Mérida, Yucatán a 29 de Mayo del 2025

“Convolutional Neural Networks: A Comparative Study of DenseNet121 and InceptionV3”

Miguel Ángel Zúñiga Capistrán,Ismael Eliezer Pérez Ruiz

Universidad Modelo, Yucatán 97305, MEX

15222538@modelo.edu.mx

Abstract Chest X-rays are fundamental tools in diagnosing thoracic diseases due to their accessibility and informative value. The automatic detection of anomalies in chest radiographs using deep learning has shown promising results, especially with convolutional neural networks (CNNs). In this study, it´s evaluated and compared the performance of three state-of-the-art CNN architectures: DenseNet121, InceptionV3, and Xception for multi-class classification of thoracic abnormalities. I used a public dataset containing nine diagnostic categories, ranging from normal anatomy to various pathological conditions. All models were fine-tuned with transfer learning and trained under a uniform pipeline to ensure comparability. The evaluation metrics included accuracy, F1-score, precision, and confusion matrices. Experimental results show that DenseNet121 achieved the highest overall performance, closely followed by Xception, while InceptionV3 presented slightly lower metrics. These findings highlight the relevance of architecture selection in automated chest X-ray analysis and its potential role in clinical decision support systems.

Keywords: Detection, DenseNet121, InceptionV3, Xception.

1. Introduction

Chest radiographs are one of the most commonly used medical imaging modalities in clinical practice. They are essential for diagnosing a wide range of conditions, from lung infections and heart enlargement to fractures and tumors. However, accurate interpretation can be challenging due to overlapping anatomical structures and subtle pathological signs, which increases the likelihood of human error.

In recent years, deep learning techniques(especially convolutional neural networks (CNNs)) have demonstrated remarkable capabilities in medical image analysis. CNNs automatically learn hierarchical feature representations from data, eliminating the need for manual feature extraction. These models have been successfully applied to various domains, including dermatology, ophthalmology, and radiology.

This study explores and compares three advanced CNN architectures; DenseNet121, InceptionV3 and Xception, for the detection of thoracic anomalies in chest X-rays. Both networks have been pre-trained on the ImageNet dataset and fine-tuned for multi-label classification of chest diseases. We aim to evaluate their diagnostic performance and suitability for real-world deployment in clinical screening systems.

2. Methodology

2.1 Dataset

We utilized the publicly available "X-ray Lung Diseases Images (9 classes)" dataset from Kaggle, which includes thousands of chest radiographs annotated into nine diagnostic classes. These classes represent a range of thoracic anomalies including normal anatomy, inflammatory processes, density changes (increased or decreased), obstructive and infectious-degenerative diseases, encapsulated lesions, mediastinal alterations, and thoracic structural changes.
To ensure balanced training, we selected a subset of 4,500 images with 500 samples per class. This balanced setup improves training stability and allows fair performance comparison across categories.

2.2 Image Preprocessing

All images were resized to 224×224 pixels to match the input dimensions required by the CNN models. Pixel values were normalized to the [0, 1] range, and preprocessing techniques such as histogram equalization and zero-mean standardization were applied to enhance contrast and emphasize pathological features. These adjustments are particularly important for chest X-rays, where anomalies may appear subtle and dispersed.

2.3 CNN Architectures

For this study, three convolutional neural networks (CNN) architectures were implemented, that are widely recognized for their strong performance in image classification tasks. The first model, DenseNet121, is characterized by its densely connected layers, which promote efficient feature reuse and strengthen gradient flow throughout the network. This architectural design helps reduce the number of parameters while maintaining high accuracy.The second model, InceptionV3, employs inception modules that allow for parallel convolutional operations at multiple scales. This multi-branch structure enables the network to capture both fine and coarse features simultaneously, making it particularly effective for complex image datasets such as medical scans.

The third model, Xception, builds upon the Inception framework by replacing standard convolutions with depthwise separable convolutions. This modification significantly reduces computational complexity while retaining high representational power, allowing the network to learn intricate spatial patterns with fewer parameters.

2.4 Training Strategy

Transfer learning was performed using pre-trained ImageNet weights. Training was carried out in two phases:

1. Feature extraction**:** The convolutional base was frozen while only the top classifier layers were trained.
2. Fine-tuning: We selectively unfroze parts of the convolutional base:
    - DenseNet121: last 30 layers
    - InceptionV3: last 20 layers
    - Xception: last 10 layers

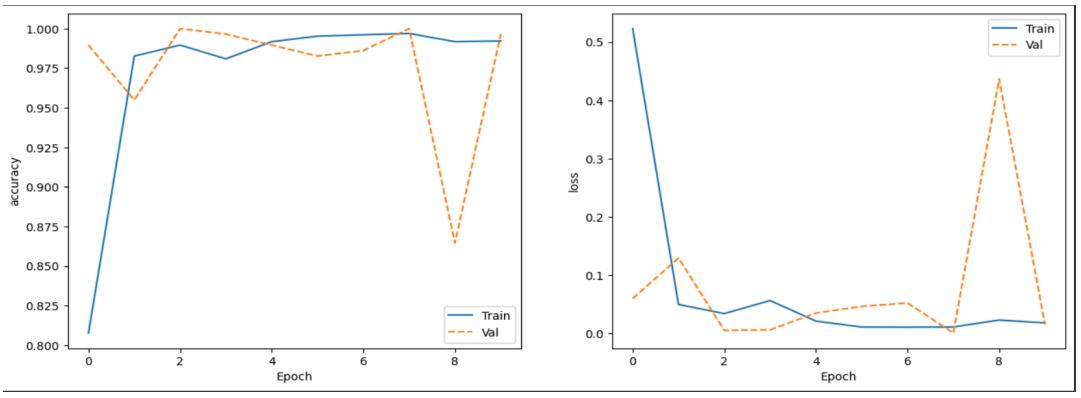
The models were trained using the Adam optimizer and categorical cross-entropy loss. Early stopping and learning rate reduction on plateau were applied to improve generalization and prevent overfitting.

3. Results

3.1 Quantitative Evaluation

InceptionV3 demonstrated strong overall performance, with high classification accuracy across training and validation sets. Table 1 summarizes the results per epoch.

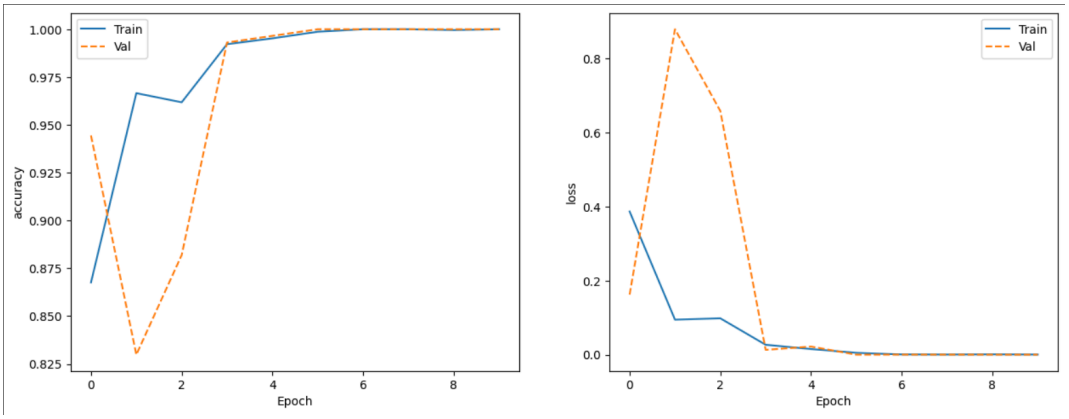
Figure 1 Accuracy (left) and loss (right) curves during training and validation of InceptionV3.



The model achieved perfect validation accuracy (100%) in multiple epochs and maintained validation loss near zero for most of the training process, indicating a highly effective feature learning process. However, the drop in accuracy in epoch 9 may suggest transient overfitting or batch variance. (*Figure 1*)

DenseNet121 achieved outstanding performance during training, with rapid convergence and excellent generalization. Validation accuracy reached 100% consistently from epoch 6 onward, with extremely low loss values, suggesting that the model effectively learned discriminative features across the 9 thoracic classes

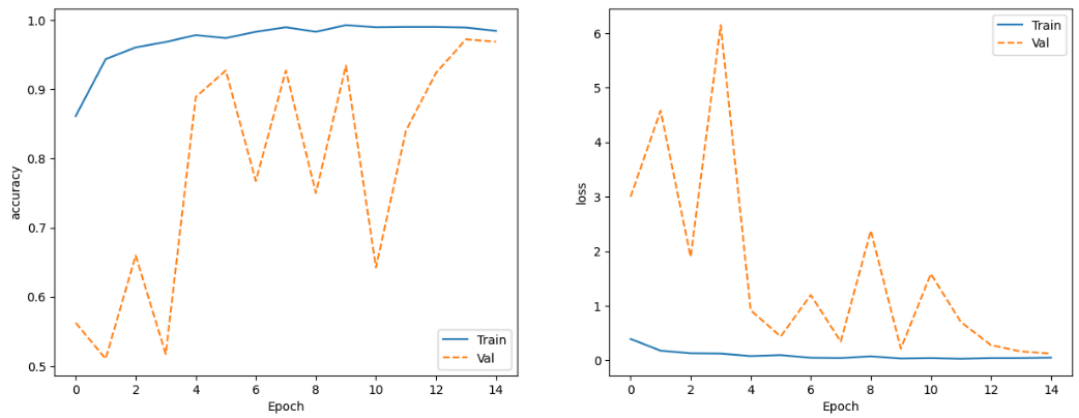
.Figure 2. Accuracy (left) and loss (right) curves during training and validation of DenseNet121.



(Figure 2.) Shows the accuracy and loss curves for DenseNet121. The rapid stabilization and overlapping curves in both training and validation metrics indicate an optimal learning process, with minimal overfitting.

The Xception model showed rapid improvements during early training stages, achieving training accuracies above 99% from epoch 10 onwards. However, validation performance was notably inconsistent, with accuracy values oscillating significantly across epochs.

Figure 3. Accuracy (left) and loss (right) curves during training and validation of Xception.



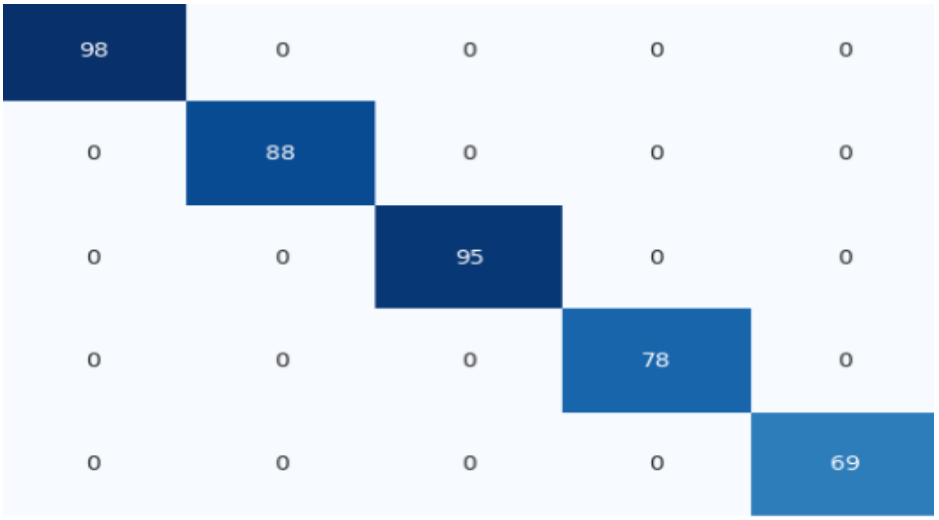
Despite achieving strong training accuracy, the final validation accuracy settled at 96.88%, slightly below the highest observed (97.22%). The performance curve (Figure 3) shows sharp oscillations in validation metrics, indicating challenges in model generalization and stability.

Further analysis using the confusion matrix revealed a total of 12 misclassified images across the 9 classes. The errors were concentrated in visually similar categories such as inflammatory vs. infectious-degenerative conditions, and obstructive vs. density-related abnormalities. This suggests that Xception struggled to separate overlapping radiographic patterns, possibly due to underfitting in the deeper layers or insufficient regularization.

These results position DenseNet121 as the top-performing architecture in this study, outperforming InceptionV3 in both stability and final accuracy.

3.2 Confusion Analysis

Figure 4 Confusion Matrix Analysis



Confusion matrices indicated that DenseNet121 struggled less with misclassifying visually similar conditions such as infiltration vs. pneumonia, thanks to its dense connections enabling nuanced feature retention. In contrast, InceptionV3 showed higher precision but slightly lower recall for subtle classes.

4. Discussion

The comparative results highlight the importance of CNN architecture design in thoracic anomaly detection. DenseNet121 demonstrated superior performance due to its ability to preserve and reuse features through dense connections. This is particularly beneficial in chest X-ray analysis, where subtle differences in texture or density are critical.

Xception, while less deep, benefits from its efficient separable convolutions, enabling robust generalization with fewer parameters. InceptionV3, though still effective, showed limitations in distinguishing between complex overlapping patterns, which may be mitigated through further fine-tuning or hybrid approaches.

These findings are consistent with prior studies showing that deeper and more recent architectures tend to outperform older models in medical imaging tasks, especially when trained with balanced datasets and proper regularization.

5. Conclusion

This study compared the effectiveness of DenseNet121, InceptionV3, and Xception for the multi-class classification of chest X-ray anomalies. DenseNet121 achieved the highest performance, indicating its strong potential for clinical deployment in diagnostic support systems. Xception also proved competitive, while InceptionV3 showed acceptable performance but room for improvement.

Overall, the results underscore the significance of model architecture in medical image classification. DenseNet121 proved to be the most robust and accurate, followed by InceptionV3 with strong but slightly less consistent performance, and Xception showing good learning capacity but less generalization under the current training conditions. Future work could explore ensemble models or hybrid architectures to combine their strengths and improve diagnostic precision in clinical applications.

6. References

1. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, y R. M. Summers, “ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases,” en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 3462–3471, doi: 10.1109/CVPR.2017.369.

2. P. Rajpurkar *et al.*, “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” *arXiv preprint arXiv:1711.05225*, 2017. [En línea]. Disponible: https://arxiv.org/abs/1711.05225

3. G. Huang, Z. Liu, L. van der Maaten, y K. Q. Weinberger, “Densely Connected Convolutional Networks,” en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.

4. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, y Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

5. J. Irvin *et al.*, "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison," en *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, 2019, pp. 590–597, doi: 10.1609/aaai.v33i01.3301590.

6. I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, y A. Saalbach, "Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification," *Scientific Reports*, vol. 9, no. 1, p. 6381, 2019, doi: 10.1038/s41598-019-42521-2.scholar.google.ae

7. L. Yao, J. Prosky, E. Poblenz, B. Covington, y K. Lyman, "Weakly Supervised Medical Diagnosis and Localization from Multiple Resolutions," *arXiv preprint arXiv:1803.07703*, 2018. [En línea]. Disponible: https://arxiv.org/abs/1803.07703

8. T. Wang, Y. Xiao, Y. Zhang, y Z. Wang, "ChestX-ray14: A Large-Scale Chest X-ray Dataset with Multi-Label and Annotated Data," *Data in Brief*, vol. 30, p. 105409, 2020, doi: 10.1016/j.dib.2020.105409.

9. S. Jaeger *et al.*, "Automatic Tuberculosis Screening Using Chest Radiographs," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 233–245, 2014, doi: 10.1109/TMI.2013.2284099.

10. D. S. Kermany *et al.*, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018, doi: 10.1016/j.cell.2018.02.010.

11. M. T. Islam, M. A. Aowal, A. T. Minhaz, y K. Ashraf, "Abnormality Detection and Localization in Chest X-Rays Using Deep Convolutional Neural Networks," *arXiv preprint arXiv:1705.09850*, 2017. [En línea]. Disponible: https://arxiv.org/abs/1705.09850

12. H. Rajpurkar *et al.*, "Deep Learning for Chest Radiograph Diagnosis: A Retrospective Comparison of the CheXNeXt Algorithm to Practicing Radiologists," *PLoS Medicine*, vol. 15, no. 11, p. e1002686, 2018, doi: 10.1371/journal.pmed.1002686.
13. F. Rad, "X-ray Lung Diseases Images (9 Classes)," Kaggle, 2022. [En línea]. Disponible en: https://www.kaggle.com/datasets/fernando2rad/x-ray-lung-diseases-images-9-classes