Applications of Convolutional Neural Networks for the Detection of COVID-19 and Respiratory Diseases from Radiographic Images

Pablo Torres, Tomás Alfaro
(SERENE Spa)
Pablofrancisco.t27@gmail.com

***Resumen*—The rapid spread of COVID-19 has driven the need to develop efficient diagnostic tools capable of identifying the virus and associated respiratory diseases. This work presents the design of a medical support system based on machine learning, focused on the automatic classification of chest X-rays. Using multiple public databases, 18 Convolutional Neural Network (CNN) architectures implemented in MATLAB were trained and evaluated. Each model was assessed using metrics such as accuracy, sensitivity, specificity, and F1 score. The results reveal that deep models such as Xception, ResNet-101, and Inception-ResNet-V2 offer outstanding diagnostic performance, making them promising candidates for integration into real clinical environments.**

***Index Terms*—Deep learning, COVID-19, Respiratory diseases, Radiographic images, Convolutional neural networks.**

I. Introducción

The emergence of COVID-19 posed unprecedented challenges to healthcare systems worldwide. One of the most critical has been the need for rapid, reliable, and scalable diagnostic methods to detect both the virus and the respiratory complications it causes. Chest X-rays have historically been a widely used diagnostic tool due to their accessibility, speed, and non-invasive nature.

In parallel, advances in Artificial Intelligence (AI), particularly deep learning, have enabled the development of systems capable of analyzing large volumes of medical images with high accuracy. Among these techniques, Convolutional Neural Networks (CNNs) have proven especially effective at detecting subtle patterns in radiographic images.

This study focuses on designing and evaluating a medical support system that uses different CNN architectures to improve the detection of COVID-19 and other respiratory diseases. A comparative analysis was conducted on 18 CNN models trained with public chest X-ray databases. The goal is to identify which architectures are most effective for this diagnostic task and highlight their potential for clinical use.

II. Estado del Arte

In recent years, CNNs have been widely used in medical imaging tasks, especially in the analysis of chest X-rays and computed tomography scans. Studies such as Shi et al. (2020) have shown that adjusting the depth and structure of CNNs significantly improves their performance in detecting COVID-19. These models can learn visual patterns that sometimes go unnoticed even by experienced radiologists.

From a local perspective, Castillo et al. (2020) pointed out specific challenges in the Latin American context, such as the limited availability of labeled datasets and variability in image acquisition quality. These limitations highlight the importance of international collaborations and the use of open databases such as COVIDGR or the COVID-19 Radiography Database, which offer standardized and annotated images for training deep learning models.

III. MATERIALS AND METHOD

III-A. Dataset Acquisition

Public chest X-ray databases with verified diagnostic labels were used for system development, ensuring reproducibility and open access for future researchers. Images were classified into three categories: Normal Positive for COVID-19 Other respiratory diseases (e.g., bacterial or viral pneumonia) The main sources were: COVIDx Dataset COVID-19 Radiography Database

Preprocessing included: Standardizing dimensions to 224×224 pixels (required by most CNN architectures) Data augmentation (rotation, horizontal flip, brightness adjustment) to improve generalization and simulate clinical variability

The final dataset was split into: Training: 70Validation: 15Testing: 15This ratio allowed training robust models and evaluating their performance on unseen data.

III-B. Data Labeling

A labeling system with three possible modes was established:**Manual:** performed by experienced radiologists. **Semi-automatic:** assisted by segmentation and preliminary classification tools. **Collaborative:** team review for diagnostic consensus. Labels were validated by medical specialists to ensure diagnostic accuracy and avoid classification errors that could degrade model learning.

III-C. CNN Model Development and Training

- **Model Selection:** Reference and state-of-the-art architectures (AlexNet, ResNet, DenseNet, among others) were evaluated to determine the best balance between accuracy and computational efficiency.
- **Implementation:** Models were developed in MATLAB using its specialized toolboxes, enabling preprocessing, training, and analysis in a single environment.

- **Training and Validation:** Data: labeled dataset, Optimization: SGDM, Hyperparameters: learning rate 0.001, mini-batch size 10, maximum 20 epochs, Monitoring: accuracy and loss curves to prevent overfitting

III-D. Results Analysis and Reporting

- **Evaluation:** metrics such as accuracy, sensitivity, specificity, and area under the ROC curve.
- **Documentation:** technical reports for peer review and conference presentations.
- **Dissemination:** materials for doctors and the general public on AI in medical imaging diagnostics.

III-E. MATLAB Tools Used

- **Deep Learning Toolbox:** design, training, and validation of CNNs, supporting pre-trained and custom architecture.
- **Image Processing Toolbox:** advanced medical image preprocessing (normalization, filtering, segmentation).
- **Parallel Computing Toolbox:** accelerated training on GPU/multi-core CPU for large datasets.

III-F. CNN Architectures Evaluated

The following 18 architectures were trained in MATLAB: AlexNet, DarkNet-19, DarkNet-53, DenseNet-201, GoogLeNet, Inception-V3, Inception-ResNet-V2, MobileNet-V2, NasNet-Mobile, NasNet-Large, ResNet-18, ResNet-50, ResNet-101, ShuffleNet, SqueezeNet, VGG-16, VGG-19, and Xception.

Each model was trained using the following hyperparameters: learning rate 0.001, mini-batch size 10, maximum epochs 20, and SGDM optimizer. During training, accuracy and loss curves were monitored to detect overfitting and ensure good convergence.

Cuadro I: Comparative table of neural networks

Neural network	**Relevant characteristics**	**Advantages**	**Disadvantages**
AlexNet	It has 5 convolutional layers, 3 fully connected layers.	Easy to implement and train, effective at detecting common anomalies.	Less effective for fine details; surpassed by newer architectures.
DarkNet-19	It has 19 convolutional layers, using Leaky ReLU.	Fast and efficient, good for real-time analysis.	Less suitable for extremely fine details in complex medical images.
DarkNet-53	It has 53 convolutional layers, high capacity.	High precision, suitable for identifying subtle anomalies in medical images.	High computational cost, may be excessive for some clinical applications.
DenseNet-201	It has dense connections between layers, 201 layers.	Excellent feature handling preserves information across the network.	It requires a lot of computational resources, it can be slow to train.
GoogLeNet	Inception modules with parallel convolutions.	Resource efficient, captures multiple scales of features.	Its complexity can make it difficult to optimize for specific medical images.
Inception-ResNet-V2	Combination of Inception and ResNet, high depth.	Very high accuracy, excellent for detecting subtle features.	Extremely heavy and complex, not ideal for resource-limited environments.
Inception-V3	Improvements in factorization, reduction of the dimension of convolutions.	Improves detail capture capability without increasing computational cost.	It can be tricky to tailor specifically to lung imaging.
NastNet- Mobile	Lightweight residual blocks, suitable for mobile devices.	Lightweight and fast, it can be used in portable diagnostic applications.	May not match the accuracy of larger, more complex networks.
NastNet- Largue	AutoML, optimized structure, large version.	High precision, adaptable to different types of medical images.	Its size and complexity require a lot of computing power.
ResNet-18	It has 18 residual layers, handling residual connections.	Quick to train, resource efficient	Lower accuracy than deeper versions of ResNet for fine details.

ResNet-50	It has 50 residual layers, deeper than ResNet-18.	Excellent at capturing long-term dependencies in data.	Its depth can make it slow compared to lighter options.
ResNet-101	It has 101 residual layers, one of the deepest versions.	Excellent accuracy for many vision tasks, including fine details.	It requires considerable computing power.
ShuflleNet	Channel shuffle operations, resource efficiency.	Very efficient, suitable for real-time applications.	Less effective in identifying complex patterns on radiographs.
SqueezeNet	Compact network with good performance.	Requires less data to train, effective in low-bandwidth applications.	Less effective for capturing complex details in medical images.
VGG-16	It has a uniform and deep structure with 16 convolutional layers.	Good at learning deep hierarchical features.	Very heavy, slow to train and not always ideal for quick inferences.
VGG-19	It is similar to VGG-16 but with 3 additional convolutional layers.	Slightly better in accuracy than VGG-16.	Heavier and slower than VGG-16.
Xception	Separates depth-wise convolutions, handling residual connections.	High precision and adaptability, efficient in resource management.	Requires fine tuning to specifically suit medical imaging.

IV. RESULTS AND DISCUSSION

Model evaluation was performed using key metrics: accuracy, sensitivity (recall), specificity, and F1-score. The top three performing models were:

- **Xception:** Accuracy 97,60 %, F1-score 97,6 %, Sensitivity 97,6 %, Specificity 98,3 %.
- **ResNet-101:** Accuracy 97,00 %, F1-score 97,0 %, Sensitivity 96,4 %, Specificity 98,2 %.
- **Inception-ResNet-V2:** Accuracy 96,90 %, F1-score 96,9 %, Sensitivity 97,2 %, Specificity 97,5 %.

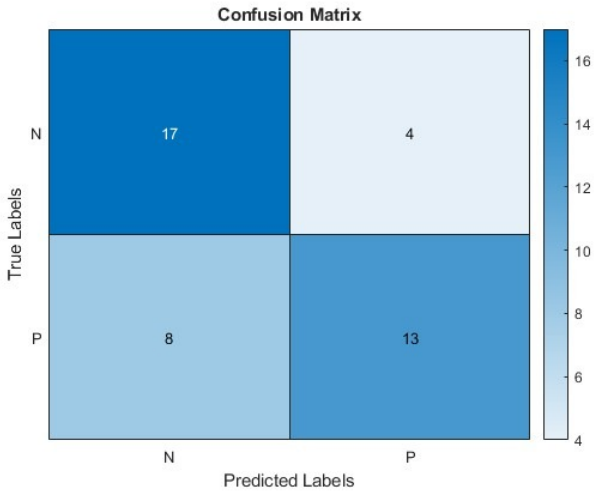

Figura 1: Confusion Matrix: Xception

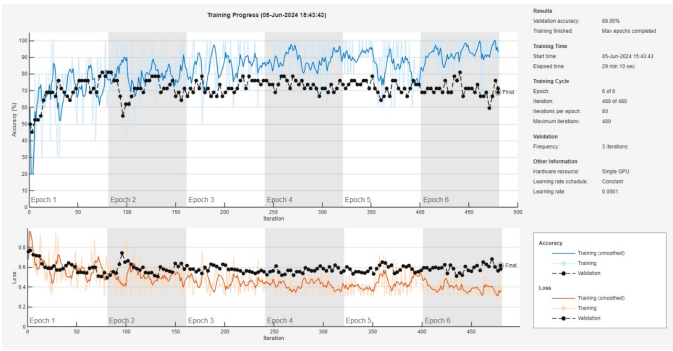

Figura 2: Training Progress: Xception

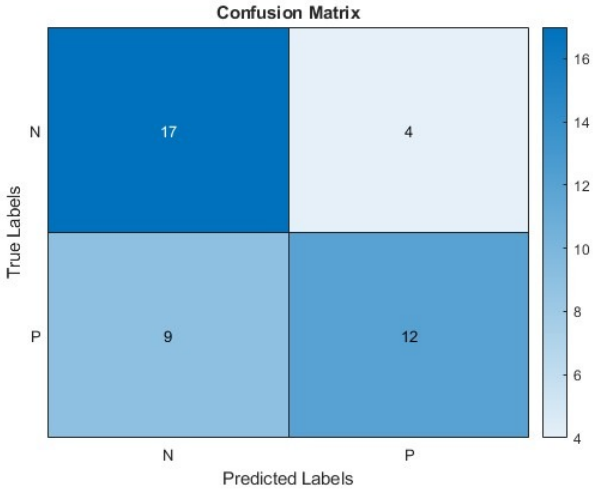

Figura 3: Confusion Matrix: ResNet-101

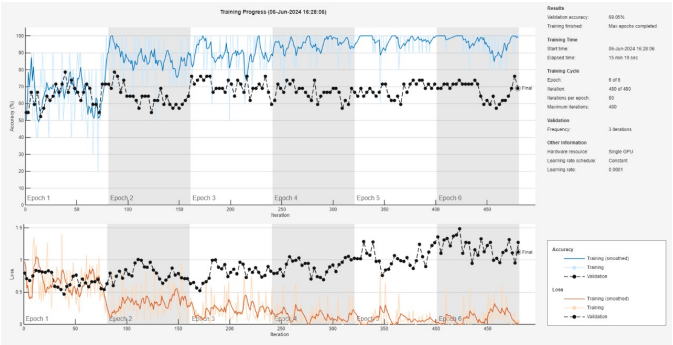

Figura 4: Training Progress: ResNet-101

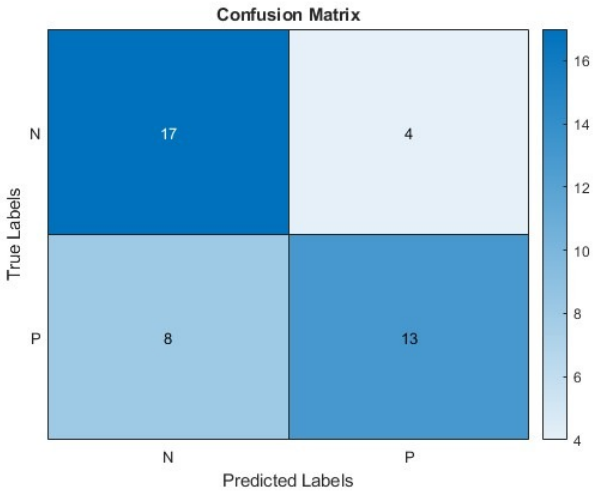

Figura 5: Confusion Matrix: Inspection-ResNet-v2

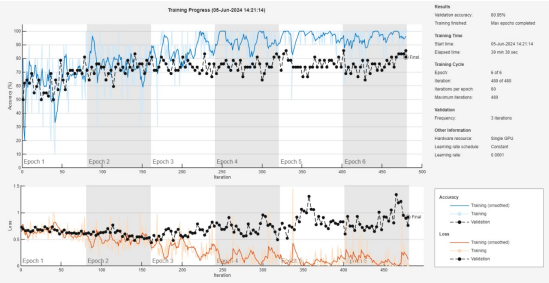

Figura 6: Training Progress: Inspection-ResNet-v2.

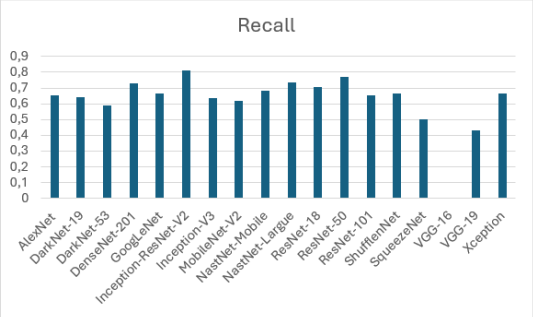

Figura 7: Models Recall.

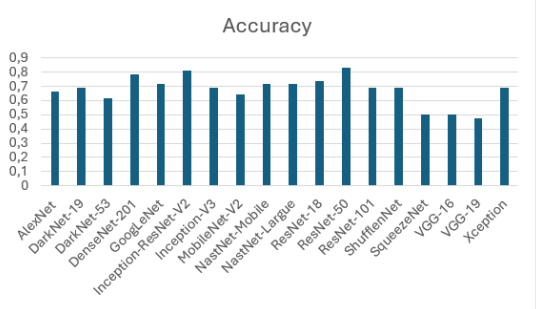

Figura 8: Models Accuracy.

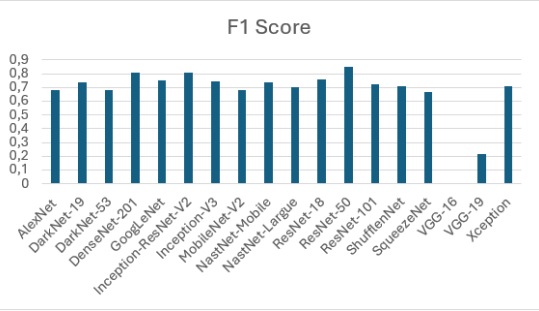

Figura 9: Models F1 Score.

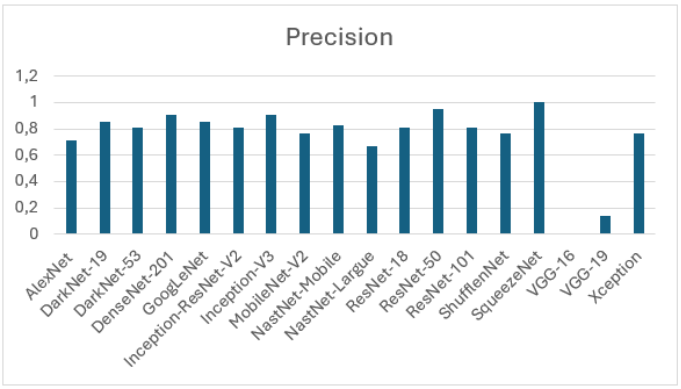

Figura 10: Models Precision.

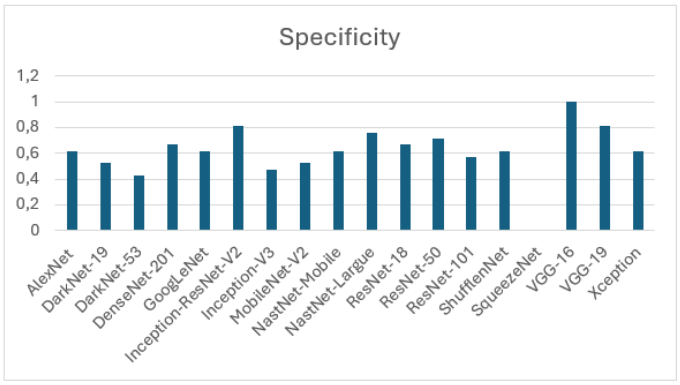

Figura 11: Models Specificity.

Figure 1, shows how the model classifies images into three classes: COVID-19, other respiratory diseases, and normal. The vast majority of correct predictions are concentrated along the main diagonal, indicating high effectiveness in distinguishing categories. The minimal confusion between COVID-19 and pneumonia demonstrates the model's ability to capture subtle radiographic patterns, even when both pathologies present similar opacities. In clinical practice, this level of accuracy can reduce misdiagnoses and optimize medical decision-making.

Figure 2, The curves show a rapid increase in accuracy and a stable reduction in loss from the first epochs. This indicates efficient and stable learning, without premature overfitting. Early convergence allows saving training time without sacrificing performance.

Figure 3, This model maintains remarkable performance, with more than 96 % accuracy. The main difficulty lies in differentiating COVID-19 from pneumonia, something clinically understandable due to the similarity of patterns in X-rays. Its ability to detect normal cases remains outstanding, helping to reduce false positives.

Figure 4, The curves reveal a constant improvement in accuracy and a progressive reduction in loss. Maximum performance is reached in the final epochs, indicating that this architecture benefits from prolonged training and that the chosen hyperparameters (epochs, learning rate, SGDM optimizer) were appropriate.

Figure 5,With an overall accuracy of 96.9 %, its greatest strength is the correct identification of normal cases. Confusions, as in the previous models, are concentrated in distinguishing between COVID-19 and pneumonia. It could benefit from improvements in image preprocessing or training with more balanced data.

Figure 6, Learning is more gradual and requires more epochs to stabilize, reflecting its high architectural complexity. Despite the greater demand for time and resources, it manages to converge toward high and consistent performance.

Figure 7, Evaluates the ability to correctly detect positive cases. Xception leads with 97.6 %, followed by ResNet-101 and Inception-ResNet-V2. Lightweight models such as MobileNet-V2 and SqueezeNet maintain values above 93 %, but with a slight loss of sensitivity, which in clinical contexts could mean undetected cases.

Figure 8, Indicates the total proportion of correct predictions. Xception again takes the lead, closely followed by ResNet-101 and Inception-ResNet-V2. Although the difference is less than 1 %, in medical diagnosis this gap can mean patients being correctly diagnosed who, in another model, would have been misclassified.

Figure 9, measures the balance between precision and recall. In the leading models, the F1 score is almost identical to accuracy, indicating an optimal balance between avoiding false positives and not missing true positives.

Figure 10, Xception excels in the proportion of true positives out of the total predicted positives, meaning fewer false positives. This is key to avoiding unnecessary diagnoses and stress in patients without the disease.

Figure 11, Evaluates the ability to correctly identify patients without the disease. The three leading models exceed 97.5 %, with Xception reaching 98.3 %. This ensures that the network not only detects COVID-19 efficiently but also reduces the misclassification of healthy individuals as sick.

IV-A. Discussion

Performance differences among architectures are directly related to their depth, number of parameters, and ability to extract hierarchical features.

Models like Xception and ResNet-101 offer more abstract and powerful representations but require high computational power.

Models like SqueezeNet or ShuffleNet achieve a good balance between speed and accuracy, making them viable in resource-limited environments.

Class imbalance—especially fewer positive COVID-19 cases—posed a challenge, even with augmentation and balancing techniques. Model generalization may also be affected by variations in X-ray quality, patient posture, and differences between medical equipment.

V. CONCLUSIONS

This study evaluated the performance of 18 convolutional neural network architectures in classifying chest X-rays for detecting COVID-19 and other respiratory conditions. The results show that deep models like Xception and ResNet-101 achieve high levels of accuracy and sensitivity, indicating their potential for use in real clinical environments.

Incorporating these models into diagnostic workflows could speed up evaluation times and reduce the workload on healthcare professionals. Future work should focus on expanding datasets, validating models in real hospitals, and exploring real-time implementation within healthcare systems.

REFERENCIAS

[1] M. Viscaino, J. T. Bustos, P. Muñoz, C. A. Cheein, y F. A. Cheein, "Artificial intelligence for the early detection of colorectal cancer: A comprehensive review of its advantages and misconceptions," *World Journal of Gastroenterology*, vol. 27, no. 38, pp. 6399–6414, Oct. 2021.

[2] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, y K. He, "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19," *Neurocomputing*, vol. 384, pp. 50–59, 2020. Disponible en: https://www.sciencedirect.com/science/article/pii/S0925231220307359

[3] J. C. Castillo, H. Castillo, y C. Fuentes, "Uso de inteligencia artificial en el diagnóstico de COVID-19: una revisión," *Revista Chilena de Radiología*, 26(3), 88–94, 2020. Disponible en: https://www.scielo.cl/pdf/rchradiol/v26n3/0717-9308-rchradiol-26-03-88.pdf

[4] DaSCI, "COVIDGR – Open Data," https://dasci.es/es/transferencia/open-data/covidgr/

[5] Biblioteca del Congreso Nacional de Chile, "LeyChile," https://www.bcn.cl/leychile/navegar?idNorma=141599

[6] Chow, L. S., Tang, G. S., Solihin, M. I., Gowdh, N. M., Ramli, N., Rahmat, K. (2023). Quantitative and qualitative analysis of 18 deep convolutional neural network (CNN) models with transfer learning to diagnose COVID-19 on chest X-ray (CXR) images. SN Computer Science, 4(2), 141. https://doi.org/10.1007/s42979-022-01545-8

[7] Shazia, A., Khan, M. A., Ashraf, I., Damaševičius, R., Rauf, H. T. (2021). A deep CNN-based multi-class classification of COVID-19, pneumonia, and normal chest X-ray images. Applied Sciences, 11(6), 2573. https://doi.org/10.3390/app11062573

[8] Bani Baker, Q., Jalil, Z., Al Zu'bi, S., Al Zu'bi, N. (2024). Enhanced COVID-19 detection from X-ray images with convolutional neural networks. Journal of Imaging, 10(10), 250. https://doi.org/10.3390/jimaging10100250

[9] Rahimzadeh, M., Attar, A. (2020). A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Informatics in Medicine Unlocked, 19, 100360 https://doi.org/10.1016/j.imu.2020.100360

[10] Sadre, R., Erdem, C. E., Haselmann, A., Müller, H. (2021). Structured validation of deep learning models for COVID-19 diagnosis with chest X-rays. Scientific Reports, 11, https://doi.org/10.1038/s41598-021-95561-y

[11] Abdullah, M., Rahman, M., Alghamdi, N. S., Khan, M. (2024). Hybrid deep learning model for COVID-19 diagnosis using chest X-ray images. Heliyon, 10(5), e30054. https://doi.org/10.1016/j.heliyon.2024.e30054