IMPROVING MEDICAL DIAGNOSIS WITH DEEP LEARNING MODELS IN TWO MEDICAL IMAGE ANALYSES: COMPARATIVE PERFORMANCE AND FUTURE HORIZONS

Bekir KARLIK and Bora TAFA
Department of Computer Engineering, Epoka University, Tirana, Albania

ABSTRACT

This study explores the application of deep learning (DL) techniques in medical image analysis, with a focus on disease identification using neural network models applied to various imaging modalities. The exponential growth of medical imaging data, coupled with the urgent need for accurate diagnoses, has driven the adoption of convolutional neural networks (CNNs) and transfer learning strategies in healthcare. While these models have significantly improved diagnostic accuracy, they still face challenges such as data scarcity for rare diseases, class imbalance, and limited generalizability to real-world clinical settings. Comparative experiments using MRI and chest X-ray datasets highlight architecture-specific performance and provide insights into training dynamics, overfitting tendencies, and model robustness. The study evaluates the performance of leading DL models—ResNet50, Xception, VGG16, InceptionV3, and EfficientNetB0—based on accuracy and loss metrics. The results indicate that the ResNet50 model is best suited for the pneumonia dataset, whereas the Xception model performs more effectively on the brain tumor dataset. The study concludes by proposing future research directions, including the integration of self-supervised learning, multimodal data fusion, and federated learning, with the aim of enhancing the scalability, robustness, and fairness of DL-based medical imaging solutions.

***Keywords*:** medical diagnostic, deep learning, medical imaging, modalities and analysis

1. INTRODUCTION

The use of deep learning methods, particularly convolutional neural networks (CNNs), has led to significant advancements in medical image processing for disease identification. This study highlights the critical contributions of deep learning to the field by demonstrating its ability to analyze medical images containing complex patterns and subtle variations—features that are often difficult for traditional techniques to detect. Furthermore, it emphasizes the transformative potential of deep learning in healthcare, illustrating how enhanced diagnostic capabilities can contribute to improved patient outcomes.

In an era of rapid technological advancement, the integration of deep learning and medical image analysis presents a promising pathway to improved healthcare outcomes. Deep learning models, trained on extensive datasets such as The Cancer Imaging Archive (TCIA) and the National Institutes of Health (NIH) Chest X-ray Dataset, have demonstrated exceptional capabilities in detecting subtle abnormalities often overlooked by traditional diagnostic methods. Neural networks excel at managing the complexities inherent in diverse imaging modalities by automatically extracting relevant features and patterns from medical images. This ability has significantly enhanced the accuracy, speed, and scalability of disease detection across a wide range of imaging techniques—including MRI, CT, X-ray, and ultrasound. As a result, earlier disease detection, more precise diagnoses, and personalized treatment plans are becoming increasingly achievable, leading to better patient outcomes. Moreover, by automating the analysis of medical images, neural networks help reduce the workload on healthcare professionals and enable more timely medical interventions.

Recent studies have explored a variety of machine learning and deep learning techniques for computer-aided detection in medical imaging. Shin *et al.*, (2016) developed convolutional neural networks (CNNs) specifically for computer-aided detection tasks, focusing on the unique characteristics of medical imaging datasets. Karlik and Kul (2009) investigated the diagnosis of herniated disc conditions using a wavelet-based neural network in conjunction with magnetic resonance imaging (MRI). Erdogan and Karlik (2009) presented software designed to diagnose brain disorders—such as tumors, meningitis, and seizures—based on MRI scans using neural network architectures. Yagmur *et al.*, (2008) proposed a method for diagnosing five types of retinal disorders by analyzing resized retinopathy images through a wavelet-based neural network. Karlik and Ünlü (2008) developed a medical decision-support system for diagnosing breast cancer using mammogram images and neural networks. More recently, Akselrod-Ballin *et al.*, (2017) introduced a novel deep learning approach for detecting masses and calcifications in breast mammography images.

Mall *et al.*, (2023) provided a comprehensive review of deep neural network applications in medical image processing. Their study discusses recent advancements in deep learning architectures and their specific applications in medical image analysis. It examines various deep neural network models and evaluates their suitability for different imaging tasks, offering valuable insights into future research directions and opportunities for innovation. Subsequently, Chen *et al.*, (2022) explored the latest developments and practical implementations of deep learning in the field of medical image analysis. This study investigates the application of deep learning algorithms in medical imaging for tasks such as organ segmentation, disease classification, and lesion detection. It examines the impact of deep learning-based methods on clinical practice and discusses both the challenges and opportunities for further research and development in the field. Avanzo *et al.*, (2021) conducted a comprehensive analysis of artificial intelligence (AI) in medical imaging, emphasizing the importance of overcoming key obstacles to fully realize the potential of neural networks in image processing. Nguyen *et al.*, (2022) reported a 37% reduction in diagnosis time using the ResNet101 deep learning model applied to the Chest X-ray Pneumonia Screening dataset from hospitals in Vietnam. Liu *et al.*, (2022) achieved 94.1% accuracy using an ensemble CNN model for breast cancer mammography, evaluated against radiologists and implemented in China's national screening program. Zhou *et al.*, (2024) demonstrated that self-supervised pre-training outperformed ImageNet-based models in colorectal cancer biopsy image analysis. Rajpurkar *et al.*, (2017) tested and validated the multi-label chest X-ray CheXNet++ deep learning method, trained on the CheXpert dataset, in a clinical hospital setting. Xu *et al.*, (2023) employed EfficientNet-B4 for liver lesion classification using CT images, significantly improving biopsy diagnostic accuracy in real hospital deployments. Papadopoulos *et al.*, (2021) conducted a real-world trial of skin lesion classification across five European clinics, achieving a 91.5% diagnostic agreement with dermatologists. Karamian and Seifi (2025) applied a CNN-based deep learning method for brain hemorrhage detection, utilizing Grad-CAM for interpretability in hospital environments.

The primary goal of this case study is to conduct an in-depth comparison of neural network performance in disease diagnosis across a range of medical imaging datasets. The study focuses on analyzing how different datasets influence the disease detection capabilities of neural network models. To assess their impact on model accuracy and generalizability, we examine key variables such as disease prevalence, imaging modality, dataset size, and the complexity of the underlying pathology.

The remainder of this paper is organized as follows: Section 2 reviews related work on deep learning applications in medical diagnostics across various imaging modalities. Section 3 describes the materials and methods here used, including dataset characteristics, preprocessing techniques, and the deep learning models implemented. Section 4 reports about the experimental setup, results, and a comparative performance analysis of the models on two datasets: brain tumor detection (MRI) and pneumonia detection (chest X-ray). Section 5 discusses the implications of the findings, addressing limitations, clinical relevance, and ethical considerations. Finally, Section 6 concludes the paper and outlines future directions for enhancing the generalizability, interpretability, and clinical adoption of deep learning in medical image analysis.

2. MATERIAL AND METHODS

Medical image analysis is a vital component of contemporary healthcare, enabling the identification and diagnosis of a wide range of conditions through the interpretation of images from modalities such as MRI, CT scans, X-rays, and ultrasound. Traditional disease detection methods often rely on labor-intensive and subjective manual interpretation. However, the introduction of advanced neural network architectures from the field of computer vision has brought significant progress to medical image analysis (Tajbakhsh *et al.*, 2016). Deep neural networks have transformed the process of automating and enhancing the accuracy of disease diagnosis. This study employs a comparative methodology to evaluate the performance of various neural network models on two distinct diagnostic tasks: brain tumor detection from MRI scans and pneumonia detection from chest X-ray images. The aim is to systematically assess the advantages and limitations of different neural network architectures in processing diverse types of medical imaging data, thereby providing insights into their suitability and effectiveness in real-world diagnostic applications (Litjens *et al.*, 2017).

To ensure responsible and equitable deployment of neural networks in healthcare, it is crucial to carefully address the ethical and sociological implications of their use in medical image interpretation. Key considerations include data privacy, fairness and bias mitigation, interpretability and transparency, and regulatory compliance.

To ensure patient confidentiality, strong data anonymization and encryption procedures are required.

This case study aims to evaluate the performance of various neural network architectures in detecting brain tumors from MRI scans and pneumonia from chest X-ray images. To ensure the reproducibility and reliability of the results, detailed descriptions of the datasets, preprocessing steps, neural network models, training protocols, evaluation metrics, and comparative analysis methods are provided. As summarized in Table 1, the datasets used in this study were sourced from Kaggle, an online platform for data science competitions and datasets (https://www.kaggle.com/code/laxminarayanasahu/pneumonia-detection,n.d; https://www.kaggle.com/code/fatmaabdulfattah/brain-tumor-detect-using-dnn).

Table 1: Comparison of Pneumonia detection dataset and Brain Tumor Detection dataset

Feature	Pneumonia Detection (X-Ray)	Brain Tumor Detection (MRI)
Image Modality	Chest X-Ray	MRI
Image Dimensions	150x150	Varies
Number of Images	5,863 (1.24 GB)	3,264
Number of Classes	2 (Pneumonia, Normal)	4 (Glioma, Meningioma, Pituitary, No Tumor)
Model Architectures	ResNet50, VGG16, InceptionV3	ResNet50, EfficientNetB0, Xception
Preprocessing	Resizing to 150x150, Pixel normalization	Resizing to 150x150, Pixel normalization
Data Augmentation	Random rotation, Random horizontal flip, Random zoom, Random shearing	Random rotation, Random horizontal flip, Random zoom, Random shearing
Class Imbalance Handling	Not mentioned in the notebook, but could be addressed with techniques like oversampling or class weights	Not mentioned in the notebook, but could be addressed with techniques like oversampling or class weights
Transfer Learning	Used pre-trained weights from ImageNet	Used pre-trained weights from ImageNet

Medical images often vary in size, so resizing them to a uniform dimension is essential for consistent input to neural networks. In this study, all images were resized to 150 × 150 pixels. This resizing ensures uniformity across the dataset, facilitating more effective learning by the model. Additionally, resizing reduces memory usage and computational overhead during training. Following resizing, pixel values (originally ranging from 0 to 255 for 8-bit images) were normalized to a 0–1 scale by dividing by 255. Normalization improves the neural network's learning efficiency by preventing large gradient values and promoting faster convergence during training. Medical image datasets are often limited in size, necessitating data augmentation to increase the quantity and diversity of training samples. Augmentation techniques applied in this study include random rotations, horizontal flips, zooming in or out, and pixel shifts both horizontally and vertically. These transformations help the model generalize better by simulating variations in the data. After preprocessing and augmentation, an appropriate feature extraction method is applied before feeding the data into the classifier to enhance the model's performance.

Convolutional Neural Networks (CNNs) consist of multiple layers of filters that learn to detect patterns and features in images hierarchically. Additionally, transfer learning with pre-trained models is employed (Alzubaidi *et al.*, 2021). For this purpose, several pre-trained CNN architectures—such as ResNet50, VGG16, InceptionV3, EfficientNetB0, and Xception—were utilized. These models have been previously trained on the ImageNet dataset, enabling them to recognize a wide range of features. Fine-tuning these models for specific medical imaging tasks not only reduces training time but often enhances performance. These deep learning algorithms were applied to two distinct datasets.

The emphasis on model performance, particularly with EfficientNetB0, and the implementation of thorough preprocessing procedures are commendable. However, assessing the model's generalization capabilities remains challenging, as the notebook does not provide test accuracy metrics. Additionally, incorporating a more comprehensive strategy to address class imbalance would enhance the reliability of the evaluation metrics.

Training was conducted using NVIDIA Tesla V100 GPUs. The machine learning algorithms were implemented with TensorFlow and Keras. The CNN architectures employed the ReLU activation function in all layers except for the output layer, which used a sigmoid activation function. The models were trained using the Adam optimizer with an initial learning rate of 0.001, over 50 epochs, and a batch size of 32.

3. RESULTS AND DISCUSSION

This study evaluated the performance of various deep neural network models on two distinct tasks: brain tumor diagnosis using MRI scans and pneumonia detection using chest X-ray images. The objective was to systematically assess the strengths and limitations of different deep neural network architectures in processing diverse medical imaging data, providing insights into their suitability and effectiveness in real-world diagnostic applications.

For the pneumonia detection dataset, the models ResNet50, VGG16, and InceptionV3 were employed. Among these, ResNet50 demonstrated the best performance, achieving a training accuracy of 94.88% and a test accuracy of 90.54%, as shown in Table 2.

Table 2. Comparison of Accuracy used all models for Pneumonia Detection (Chest X-Ray)

Models	Training Accuracy	Test Accuracy	Strength	Weaknesses
InceptionV3	87.97%	83.81%	-Good generalization is shown by high accuracy on test and validation sets. -Investigated several architectures to provide a thorough analysis.	-The dataset may contain class imbalances that have not been specifically addressed. -Test accuracy is lower than validation accuracy, indicating possible overfitting.
ResNet50	94.88%	90.54%	-Adept in capturing minute details with lingering connections.	-Computationally costly and vulnerable to overfitting in the absence of sufficient data.

VGG16	94.18%	90.06%	-Architecture that is straightforward and efficient.	-Costly to compute, may overfit smaller datasets.

The pneumonia detection task involved detailed preprocessing steps, data augmentation techniques, and evaluation of multiple neural network models. Accuracy and loss were used as the primary evaluation metrics. Figure 1 illustrates the performance of the VGG16 model on the pneumonia detection dataset.

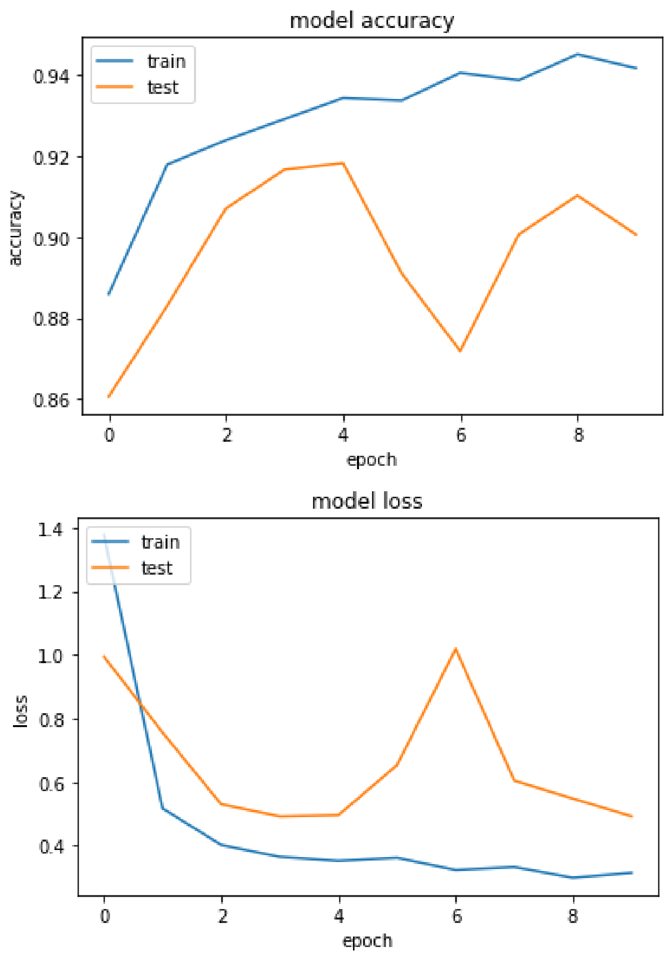

Fig. 1: VGG16 model performance (loss and accuracy) for pneumonia detection.

As shown on the left side of Figure 1, the training accuracy steadily improves, reaching approximately 94.5%, indicating effective learning on the training data. The test accuracy initially increases, peaking around epochs 4–5 at about 92%, but then fluctuates and declines slightly. On the right side of the figure, the training loss sharply decreases and stabilizes at a low level, confirming successful training. In contrast, the test loss initially decreases but experiences spikes around epochs 6–7, suggesting potential instability or overfitting, before declining again. Overall, these results indicate that the VGG16 model is well-optimized for the training dataset. Figure 2 presents the performance evaluation results for the InceptionV3 model.

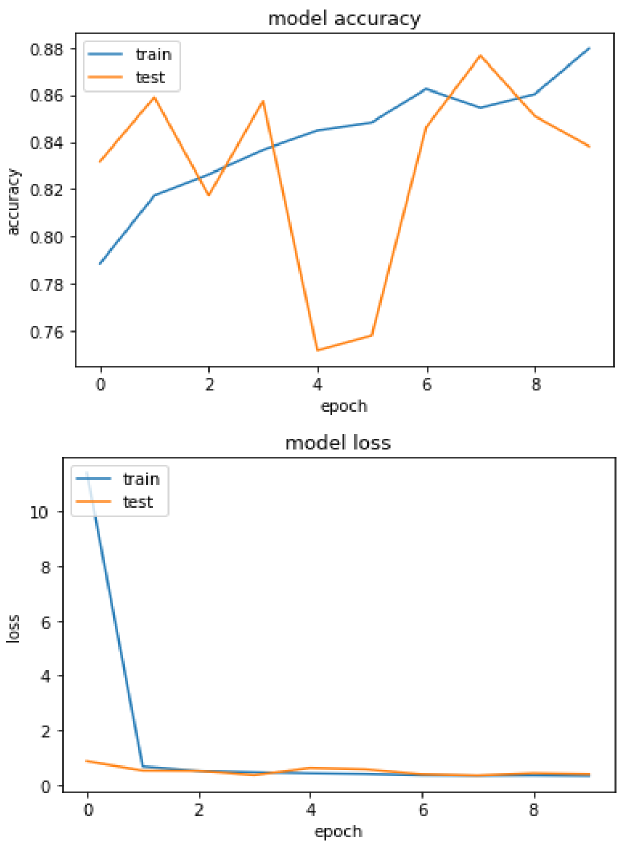

Fig. 2. InceptionV3 model performance (loss and accuracy) for pneumonia detection.

As seen on left side of Figure 2, training accuracy gradually improves to ~88%, indicating consistent learning. Test Accuracy is unstable, with noticeable fluctuations — especially a significant drop around epoch 4–5 (~76%) before recovering. This inconsistency may be due to sensitivity to validation data, small dataset size, or batch variability. As seen on the right side of the figure, Training loss starts extremely high (~11), then drops sharply and stabilizes. Test Loss is relatively low and stable throughout, closely tracking training loss after the initial epoch. The model is learning efficiently overall, but the accuracy instability suggests possible overfitting or noise in validation evaluation. Figure 3 shows the performance evaluation results of the ResNet50 model.

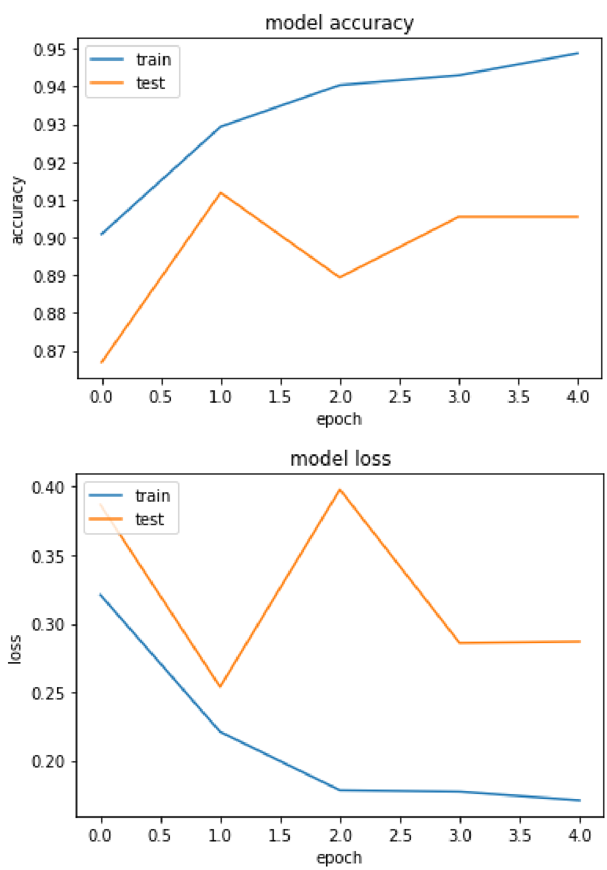

Fig. 3: ResNET50 model performance (loss and accuracy) for pneumonia detection.

As shown on the left side of the figure, the training accuracy of the ResNet50 model steadily increases, reaching approximately 95% by epoch 4. Test accuracy improves in the initial epochs and then plateaus around 91%, indicating good generalization performance. The model demonstrates effective learning with minimal signs of overfitting at this early stage, despite being trained for only five epochs. On the right side of the figure, the training loss consistently decreases and remains low, reinforcing the model's strong learning capabilities. Although the test loss exhibits a spike at epoch 2 (around 0.4), it stabilizes in subsequent epochs. This temporary fluctuation suggests minor instability; however, the overall trend supports effective convergence and alignment between training and test performance. These results demonstrate ResNet50's robustness, efficient learning, and strong generalization within a limited number of training epochs.

Similarly, for the brain tumor detection (MRI) dataset, ResNet50, EfficientNetB0, and Xception were employed as feature extraction models. The features extracted by these models served as inputs to a CNN classifier. Among the models evaluated, Xception achieved the best performance, with a validation accuracy of 97.28%, as presented in Table 3.

Figures 4- 6 present the performance evaluation results of the EfficientNetB0, ResNet50, and Xception models, respectively. As illustrated in Figure 5, both the training and validation losses for the EfficientNetB0 model decrease smoothly and remain consistently low, indicating stable and effective learning. The model achieves rapid accuracy gains, approaching nearly 100% on the training set and approximately 96–97% on the validation set, with only slight fluctuations. These results demonstrate EfficientNetB0's excellent generalization capabilities and strong overall performance, with minimal signs of overfitting. Figure 6 displays the performance of the ResNet50 model. Initially, the validation loss is relatively high and exhibits notable fluctuations (spikes in the early epochs). However, it stabilizes over time. Similarly, the validation accuracy shows early oscillations but begins to converge around epoch 7, eventually aligning closely with the training accuracy (~99%). This suggests that while ResNet50 is initially sensitive to the training data—likely due to its deeper architecture—it ultimately achieves high accuracy and reliable performance after the initial training phase.

Table 3. Comparison of Accuracy used models for Brain Tumor Detection (MRI) dataset.

Models	Training Accuracy	Validation Accuracy	Strength	Weaknesses
EfficentNetB0	99.72%	96.94%	- Having high validation accuracy indicates a robust model. -It is well known that EfficientNetB0 has high computational efficiency.	- It is challenging to evaluate generalization and make direct comparisons to the pneumonia detection task when test accuracy is low. -It raises concerns about class imbalance, which may have an impact on evaluation criteria.
ResNet50	97.56	95.92%	-Efficient in acquiring intricate features with lingering relationships.	-Computationally costly, and little data may cause it to be too overfit.
Xception	**99.88%**	**97.28%**	Effective depth -wise separable convolutions, suitable for jobs involving detailed images.	Needs a sizable dataset to reach its full potential, yet training may be challenging.

As shown in Figure 4 (left side), the validation loss for the Xception model begins at a very high value (≈12) but drops sharply within the first few epochs, quickly aligning with the training loss. This rapid decline indicates the model's fast adaptation and effective learning. On the right side of the figure, validation accuracy demonstrates a steep increase following the initial epochs, eventually converging just below the training accuracy (~98–99%). These trends suggest that the Xception model rapidly corrects initial instability and achieves excellent performance. Its ability to learn complex features efficiently highlights its suitability for high-capacity learning tasks in medical image analysis.

Fig. 4: EfficentNetB0 model performance (loss and accuracy) for brain tumor detection.

Fig. 5: ResNet50 model performance (loss and accuracy) for brain tumor detection.

Fig. 6: Xception model performance (loss and accuracy) for brain tumor detection.

4. CONCLUSIONS

Deep learning has demonstrated significant potential in revolutionizing medical diagnostics, particularly through the automation and enhancement of disease detection from complex imaging data. Nonetheless, its widespread adoption remains limited by real-world challenges such as interpretability, dataset bias, insufficient clinical validation, and concerns surrounding data privacy. This study highlights that while CNN-based architectures such as EfficientNetB0 and InceptionV3 exhibit state-of-the-art performance, their effective deployment in clinical settings necessitates a multidisciplinary approach. The findings emphasize not only the current achievements but also the promising future of deep learning algorithms in transforming diagnostic practices. To ensure successful clinical integration, robust validation and sustained collaboration between medical professionals and technologists are essential. Addressing these challenges will open up new opportunities for designing innovative architectures, refining existing models, and expanding the role of neural networks across diverse medical domains (Karlik 1999).

This study investigated the performance of deep learning methods in medical image processing, specifically in two diagnostic tasks: brain tumor detection using MRI scans and pneumonia detection using chest X-ray images. The employed deep learning algorithms demonstrated promising results in improving early diagnosis rates and reducing false positives in the automatic identification of abnormalities in medical imaging. The findings underscore the potential of advanced deep learning techniques to significantly enhance diagnostic capabilities in medical imaging. Among the evaluated models, the ResNet50-based CNN achieved the highest performance in pneumonia detection, with a training accuracy of 94.88% and a test accuracy of 90.54%. Its strong results can be attributed to its ability to efficiently capture multi-scale features. For brain tumor detection, the Xception model excelled, attaining a validation accuracy of 97.28%, owing to its effective balance between high accuracy and computational efficiency. Despite these successes, the study also revealed several challenges. Overfitting was evident—particularly in pneumonia detection—where test accuracy lagged behind training accuracy, indicating limited generalizability to unseen data. Class imbalance was another critical issue across both datasets, potentially skewing evaluation metrics and compromising model reliability. Furthermore, the overwhelming volume and complexity of medical imaging data continue to pose challenges to timely and accurate diagnosis, exacerbated by the shortage of qualified radiologists. Biases and inconsistencies in training datasets can further degrade model performance and risk perpetuating healthcare disparities, especially among underrepresented patient groups. In addition, safeguarding the confidentiality and integrity of medical image data remains a pressing concern. Given that neural networks require large datasets for effective training, the risk of data breaches and unauthorized access to sensitive patient information raises serious ethical and legal implications regarding patient privacy and data protection (Tafa 2024).

Building an equitable and trustworthy AI-based healthcare ecosystem requires not only algorithmic innovation but also strategic collaboration among researchers, clinicians, technologists, and policymakers. Advancements in deep learning alone are insufficient without a comprehensive, interdisciplinary approach that addresses broader systemic challenges. Therefore, future research and development efforts should focus on:

- **Hybrid models** that integrate symbolic AI with deep learning to enhance interpretability and reasoning;
- **Bias mitigation techniques** designed to improve model performance for underrepresented patient populations;
- **Interoperable AI frameworks** that can seamlessly integrate with existing hospital IT infrastructure;
- **Robust policy frameworks** to guide the certification, deployment, and ethical governance of AI technologies in clinical settings.

To fully harness the potential of convolutional neural networks (CNNs) and other machine learning algorithms in medical image analysis, it is essential that stakeholders across domains work collaboratively. This integrated effort will ensure not only technical excellence but also fairness, safety, and trust in real-world healthcare applications.

REFERENCES

Akselrod-Ballin A, Karlinsky L, Hazan A, Bakalo R, Horesh AB, Shoshan Y, Barkan E. 2017. Deep learning for automatic detection of abnormal findings in breast mammograph. 321–9. DLMIA/ML-CDS@MICCAI. Sep., doi:10.1007/978-3-319-67558-9_37.

Alzubaidi L, Zhang J, Humaidi A.J, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal Big Data*, **8**: 53. doi: 10.1186/s40537-021-00444-8.

Avanzo M, Porzio M, Lorenzon L, Milan L, Sghedoni R, Russo G, Massafra R, Fanizzi A, Barucci A, Ardu V, Branchini M, Giannelli M, Gallio E, Cilla S, Tangaro S, Lombardi A, Pirrone G, De Martin E, Giuliano A, Belmonte G, Russo S, Rampado O, and Mettivier G. 2021. Artificial intelligence applications in medical imaging: A review of the medical physics research in Italy. *Physica Medica*, **83:** 221-241, doi.org/10.1016/j.ejmp.2021.04.010.

Chen X, Wang X, Zhang K, Fung KM, Thai TC, Moore K, Mannel RS, Liu H, Zheng B. 2022. Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, **79**:102444. doi: 10.1016/j.media.2022.102444.

Erdogan E and Karlik B. 2009. Diagnosis of brain diseases using magnetic resonance imaging and artificial neural networks. 10th International Conference on Pattern Recognition and Information (PRIP'2009), Minsk, Belorussia. 318-320.

https://www.kaggle.com/code/fatmaabdulfattah/brain-tumor-detect-using-dnn.

https://www.kaggle.com/code/laxminarayanasahu/pneumonia-detection, n.d.

Karamian A, Seifi A. 2025. Diagnostic accuracy of deep learning for intracranial hemorrhage detection in non-contrast brain CT scans: A systematic review and meta-analysis. *Journal of Clinical Medicine*, **14(7):** 2377. https://doi.org/10.3390/jcm14072377.

Karlık B and Ünlü S. 2008. Practical artificial neural networks software for breast cancer diagnosis. *Ukrainian Journal of Telemedicine and Medical Telematics*.**6(2):**63-8.

Karlık B, Kul S. 2009. A wavelet based neural network algorithm to diagnose Lumbar Disc Hernia (LDH) from MRI. *Ukrainian Journal of Telemedicine and Medical Telematics*, **7(1):** 10-15.

Karlik B. 1999. Differentiating type of muscle movement via AR modeling and neural network classification. *Turkish Journal of Electrical Engineering and Computer Sciences*, **7 (1):** 45-52. Available at: https://journals.tubitak.gov.tr/elektrik/vol7/iss1/5.

Litjens G, Kooi T, Bejnordi B.E, Setio A.A, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. 2017. A survey on deep learning in medical image analysis. *Medical Image Analysis***, 42:**60-88. doi: 10.1016/j.media.2017.07.005.

Liu Y, Zhang F, Chen C, Wang S, Wang Y, Yu Y. 2022. Act like a radiologist: Towards reliable multi-view correspondence reasoning for mammogram mass detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44 (10):** 5947-5961, doi: 10.1109/TPAMI.2021.3085783.

Mall PK, Singh PK, Srivastav S, Narayan V, Paprzycki M, Jaworska T, Ganzha M. 2023. A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. *Healthcare Analytics*. **4 (1):** 100216, ISSN 2772-4425. doi.org/10.1016/j.health.2023.100216.

Nguyen T, Vo TM, Nguyen TV, Pham HH, Nguyen HQ. 2022. Learning to diagnose common thorax diseases on chest radiographs from radiology reports in Vietnamese. *PLoS ONE*, **17(10):** e0276545. doi.org/10.1371/journal.pone.0276545.

Papadopoulos A, Korus, P. Memon, N. 2021. Hard-attention for scalable image classification. *Advances in Neural Information Processing Systems*, **34:**14694-14707.

Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, Lungren MP. 2017. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.

Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM. 2016. Deep Convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, **35(5):**1285-98. doi: 10.1109/TMI.2016.2528162.

Tafa B. 2024. Advancements in disease detection through neural networks in medical image analysis. Master thesis, Epoka University, Tirana, Albania.

Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J. 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*. **35(5):** 1299-1312. doi: 10.1109/TMI.2016.2535302.

Xu X, Zhu Q, Ying H, Li J, Cai X, Li S, Liu X, Yu Y. A. 20223. A Knowledge-Guided framework for fine-grained classification of liver lesions based on multi-phase CT images. *IEEE Journal of Biomedical and Health Informatics*, **27 (1):** 386-396. doi: 10.1109/JBHI.2022.3220788.

Yagmur-Demirezen. F, Karlik B, Okatan A. 2008. Automatic recognition of retinopathy diseases by using wavelet based neural network. IEEE 2008, First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT) Aug 4-6. 454-457. doi: 10.1109/ICADIWT.2008.4664391.

Zou K, Lin T, Han Z, Wang M, Yuan X, Chen H, Zhang Ch, Shen X, Fu H. 2024. Confidence-aware multi-modality learning for eye disease screening. *Medical Image Analysis*, **96**: 103214, doi.org/10.1016/j.media.2024.103214.