VIVE. Revista de Investigación en Salud
https://revistavive.org
Volumen 4 No. 12 septiembre-diciembre 2021
https://doi.org/10.33996/revistavive.v4i12.119
ISSN: 2664-3243
pp. 624 - 633

Detección de la tuberculosis con algoritmos de Deep Learning en imágenes de radiografías del tórax

Tuberculosis detection with Deep Learning algorithms in chest X-ray images

Detecção de tuberculose com algoritmos de aprendizagem profunda em imagens de raio-X do tórax

Juan Carlos Valero Gómez
jvalerog@unam.edu.pe
https://orcid.org/0000-0002–7077–7531
Universidad Nacional de Moquegua, Ilo - Perú

Alex Peter Zúñiga Incalla
azunigai@unam.edu.pe
https://orcid.org/0000-0003–4734–2262
Universidad Nacional de Moquegua, Ilo - Perú

Juan Carlos Clares Perca
jclaresp@unam.edu.pe
https://orcid.org/0000-0001–5146–2879
Universidad Nacional de Moquegua, Ilo - Perú

Recibido 13 de octubre 2021 / Arbitrado y aceptado 16 de noviembre 2021 / Publicado 30 de diciembre 2021

RESUMEN

Se estima que en 2019 murieron alrededor de 1,4 millones de personas infectadas por tuberculosis, gran parte de ellos en países en desarrollo. Si la tuberculosis se hubiera diagnosticado oportunamente se habría evitado la muerte de las personas infectadas. Uno de los métodos de detección de tuberculosis más relevante es el análisis de radiografías del tórax; sin embargo, contar con profesionales altamente capacitados para el diagnóstico de la tuberculosis en todos los centros de salud es imposible en los países emergentes, este es uno de los principales motivos de que este método no sea ampliamente usado. En las últimas décadas las redes neuronales han tenido un papel muy relevante en la solución de problemas de la sociedad y en especial en el sector salud. Se ha empleado tres algoritmos de aprendizaje profundo reconocidos en el desarrollo de visión computacional que son VGG19, MobileNet e InceptionV3, se ha logrado obtener resultados muy auspiciosos para la detección de la tuberculosis. Caso especial ha sido MobileNet que ha destacado entre las demás, dando resultados importantes en las diferentes métricas de evaluación empleadas. Además, MobileNet cuenta con una arquitectura menos compleja y los pesos obtenidos después del entrenamiento son muy menores en comparación de los otros dos algoritmos. Se concluye que MobileNet es el algoritmo de Deep Learning más eficiente a comparación de VGG19 e InceptionV3, cuenta con mejor precisión para la detección de la tuberculosis y; el costo computación y tiempo de procesamiento es significativamente menor.

Palabras clave: Redes neuronales; aprendizaje profundo; tuberculosis; radiografías

ABSTRACT

It is estimated that around 1.4 million people infected with tuberculosis died in 2019, most of them in developing countries. If tuberculosis had been diagnosed in time, the death of infected people would have been prevented. One of the most relevant tuberculosis detection methods is the analysis of chest radiographs; However, having highly trained professionals for the diagnosis of tuberculosis in all health centers is impossible in emerging countries, this is one of the main reasons why this method is not widely used. In recent decades, neural networks have played a very relevant role in solving problems in society and especially in the health sector. Three recognized Deep Learning algorithms have been used in the development of computational vision that are VGG19, MobileNet and InceptionV3, it has been possible to obtain very auspicious results for the detection of tuberculosis. MobileNet has been a special case, which has stood out among the others, giving important results in the different evaluation metrics used. In addition, MobileNet has a less complex architecture and the weights obtained after training are very less compared to the other two algorithms. It is concluded that MobileNet is the most efficient Deep Learning algorithm compared to VGG19 and InceptionV3, it has better precision for the detection of tuberculosis and the computational cost and processing time is significantly lower.

Key words: Neural networks; Deep Learning; tuberculosis; x-ray

RESUMO

Estima-se que cerca de 1,4 milhão de pessoas infectadas com tuberculose morreram em 2019, a maioria delas em países em desenvolvimento. Se a tuberculose tivesse sido diagnosticada a tempo, a morte de pessoas infectadas teria sido evitada. Um dos métodos de detecção de tuberculose mais relevantes é a análise de radiografias de tórax; No entanto, ter profissionais altamente capacitados para o diagnóstico da tuberculose em todos os centros de saúde é impossível nos países emergentes, esse é um dos principais motivos pelo qual esse método não é amplamente utilizado. Nas últimas décadas, as redes neurais têm desempenhado um papel muito relevante na resolução de problemas na sociedade e principalmente no setor da saúde. Três algoritmos de aprendizado profundo reconhecidos foram usados no desenvolvimento da visão computacional que são VGG19, MobileNet e InceptionV3, sendo possível obter resultados muito auspiciosos para a detecção da tuberculose. A MobileNet tem sido um caso especial, que tem se destacado entre os demais, apresentando resultados importantes nas diferentes métricas de avaliação utilizadas. Além disso, o MobileNet possui uma arquitetura menos complexa e os pesos obtidos após o treinamento são muito menores em comparação com os outros dois algoritmos. Conclui-se que o MobileNet é o algoritmo de aprendizado profundo mais eficiente em comparação ao VGG19 e ao InceptionV3, possui melhor precisão para detecção de tuberculose e; o custo de computação e o tempo de processamento são significativamente menores.

Palavras-Chave: redes neurais; aprendizagem profunda; tuberculose; raio X

INTRODUCCIÓN

La tuberculosis es una enfermedad que afecta a gran parte del mundo, y con mayor grado de incidencias en países emergentes (1). La tuberculosis es una enfermedad infecciosa difícilmente de detectar, se debe realizar varias pruebas para obtener un diagnóstico certero. Las radiografías del tórax es una de las herramientas principales para la detección de la tuberculosis, la cual debe ir acompañada por exámenes biológicos hasta genéticos antes de que se pueda hacer un diagnóstico real; de este modo poder determinar el tratamiento correspondiente al paciente para su recuperación (2). La Organización Mundial de la Salud (OMS) recomiendan a las entidades de salud de todos los países a realizar diagnósticos tempranos y oportunos para tratar a los pacientes, para tal fin aconsejan la utilización de las radiografías del tórax como una herramienta fundamental en la detección de la tuberculosis en una primera instancia, de este modo se puede tratar oportunamente al paciente (3). No obstante, el personal médico de países en desarrollo, en los centros de salud en las áreas rurales no cuentan con el personal especializado o no tienen la experiencia necesaria para interpretar los patrones característicos de enfermedades pulmonares en una imagen de radiografía del tórax, esto puede conllevar a un diagnóstico erróneo y por consecuencia a un tratamiento tardío o equivocado en perjuicio en la salud del paciente (4).

Los algoritmos de Deep Learning en especial los desarrollados a base de redes neuronales convolucionales son los más idóneos para la detección de enfermedades a través de imágenes microscópicas, radiográficas, tomográficas, entre otros (5-7). Las redes neuronales convolucionales ha demostrado un enorme potencial en la telemedicina, en la última década recién se ha podido aprovechar y desarrollar redes neuronales complejas, debido a que no se contaba con la capacidad computacional para poder desplegar algoritmos con costo computacional muy alto; y que la tecnología de hardware de décadas pasadas no podía afrontar. Esto se debe a que los algoritmos de Deep Learning necesitan ser entrenados con grandes cantidades de datos, en la telemedicina se utiliza generalmente imágenes, cada imagen cuenta con miles de pixeles y la distribución de estos pixeles denotan un patrón propio de

una determinada clase, cada pixel toma un valor que se procesará a través de matrices o se convierten en vectores para poder llevar a cabo los cálculos computacionales según la arquitectura de los algoritmos (8).

La detección de la tuberculosis asistida por sistemas informáticos como CAD4TB, semanticMD y Qure.ai son algunos de los ejemplos de software de detección asistida por computadora (CAD) que están disponibles en el mercado. Si bien obtienen resultados positivos, estos no superan a un profesional médico experto, esto debido a que no cuentan con algoritmos complejos de Deep Learning para obtener mejores resultados (9).

En ese sentido, en el presente artículo se plantea desarrollar tres algoritmos de Deep Learning que puedan superar las expectativas de los expertos, para ellos se empleará algoritmos pre-entrenados y se importará los pesos de ImageNet (10) para lograr un mejor desempeño. Los resultados serán evaluados a través de métricas que determinarán si los algoritmos propuestos tienen resultados óptimos.

MATERIALES Y MÉTODOS

Dataset

En el presente estudio se trabajó con dos conjuntos de datos publicados en la comunidad de ciencias de datos KAGGLE por Rahman *et al.* (11) y Jaeger *et al.* (12), al combinar los dos conjuntos de datos dan un total de 7662 imágenes de radiografías del tórax de pacientes anonimizados, el objetivo de unificar los dos datasets es de tener una mayor variabilidad en los datos, al momento de entrenar los algoritmos puedan generalizar adecuadamente ante diferentes fuentes de imágenes en la etapa de validación y test. Se realizó un análisis de cada imagen para descartar las que contaban con ruido; es decir, no correspondía a una de las clases, eran ilegibles, tenían textos o cuadros blancos o negros sobrepuestos, mal posicionamiento entre otros aspectos que hacían que estas imágenes no cumplieran las características de cada clase del dataset. Después de la depuración de los datos inválidos que podrían generar ruido al momento de entrenar los algoritmos, quedó un dataset de un total de 5748 imágenes, 2905 imágenes de pacientes sanos y 2843 imágenes de pacientes con tuberculosis. El dataset se dividió en dos conjuntos, el conjunto de entrenamiento con el 80% y el conjunto de test con el 20% del total. Del total de los datos del entrenamiento se divido nuevamente en datos para entrenar y datos para la validación, en este caso del total de los datos de entrenamiento un 20% se destinará para la validación cruzada en la etapa de entrenamiento. Se puede ver una muestra en la Figura 1.

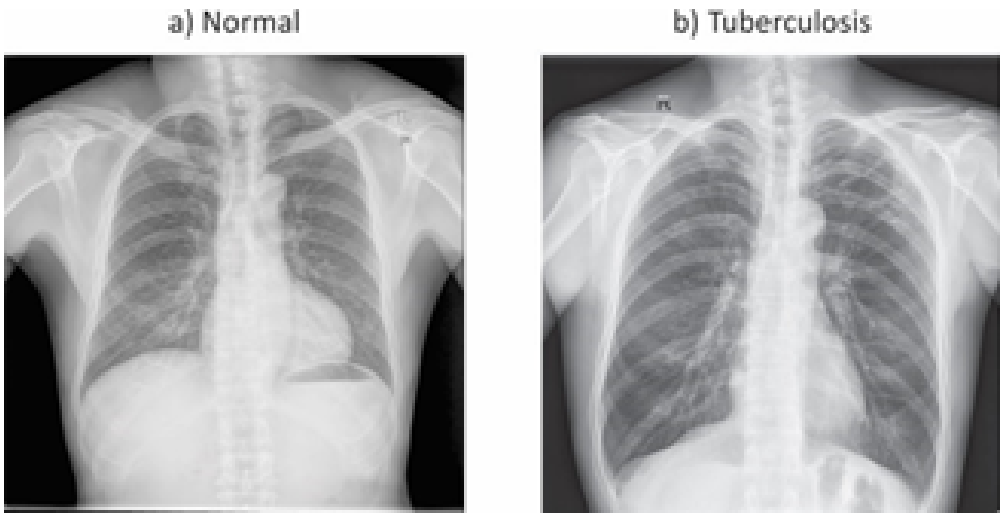

***Figura* 1.** Imágenes de paciente sano y paciente con tuberculosis.

Se emplearon tres algoritmos de Deep Learning para el estudio, InceptionV3 (13), MobileNet (14) y VGG19 (15). Para hacer uso de los algoritmos propuestos se utilizó la librería de Tensorflow y Keras. Para recortar los tiempos de entrenamiento de los algoritmos y una mejor eficacia de los mismo se utilizó la técnica de Transfer Learning con ImageNet. El hecho de usar la misma arquitectura con los mismos parámetros que el modelo pre-entrenado, otorgará a los algoritmos la capacidad de reconocer bastantes clases desde el comienzo, lo que además se traducirá en un tiempo de entrenamiento muy reducido. Para tener la capacidad de computo necesaria para ejecutar los algoritmos InceptionV3, MobileNet y VGG19 se empleó la plataforma de Google Colab que proporciona el uso de GPU's necesarios para la experimentación.

A la arquitectura del algoritmo VGG19 se ha adicionado una capa pooling de 4x4, se vectoriza los datos con la técnica flatten, se agregaron cuatro capas dense de 512, 256, 128 y 64 de valor de salida y cada una con activación relu; se aplica un BatchNormalization, finalmente se agrega una capa dense de 2 con una activación softmax para la clasificación.

A la arquitectura del algoritmo MobileNet se ha adicionado una capa pooling de 4x4, se vectoriza los datos con la técnica flatten, se agregaron cuatro capas dense de 1024, 512, 256, 128, 64 y 32 de valor de salida y cada una con activación relu; se aplica un BatchNormalization, finalmente se agrega una capa dense de 2 con una activación softmax para la clasificación.

A la arquitectura del algoritmo InceptionV3 se ha adicionado una capa pooling de 4x4, se vectoriza los datos con la técnica flatten, se agregaron cuatro capas dense de 1024, 512, 256, 128, 64 y 32 de valor de salida y cada una con activación relu; se aplica un BatchNormalization, finalmente se agrega una capa dense de 2 con una activación softmax para la clasificación.

Los hiperparametros empleados para los tres algoritmos fueron epochs = 100, batch_ size = 20 y learning_rate a razón de 0.001. Esta configuración permitió obtener el mejor desempeño que pueden ofrecer InceptionV3, MobileNet y VGG19.

RESULTADOS Y DISCUSIÓN

En la etapa de entrenamiento se puede observar el desempeño de cada algoritmo empleado para el presente estudio, se logró evaluar a cada algoritmo a través de las métricas de precisión (accuracy) y la perdida de entropía cruzada (cross entropy loss). En el caso de VGG19 se visualiza que en las primeras 20 épocas sufre de sobreajuste (overfiting), pero al pasar más épocas se corrige y muestra una tendencia estable, ver Figura 2. Para MobileNet se visualiza que desde las primeras épocas obtiene buenos resultados sin caer en el sobreajuste, con pocas variaciones y manteniéndose constante, ver Figura 3. En el caso de InceptionV3 al igual que MobileNet, se obtiene buenos resultados desde el inicio, no obstante, los resultados de entrenamiento y validación son más distantes en comparación de los resultados de VGG19 y MobileNet, ver Figura 4.

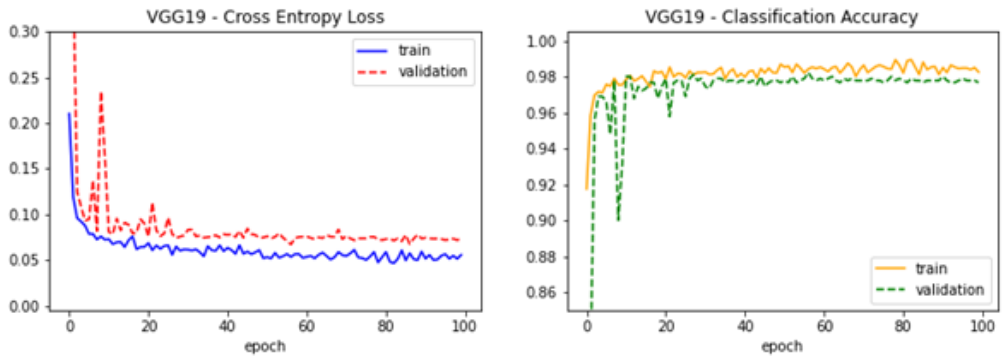

***Figura* 2.** Evaluación del desempeño en la etapa de entrenamiento y validación de las métricas de precisión y la entropía cruzada para VGG19.

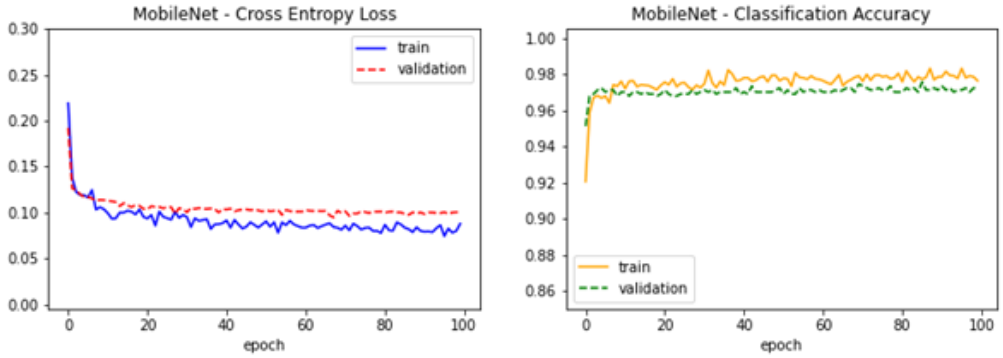

***Figura* 3.** Evaluación del desempeño en la etapa de entrenamiento y validación de las métricas de precisión y la entropía cruzada para MobileNet.

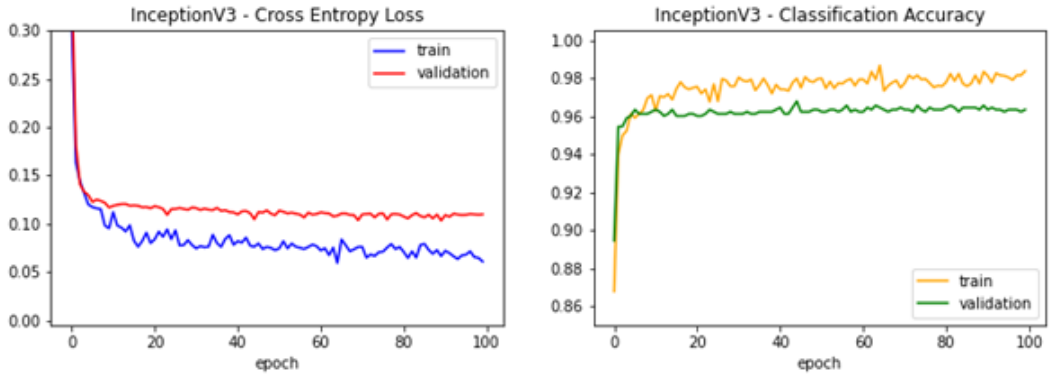

Figura 4. Evaluación del desempeño en la etapa de entrenamiento y validación de las métricas de precisión y la entropía cruzada para InceptionV3.

Para la etapa de test se separó 1140 imágenes, a este grupo de datos se aplicó el método predict que otorgó una estimación para cada clase, la estimación estará en el rango de 0 a 1, posteriormente se aplicará el método argmax para que devuelva la clase con mayor estimación. De este modo, solo obtiene 0 (normal) y 1 (tuberculosis).

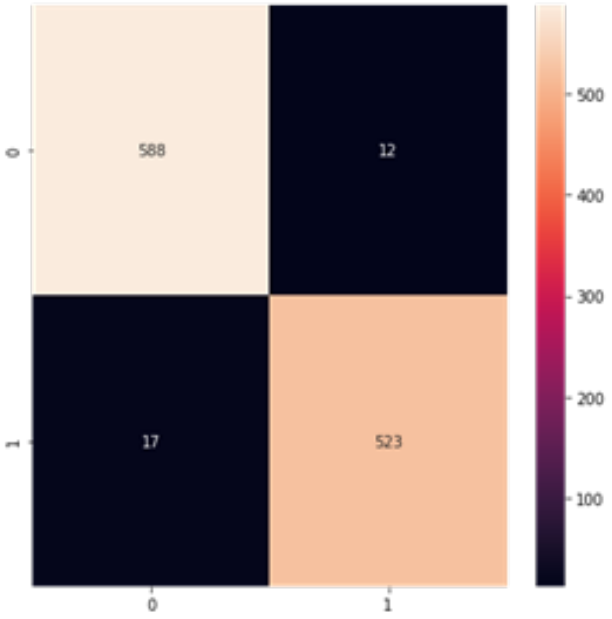

Figura 5. Matriz de Confusión para los resultados con VGG19, donde 0 = Normal y 1 = Tuberculosis.

En la Figura 5, se visualiza los resultados de la Matriz de Confusión para VGG19, acierta correctamente 588 y 523 veces para Normal y Tuberculosis respectivamente. No obstante, también obtiene 17 falsos positivos y 12 falsos negativos; es decir, 17 veces predijo que era Normal cuando realmente era tuberculosis, y 12 veces predijo que era tuberculosis cuando era Normal.

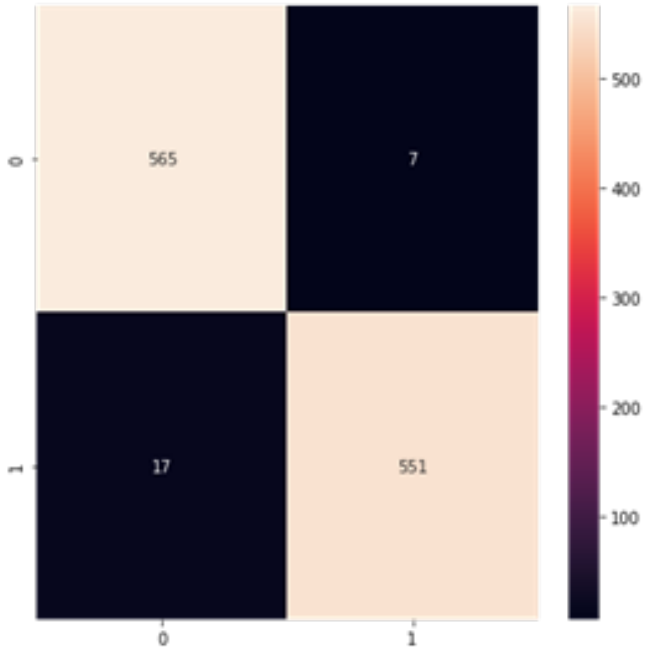

Figura 6. Matriz de Confusión para los resultados con MobileNet, donde 0 = Normal y 1 = Tuberculosis.

En la Figura 6, se visualiza los resultados de la Matriz de Confusión para MobileNet, acierta correctamente 565 y 551 veces para Normal y Tuberculosis respectivamente. No obstante, también obtiene 17 falsos positivos y 7 falsos negativos; es decir, 17 veces predijo que era Normal cuando realmente era tuberculosis, y 7 veces predijo que era tuberculosis cuando era Normal.

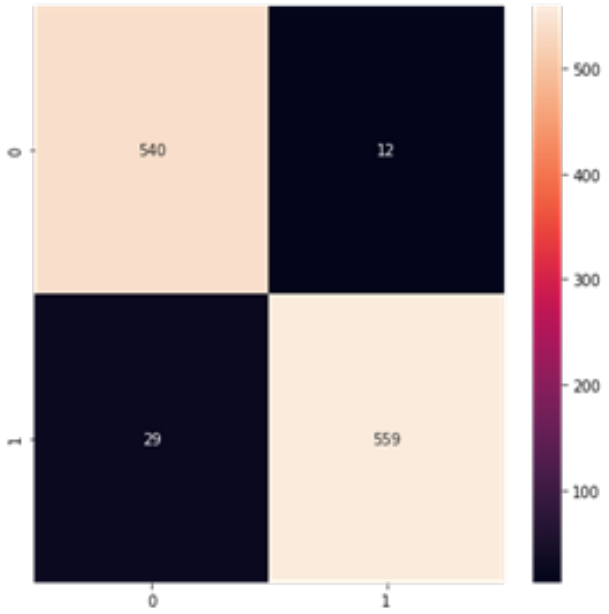

Figura 7. Matriz de Confusión para los resultados con InceptionV3, donde 0 = Normal y 1 = Tuberculosis

En la Figura 7, se visualiza los resultados de la Matriz de Confusión para InceptionV3, acierta correctamente 540 y 559 veces para Normal y Tuberculosis respectivamente. No obstante, también obtiene 29 falsos positivos y 12 falsos negativos; es decir, 29 veces predijo que era Normal cuando realmente era tuberculosis, y 12 veces predijo que era tuberculosis cuando era Normal.

A partir de la Matriz de Confusión podemos calcular las métricas de accuracy, precisión, recall y F1-score de cada algoritmo empleado para el presente estudio. Los resultados podemos verlos en la Tabla 1, cada algoritmo ha demostrado un alto desempeño, pero MobileNet ha obtenido mejores resultados para cada métrica, le sigue VGG19 que por poco margen dista de los resultados obtenidos por MobileNet, el caso de InceptionV3 obtuvo resultados por debajo de MobileNet y VGG19, esto debido a que sufrió de un sobreajuste más pronunciado en la etapa de entrenamiento como se puede visualizar en la Figura 4.

Tabla 1. Métricas de evaluación para cada algoritmo.

Algoritmo	Clase	Accuracy	Precision	Recall	F1-Score
VGG19	Normal	0.974	0.97	0.98	0.98
	Tuberculosis		0.98	0.97	0.97
MobileNet	Normal	0.978	0.97	0.99	0.98
	Tuberculosis		0.99	0.97	0.98
InceptionV3	Normal	0.964	0.95	0.98	0.96
				0.95	0.96

Se presentaron tres algoritmos diferentes pre-entrenados para la aplicación en la detección de tuberculosis. VGG19 obtuvo un 97.4% en la métrica de accuracy, 97.5% en la métrica de precision, 97.5% en la métrica de recall, 97.5% en la métrica de F1-score. MobileNet obtuvo un 97.8% en la métrica de accuracy, 98.0% en la métrica de precision, 98.0% en la métrica de recall, 98.0% en la métrica de F1-score. Por último, InceptionV3 obtuvo un 96.4% en la métrica de accuracy, 96.5% en la métrica de precision, 96.5% en la métrica de recall, 96.0% en la métrica de F1-score. De los resultados se puede concluir que MobileNet ha demostrado un mejor desempeño ante VGG19 e InceptionV3, en las diferentes métricas empleadas.

Además, se guardó los pesos (weights) obtenidos después de entrenar cada algoritmo, de este modo se puede evaluar la complejidad y el costo computacional que conlleva. Ver Tabla 2.

Tabla 2. Parámetros y Weights de cada algoritmo.

Algoritmo	N° Parámetros	Weights (hdf5)
VGG19	21.2 M	162 MB
MobileNet	8.1 M	62 MB
InceptionV3	24.6 M	189 MB

MobileNet cuenta con una menor cantidad de parámetros y peso (weights) en comparación a InceptionV3 y VGG19. Los resultados demuestran en relación al costo computacional es relativamente bajo con MobileNet, además de contar con mejores resultados en las métricas de evaluación; y su complejidad es muy inferior a los otros algoritmos con 8.1 millones de parámetros y un peso de 62 MB, conllevando a la utilización de menos recursos computacionales y la reducción en los tiempos de ejecución del algoritmo. Los resultados indican que MobileNet cuenta con un nivel muy alto de eficiencia en relación a VGG19 e InceptionV3.

CONCLUSIONES

En este trabajo, se ha demostrado que los algoritmos de Deep Learning pueden ser una herramienta importante para la detección de la tuberculosis, otorgando al personal sanitario un nivel de precisión muy alto del 97.80% en el caso del algoritmo de Mobilnet para clasificar si el paciente padece de tuberculosis o no, el médico de acuerdo al cuadro clínico del paciente y los resultados de la clasificación del algoritmo podrá realizar un diagnóstico certero, determinando si el paciente padece de tuberculosis, de este modo se podrá dar un oportuno tratamiento y evitar que el paciente llegue a un estadio de la enfermedad avanzado con una posible mortandad.

REFERENCIAS BIBLIOGRÁFICAS

1. Taye H, Alemu K, Mihret A, Wood JLN, Shkedy Z, Berg S, et al. Factors associated with localization of tuberculosis disease among patients in a high burden country: A health facility-based comparative study in Ethiopia. J Clin Tuberc Other Mycobact Dis. 2021;23(March):1–6.

2. National Institute for Research in Tuberculosis, Tripathy S. Tuberculosis research conducted over the years at the ICMR-National Institute for Research in Tuberculosis (ICMR-NIRT). Indian J Tuberc [Internet]. 2020;67(4):S7–15. Available from: https://doi.org/10.1016/j.ijtb.2020.12.001

3. World Health Organization. Chest Radiography in Tuberculosis. World Heal Organ. 2016;1–44.

4. Huicho L, Canseco FD, Lema C, Jaime Miranda J, Lescano AG. Incentivos para atraer y retener personal de salud de zonas rurales del Perú: Un estudio cualitativo. Cad Saude Publica. 2012;28(4):729–39.

5. Agarwala S, Kale M, Kumar D, Swaroop R, Kumar A, Kumar Dhara A, et al. Deep Learning for screening of interstitial lung disease patterns in high-resolution CT images. Clin

Radiol [Internet]. 2020;75(6):481.e1-481.e8. Available from: https://doi.org/10.1016/j.crad.2020.01.010

6. Kerkech M, Hafiane A, Canals R. Vine disease detection in UAV multispectral images using optimized image registration and Deep Learning segmentation approach. Comput Electron Agric. 2020;174(April)

7. Xiao W, Huang X, Wang JH, Lin DR, Zhu Y, Chen C, et al. Screening and identifying hepatobiliary diseases through Deep Learning using ocular images: a prospective, multicentre study. Lancet Digit Heal [Internet]. 2021;3(2):e88–97. Available from: http://dx.doi.org/10.1016/S2589-7500(20)30288-0

8. Thompson NC, Greenewald K, Lee K, Manso GF. The Computational Limits of Deep Learning. arXiv. 2020

9. Khan FA, Majidulla A, Tavaziva G, Nazish A, Abidi SK, Benedetti A, et al. Chest x-ray analysis with Deep Learning-based software as a triage test for pulmonary tuberculosis: a prospective study of diagnostic accuracy for culture-confirmed disease. Lancet Digit Heal [Internet]. 2020;2(11):e573–81. Available from: http://dx.doi.org/10.1016/S2589-7500(20)30221-1

10. Deng J, Russakovsky O, Krause J, Bernstein M, Berg A, Fei-fei L. Scalable Multi-label Annotation. 2014;1–4

11. Rahman T, Khandakar A, Kadir MA, Islam KR, Islam KF, Mazhar R, et al. Reliable tuberculosis detection using chest X-ray with Deep Learning, segmentation and visualization. IEEE Access. 2020;8:191586–601

12. Jaeger S, Candemir S, Antani S, Wáng Y-XJ, Lu P-X, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg. 2014;4(6):475–7

13. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2016;2016-Decem:2818–26

14. Howard A, Wang W, Chu G, Chen L, Chen B, Tan M. Searching for MobileNetV3 Accuracy vs MADDs vs model size. Int Conf Comput Vis. 2019;1314–24

15. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc. 2015;1–14

Conflicto de intereses: Ninguno declarado por los autores.

Financiación: Este estudio es autofinanciado por los autores.

Agradecimiento: Ninguno manifestado por los autores

ACERCA DE LOS AUTORES

Juan Carlos Valero Gómez. Ingeniero de Sistemas e Informática de profesión. Experiencia en ciencias de datos, visión computacional, administración de sistemas operativos Linux, desarrollador de aplicaciones web en backend y frondend. Universidad Nacional de Moquegua, Perú.

Alex Peter Zúñiga Incalla. Ingeniero de Sistemas. Maestro en Ciencias de la Educación con mención en Tecnologías de la Información e Informática Educativa. Docente de la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional de Moquegua, Perú.

Juan Carlos Clares Perca. Ingeniero en Informática y Sistemas. Magister en Administración de la Educación. Docente en la Universidad Nacional de Moquegua, Perú.