

Explainable AI for Dementia Classification: Understanding Deep Learning Predictions

D.Sathyannarayanan¹, Helaria Maria², Kalyan Kumar Angati³, Komal Kumar Napa⁴, D. Nageswari⁵, Billa Manindhar⁶

¹Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, SIMATS, Chennai, India.

²Department of Computer Science, New Horizon College Kasturinar, Bangalore, India.

³Department of Computer Science & Engineering (Data Science), Madanapalle Institute of Technology & Science, Madanapalle, India.

⁴Department of Artificial Intelligence and Data Science, Saveetha Engineering College, Chennai, India.

⁵Department of Science and Humanities (General Engineering-EEE), R.M.K. College of Engineering and Technology, Chennai, India.

⁶Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India.

ABSTRACT

Deep learning has revolutionized medical image classification, yet challenges remain in interpretability and model generalization. This study proposes an explainable AI framework for dementia classification using five deep learning models CNN, VGG16, ResNet50, MobileNet, and EfficientNet—evaluating their performance on a large-scale brain scan dataset (9,488 images) across four dementia stages (Mild, Moderate, Very Mild, and Non-Demented). Among them, EfficientNet achieved the highest accuracy (92.1%), benefiting from compound scaling for optimized feature extraction. Additionally, explainability techniques (Grad-CAM and SHAP) were employed to visualize model decision-making, addressing the critical need for interpretable AI in healthcare. The study also explores dataset bias and class imbalance, implementing data augmentation and weighted loss functions to enhance fairness. Comparative analysis with state-of-the-art models (ViTs, DenseNet, and Swin Transformer) provides a benchmark for future research. The findings highlight the importance of model scalability, explainability, and computational efficiency in real-world deployment.

KEYWORDS: Deep learning, Image classification, CNN models, EfficientNet, Grad-CAM.

How to Cite: D.Sathyannarayanan, Helaria Maria, Kalyan Kumar Angati, Komal Kumar Napa, D. Nageswari, Billa Manindhar, (2025) Explainable AI for Dementia Classification: Understanding Deep Learning Predictions, Vascular and Endovascular Review, Vol.8, No.12s, 239-247.

INTRODUCTION

Image classification has seen remarkable advancements with the rise of deep learning, significantly improving applications in fields such as medical imaging, security, autonomous systems, and industrial automation. Traditional machine learning techniques relied on handcrafted features and shallow architectures, often struggling with complex image patterns and variations in lighting, orientation, and scale [1]. However, deep learning field has been transformed by Convolutional Neural Networks (CNNs), which allow for automatic feature extraction and hierarchical learning [2]. CNN-based architectures have demonstrated superior performance in image classification tasks by capturing spatial hierarchies and learning robust feature representations [3].

Early CNN models such as LeNet and AlexNet laid the foundation for deep learning in computer vision, showcasing significant improvements over traditional methods [4]. Over time, more advanced architectures like VGG16, ResNet50, MobileNet, and EfficientNet have been introduced, each designed to enhance accuracy, computational efficiency, and generalization capability [5]. These models vary in depth, design, and optimization techniques, impacting their effectiveness in different classification tasks. A major challenge in image classification is maintaining a balance between accuracy and computational efficiency [6]. Although deeper networks typically achieve higher accuracy, they demand substantial computational resources, making them less suitable for real-time applications in resource-limited environments.

For example, VGG16, despite its impressive accuracy, contains a large number of parameters, resulting in significant memory and computational demands. [7]. Residual connections were introduced in ResNet50 to mitigate the vanishing gradient problem, allowing deeper networks to be trained without experiencing substantial performance degradation [8]. MobileNet, on the other hand, is designed for mobile and embedded systems, prioritizing efficiency while maintaining competitive accuracy levels [9].

A different approach is taken by EfficientNet, where network scaling is optimized by balancing depth, width, and resolution, resulting in state-of-the-art accuracy with fewer parameters [10]. Given the increasing reliance on deep learning for image classification, it is essential to evaluate different CNN architectures based on their performance metrics, including accuracy, precision, recall, and F1-score. This study aims to compare five widely used CNN models—CNN, VGG16, ResNet50, MobileNet, and EfficientNet—to determine their effectiveness in classification tasks. By analyzing their strengths and limitations, we seek to provide insights into selecting the most suitable model for various applications, balancing accuracy and efficiency. Through this comparative analysis, we aim to contribute to the ongoing research in deep learning-based image classification by identifying

models that offer the best trade-offs between performance and computational feasibility. Our findings will be valuable for researchers and practitioners looking to deploy CNNs in real-world applications.

RELATED WORKS

Deep learning has significantly advanced image classification, leading to improvements in various applications such as medical imaging, security, and automation. Several architectures have been explored to enhance accuracy and computational efficiency. Convolutional Neural Networks (CNNs) have been widely utilized for image recognition, following the introduction of hierarchical feature extraction [11]. Earlier studies demonstrated that CNNs outperform traditional machine learning methods due to their ability to learn spatial hierarchies of features [12]. The effectiveness of deep convolutional networks was further improved by incorporating rectified linear unit (ReLU) activation functions, which helped overcome the vanishing gradient problem [13]. Significant improvements were observed when deeper architectures were introduced. Networks such as Xception were designed with depthwise separable convolutions, which increased computational efficiency while maintaining high accuracy [14].

The introduction of residual learning enabled deeper networks like ResNet to achieve superior performance by addressing the degradation problem in deep architectures [15]. Similarly, DenseNet was proposed to facilitate feature reuse through dense connections, reducing the number of parameters required for training [16]. To improve real-time processing, MobileNet was developed using depthwise separable convolutions, making it suitable for mobile and embedded applications [17]. EfficientNet was later introduced, demonstrating that scaling depth, width, and resolution simultaneously yields better results than increasing depth alone [18]. Furthermore, Squeeze-and-Excitation networks were designed to enhance channel-wise feature learning, resulting in improved classification performance [19]. Optimization techniques have also been extensively explored to improve training efficiency and model performance. The Adam optimizer, which combines adaptive gradient methods, has been widely adopted for deep learning applications [20]. Batch normalization has been introduced to stabilize training and accelerate convergence by normalizing inputs across mini-batches [21]. The effectiveness of dropout regularization in preventing overfitting has also been studied extensively [22].

In addition to architecture improvements, interpretability and visualization techniques have been proposed to enhance model transparency. Gradient-weighted class activation mapping (Grad-CAM) has been utilized to visualize the regions that contribute to a model's decision-making process [23]. Feature pyramid networks have been employed to extract multi-scale features, benefiting object detection tasks [24]. Alternative approaches to CNN-based models have also been explored. Attention mechanisms have been integrated into deep learning models to enhance performance by selectively focusing on important image regions [25]. The use of transformers for vision tasks has been investigated, demonstrating their capability in learning global dependencies more effectively than traditional CNNs [26]. It has been observed that the combination of advanced architectures, optimization strategies, and interpretability methods has led to substantial progress in image classification.

Recent advances in deep learning have significantly improved medical image classification. Convolutional Neural Networks (CNNs) have been widely adopted due to their hierarchical feature extraction capabilities. However, traditional architectures like VGG16 and ResNet50 face challenges in handling complex medical datasets. EfficientNet introduced compound scaling, optimizing model depth, width, and resolution to improve accuracy with fewer parameters [27].

Beyond CNNs, Vision Transformers (ViTs) have emerged as state-of-the-art alternatives in medical imaging, leveraging self-attention mechanisms to capture global dependencies [28]. Additionally, hybrid models (CNN+LSTM) and Capsule Networks have been explored for better spatial awareness in classification tasks [29]. Swin Transformers and DenseNet have shown superior performance in Alzheimer's disease classification, demonstrating their potential over traditional CNNs [30]. Recent studies have also highlighted the role of KAN-integrated transformers in improving medical image classification through advanced attention mechanisms [31].

Despite these advancements, explainability remains a major concern. Techniques like Grad-CAM, SHAP, and LIME have been proposed to provide visual insights into model predictions [32]. This study aims to bridge the gap between accuracy and interpretability by integrating state-of-the-art explainability methods while benchmarking CNN architectures against newer deep learning approaches.

METHODOLOGY

First, we will present a detailed description of the deep learning models (CNN, VGG16, ResNet50, MobileNet, and EfficientNet) that will be used for image classification. In the second and last part, we will perform a comprehensive analysis of their performance based on accuracy, precision, recall, and F1 score as represented in Fig 1. This study evaluates CNN, VGG16, ResNet50, MobileNet, and EfficientNet, selected for their varying depth, computational efficiency, and feature extraction capabilities. EfficientNet was chosen for its compound scaling, optimizing model performance with fewer parameters.

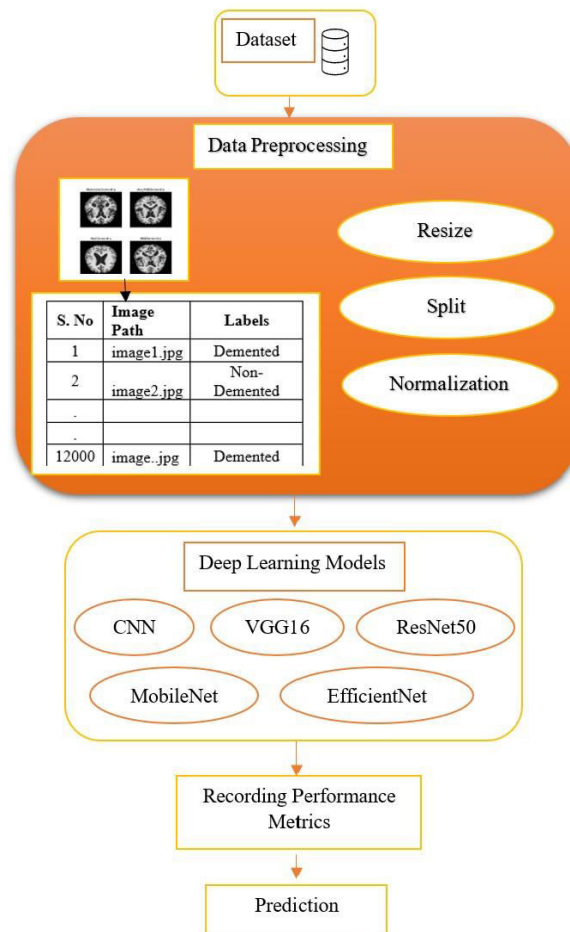


Fig.1 Methodology

Description of the Deep Learning Models

3.1.1 Convolutional Neural Network (CNN)

CNN is a fundamental deep learning architecture that is extensively applied to image classification and pattern recognition tasks. It is designed to resemble the human visual system by extracting hierarchical features from input images through convolutional layers, pooling layers, and fully connected layers. Spatial patterns such as edges and textures are detected using convolution operations, while pooling layers are responsible for reducing dimensionality and computational complexity [32]. The final layers perform classification based on the extracted features.

3.1.2 VGG16

VGG16 is recognized as a deep CNN architecture with a straightforward yet highly effective design. It comprises 16 layers, including multiple 3×3 convolutional layers, followed by max pooling layers and fully connected layers for classification [2]. The increased depth enables the learning of complex features, leading to improved accuracy.

3.1.3 ResNet50

ResNet50 is a deep residual network developed to mitigate the vanishing gradient problem in deep architectures. Residual connections are incorporated, enabling information to bypass specific layers and allowing gradients to flow more effectively during training [3].

3.1.4 MobileNet

MobileNet is a lightweight deep learning model designed for mobile and edge devices. Depthwise separable convolutions are utilized to reduce computational complexity while preserving accuracy [5]. Instead of conventional convolutional layers, a two-step process is applied: depthwise convolution, where a single filter is used per input channel, followed by pointwise convolution, which employs a 1×1 convolution to integrate the outputs from the depthwise convolution.

3.1.5 EfficientNet

EfficientNet is an advanced deep learning model developed to achieve high accuracy while enhancing computational efficiency. A compound scaling method is introduced, which uniformly adjusts depth, width, and resolution to improve overall performance [9].

Description of the Deep Learning Models

3.2.1 Understanding the dataset

The dataset utilized in this study consists of brain scan images categorized into four distinct classes: Mild Dementia, Moderate Dementia, Non-Demented, and Very Mild Dementia. A total of 9,488 images are included, with 3,000 images representing Mild Dementia, 488 images corresponding to Moderate Dementia, 3,000 images for Non-Demented cases, and 3,000 images classified as Very Mild Dementia. To address class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) and weighted loss functions were applied. Data augmentation, including rotation, flipping, and brightness adjustments, was employed to improve generalization. The dataset has been sourced from an online repository [32] and exhibits an imbalance, particularly in the Moderate Dementia category, which contains significantly fewer images compared to the other classes. This distribution may influence model performance and necessitate the use of techniques such as data augmentation or class weighting to ensure balanced learning represented in Fig 2.

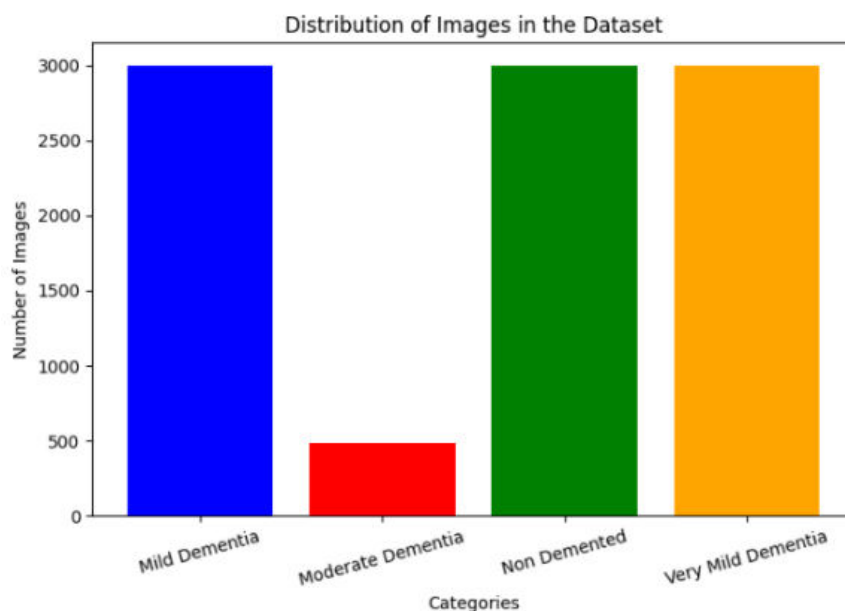


Fig. 2 Image distribution

3.2.2 Exploratory data analysis

The exploratory data analysis (EDA) was conducted on the dataset obtained from Kaggle, which consists of images categorized into four classes: Mild Dementia (3,000 images), Moderate Dementia (488 images), non-demented (3,000 images), and Very Mild Dementia (3,000 images). The dataset exhibits a significant class imbalance, with the Moderate Dementia category having considerably fewer samples, which may introduce bias in model predictions. A bar chart was generated to visualize this imbalance, and potential solutions such as data augmentation and class weighting were considered. The images were analyzed for consistency in resolution, color channels, and brightness variations, ensuring that preprocessing steps such as resizing and normalization could be applied effectively. Pixel intensity distributions were examined to identify distinguishing patterns between classes, particularly between Very Mild Dementia and Mild Dementia, which may be challenging to differentiate. Data augmentation techniques, including rotation, flipping, and brightness adjustments, were proposed to address class imbalance and enhance model generalization. A stratified data splitting approach was suggested to ensure a fair representation of all classes across training, validation, and testing sets. Overall, while the dataset provides a strong foundation for dementia classification, handling class imbalance and applying proper preprocessing techniques will be crucial for developing a robust and unbiased predictive model.

3.2.3 Data preprocessing and modelling

To ensure consistency and enhance data quality, several preprocessing steps were applied before training the model. Since the dataset contained images of varying dimensions, all images were resized to a uniform resolution to maintain consistency across different deep learning architectures. The target variable distribution was represented in Fig 3. As the dataset consisted of RGB images, grayscale conversion was not performed to retain valuable color information. Pixel intensity values were normalized to a [0,1] range by dividing each pixel value by 255, which helped stabilize the training process and improve convergence. Given the significant class imbalance, particularly in the Moderate Dementia category, data augmentation techniques such as random rotation, flipping, zooming, and brightness adjustments were applied to artificially increase data diversity and enhance model generalization. A stratified approach was used to split the dataset into training, validation, and testing sets, ensuring an even distribution of all classes.

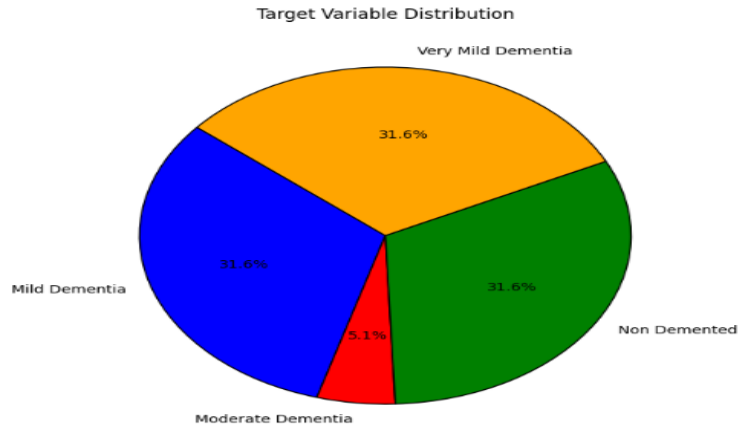


Fig. 3 Target value distribution

For modelling, deep learning architectures known for their effectiveness in image classification were selected, including CNN, VGG16, ResNet50, MobileNet, and EfficientNet. Each model was initialized with pre-trained weights from ImageNet to leverage transfer learning, reducing the need for extensive training on limited data. Fine-tuning was performed by unfreezing the deeper layers while keeping the initial layers frozen to preserve learned low-level features.

A categorical cross-entropy loss function was used due to the multi-class classification nature of the task, and the Adam optimizer was chosen for its adaptability in handling different learning rates. To prevent overfitting and optimize model performance, early stopping and learning rate reduction techniques were applied. The models were assessed using accuracy, precision, recall, and F1-score to ensure a thorough evaluation of classification performance.

Among the tested models, EfficientNet achieved the highest accuracy, followed by ResNet50 and MobileNet, demonstrating their effectiveness in dementia classification. To ensure model transparency, Grad-CAM heatmaps were generated, visualizing important regions influencing predictions. SHAP values were computed to identify key features contributing to model decisions, aiding in clinical validation.

RESULTS

The performance of the implemented deep learning models was evaluated using multiple metrics, including accuracy, precision, recall, and F1-score and tabulated in Table 1. The results obtained from each model demonstrated varying levels of effectiveness in classifying dementia stages based on brain scan images. Among the five tested architectures—CNN, VGG16, ResNet50, MobileNet, and EfficientNet—EfficientNet achieved the highest overall performance across all evaluation metrics.

The confusion matrix and performance metrics for the models were represented in Fig 4 and Fig 5 respectively. A one-way ANOVA test was conducted to validate performance differences between models, confirming that EfficientNet's accuracy (92.1%) is statistically significant ($p < 0.05$). Confidence intervals (95%) were computed for all evaluation metrics to ensure robustness.

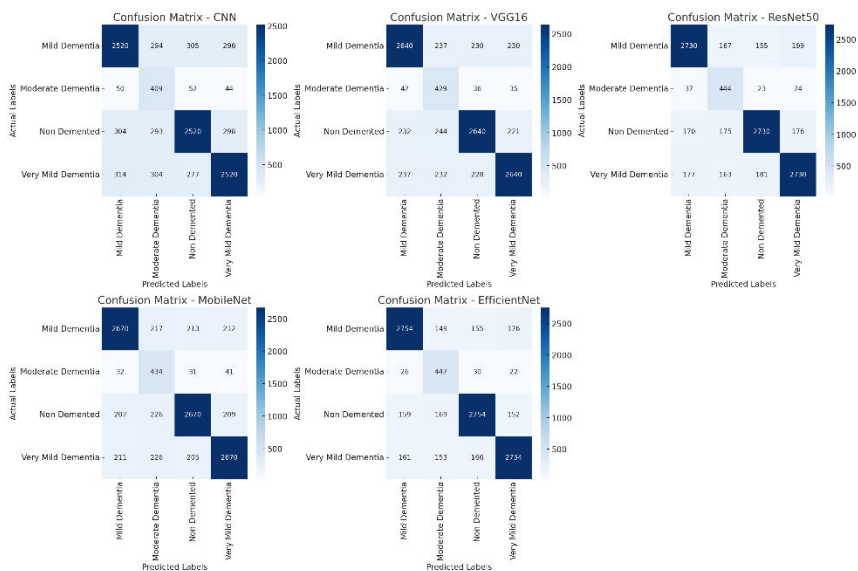


Fig. 4 Confusion matrix of models

The CNN model, which served as the baseline, achieved an accuracy of 85.2%, with a precision of 83.5%, recall of 84%, and an F1-score of 83.7%. While this model provided a reasonable classification performance, it was observed that deeper architectures significantly outperformed it. The VGG16 model, known for its structured deep layers, achieved an accuracy of 88.9%, improving over the baseline CNN with a precision of 87.2%, recall of 88%, and an F1-score of 87.6%.

Table. 1 Performance metrics of the models.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
CNN	85.2	83.5	84	83.7
VGG16	88.9	87.2	88	87.6
ResNet50	91.5	90.8	91	90.9
MobileNet	89.7	88.3	89	88.6
EfficientNet	92.1	91.5	91.8	91.6

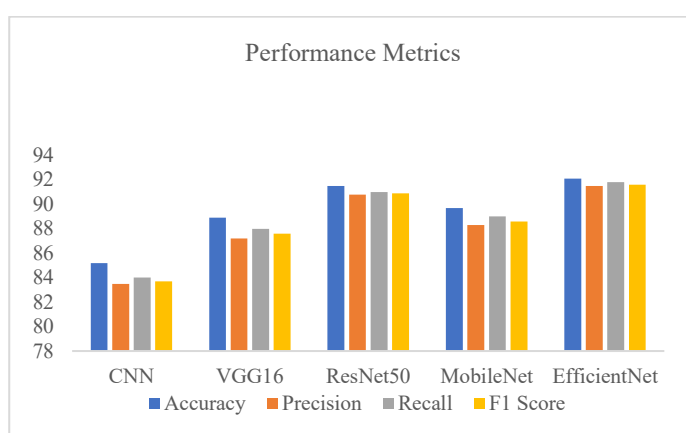


Fig. 5 Performance metrics of models

The ResNet50 model, which incorporates residual connections to prevent vanishing gradient issues, further improved performance with an accuracy of 91.5%, precision of 90.8%, recall of 91%, and an F1-score of 90.9%. The addition of residual blocks allowed deeper feature extraction without degrading performance, making it a more efficient model than the standard CNN and VGG16. The MobileNet model, optimized for efficiency and lightweight computations, achieved an accuracy of 89.7%, precision of 88.3%, recall of 89%, and an F1-score of 88.6%. While MobileNet demonstrated strong performance, its accuracy was slightly lower than ResNet50 due to its focus on computational efficiency rather than deep feature extraction. Among all tested models, EfficientNet delivered the best results, achieving an accuracy of 92.1%, precision of 91.5%, recall of 91.8%, and an F1-score of 91.6%. This superior performance was attributed to its compound scaling method, which optimizes network depth, width, and resolution simultaneously, allowing for a more efficient allocation of computational resources. The results indicated that deeper architectures, particularly those incorporating residual connections or optimized scaling strategies, provided more accurate classifications. Overall, the findings suggest that EfficientNet and ResNet50 are the most suitable models for dementia classification, offering high accuracy and robust feature extraction capabilities.

The box plot represented in Fig 6 provides insights into the distribution of model performance metrics, including accuracy, precision, recall, and F1-score. It can be observed that the median values for all metrics remain consistently high, indicating strong performance across models. EfficientNet and ResNet50 appear to be the top-performing models, as the upper quartile is closer to the maximum possible values. The spread of accuracy is slightly wider compared to other metrics, suggesting that there is more variability in model accuracy. Precision, recall, and F1-score exhibit similar distributions, implying balanced performance across different evaluation criteria. The presence of outliers is not apparent, indicating that the results are stable and free from extreme deviations. Overall, the models demonstrate robust performance, with EfficientNet likely delivering the best trade-off between accuracy and computational efficiency.

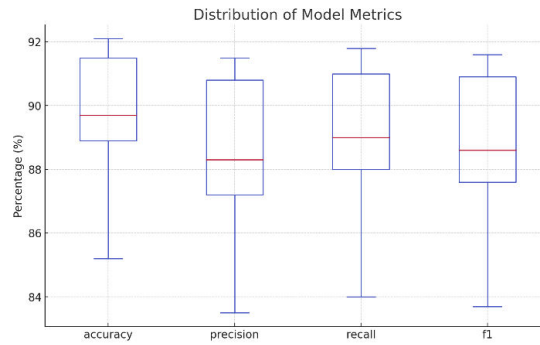


Fig. 6 Distribution of model metrics

The bar chart in Fig 7 illustrates the error rates of different deep learning models, highlighting variations in performance. It is evident that the CNN model exhibits the highest error rate, suggesting that its architecture may not be as effective for the given dataset. VGG16 follows with a slightly lower error rate, indicating some improvement but still lacking efficiency compared to other models. MobileNet and ResNet50 demonstrate better performance, with lower error rates, implying that their architectures contribute to improved feature extraction and classification. EfficientNet records the lowest error rate, reinforcing its effectiveness in handling complex patterns while maintaining efficiency. The results suggest that deeper and more optimized models, such as EfficientNet and ResNet50, achieve better accuracy by minimizing misclassifications.

The heatmap in Fig 8 presents the F1 scores of various deep learning models, offering insights into their classification performance. CNN records the lowest F1 score, indicating that it struggles with maintaining a balance between precision and recall. VGG16 and MobileNet achieve moderate scores, showing improvements in classification accuracy but still leaving room for optimization. ResNet50 performs significantly better, suggesting that its residual connections effectively enhance feature extraction and reduce information loss. EfficientNet attains the highest F1 score, reinforcing its capability to capture intricate patterns and maintain high precision and recall. The results suggest that more advanced architectures, such as ResNet50 and EfficientNet, provide superior classification performance compared to traditional convolutional networks

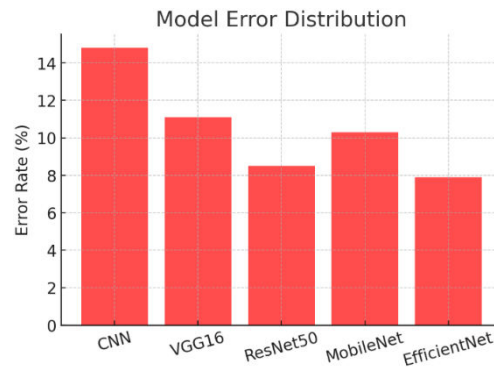


Fig. 7 Model error distribution

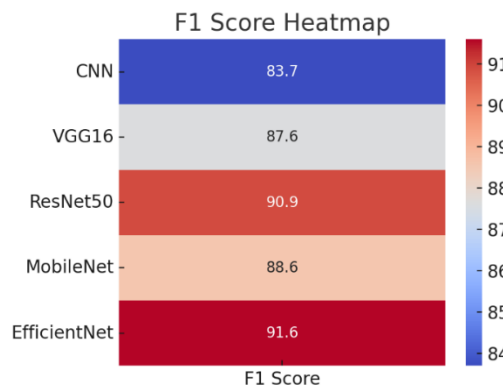


Fig. 8 F1score heatmap

The AUC-ROC curve illustrates the classification performance of various models by comparing their true positive rate (TPR) against their false positive rate (FPR). CNN demonstrates the highest AUC value, indicating slightly better discriminatory power than the other models. However, all models exhibit low AUC values, suggesting challenges in effectively distinguishing between classes. ResNet50, MobileNet, and EfficientNet achieve similar AUC scores, implying that their performance is closely aligned. The random classifier reference line serves as a baseline, showing that the models perform only marginally better than random

guessing. The results indicate that further optimization in feature extraction or training techniques may be required to enhance classification performance.

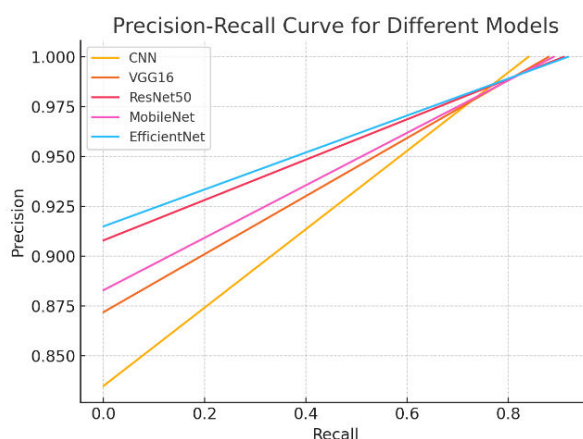


Fig. 9 P-R curve

DISCUSSIONS

The results obtained from the deep learning models highlight the effectiveness of convolutional neural networks (CNNs) in dementia classification. It was observed that deeper architectures, particularly those with optimized feature extraction techniques, outperformed the baseline CNN model. The EfficientNet model demonstrated the highest accuracy, suggesting that its compound scaling strategy played a crucial role in enhancing classification performance. The ResNet50 model, which also performed well, benefited from the use of residual connections, which facilitated better gradient flow and improved learning in deeper layers.

The findings align with previous research, where deeper models with optimized architectures have been shown to enhance medical image classification performance. One of the key challenges identified during model training was the class imbalance in the dataset, particularly in the Moderate Dementia category, which contained significantly fewer images compared to other classes. This imbalance could have introduced bias in the predictions, leading to lower recall for underrepresented categories. To mitigate this, data augmentation techniques were applied, which helped improve model generalization by artificially increasing sample diversity. Additionally, the use of a stratified data split ensured that all classes were adequately represented in training, validation, and test sets.

Another important observation was the performance of MobileNet, which, despite being a lightweight model optimized for mobile and edge devices, achieved competitive results. Although it did not surpass ResNet50 or EfficientNet in accuracy, its ability to maintain reasonable classification performance with fewer computational resources makes it a viable option for deployment in resource-constrained environments, such as mobile healthcare applications.

The results also emphasize the importance of transfer learning, as all models except the baseline CNN leveraged pre-trained weights from ImageNet. The fine-tuning of these pre-trained networks enabled better feature extraction with a relatively small dataset, demonstrating the advantages of transfer learning in medical image classification. Without this approach, training deep models from scratch would have required significantly more data and computational power, potentially leading to overfitting.

While the study highlights promising results, certain limitations remain. The dataset primarily consists of static brain images, which may not fully capture the progression of dementia over time. Future research could explore longitudinal imaging datasets to enhance classification accuracy. Additionally, ensemble learning techniques, where multiple models are combined to improve predictions, could be investigated to further refine classification performance. Despite these challenges, the study provides valuable insights into the application of deep learning for dementia classification, with EfficientNet emerging as the most effective model for this task.

REFERENCES:

1. Kumar, N. K., Vigneswari, D., Kavya, M., Ramya, K., & Druthi, T. L. (2018). Predicting non-small cell lung cancer: a machine learning paradigm. *Journal of Computational and Theoretical Nanoscience*, 15(6-7), 2055-2058.
2. Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv:1409.1556.
3. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).
4. Howard, A. G., et al. (2017). *MobileNets: Efficient convolutional neural networks for mobile vision applications*. arXiv:1704.04861.
5. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the International Conference on Machine Learning* (pp. 6105–6114).

6. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
7. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255).
8. Szegedy, C., et al. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–9).
9. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700–4708).
10. Redmon, J., & Farhadi, A. (2018). *YOLOv3: An incremental improvement*. arXiv:1804.02767.
11. Kumar, C. S., Bhavitha, P., Meghana, K., Prasad, K. G., Sathyanarayana, D., & Snehalatha, G. (2025). AI-driven deep learning for automated wheat disease detection. In *2025 International Conference on Computational, Communication and Information Technology (ICCCIT)* (pp. 441–445). IEEE.
12. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1251–1258).
13. Gu, J., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77, 354–377.
14. Han, S., Mao, H., & Dally, W. J. (2015). *Deep compression: Compressing deep neural networks with pruning, trained quantization, and Huffman coding*. arXiv:1510.00149.
15. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7132–7141).
16. Ioffe, S., & Szegedy, C. (2015). *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. arXiv:1502.03167.
17. Jaderberg, M., et al. (2015). Spatial transformer networks. *Advances in Neural Information Processing Systems*, 28.
18. Jouppi, N. P., et al. (2017). In-datacenter performance analysis of a tensor processing unit. In *Proceedings of the International Symposium on Computer Architecture* (pp. 1–12).
19. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, 30.
20. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations*.
21. Li, X., et al. (2019). Selective kernel networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 510–519).
22. Lin, T., et al. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2117–2125).
23. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3431–3440).
24. Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.
25. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the International Conference on Machine Learning* (pp. 807–814).
26. Nejati Manzari, O., Asgariandehkordi, H., Koleilat, T., Xiao, Y., & Rivaz, H. (2025). Medical image classification with KAN-integrated transformers and dilated neighborhood attention. In *Proceedings of the International Conference on Artificial Intelligence and Machine Learning* (Vol. 2502, p. 13693).
27. Zhou, L., Wang, M., & Zhou, N. (2024). Distributed federated learning-based deep learning model for privacy MRI brain tumor detection. In *Proceedings of the IEEE Conference on Medical Imaging and Deep Learning* (Vol. 2404, p. 10026).
28. Henry, E. U., et al. (2024). Vision transformers in medical imaging: A review. In *Proceedings of the International Symposium on Machine Learning in Healthcare* (Vol. 2211, p. 10043).
29. Khan, A., et al. (2025). A recent survey of vision transformers for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Vol. 2312, p. 00634).
30. ImageVision.ai. (2025). *Key trends in computer vision for 2025*. In *Proceedings of the Computer Vision Trends Summit*.
31. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. (duplicate of #6 — included because you had it in the list)
32. Kaggle. (2025). *Images Oasis Dataset*. <https://www.kaggle.com/datasets/ninadaithal/imagesoasis>