of machine learning (ML) and DL in medical research, with a specific focus on AD diagnosis. It underscores the growing integration of advanced DL techniques in different phases of AD identification, particularly in the analysis of medical imaging. Recent researches have witnessed a surge in efforts towards early diagnosis and prognosis of AD, bolstered by advancements in DL techniques [7]. Illustrating this advancement, a study introduced a DL-based method for effectively classifying Alzheimer's and healthy brains, utilizing a CNN model to create a well-trained predictive model for AD diagnosis [6]. A study employing a combination of a stacked autoencoder, softmax regression layer, labeled learning samples has been conducted, aiming to reduce the necessity for extensive prior experience. The evaluation utilized neuroimaging data sourced from the AD neuroimaging initiative (ADNI) database. Results of the study indicate an achievement of 88.58% accuracy for binary classification (utilizing MRI and PET images) and a 47.42% accuracy for 4-classes classification [3].

In another review study of AD diagnosis, Bhatkoti and Paul [9] introduced a novel DL framework aimed at early Alzheimer's diagnosis, emphasizing the effectiveness of a superior k-sparse autoencoder. The study conducted experiments using 150 images from the ADNI dataset, reporting an accuracy of 83.14% with 100 classifiers, surpassing the 71.327% achieved with 50 classifiers. Additionally, the Feng *et al.* [10] proposed a novel method utilizing 3D-CNN and fully stacked bidirectional long short-term memory (FSBi-LSTM) for Alzheimer's diagnosis using MRI and PET data. This approach achieved high accuracies of 86.36% and 65.35% in distinguishing AD and different MCI stages, respectively, outperforming existing algorithms on the ADNI dataset.

Subsequently, Liu *et al.* [11] introduced a multi-model deep CNN framework designed for automatic AD classification. Their method, integrating a deep CNN for segmentation and a 3D DenseNet for feature learning, achieved an accuracy: 88.90% (AD vs. NC) and 76.2% (MCI vs. NC), which is appreciable according to the that time frame. Following that, Ibrahim *et al.* [12] introduced a novel hybrid model merging particle swarm optimization (PSO) with CNNs for AD detection and claimed the successful detection of AD disease in 2022. Authors used an approach based on CNN with transfer learning for AD detection to accuracy enhancement. This study used generative adversarial networks (GANs) to augment training data to obtain the accurate classification of AD disease also utilizing ADNI, Kaggle, and brain tumor datasets, the model achieved over 95% accuracy in AD detection.

A novel approach using CNN, transfer learning, and generative adversarial network (GAN) enhances AD detection accuracy on MRI scans, improving by 2.85–3.88% with TL and 2.43–2.66% with GANs, and outperforming existing methods by 1.8–40.1% Chui *et al.* [13]. Shamrat *et al.* [14] introduces AlzheimerNet, a CNN fine-tuned for classifying Alzheimer's stages from MRI scans. After comparing five pre-trained models, InceptionV3 is selected for further optimization, utilizing an RMSprop optimizer with a learning rate of 0.00001. AlzheimerNet achieves a remarkable 98.67% test accuracy, outperforming conventional methods, thus advancing early diagnosis and treatment approaches for Alzheimer's disease. Additionally, Borkar *et al.* [15] proposed a non-invasive and cost-effective approach using DL on MRI scans in 2023. The model, combining CNN and LSTM with Adam optimization. This study concentrates on a DL-based framework for diagnosing AD stages, utilizing CNN, a prevalent technique in brain image processing. This approach effectively classified well-known four AD classes. While the proposed model achieved an accuracy of 99.70%, there is a possibility of overfitting during training. Nonetheless, this achievement is remarkable.

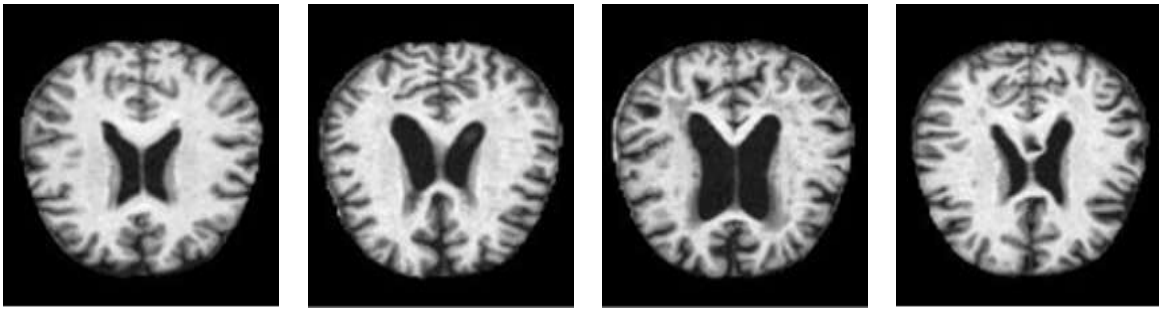
Accurate detection of AD is still a challenging task. The classification of this disease becomes more tedious task when objective is to categorize AD among various class. The literature indicated that various different studies explored a variety of techniques to improve the classification results of AD detection. Therefore, this study explores the potential of robust DL techniques and optimize these models to achieve more accurate classification outcomes.

3. USED DATASET

The extensive dataset employed in this study is Kaggle's Alzheimer's dataset, primarily focusing on the 2D MRI modality. Comprising a total of 6,400 images, this dataset is meticulously organized to intricately capture the nuances of Alzheimer's severity. Each image, presented in a 2D format as shown in Figure 1. The dataset is intelligently categorized into four distinct categories, each representing a specific stage of Alzheimer's progression:

- MD (896 images): this category serves as a snapshot of the early stages of AD, providing crucial insights into the initial manifestations of the disease.
- MoD (64 images): the smallest set in the dataset, this category illustrates a more pronounced and advanced stage of AD, offering a comprehensive view of the disease spectrum.
- ND (3,200 images): as the largest category, this section represents a state of cognitive health, serving as a benchmark for the AD symptoms.
- VMD (2,240 images): encompassing a larger set of images, this segment reflects a more advanced yet still subtle progression of Alzheimer's, capturing the evolving nature of the condition.

In our research, we acknowledge the issue of deterioration of MRI images during data collection, often due to inadequate brightness in the optical equipment, resulting in low variation. To enhance image quality, we have employed image improvement techniques, focusing initially on normalization. The refinement process involves addressing pixel intensity values, thereby mitigating the effects of reducing and impulse noise, ensuring a refined and consistent intensity distribution.



(a) (b) (c) (d)

Figure 1. Showcases MR image samples representing various stages of dementia: (a) ND, (b) MoD, (c) MD, and (d) VmD

This early-stage intervention is pivotal, laying the foundation for subsequent analyses and optimizing our dataset for comprehensive investigation. Normalization process is carried out using (1), in the similar manner with earlier study [6]. To normalize the image, pixel values are adjusted to fall within the range of (-1, 1) by performing a pixel-wise multiplication with a factor of 0.007843 (which corresponds to 1/255), as illustrated in (1).

$$I\hat{N} = \left(1 - \hat{I}^T_{Min}\right) + \frac{\hat{I}^T_{Max}-\hat{I}^T_{Min}}{\text{Max}-\text{Min}}\hat{I}^T_{Miv} \quad (1)$$

Where, I denote the input brain image, and $I\hat{N}$ represent the normalized brain image. Additionally, $\hat{I}^T_{Max}, \hat{I}^T_{M/}$ in Denote intensity of normalized image, in this case, Min=0 and Max=255, and the pixel-wise multiplication factor is 0.007843 as stated earlier, representing 1/255 [6]. Ensuring the robustness of our DL model necessitates sufficient training dataset to prevent overfitting and foster generalization. However, acquiring extensive datasets in medical research encounters challenges, especially in AD research, where obtaining a substantial number of scans is particularly challenging in neuroimaging [4]. Furthermore, dealing with a small imbalanced dataset exacerbates overfitting issues, impacting model efficiency. Consequently, our proposed framework incorporates data augmentation for each available MRI image, generating new images to tackle data availability and class imbalance issues [16].

4. PROPOSED METHOD

The proposed framework's flowchart is depicted in Figure 2, encompasses several processing stages. The initial stage involves pre-processing and data augmentation. Subsequently, the workflow proceeds to the CNN architecture [17], [18]. This comprehensive approach involves several components, including transfer learning, model training, and fine-tuning of parameters, leading up to the classification stage. Further elaboration on each of these stages follows in subsequent sections. In the CNN framework for AD detection training and testing phases are supplied with input MRI data with appropriate samples of AD classes, within the architecture of the CNN model, key components include the input layer, convolution layer, max-pooling layer, and output layer. Convolutions perform essential linear operations between input data and filters, acting as adept feature detectors. These filters undergo training to extract specific information from images, focusing on narrow receptive fields. This strategic design highlights the convolution layer's pivotal role in capturing crucial features inherent in the dataset.

In the CNN model, convolution operation is the key step, which is responsible to extract the feature maps consisting more valuable features contributes led to better training which results more effective classification process. In CNN models, rectified linear unit (ReLU), max pooling, flatten, and dense layers, excels in complex tasks like classification. ReLU is a potent activation function used in many CNN architectures, prevents linearity, enhancing computational efficiency. $f(x) = x \; if \; x >= 0$ or expressed as a piece-wise defined function.

$$\text{f(x)}=\begin{cases} 1 & if \; x > 0 \\ 0 & otherwise \end{cases} \quad (2)$$

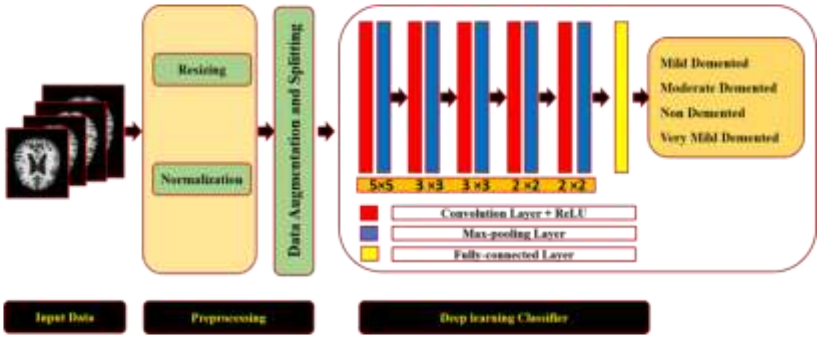


Figure 2. The workflow of the proposed methodology

In the proposed framework for AD classification three DL architectures namely VGG16, ResNet50, and Inception V3 have been customized. VGG16 Developed by Simonyan and Zisserman [19]. This model VGG16 secured second place in the ILSVRC-2014 competition with a 92.7% classification accuracy. It comprises 13 convolutional, 5 max-pooling, and 3 fully connected layers, VGG16 employs a 3×3 filter size and concludes with an output layer containing 1,000 neurons. ResNet50 designed by He *et al.* [20], ResNet50 addresses the vanishing gradient problem with its unique architecture. It features 5 stages, 3 residual blocks in the initial stage, and concludes with an average pooling layer and a fully connected layer containing 1,000 neurons. The third model i.e., Inception V3 developed by Szegedy *et al.* [21] in year 2016, Inception V3 introduced a multi-branch architecture and Inception modules. These modules efficiently process images with varying details using parallel convolutional layers with different filter sizes (1×1, 3×3, and 5×5). Factorized convolutions enhance computational efficiency.

In this study, in order to obtain the best classification outcomes for AD classes, DL models have been customized. Extensive experiments have been conducted to obtain the optimal values of the parameters. Here, three learning rates of 0.01, 0.001, and 0.001 have been tested. In the similar manner two well-known optimizers namely Adam and stochastic gradient decent (SGD) have been tested. In this context, the training process of customized DL models extended to a maximum of 100 epochs, utilizing a learning rate set at 0.0001 and chosen optimizer is Adam. All three DL models are implemented using these specifications.

5. RESULTS AND DISCUSSION

In this study, DL based end-to-end framework has been proposed for the classification of AD using MRI. The experimental configuration demonstrated the performance of the model by utilizing various accuracy parameters. The network undergoes training for 100 epochs. Data is partitioned into 80% as training data, 10% as validation data and 10% as test data. We used Keras library from Tenserflow module in Google Colab, necessitating parallel processing for training deep neural networks. To achieve this, we have utilized the Python 3.11.6 and Kaggle dataset. This experimental design is meticulously crafted to ensure a robust training process and dependable evaluation of the model's performance. The primary objective of model evaluation is to gauge the extent to which a given model generalizes to new data, facilitating effective comparison and analysis among different models. To achieve this, we employed key metrics to quantify the DL model's performance. The accuracy measurements used for the DL model are as follows: the overall accuracy, precision, recall, and F1-score are computed using the confusion matrix, which encompasses various measures and their contributing factors.

The experimental results are thoroughly analyzed concerning the performance metrics. As discussed in the methodology section, different values of learning rate, epochs and optimizers have been tested and final DL models (VGG16, ResNet50, and Inception V3) are customized by adopting the optimal values of parameters. A comprehensive quantitative analysis has been carried out by utilized the confusion matrix, which is obtained by implementing these DL models for AD classification. Confusion matrix offering insights of the misclassifications among various classes. This matrix serves as a base for model evaluation and used by many different studies to assess the performance of model [6], [22]-[26]. Various accuracy metrics, including precision, F1-score, and recall, were calculated based on the confusion matrix, providing a systematic assessment of the classification model. In order to demonstrate the systematic evaluation of implemented all the DL models first of all the training and validation loss plots across 100 epochs have been plotted and resultants graphs are shown in Figure 3. Where, Figure 3(a) represents the variation in training accuracy and loss percentage with respect to the changing the number of epochs for VGG16, Figure 3(b) demonstrating for ResNet50 model, and Figure 3(c) for the InceptionV3 model respectively.

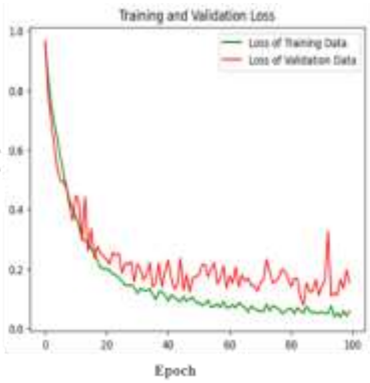
Roughly, it can be observed from all three graphs that there is a consistent trend of improving performance in terms of reducing loss with the increasing the number of epochs during training phase of all three models: VGG16, ResNet50, and Inception V3. For the VGG16 model, initially, the training loss decreases steadily, indicating effective training of the model. However, distinctive patterns emerge in the validation loss. Furthermore, for VGG16 model it can be observing that the training and validation loss closely align during the initial epochs, however, a subtle divergence occurs in the later stages (epochs 80-100) Figure 3(a). It can be seen that the validation loss slightly increases compared to the training loss.

In case of ResNet50 model, a different pattern has been observed. While both training and validation losses decreased initially, there is a noticeable increase in validation loss during the 60 to 80 epochs and again there is a decline in validation loss after 80 epochs Figure 3(b). On the other hand, it can be clearly seen that there is a significant impact of decreasing in training loss as we increase the epochs. However, it can be seen from the graph that training loss becomes stable at nearly 90 to 100 epochs. In contrast InceptionV3 exhibits a higher validation loss as compared to both VGG16 and ResNet50 model throughout the 100 epochs Figure 3(c). This suggests that InceptionV3 faces challenges in generalizing to the validation data. It can also be observed from the graph of InceptionV3 that there is a significant rise in training loss at initial epochs as compared training loss plot of VGG16 and ResNet50. It is also noted that during 90 to 100 epochs this model's training loss significantly decline. This pattern demonstrates the need to train the models for sufficient number of epochs. These distinctions in training and validation loss across the models provide valuable insights into their performance and potential areas for optimization [8], [10].

The resultant training and validation accuracy plots spanning over 100 epochs for all three models i.e., VGG16, ResNet50 and InceptionV3 are shown in Figures 4 respectively. Roughly, a consistent pattern has been observed across VGG-16, ResNet50, and InceptionV3. Initially, both training and validation accuracy closely follow each other, demonstrating effective learning during the early stages shown in Figure 4. For VGG16 model, this alignment continues throughout the entire 100 epochs. This indicates a stable and consistent learning process with minimal overfitting, as both the training and validation accuracy remain close Figure 4(a).

In case of ResNet50 model, a similar pattern has been observed for the initial 20 epochs, where, training and validation accuracies are closely aligned. However, a notable divergence occurs after 20 epochs, where the training accuracy continues to increase while the validation accuracy lags behind as the epochs increases. This widening gap indicated the overfitting of the model and complexities of the training samples of various classes Figure 4(b). Third DL model i.e., InceptionV3 exhibits a comparable pattern to ResNet50 model, with close alignment between training and validation accuracy for the initial 20 epochs. However, in the later stages (epochs 80-100), a noticeable difference emerges, with the training accuracy touching and validation accuracy reaching. Such changes in the graphs demonstrates the challenge in generalization process during the training phase of the DL models Figure 4(c). The observed findings in these accuracy and loss plots offer valuable insights of the models' learning dynamics and generalization capabilities, providing essential information for model evaluation and potential optimization. The trends observed in the training data accuracy and loss plots, notably the consistent training accuracy between 85% and 96% for VGG16, ResNet50, and Inception V3, are prominently depicted in Figures 3 and 4. These figures vividly illustrate the stable learning patterns of the models throughout the 100 epochs, showcasing training accuracies within the mentioned range. The associated loss plots, maintained between 4% to 15% across all three models, further highlight the steady optimization and learning dynamics observed in the training phase. This consistency in accuracy and loss patterns, as visibly depicted in Figures 3 and 4, emphasizes the models' resilience in maintaining effective learning and model refinement over the training epochs. The confusion matrix for the VGG16 model is presented in Table 1. Results indicated that VGG16 model obtained a classification accuracy of 95.84% F1-score for various classes are as follows: for MD 94.44% (precision=93.40%, recall=95.52%), 100% for MoD (precision=100%, recall=100%), 96.45% for ND (precision=96.01%, recall=96.90%), and 95.40% for VmD (precision=96.46%, recall=94.37%). It is found that F1-score exceeded 95%, for all AD classes emphasizing its robust in discriminatory capabilities.

ResNet50, known for its innovative residual blocks, demonstrated remarkable performance in various studies [16], [20], [22], [24]. The accuracy measures obtained for ResNet50 model has shown in Table 1. Results indicated that ResNet50 model obtained a classification accuracy of 89.38%. The obtained F1-score for various classes is as follows: for MD 81.48% (precision=77.01%, recall=86.51%), 92.31% for MoD (precision=100%, recall=85.71%), 91.39% for ND (precision=92.41%, recall=90.40%), and 89.76% for VmD (precision=90.35%, recall=89.18%). It can be seen that this model shows more misclassification among various classes as compared to VGG16 model. It is also found that obtained F1-score for all the classes are less as compared to the results of VGG16 model.

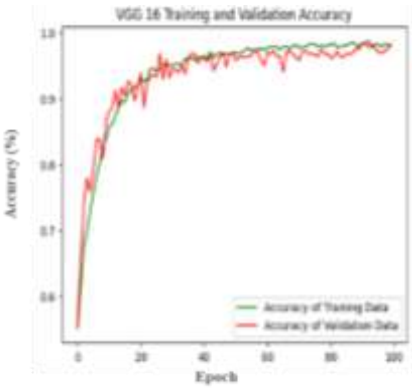


(a)

(b)

(c)

Figure 3. Training and validation loss against the 100 epochs: (a) VGG16, (b) ResNet50, and (c) InceptionV3



(a)

(b)

(c)

Figure 4. Training, validation accuracy with the 100 epochs: (a) VGG16, (b) ResNet 50, and (c) Inception V3

Inception V3, distinguished by its multi-branch architecture. The model has been utilized for various applications in the domain of medical image processing [6], [21]-[23]. It achieved an overall accuracy of 87.23% see in Table 1. It's noteworthy that for the AD classification problem, this model demonstrated the lowest accuracy compared to the VGG16 and ResNet50 models. The obtained F1-score for various classes is as follows: for MD 77.72% (precision=72.12%, recall=84.27%), 83.33% for MoD (precision=100%, recall=71.42%), 89.94% for ND (precision=91.37%, recall=88.54%), and 89.94% for VmD (precision=88.15%, recall=87.58%). In this study, VGG16 reported as the top performer among all DL model, followed by ResNet50 and Inception V3. For in-depth evaluation of the customized DL models, a comparative analysis was conducted with previous studies carried out for AD classification based on CNN and ML classifiers. Table 2 presents a comparison of classification accuracies, F1-score, precision, and recall. showcasing the superiority of customized models.

Table 1. Evaluation of confusion matrix for the customized VGG16, ResNet50 and InceptionV3 model. All the accuracy measures are shown in percentage

VGG16			ResNet50			InceptionV3		
Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
93.40	95.52	94.44	77.01	86.51	81.48	72.12	84.27	77.72
100	100	100	100	85.71	92.31	100	71.42	83.33
96.01	96.90	96.45	92.41	90.40	91.39	91.37	88.54	89.94
96.46	94.37	95.40	90.35	89.18	89.76	88.15	87.01	87.58
OA=95.84			OA=89.38			OA=87.23		

The comparison with existing studies in AD detection reveals a significant contribution of the proposed framework, by customizing VGG16, ResNet50, and InceptionV3 models. A summary of previous research work is demonstrated in Table 2. Hussain *et al.* [22], used OASIS dataset for AD detection by employing VGG16 for binary classification and achieved an accuracy of nearly 50%. Whereas, Cui *et al.* [23], targeted MCI detection with InceptionV3 and achieved an accuracy of 85.7%. Previous research work [16] employed VGG16 and ResNet50 for SMC, MCI, EMC classification and obtained 78.84% (VGG-16) and 80.98% (ResNet50) accuracy. Shehri [24] used ResNet50 for ND, VmD, MD, MaD classification and reported an accuracy of 81.92%. Jo *et al.* [25] used a CNN-based classifier and demonstrated an accuracy <90% for MCI detection. Moreover, Ajagbe *et al.* [26] utilized VGG16 and CNN for ND, MoD, MD, VmD classification and obtained an accuracy of 77.04% and 71.02% respectively. One another recent research Inan *et al.* [27] in 2024 proposed architecture using benchmark datasets from ADNI and OASIS, demonstrating superior performance in Alzheimer's classification. Significant results have been achieved, including cross-validation accuracy of 83.64% for CN vs AD, 82.69% for CN vs MCIc, and 71.40% for CN vs MCInc on the ADNI dataset. Additionally, an accuracy of 91.54% for CN vs AD has been attained on the OASIS dataset. One more study by Allada *et al.* [28] proposed a novel approach in 2024 for AD classification, CSCOOT_CNN, combines competitive swarm coot optimizations with CNNs trained via transfer learning. Utilizing ADNI photographs, pre-processing, feature extraction, and CNN classification yield improved results, with 92.6% accuracy, 97.9% specificity, and 90.9% sensitivity. Remarkably, our models performed multiclass classification and achieved higher accuracy (95.85% (VGG16), 89.38% (ResNet-50), and 87.23% (InceptionV3)) and robust performance across precision, recall, and F1-scores for testing and training. The results of our study clearly demonstrated the superior performance of our customized DL models for accurate AD classification.

Table 2. Comparison of the proposed work in this study with state of arts methods

Ref.	Year	Dataset	Model	AD classes	Accuracy	Precision	Recall	F1-score
[14]	2023	ANDI	VGG16 and ResNet50	SMC, MCI, EMC	78.84 and 80.98	- -	- -	- -
[20]	2020	OASIS	VGG16	Binary classification	50%	25	50	50
[21]	2019	ANDI	InceptionV3	MCI	85.7	- -	- -	- -
[22]	2022	ANDI	ResNet50	ND, VMD, MD, MaD	81.92	- -	- -	- -
[23]	2020	ANDI	CNN-based classifier	MCI	<90%	- -	- -	- -
[24]	2021	ANDI	VGG16, CNN	ND, MoD. MD, VmD	77.04 and 71.02	57.08 and 50.04	38.78 and 50.09	81.22 and 78.06
Proposed models	2024	ANDI	VGG16, ResNet 50, Inception V3	ND, MoD MD and VmD	95.85, 89.38, and 87.23	96.25, 89.75, and 87.75	96.75, 88, 82.75	96.25, 88.5, and 85.75

6. CONCLUSION

This research work developed an end-to-end framework for AD detection by employing VGG16, ResNet50, and InceptionV3. This study emphasized on the multi-class classification of AD in the following categories: MD, VmD, MoD, and ND. Results indicated that VGG16 emerged as the top performer with an accuracy of 95.85%, which is 6.47% higher than ResNet50, and 8.62% higher than InceptionV3. It is found that VGG-16 model able to achieve the F1-score greater than 95%. It has been observed that MD class is classified with lowest F1-score by all three DL models. Results indicated that a well-trained DL model exhibits excellent classification results. Outcome of this study indicated that a significant improvement can be achieved in the accuracy of DL model by selecting the optimal values of the parameters. These findings will serve as a guide to researchers and practitioners with insightful distinctions and findings of these models. The in-depth quantitative evaluations, alongside comparisons with state-of-the-art results, affirm the effectiveness of the proposed framework for Alzheimer's detection and classification using brain MR image.

REFERENCES

[1] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review," *Computer Methods and Programs in Biomedicine*, vol. 187, p. 105242, Apr. 2020, doi: 10.1016/j.cmpb.2019.105242.

[2] G. S. Babu, S. N. T. Rao, and R. R. Rao, "Automated assessment for Alzheimer's disease diagnosis from MRI images: meta-heuristic assisted deep learning model," *International Journal of Imaging Systems and Technology*, vol. 32, no. 2, pp. 544–563, Mar. 2022, doi: 10.1002/ima.22650.

[3] F. Zhang, Z. Li, B. Zhang, H. Du, B. Wang, and X. Zhang, "Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease," *Neurocomputing*, vol. 361, pp. 185–195, Oct. 2019, doi: 10.1016/j.neucom.2019.04.093.

[4] S. E. Sorour, A. A. A. El-Mageed, K. M. Albarrak, A. K. Alnaim, A. A. Wafa, and E. El-Shafeiy, "Classification of Alzheimer's disease using MRI data based on deep learning techniques," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 2, p. 101940, Feb. 2024, doi: 10.1016/j.jksuci.2024.101940.

[5] B. Valsala and K. H. Kariputtaiah, "Alzheimer's detection through neuro imaging and subsequent fusion for clinical diagnosis," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, pp. 1491–1498, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1491-1498.

[6] M. EL-Geneedy, H. E. D. Moustafa, F. Khalifa, H. Khater, and E. AbdElhalim, "An MRI-based deep learning approach for accurate detection of Alzheimer's disease," *Alexandria Engineering Journal*, vol. 63, pp. 211–221, Jan. 2023, doi: 10.1016/j.aej.2022.07.062.

[7] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, and V. Calhoun, "Deep residual learning for neuroimaging: an application to predict progression to Alzheimer's disease," *Journal of Neuroscience Methods*, vol. 339, p. 108701, Jun. 2020, doi: 10.1016/j.jneumeth.2020.108701.

[8] R. Saini, P. Semwal, and T. H. Jaware, "Brain tumor classification using VGG-16 and MobileNetV2 deep learning techniques on magnetic resonance images (MRI)," in *Communications in Computer and Information Science*, vol. 1788 CCIS, 2023, pp. 300–313.

[9] P. Bhatkoti and M. Paul, "Early diagnosis of Alzheimer's disease: a multi-class deep learning framework with modified k-sparse autoencoder classification," in *2016 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Nov. 2016, vol. 0, pp. 1–5, doi: 10.1109/IVCNZ.2016.7804459.

[10] C. Feng *et al.*, "Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM," *IEEE Access*, vol. 7, pp. 63605–63618, 2019, doi: 10.1109/ACCESS.2019.2913847.

[11] M. Liu *et al.*, "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *NeuroImage*, vol. 208, p. 116459, Mar. 2020, doi: 10.1016/j.neuroimage.2019.116459.

[12] R. Ibrahim, R. Ghnemat, and Q. A. Al-Haija, "Improving Alzheimer's disease and brain tumor detection using deep learning with particle swarm optimization," *AI (Switzerland)*, vol. 4, no. 3, pp. 551–573, Jul. 2023, doi: 10.3390/ai4030030.

[13] K. T. Chui, B. B. Gupta, W. Alhalabi, and F. S. Alzahrani, "An MRI scans-based Alzheimer's disease detection via convolutional neural network and transfer learning," *Diagnostics*, vol. 12, no. 7, p. 1531, Jun. 2022, doi: 10.3390/diagnostics12071531.

[14] F. M. J. M. Shamrat *et al.*, "AlzheimerNet: an effective deep learning based proposition for Alzheimer's disease stages classification from functional brain changes in magnetic resonance images," *IEEE Access*, vol. 11, pp. 16376–16395, 2023, doi: 10.1109/ACCESS.2023.3244952.

[15] P. Borkar, V. A. Wankhede, D. T. Mane, S. Limkar, J. V. N. Ramesh, and S. N. Ajani, "Deep learning and image processing-based early detection of Alzheimer disease in cognitively normal individuals," *Soft Computing*, Jun. 2023, doi: 10.1007/s00500-023-08615-w.

[16] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3573–3587, Aug. 2018, doi: 10.1109/TNNLS.2017.2732482.

[17] P. Semwal and R. Saini, "Deep-learning models for COVID-19 detection using chest X-Ray images," in *2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development, OTCON 2022*, Feb. 2023, pp. 1–6, doi: 10.1109/OTCON56053.2023.10113965.

[18] M. M. Attia, N. F. F. Areed, H. M. Amer, and M. El-Seddek, "A deep learning framework for accurate diagnosis of colorectal cancer using histological images," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 2, p. 2167, Apr. 2024, doi: 10.11591/ijece.v14i2.pp2167-2180.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint, 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-December, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[22] E. Hussain, M. Hasan, S. Z. Hassan, T. Hassan Azmi, M. A. Rahman, and M. Z. Parvez, "Deep learning based binary classification for Alzheimer's disease detection using brain MRI images," in *Proceedings of the 15th IEEE Conference on Industrial Electronics and Applications, ICIEA 2020*, Nov. 2020, pp. 1115–1120, doi: 10.1109/ICIEA48937.2020.9248213.

[23] Z. Cui, Z. Gao, J. Leng, T. Zhang, P. Quan, and W. Zhao, "Alzheimer's disease diagnosis using enhanced inception network based on brain magnetic resonance image," in *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, Nov. 2019, pp. 2324–2330, doi: 10.1109/BIBM47256.2019.8983046.

[24] W. Al Shehri, "Alzheimer's disease diagnosis and classification using deep learning techniques," *PeerJ Computer Science*, vol. 8, p. e1177, Dec. 2022, doi: 10.7717/PEERJ-CS.1177.

[25] T. Jo, K. Nho, S. L. Risacher, and A. J. Saykin, "Deep learning detection of informative features in tau PET for Alzheimer's disease classification," *BMC Bioinformatics*, vol. 21, no. S21, p. 496, Dec. 2020, doi: 10.1186/s12859-020-03848-0.

[26] S. A. Ajagbe, K. A. Amuda, M. A. Oladipupo, O. F. AFE, and K. I. Okesola, "Multi-classification of alzheimer disease on magnetic resonance images (MRI) using deep convolutional neural network (DCNN) approaches," *International Journal of Advanced Computer Research*, vol. 11, no. 53, pp. 51–60, Mar. 2021, doi: 10.19101/ijacr.2021.1152001.

[27] M. S. K. Inan *et al.*, "A slice selection guided deep integrated pipeline for Alzheimer's prediction from structural brain MRI," *Biomedical Signal Processing and Control*, vol. 89, p. 105773, Mar. 2024, doi: 10.1016/j.bspc.2023.105773.

[28] A. Allada, R. Bhavani, K. Chaduvula, and R. Priya, "CSCOOT: competitive swarm coot optimization-based CNN transfer learning for Alzheimer's disease classification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 7s, pp. 337–349, 2024.

BIOGRAPHIES OF AUTHORS

Dr. Rashmi Saini is currently an Associate Professor in the Department of Computer Science and Engineering (CSE) at G. B. Pant Institute of Engineering and Technology, Pauri Garhwal, Uttarakhand. She received M. Tech and B.Tech. degrees in computer science and Engineering from the National Institute of Technical Teachers Training and Research, Chandigarh, and Guru Gobind Singh Indraprastha University Delhi respectively. She has received her Ph.D. degree from Indian Institute of Technology (IIT) Roorkee. She is the recipient of the ISPRS Foundation (TIF) Grant to attend the ISPRS Technical Commission III Symposium in Beijing, China. She has worked for various international conferences as a PC member and session chair. Her field of interest is medical image processing (MIP), object-based image analysis (OBIA), remote sensing, ML, and DL classification techniques. She can be contacted at email: 2rashmisaini@gmail.com.

Suraj Singh a skilled professional in Computer Science and Engineering, earned his B.Tech and M.Tech degrees from Uttarakhand Technical University. Presently, he serves as an Assistant Professor at Govind Ballabh Pant Institute of Engineering and Technology in Pauri, Uttarakhand. His expertise spans across MIP, ML, DL, and remote sensing. His passion for leveraging cutting-edge technologies to address real-world challenges, particularly in healthcare applications and environmental monitoring, underscores his valuable contributions to the field. As a vital member of the academic community, Suraj Singh continues to leave a lasting impact in the realms of computer science and engineering. He can be contacted at email: surajfarswan0@gmail.com.

Prabhakar Semwal an Assistant Professor at THDC Institute of Hydropower Engineering and Technology (THDC-IHET), possesses a commendable three years of experience in research fields such as DL and artificial intelligence. His expertise likely encompasses a range of topics within these domains, including neural networks, ML algorithms, natural language processing, computer vision, and possibly other advanced AI techniques. He can be contacted at email: prabhakarsemwal1744@gmail.com.