

## Article

# Diagnóstico de COVID-19 en imágenes de rayos X de tórax mediante de Redes Neuronales Convolucionales

Mirko Jerber Rodríguez Mallma<sup>1,\*</sup>

**Citation:** To be added by editorial staff during production.

Academic Editor: Firstname  
Lastname

Received: date  
Revised: date  
Accepted: date  
Published: date



**Copyright:** © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

<sup>1</sup> Facultad de Ingeniería Industrial y de Sistemas, Universidad Nacional de Ingeniería, Lima, Perú.

\* Correspondence: [mjrodriguezm@uni.pe](mailto:mjrodriguezm@uni.pe)

**Abstract:** Los Rayos X de tórax son una prueba rápida y económica que puede ser usada para diagnosticar el COVID-19, pero aún a la fecha éstas no son utilizadas como una prueba de primera mano para el diagnóstico de esta condición médica. Investigaciones sugieren que el uso del aprendizaje profundo (DL) puede ayudar a explotar el uso de las radiografías de tórax ya que las redes neuronales convolucionales (CNN) han demostrado una alta precisión en el diagnóstico de enfermedades a partir de imágenes médicas. Por lo tanto, en el presente trabajo se desarrolla un modelo capaz de predecir el diagnóstico de COVID-19 en pacientes a partir de imágenes de Rayos X de tórax, a través del desarrollo de un modelo de Deep Learning (DL) basado en Redes Neuronales Convolucionales (CNN) pre entrenadas como VGG16, ResNet50, InceptionV3 y DenseNet201, y aplicando técnicas de preprocesamiento de datos como Data Augmentation y Normalization para evitar el sobreajuste y obtener resultados más eficientes. A partir de dicho dataset de imágenes se calcularon métricas de performance para cuatro escenarios de experimentación utilizados en este estudio, a saber, Accuracy, Specificity, Sensitivity y AUC. El mejor modelo para el diagnóstico de COVID-19 se consiguió con la red convolucional pre entrenada de arquitectura tipo DenseNet201, la cual consiguió un Accuracy de 89.8%, Sensitivity de 84.5%, Specificity de 95.0% y AUC de 89.10%.

**Keywords:** COVID-19; diagnosis; Rayos X; deep learning; prediction

## 1. Introduction

El COVID-19 es una enfermedad infecciosa causada por el coronavirus 2 (SARS-CoV-2) [1]. El virus se observó por primera vez en Wuhan, China, en diciembre de 2019 y luego se propagó por todo el mundo. El 11 de marzo de 2020, la Organización Mundial de la Salud (OMS) declaró al COVID-19 una pandemia [1]. El virus se propaga principalmente a través de gotículas respiratorias o aerosoles que se transmiten a la boca, la nariz o los ojos de las personas que están muy cerca. Los síntomas más comunes incluyen fiebre alta, tos continuada y dificultad para respirar [2].

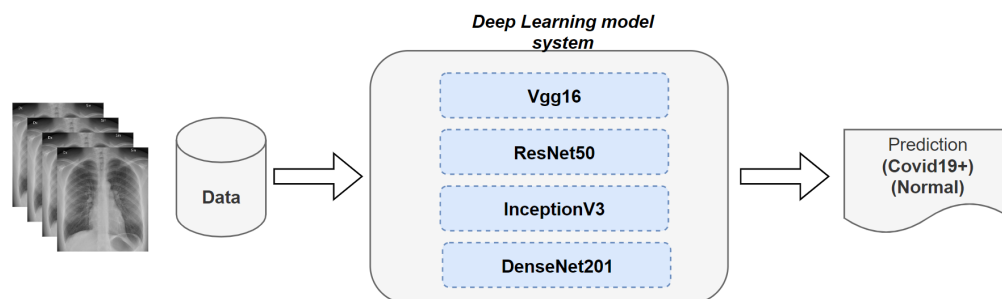
El COVID-19 generalmente se diagnostica mediante una prueba de RT-PCR [3] y, a menudo, se complementa con radiografías de tórax, incluidas imágenes de Rayos X y tomografías computarizadas (TC) [4]. Las máquinas de Rayos X están ampliamente disponibles en todo el mundo y proporcionan imágenes rápidamente, por lo que algunos investigadores [5] han recomendado exploraciones de tórax para la detección durante la pandemia. A diferencia de la prueba RT-PCR, las exploraciones de tórax brindan información sobre el estado de la infección (es decir, presencia o ausencia de la enfermedad) y la gravedad de la enfermedad. Para realizar un diagnóstico por Rayos X se requiere un equipo relativamente barato y se puede realizar rápidamente en salas aisladas con un dispositivo portátil de radiografía de tórax (CXR), lo que reduce el riesgo de infección dentro de los hospitales [6, 7].

El Machine Learning (ML) y el Aprendizaje Profundo (DL) son un subconjunto de algoritmos de Inteligencia Artificial (AI) que “aprenden” de forma automática a identificar categorías y son utilizados para realizar pronósticos a partir de datos [8]. El uso de estos algoritmos como herramientas de soporte a la toma de decisiones médicas y su aplicación en el pronóstico, diagnóstico y tratamiento de enfermedades se ha desarrollado recientemente [9]; sin embargo, aún existen determinadas condiciones que hacen difícil una adopción mucho más generalizada [10-15]. Varios estudios han utilizado redes neuronales convolucionales (CNN) para el problema de la clasificación de imágenes, la mayoría de los cuales crean diferentes arquitecturas para las redes neuronales. Las redes neuronales convolucionales son uno de los mejores algoritmos de aprendizaje profundo. En los últimos tiempos utilizamos con frecuencia las Redes Neuronales Convolucionales en coches autónomos, reconocimiento facial, clasificaciones de imágenes para detección, etc. La evolución de las Redes Neuronales Convolucionales está aumentando paulatinamente en los últimos años. En todas las categorías, como medicina, automóviles, tecnología móvil, etc., utilizamos redes neuronales convolucionales para el procesamiento de imágenes, la clasificación de imágenes y la detección de objetos [16]. Existen diferentes tipos de arquitecturas en la CNN como VGGNet (Visual Geometry Group), ResNet (Residual Network), Dense Net, Inception Net que han sido utilizadas en el análisis de imágenes médicas para el diagnóstico de enfermedades [17].

Tomando en cuenta los estudios previos [10-17], en los que se evidencia que los algoritmos de IA, específicamente las CNN, son una poderosa herramienta que puede utilizarse en el área médica para el diagnóstico de COVID-19 basado en el procesamiento de imágenes de Rayos X de tórax, en el presente estudio se utiliza un dataset de 2400 pacientes para analizarlos, procesarlos y (1) pronosticar el diagnóstico de COVID-19 de los pacientes y (2) extender el uso de nuevas herramientas de IA como soporte a la toma de decisiones médicas.

## 2. Materials and Methods

Se propone la construcción de un modelo de Deep Learning (DL) para el reconocimiento de imágenes médicas de Rayos X de tórax a través del uso de modelos de Redes Neuronales Convolucionales y mediante la técnica de Aprendizaje por Transferencia para el diagnóstico de COVID-19. El objetivo es pronosticar si un paciente puede ser diagnosticado como positivo para COVID-19 o no, por lo cual, se optó por un enfoque de aprendizaje supervisado vía una clasificación binaria. La **Figura 1** muestra el esquema del sistema de predicción a construir planteado.

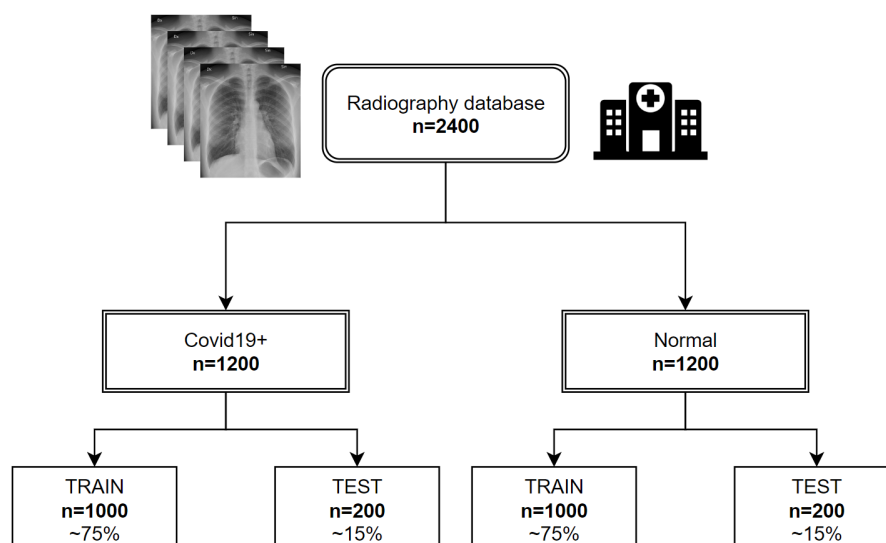


**Figure 1.** Sistema de DL propuesto.

### 2.1. Dataset

Para este estudio, se ha utilizado un dataset de imágenes del “*Italian Society of Medical and Interventional Radiology (SIRM) COVID-19 DATABASE, Novel Coronavirus 2019 Dataset*” desarrollado por Joseph Paul Cohen las cuales, a su vez fueron recolectadas de diferentes publicaciones. actualmente, el dataset es de uso público y puede ser accedido accediendo a los servidores de Kaggle vía una cuenta de usuario: <https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database>.

Para los experimentos de este estudio se han tomado 2400 imágenes en formato png de 256x256 píxeles, de las cuales 1200 pertenecen a pacientes diagnosticados con COVID-19 y 1200 a pacientes que no tienen la enfermedad, llamados pacientes de control, a los cuales se les etiquetó con de tipo Normal. Adicionalmente, para la construcción de nuestros modelos de DL, se ha considerado dividir el dataset en 2000 imágenes de entrenamiento y 400 para pruebas, tal y como se muestra en la **Figura 2**.



**Figure 2.** División del dataset para la construcción de los modelos.

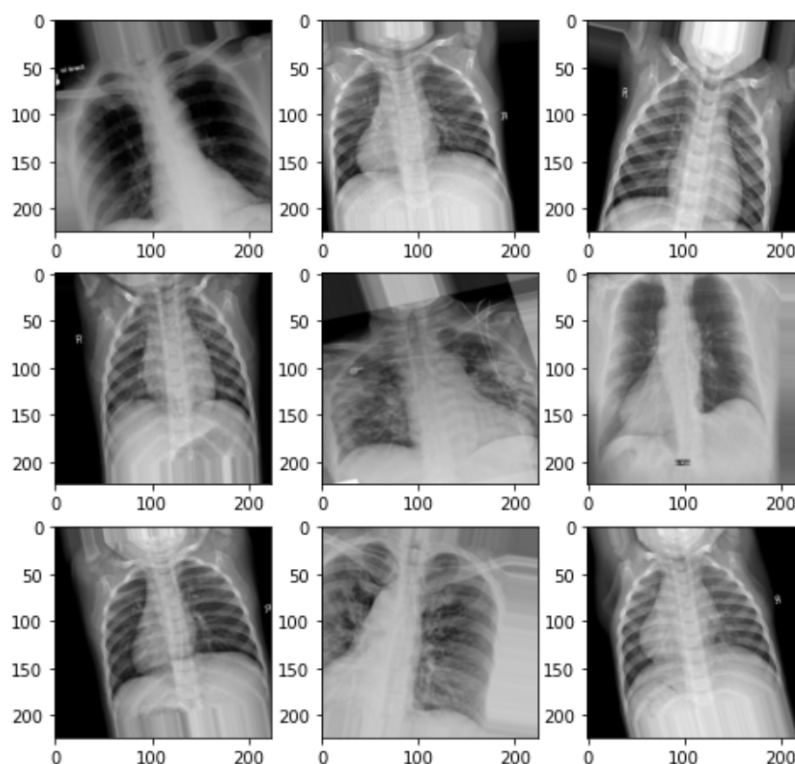
El dataset está compuesto por imágenes (matrices de datos) que corresponden a los datos de la imagen de Rayos X de cada paciente, siendo éstas las variables de entrada al sistema (variables independientes); y una variable de salida (variable dependiente) que representa la categorización de cada paciente como: 1: el paciente tiene COVID-19, y 0: el paciente NO tiene COVID-19 o Normal.

## 2.2. Preprocesamiento de datos

Debido a que la cantidad de imágenes en el dataset es la misma para ambas clases (0 y 1), no se requirió realizar un proceso de balanceo de datos; sin embargo, con el objetivo evitar optimizar y mejorar los resultados del pronóstico y construir el sistema de DL de forma efectiva, se ha realizado los siguientes procedimientos como parte de preprocesamiento de datos:

### 2.2.1. Incremento de datos (data augmentation)

Es ideal que podamos aprovechar el hecho de que las imágenes se pueden preprocesar mediante mecanismos de giro, volteo, zoom, etc., para aumentar los datos de entrenamiento y evitar el sobreajuste. Primero, cambiamos el tamaño de las imágenes a  $224 \times 224$  píxeles para reducir el volumen de datos porque las redes neuronales pre entrenadas a utilizarse soportan dicha dimensión solamente. Luego, aumentamos las imágenes de entrenamiento a través de transformaciones aleatorias para aumentar la variedad de imágenes. Específicamente, manipulamos los parámetros (a) volteando la imagen sobre su eje vertical (flip), (b) girando la imagen 15 grados en el sentido de las manecillas del reloj, (c) desplazando aleatoriamente la imagen de forma horizontal o verticalmente en un 15 % del tamaño inicial de la imagen, y (d) acercar o alejar aleatoriamente un 20 %. El aumento de datos es esencial para evitar el sobreajuste en muestras pequeñas porque las imágenes variadas adicionales evitan que la CNN se exponga dos veces a la misma imagen. La **Figura 3** muestra el resultado del proceso de aumentación de datos (data augmentation).



**Figure 3.** Imágenes resultado del proceso de Aumentación de Datos (data augmentation).

### 2.2.2. Data normalization

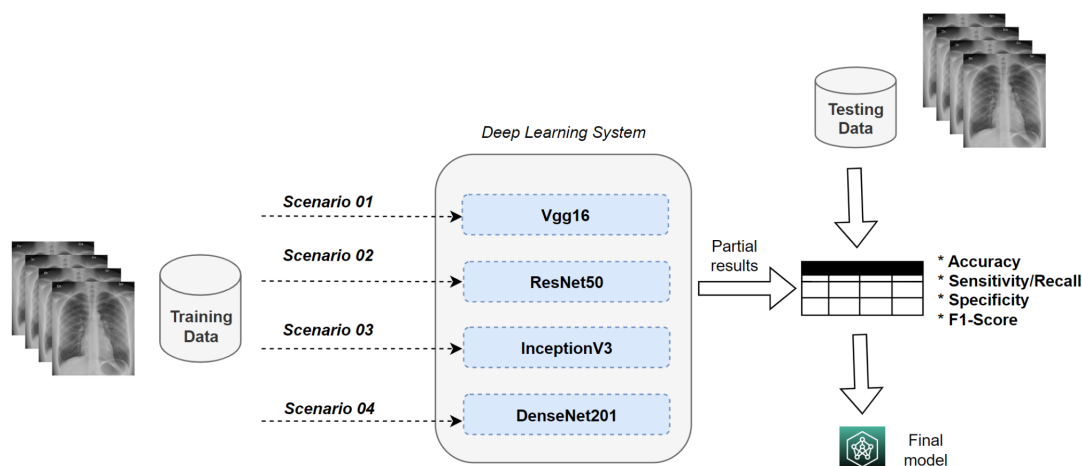
Las imágenes del dataset han sido reescaladas y normalizadas para poder ser procesadas eficientemente por las Redes Neuronales Convolucionales. El proceso consiste en dividir por 255 el valor de cada píxel en las imágenes, de este modo los nuevos valores de los píxeles están en el rango de 0 y 1. La **Figura 4** intenta mostrar este procedimiento.



**Figure 4.** Proceso de normalización de datos aplicado a cada valor de los píxeles.

### 2.3. Modelos de Deep Learning

Para la construcción, validación y evaluación del sistema de DL se utilizó el proceso que se muestra en la **Figura 5**, el cual consistió en utilizar el dataset de entrenamiento para construir y validar cuatro modelos de DL bajo cuatro escenarios de experimentación y de cuyos resultados se eligió aquel con las métricas de desempeño (performance) con los valores más cercanos a 1.



**Figure 5.** Proceso de construcción, validación y evaluación del modelo..

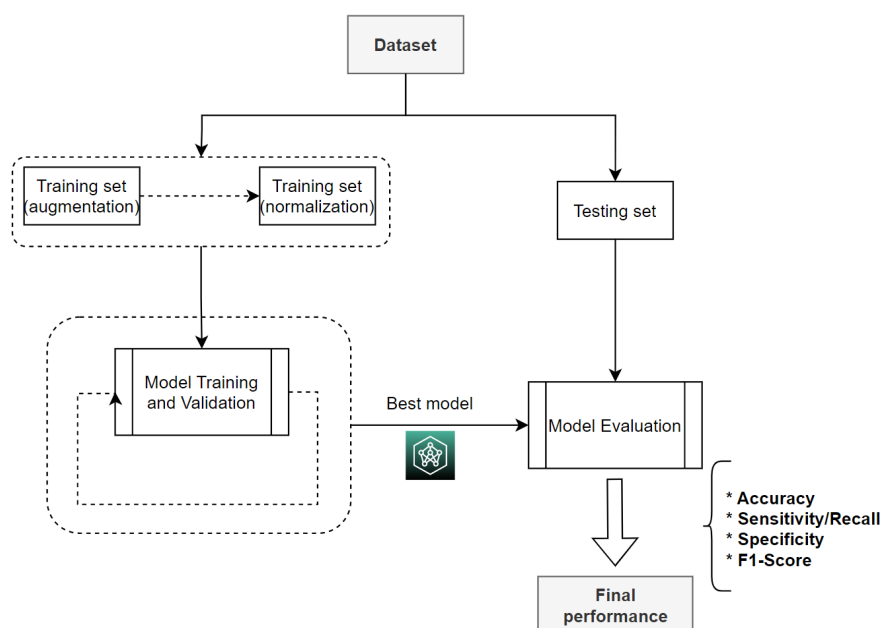
Para el proceso de entrenamiento se utilizó la técnica de Aprendizaje por Transferencia (Transfer Learning) que consiste en reutilizar modelos pre entrenados en otra tarea de reconocimiento de imágenes, en nuestro caso utilizamos cuatro modelos, a saber, VGG16, ResNet50, InceptionV3 , y DenseNet201, los cuales fueron entrenados previamente en reconocer 1000 objetos (clases) del dataset de imágenes denominado Imagenet [18].

En el proceso de Aprendizaje por Transferencia, se toma el Feature Extractor (FE) del modelo original, se crea un nuevo modelo tomando como base el modelo pre entrenado y se entrena solo el módulo de clasificación del nuevo modelo creado, de este modo se logra obtener resultados de pronóstico más óptimos con menos tiempo de entrenamiento y menor número de imágenes. El proceso de Aprendizaje por Transferencia se muestra en la Figura 6.

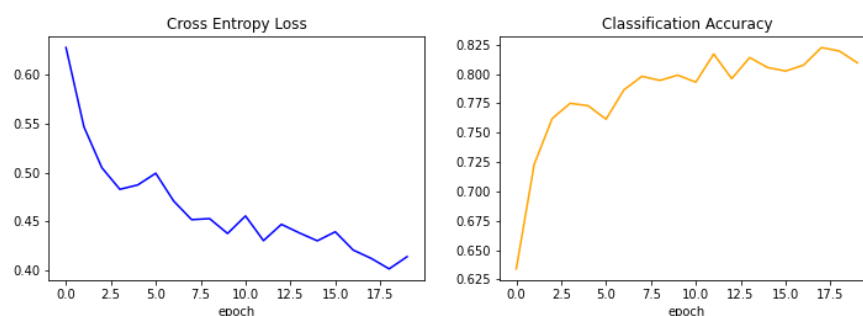


**Figure 6.** Proceso de Aprendizaje por Transferencia.

Finalmente, se utilizaron las métricas de Accuracy, Sensitivity, Specificity, Precision, Balanced Accuracy, F1-Score y Área bajo la curva (AUC) para evaluar la performance de predicción de cada modelo por cada uno de los cuatro escenarios del sistema de DL, tal y como se muestra en la **Figura 7**. Además, la **Figura 8** muestra el Error de entrenamiento versus y el Accuracy logrado para cada una de las 20 épocas consideradas en el entrenamiento.



**Figure 7.** Proceso de entrenamiento del modelo de DL.



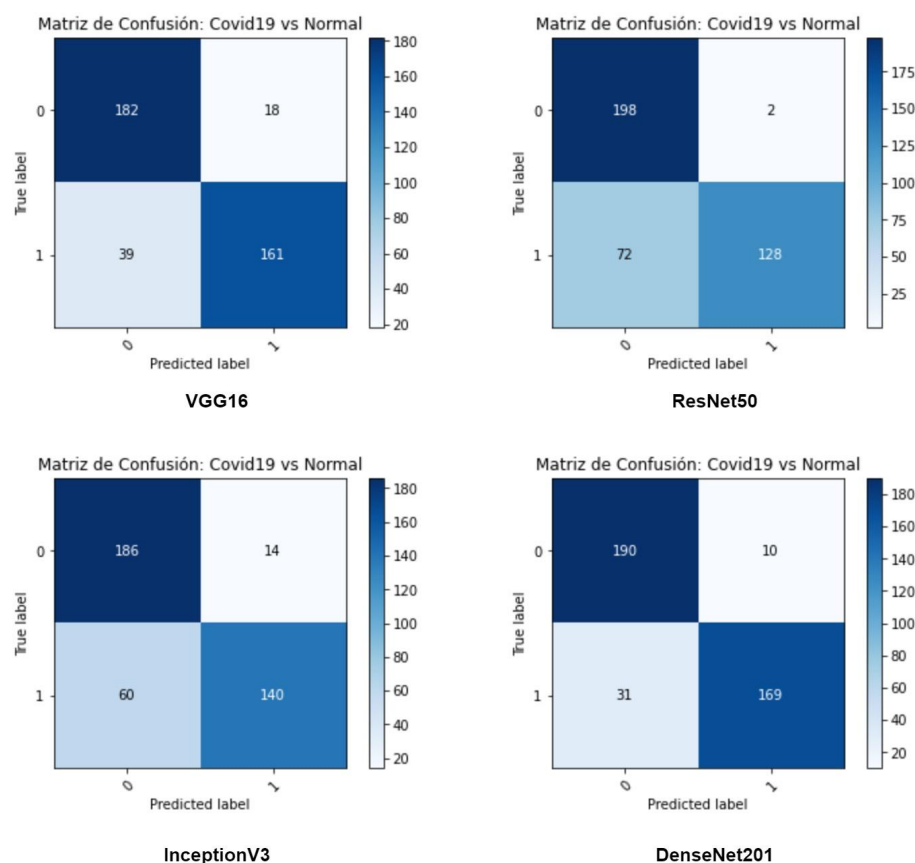
**Figure 8.** Error vs Accuracy en el proceso de entrenamiento de los modelos de DL.

Para construir los modelos de ML se usaron las librerías scikit-learn 1.0.2 [19] y tensorflow 2.9.2 en Python versión 3.8.16 bajo el entorno de Google Collab. Los algoritmos y recursos se pueden encontrar en la siguiente dirección: <https://github.com/mirkorodriguez/covid19-pretrained-models>.

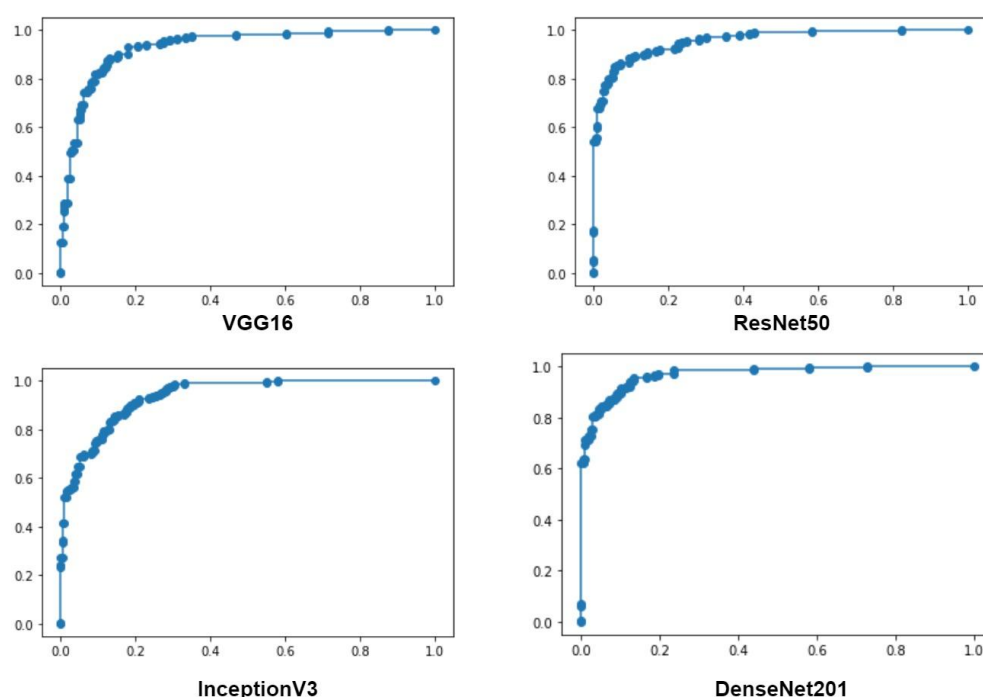
### 3. Results

#### 3.1. Performance de los modelos

La **Figura 9** muestra las matrices de confusión obtenidas como resultado de evaluar los modelos de DL, en los cuatro escenarios, con en el conjunto de datos de prueba. Además, la **Figura 10** muestra las curvas AUC para cada uno de los escenarios considerados en el presente estudio.



**Figure 9.** Matriz de Confusión de los modelos de DL evaluados con el dataset de test.



**Figure 10.** Curvas ROC de los modelos de DL evaluados con el dataset de test.

La **Tabla 1** muestra el resultado de la comparación de los resultados de los cuatro modelos de predicción usados en esta investigación.

**Table 1.** Resumen de la performance de los modelos.

Model Name	Accuracy	Sensitivity	Specificity	Precision	Balanced Accuracy	F1-Score	AUC
<b>VGG16</b>	0.856	0.805	0.910	0.899	0.824	0.858	0.800
<b>ResNet50</b>	0.815	0.640	<b>0.990</b>	<b>0.985</b>	0.733	0.815	0.776
<b>InceptionV3</b>	0.815	0.700	0.930	0.909	0.756	0.815	0.791
<b>DenseNet201</b>	<b>0.898</b>	<b>0.845</b>	0.950	0.944	<b>0.860</b>	<b>0.898</b>	<b>0.891</b>

Se observa que el mejor modelo de acuerdo a las métricas de performance, definidas previamente, es el modelo deconstruido con el modelo pre entrenado DenseNet201.

#### 4. Discussion

El uso de las técnicas de ML y DL en el área médica se ha incrementado en los últimos años y su aplicación como sistema de apoyo a la toma de decisiones médicas para el diagnóstico y pronosis de enfermedades sería de mucha utilidad [20]. Inspirados por el uso de estas técnicas en el dominio médico, este estudio planteó un método que permite pronosticar el diagnóstico de COVID-19 a partir de imágenes de Rayos X, demostrando que el uso de técnicas de DL para el diagnóstico de pacientes sería posible. Nuestro enfoque consistió en evaluar cuatro escenarios en los que se



utilizaron 4 modelos de DL pre entrenados. Después de seguir un proceso para la construcción de modelos de DL, en el que se aplicó además técnicas de data augmentation y normalization, se encontró que el mejor modelo para pronosticar el COVID-19, se produjo con el modelo de basado en la CNN de arquitectura DenseNet201 (Accuracy de 89.8%, Sensitivity de 84.5%, Specificity de 95.0% y AUC de 89.10%).

De los resultados obtenidos en este estudio, se evidencia que el modelo construido utilizando como base el modelo pre entrenado DenseNet201 es superior a los otros modelos basados en VGG16, ResNet50 e InceptionV3, tal como se muestra los resultados de la **Tabla 1**. Es pertinente resaltar que el preprocesamiento de datos, aplicado en este estudio a través de data augmentation y normalization, es muy importante para obtener resultados de pronóstico óptimos, tal como se evidencia en la **Tabla 1**.

La importancia de los resultados de este estudio va más allá del hecho de que el pronóstico médico de pacientes con COVID-19 es posible, si no que, también permite confirmar que un enfoque de DL, entendido como un marco de trabajo generalizable, aplicado en el dominio médico mediante el uso de datos históricos para pronosticar el futuro sería posible. En este caso, utilizamos imágenes de Rayos X histórica de pacientes con COVID-19 y de control para construir un sistema de DL que logró obtener muy buenos resultados de pronóstico del futuro y que puede ser usado por los profesionales médicos como herramienta de apoyo a la toma de decisiones médicas al momento de enfrentarse a nuevos casos de este tipo.

A pesar de los resultados alentadores obtenidos en esta investigación, se requiere mayor análisis antes de utilizar sus resultados directamente en la práctica médica; por ejemplo, evaluar la posibilidad de incluir otras variables y estudiar el impacto sobre los resultados en una serie de experimentos que permitan maximizar la probabilidad de éxito de la radiocirugía estereotáctica. Además, dada la naturaleza de los modelos de DL que requieren datos de calidad y en suficiente cantidad para un entrenamiento, validación y evaluación óptimos [21], se aconseja realizar nuevos experimentos con mayor cantidad de imágenes, para lo cual se sugiere incrementar casos de pacientes de otras instituciones o complementar con otras bases de datos para poder generalizar los resultados a una población más amplia y tener resultados más consistentes.

Por otro lado, se requiere realizar un análisis más exhaustivo acerca de la explicabilidad de los resultados obtenidos y sería importante verificar la interpretabilidad con técnicas como Grad-CAM, LIME, SHAP y otras, que están enfocadas en identificar los predictores más importantes para cualquier tipo de modelo de DL.

## References

- COVID-19 pandemic. [https://en.wikipedia.org/wiki/COVID-19\\_pandemic](https://en.wikipedia.org/wiki/COVID-19_pandemic). Accedido el 14 de enero del 2022.
- COVID-19 pandemic. [https://en.wikipedia.org/wiki/COVID-19\\_pandemic#Transmission](https://en.wikipedia.org/wiki/COVID-19_pandemic#Transmission). Accedido el 14 de enero del 2022.
- Reverse transcription polymerase chain reaction. [https://en.wikipedia.org/wiki/Reverse\\_transcription\\_polymerase\\_chain\\_reaction](https://en.wikipedia.org/wiki/Reverse_transcription_polymerase_chain_reaction). Accedido el 14 de enero del 2022.
- COVID-19 pandemic. [https://en.wikipedia.org/wiki/COVID-19\\_pandemic#Diagnosis](https://en.wikipedia.org/wiki/COVID-19_pandemic#Diagnosis). Accedido el 14 de enero del 2022.
- Ng M-Y, Lee EYP, Yang J, et al. *Imaging profile of the COVID-19 infection: radiologic findings and literature review*. Radiology. 2020;2(1):e200034.
- Baratella E, Crivelli P, Marrocchio C, et al. *Severity of lung involvement on chest X-rays in SARS-coronavirus-2 infected patients as a possible tool to predict clinical progression: an observational retrospective analysis of the relationship between radiological, clinical, and laboratory data*. J Brasil Pneumol. 2020;46(5):e20200226.
- Rubin GD, Ryerson CJ, Haramati LB, et al. *The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner society*. Chest. 2020;296(1):172–180.
- Alice Segato, Aldo Marzullo, Francesco Calimeri, and Elena De Momi. *Artificial intelligence for brain diseases: A systematic review featured APL Bioengineering* 4, 041503 (2020); <https://doi.org/10.1063/5.0011697>
- Basu, K., Sinha, R., Ong, A., & Basu, T. (2020). Artificial Intelligence: How is It Changing Medical Sciences and Its Future?. *Indian journal of dermatology*, 65(5), 365–370. [https://doi.org/10.4103/ijdr.IJD\\_421\\_20](https://doi.org/10.4103/ijdr.IJD_421_20)
- Vega-Huerta, H., Villanueva-Alarcón, R., Mauricio, D., Gamarra Moreno, J., Calderon Vilca, H. D. ., Rodriguez, D., & Rodriguez, C. (2022). *Convolutional Neural Networks on Assembling Classification Models to Detect Melanoma Skin Cancer*. *International Journal of Online and Biomedical Engineering (iJOE)*, 18(14), pp. 59–76. <https://doi.org/10.3991/ijoe.v18i14.34435>
- N. S. Silva, J. F. Piñas and D. Mauricio, "ProstaTest: A System for the Diagnosis of Prostate Cancer Based on IPSS, PSA, and Ultrasound Scan," 2021 IEEE Sciences and Humanities International Research Conference (SHIRCON), 2021, pp. 1-4, doi: 10.1109/SHIRCON53068.2021.9652012.
- Marcus, J.L., Sewell, W.C., Balzer, L.B. et al. *Artificial Intelligence and Machine Learning for HIV Prevention: Emerging Approaches to Ending the Epidemic*. *Curr HIV/AIDS Rep* 17, 171–179 (2020). <https://doi.org/10.1007/s11904-020-00490-6>
- Senders, J. T., Staples, P. C., Karhade, A. V., Zaki, M. M., Gormley, W. B., Broekman, M., Smith, T. R., & Arnaout, O. (2018). *Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review*. *World neurosurgery*, 109, 476–486.e1. <https://doi.org/10.1016/j.wneu.2017.09.149>
- Panesar, S. S., D'Souza, R. N., Yeh, F. C., & Fernandez-Miranda, J. C. (2019). *Machine Learning Versus Logistic Regression Methods for 2-Year Mortality Prognostication in a Small, Heterogeneous Glioma Database*. *World neurosurgery*: X, 2, 100012. <https://doi.org/10.1016/j.wnsx.2019.100012>
- Yang, Y. J., & Bang, C. S. (2019). *Application of artificial intelligence in gastroenterology*. *World journal of gastroenterology*, 25(14), 1666–1683. <https://doi.org/10.3748/wjg.v25.i14.1666>
- Rahul Chauhan; Kamal Kumar Ghanshala; R.C Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition", Graphic Era University, Dehradun, December 2018, India
- Nikolaou, V., Massaro, S., Fakhimi, M., Stergioulas, L., & Garn, W. (2021). *COVID-19 diagnosis from chest x-rays: developing a simple, fast, and accurate neural network*. *Health information science and systems*, 9(1), 36. <https://doi.org/10.1007/s13755-021-00166-4>
- J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- Brnabic, A., Hess, L.M. *Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making*. *BMC Med Inform Decis Mak* 21, 54 (2021). <https://doi.org/10.1186/s12911-021-01403-2>
- Abhinav Jain, Hima Patel, Lokesh Nagalapatti, Nitin Gupta, Sameep Mehta, Shanmukha Guttula, Shashank Mujumdar, Shazia Afzal, Ruhi Sharma Mittal, and Vitobha Munigala. 2020. *Overview and Importance of Data Quality for Machine Learning Tasks*. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20)*. Association for Computing Machinery, New York, NY, USA, 3561–3562. <https://doi.org/10.1145/3394486.3406477>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.