

Recibido 05 de noviembre 2021. Aceptado 10 de marzo de 2022. Publicado 01 de agosto de 2022.

ISSN: 2448-7775

Clasificación automática de radiografías de tórax usando inteligencia artificial

ARÓN HERNÁNDEZ TRINIDAD^{1*}, T. CÓRDOVA-FRAGA¹, R. GUZMÁN-CABRERA².

¹Universidad de Guanajuato, Campus León, División de Ciencias e Ingenierías.

²Universidad de Guanajuato, Campus Irapuato-Salamanca, División de Ingenierías.

*Autor de correspondencia: a.hernandeztrinidad@ugto.mx

RESUMEN La neumonía es una afección inflamatoria del pulmón que afecta a los alvéolos. El diagnóstico se basa en síntomas y examen físico. Las radiografías de tórax son frecuentemente utilizadas como alternativa para validar el diagnóstico. En el presente trabajo se presenta una metodología para llevar a cabo el procesamiento de imágenes basado en aprendizaje automático e inteligencia artificial con la finalidad de llevar a cabo una clasificación automática de dichas imágenes. Se presentan resultados de experimentos realizados en dos escenarios de clasificación: validación cruzada y conjuntos de entrenamiento y prueba. Se utilizan cinco métodos de aprendizaje automático distintos en cada escenario de clasificación. La configuración del experimento fue planteada de manera que permitiera identificar las mejores condiciones de clasificación, permitiendo, además, observar de una manera clara en las métricas de evaluación utilizadas el impacto del tamaño del conjunto de entrenamiento. Los resultados obtenidos permiten ver la efectividad de la metodología implementada, en virtud de que los resultados son muy competitivos con los reportados en el estado de arte.

PALABRAS CLAVE— Neumonía, procesamiento de imágenes, inteligencia artificial, aprendizaje automático.

I. INTRODUCCIÓN

Durante años, las discusiones acerca de la Inteligencia Artificial (IA) en muchas áreas de la ciencia, han sido el centro de conferencias, simposios, congresos, etc. En años recientes, el interés científico ha aumentado en relación con las imágenes médicas, donde existen diferentes trabajos con imágenes médicas que determinan las aplicaciones basadas en la Inteligencia Artificial que generan valores importantes en diagnósticos, y ayudan a optimizar el tiempo y el trabajo de una manera eficaz.

Por tal motivo, este trabajo se ha enfocado en obtener resultados a partir de una metodología que pueda ser capaz de procesar y clasificar imágenes médicas torácicas, que es uno de los exámenes que se realizan con más frecuencia, por lo que se ha aplicado la IA para crear mejoras en el examen de diagnóstico. Este reporte fue posible, usando redes neuronales convolucionales (CNN), que es un tipo de red neuronal artificial, formadas por multicapas que ayudan por medio de la visión artificial, la clasificación y segmentación de imágenes de manera efectiva. Este modelo computacional aprende a partir de un conjunto de entrenamiento, y va evolucionando poco a poco, para obtener valores que aporten sustentos de experimentos científicos. Estas CNN aprenden y se forman a sí mismos, en lugar de ser programados de forma explícita.

1. Neumonía

La neumonía es una afección inflamatoria del pulmón que afecta principalmente a los alvéolos (pequeños sacos de aire) [1]. Los síntomas típicamente incluyen una combinación de tos productiva o seca, dolor de pecho, fiebre y dificultad para respirar [2]. Generalmente es causada por una infección con virus o bacterias. El diagnóstico a menudo se basa en los síntomas y el examen físico [3]. La radiografía de tórax, los análisis de sangre, entre otros, pueden ayudar a confirmar el diagnóstico [4]. El examen físico con el diagnóstico por imágenes es vital para la detección de neumonía [5]. Las radiografías de tórax se utilizan con frecuencia en los procedimientos de validación y representan una alternativa rápida para conocer las características y la extensión de las inflamaciones pulmonares [6]. Las áreas de opacidad de la imagen de radiografía de Rayos X se correlacionan comúnmente con las regiones afectadas por neumonía [7]. En la Fig. 1. se puede ver un ejemplo de ello.

La identificación de opacidades en las imágenes de Rayos X de tórax a veces es un desafío, incluso para los médicos experimentados [8]. Las técnicas de Inteligencia Artificial, visión por computadora, se han utilizado ampliamente en los últimos años, como herramientas auxiliares en el diagnóstico del padecimiento [9]. Un sistema de clasificación adecuado y un número suficiente de muestras de formación son requisitos previos para una clasificación exitosa.

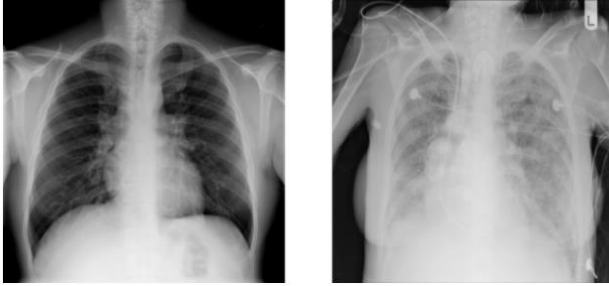


Fig. 1. Radiografías de tórax normal (izquierda) y anormal (derecha). El aumento de la opacidad sugiere neumonía.

2. Procesamiento de clasificación de imágenes

La clasificación de imágenes ha atraído durante mucho tiempo la atención de la comunidad científica debido a los resultados que se han obtenido [10]. Éstos han servido de base para muchas aplicaciones: científicas, médicas, etc. Por tal motivo, se han realizado grandes esfuerzos para desarrollar métodos y técnicas de clasificación avanzadas para mejorar los valores reportados actualmente en las investigaciones [11]. El preprocesamiento de imágenes puede incluir la detección y restauración de líneas defectuosas, rectificación geométrica o registro de imágenes, calibración radiométrica, corrección atmosférica y topográfica [12]. Si se utilizan diferentes datos auxiliares, la conversión de datos entre diferentes fuentes o formatos y la evaluación de la calidad de estos datos también son necesarias antes de que puedan incorporarse en un procedimiento de clasificación. La rectificación geométrica precisa o el registro de imágenes de datos de detección es un requisito previo [13]. En la Fig. 2. se muestra las secciones correspondientes del procesamiento de imágenes.

3. Objetivo

El papel de las técnicas de aprendizaje automático por computadora, en particular las redes neuronales convolucionales (CNN) [14], se evalúa en detalle en este documento. Se presta especial atención a aspectos particulares de una solución basada en CNN para la interpretación de la opacidad de la radiografía de tórax, incluido las métricas de medición como resultados del modelo propuesto y la relevancia de los procedimientos de aumentar el número de imágenes.

4. Conjunto de datos

El conjunto de datos escogidos para este ejercicio de aprendizaje automático es una versión modificada de Paul Mooney presentado en Kaggle, “Imágenes de Rayos X de tórax (neumonía). De acuerdo con el autor, las imágenes se encuentran etiquetadas manualmente por expertos en el área, además, del conjunto de datos, “... las imágenes de Rayos X de tórax (anterior-posterior) se seleccionaron de cohortes

retrospectivas de pacientes pediátricos de uno a cinco años del Centro Médico de Mujeres y Niños... Todas las imágenes de radiografías de tórax que se realizaron fueron parte de la atención clínica de rutina de los pacientes...”¹. El conjunto de datos original contiene 5,863 observaciones, divididas en carpetas de entrenamiento (1,341 casos normales, 3,875 casos de neumonía), validación (8 casos normales, 8 casos de neumonía) y pruebas (234 casos normales, 390 casos de neumonía). El conjunto de datos modificados contiene 2,224 observaciones, donde se propone una división más equilibrada entre las imágenes de entrenamiento y de prueba:

- Observaciones de entrenamiento: 1,600 (800 casos normales, 800 casos de opacidad pulmonar).
- Observaciones de prueba: 624 (234 casos normales, 390 casos de opacidad pulmonar).

II. DESARROLLO

La configuración inicial del experimento fue la decisión de realizar dos tipos de métodos de evaluación: conjunto de entrenamiento/prueba [15] y validación cruzada [16]. Las observaciones en el conjunto de entrenamiento forman la experiencia que el algoritmo usa para aprender. El conjunto de prueba es un conjunto de observaciones utilizadas para evaluar el rendimiento del modelo utilizando alguna medida de rendimiento. Las observaciones del conjunto de entrenamiento no se incluyen en el conjunto de prueba [17]. La validación cruzada es una técnica para evaluar modelos de *Machine Learning* mediante el entrenamiento de modelos de aprendizajes como: máquinas de vectores de soporte, vecinos más cercanos, entre otros; en subconjuntos de los datos de entrada disponibles y evaluarlos con el subconjunto complementario de los datos [18]. En la Fig. 3. se muestra un ejemplo de subconjuntos de evaluación y subconjuntos de entrenamiento. Contiene cinco modelos que se crean y se entrenan durante una validación cruzada de cinco iteraciones.

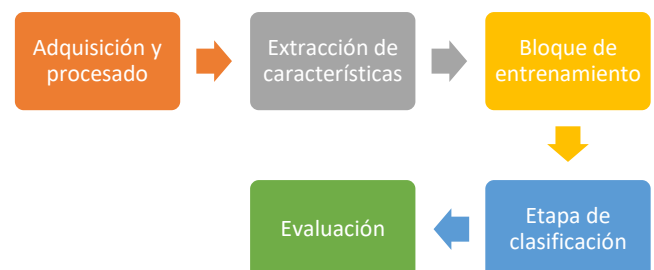


Fig. 2. Pasos por seguir para realizar de forma completa una clasificación de imágenes.



Fig. 3. (Azul) Fuentes de datos de evaluación. (Blanco) Fuentes de datos de entrenamiento (complemento de la evaluación).

¹ <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>

El modelo uno, utiliza el 20% de los datos para la evaluación y el 80% restante para el entrenamiento. El segundo modelo utiliza el 40% para evaluación y el restante para el entrenamiento, y así sucesivamente.

En la [Tabla I](#), se muestran los pasos que se realizaron para la obtención de los resultados en las pruebas antes citadas.

A. MÉTRICAS DE EVALUACIÓN DE MODELOS EN EL APRENDIZAJE AUTOMÁTICO

Las métricas de evaluación nos ayudan a cuantificar y evaluar aspectos de tendencia, comportamiento y resultados de un experimento en la Inteligencia Artificial [19]. Entonces, éstas tienen como objetivo estimar la precisión de la generalización del modelo sobre datos futuros (no visto o fuera de la muestra), de este modo, los modelos van adquiriendo una mejor evaluación. Además, las métricas de evaluación nos proporcionan los valores de rendimiento de un algoritmo de aprendizaje automático, que representa una componente integral en la ciencia de datos [20].

Para problemas de clasificación, la evaluación se puede reportar en función de la matriz de confusión [21]. Como se puede observar en la [Tabla II](#). Las filas de la tabla representan la *clase predicha*, mientras que las columnas la *clase real*. De esta matriz de confusión, *tp* y *tn* denotan el número de instancias positivas y negativas que están correctamente clasificados. Mientras tanto, *fp* y *fn*, denotan el número de instancias positivas y negativas que están incorrectamente clasificados. A partir de la [Tabla II](#), se pueden generar varias métricas para evaluar el desempeño del clasificador. En este trabajo, se usaron cinco métricas de evaluación con diferentes enfoques de evaluaciones [22].

TABLA I. METODOLOGÍA DE LOS MODELOS PROPUESTOS.

Conjunto de entrenamiento/prueba	Validación cruzada
Conjunto de datos.	
Extracción de las características de las imágenes.	
Algoritmo que otorga el porcentaje de entrenamiento y de prueba.	Creación de dos archivos: todas las características y todas las etiquetas.
Códigos creados en el lenguaje de programación Python, que contiene los modelos con sus respectivos clasificadores.	
Se obtienen las métricas de evaluación que proporcionan los valores de precisión de los modelos, así como su rendimiento.	

TABLA II. MATRIZ DE CONFUSIÓN.

	Clase positiva real	Clase negativa real
Clase positiva predicha	Verdadero positivo (<i>tp</i>)	Falso positivo (<i>fp</i>)
Clase negativa predicha	Falso negativo (<i>fn</i>)	Verdadero negativo (<i>tn</i>)

1. Exactitud

La exactitud es una métrica de evaluación común para los problemas de clasificación. Es el número de predicciones correctas hechas como una proporción de todas las predicciones hechas, en la [Ec. \(1\)](#), se muestra cómo se obtiene el valor de la métrica. En general, cuando las clases son aproximadamente iguales en tamaño, usar esta métrica, nos dará valores clasificados correctamente.

$$Exactitud = \frac{tp+tn}{total} = \frac{tp+tn}{tp+tn+fp+fn} \quad (1)$$

2. Precisión

La precisión indica con qué frecuencia es correcto los valores y se obtiene a partir de la [Ec. \(2\)](#). Esta métrica representa el número de verdaderos positivos que son realmente positivos en comparación con el número total de valores positivos predichos.

Cuando se tiene un desequilibrio de clase, la precisión puede convertirse en una métrica poco fiable para medir el desempeño del modelo de evaluación.

$$Precisión = \frac{tp}{tp+fp} \quad (2)$$

3. Sensibilidad

La métrica de sensibilidad, también conocida como exhaustividad, muestra la cantidad de verdaderos positivos, que el modelo ha clasificado en función del número total de valores positivos, como establece la [Ec. \(3\)](#).

$$Sensibilidad = \frac{tp}{tp+fn} \quad (3)$$

4. Puntuación F1

Esta métrica es la combinación de las métricas de precisión y sensibilidad, [Ec. \(4\)](#) [23]. La mejor puntuación es igual a 1 y la peor a 0.

$$F1 = 2 \frac{precisión * sensibilidad}{precisión + sensibilidad} \quad (4)$$

5. Kappa de Cohen

Es una medida de concordancia que se basa en comparar la concordancia observada en el conjunto de datos, respecto a la que podría ocurrir por mero azar.

En la [Tabla III](#), se muestra el grado de estimación con el resultado del cálculo de Kappa [24].

TABLA III. ESCALA DE VALORES DE KAPPA Y SU CONCORDANCIA.

Kappa	Estimación del grado de acuerdo
< 0	No acuerdo
0.0 – 0.2	Insignificante
0.2 – 0.4	Bajo
0.4 – 0.6	Moderado
0.6 – 0.8	Bueno
0.8 – 1.0	Muy bueno

B. METODOLOGÍA DEL MODELO PROPUESTO

La clasificación de imágenes de las radiografías de tórax, se hicieron para dos tipos de modelos, como se ha mencionado. Sin embargo, en ambos modelos propuestos, se utilizó la misma red neuronal convolucional ResNet50 [25] para la obtención del vector de características, que son utilizados para el entrenamiento de los clasificadores implementados. Esta red proporciona una mejor comprensión de los diferentes enfoques y sus aplicaciones a la clasificación de Rayos X de tórax [26].

1. Conjunto de entrenamiento y prueba

La arquitectura de este sistema se dividió en tres categorías:

1. Preprocesamiento de las imágenes.
2. Extracción de las características, usando ResNet50.
3. Clasificación.

Los experimentos fueron realizados en el lenguaje de programación Python, donde se realizaron los algoritmos correspondientes, para obtener los rendimientos de los modelos de evaluación.

Como primer paso, se declararon las librerías que contiene la red neuronal convolucional a usar [26]. Una vez declaradas, se cargan la red ResNet50. En el algoritmo, se colocan las instrucciones específicas para recolectar las imágenes y cargarlas al sistema. El objetivo principal de usar la red neuronal, en la mayoría de las tareas de clasificación de imágenes, es reducir la complejidad computacional del modelo que probablemente aumenta si la entrada son muchas imágenes. Las imágenes se redimensionaron a 224×224 píxeles, ya que el cálculo de la red se realiza con esta dimensión para realizar un procesamiento rápido. Todas las técnicas adicionales, se han aplicado sobre estas imágenes reducidas.

Para la extracción de las características, se procedió a aplicar la CNN, por lo tanto, esta etapa se ocupa de la descripción de la arquitectura de la red ResNet50 y su contribución en la extracción de características de las imágenes. En la Fig. 4. se muestra la arquitectura [27].

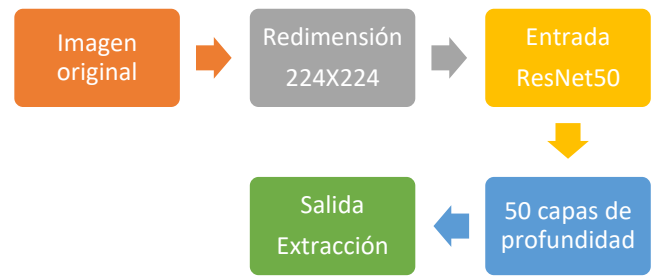


Fig. 4. Arquitectura ResNet50 para la clasificación de imágenes de Rayos X de tórax.

Se utilizaron cinco escenarios de clasificación para los dos modelos propuestos, estos clasificadores forman parte del aprendizaje profundo, los utilizados fueron:

- 1) Máquinas de Vectores de Soporte (SVM).
- 2) Regresión Logística (LR).
- 3) Vecinos más cercanos (KNC).
- 4) Estadística Naive-Bayes (NB).
- 5) Centroide más cercano (Centroid).

Ahora bien, al tener los clasificadores en los algoritmos de los modelos, se procede a obtener las métricas de evaluación, para medir el desempeño de los sistemas de clasificación implementados y que muestren los resultados correspondientes a los modelos de evaluación. Las métricas de evaluación fueron:

- a. Exactitud.
- b. Precisión.
- c. Sensibilidad.
- d. F1.
- e. Kappa.

En la Tabla IV, se reporta el número de imágenes de entrenamiento y de pruebas utilizado en cada configuración del experimento, como se puede observar, el conjunto de entrenamiento aumenta al doble en cada caso, mientras que el conjunto de prueba permanece constante, intuitivamente se puede pensar que un conjunto de entrenamiento con más instancias para aprender tendrá un mejor desempeño, esto precisamente es lo que se pretende probar con el experimento propuesto.

TABLA IV. CONJUNTO DE ENTRENAMIENTO Y PRUEBA.

Imágenes de entrenamiento	Imágenes de prueba
200	624
400	624
800	624
1,600	624

2. Validación cruzada – Amazon Machine Learning

Para este experimento, tanto las imágenes de entrenamiento como las de prueba, se utilizaron de manera conjunta, es decir, la primera configuración era de 824 imágenes, la segunda 1,024, y así consecutivamente hasta obtener el número total de imágenes 2,224.

III. RESULTADOS

A partir del conjunto de datos y de los clasificadores, tanto en el modelo de conjunto de entrenamiento/prueba y de validación cruzada, en la [Tabla V](#) y en la [Tabla VI](#) se registran los valores obtenidos de ambos modelos en cada configuración. En las columnas se muestran los clasificadores con las métricas de evaluación respectivas.

Se puede observar que, a medida que el número de imágenes aumenta, la exactitud y la precisión obtienen mejores valores, se puede suponer, que a medida que se entrena el modelo (con suficientes imágenes en el conjunto de entrenamiento) y vaya aumentando dichas instancias, el modelo va adquiriendo mejores valores, y el aprendizaje cada vez es más preciso, teniendo como resultados valores competitivos.

De igual manera, sucede con todas las métricas utilizadas en cada clasificador, aumentan los valores gradualmente, a medida que aumentan las imágenes de entrenamiento, lo que sugiere una clasificación cada vez más exacta.

TABLA V. VALORES OBTENIDOS EN EL CONJUNTO DE ENTRENAMIENTO Y PRUEBA.

Número de imágenes	Clasificador	Exactitud	Precisión
200 + 624	SVM	0.79	0.80
	LR	0.80	0.81
	KNC	0.81	0.80
	NB	0.80	0.81
	Centroid	0.78	0.78
400 + 624	SVM	0.83	0.86
	LR	0.84	0.85
	KNC	0.84	0.84
	NB	0.85	0.85
	Centroid	0.82	0.82
800 + 624	SVM	0.81	0.86
	LR	0.83	0.86
	KNC	0.84	0.85
	NB	0.84	0.85
	Centroid	0.81	0.82
1, 600 + 624	SVM	0.91	0.91
	LR	0.88	0.89
	KNC	0.89	0.89
	NB	0.86	0.85
	Centroid	0.83	0.81

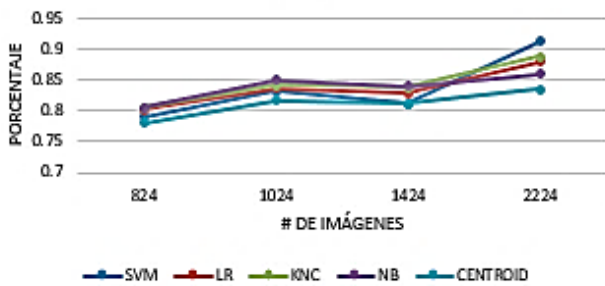
TABLA VI. VALORES OBTENIDOS EN EL MODELO DE VALIDACIÓN CRUZADA.

Número de imágenes	Clasificador	Exactitud	Precisión
824	SVM	0.94	0.94
	LR	0.91	0.91
	KNC	0.92	0.92
	NB	0.85	0.86
	Centroid	0.86	0.86
1, 024	SVM	0.94	0.94
	LR	0.91	0.92
	KNC	0.93	0.93
	NB	0.88	0.88
	Centroid	0.89	0.88
1, 424	SVM	0.95	0.95
	LR	0.92	0.92
	KNC	0.93	0.93
	NB	0.88	0.89
	Centroid	0.88	0.88
2, 224	SVM	0.95	0.95
	LR	0.92	0.92
	KNC	0.92	0.92
	NB	0.88	0.88
	Centroid	0.88	0.89

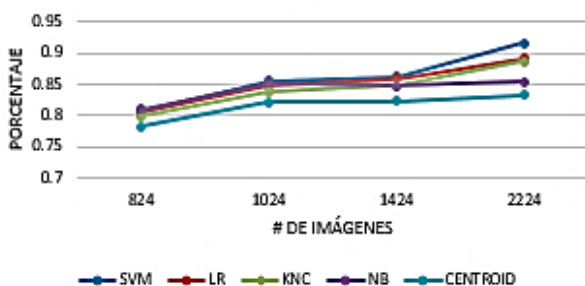
En el registro de los resultados, se observa que el mejor clasificador que nos proporciona un porcentaje mayor, al momento de clasificar, es el clasificador de Máquina de Vectores de Soporte. Como se ha mencionado, a medida que aumenta el número de imágenes, aumenta la efectividad de ambos modelos, con el clasificador SVM. En la [Fig. 5](#) se exponen las gráficas de las métricas de evaluación: exactitud y precisión en los dos modelos.

Además, se puede verificar que efectivamente, el SVM, es el clasificador con mayor porcentaje a medida que aumenta el número de imágenes. Lo anterior indica que el conjunto de entrenamiento aumenta, y, por lo tanto, la precisión y la exactitud de clasificar las instancias de radiografías de tórax, es efectivo.

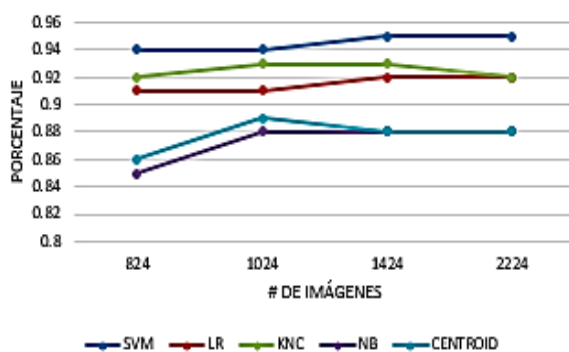
Al comparar ambos modelos con el clasificador SVM, se puede constatar, que el modelo de validación cruzada nos proporciona un mayor porcentaje de exactitud y precisión. De igual manera, con las métricas restantes. En la [Tabla VII](#) y en la [Tabla VIII](#) se registran los datos correspondientes.



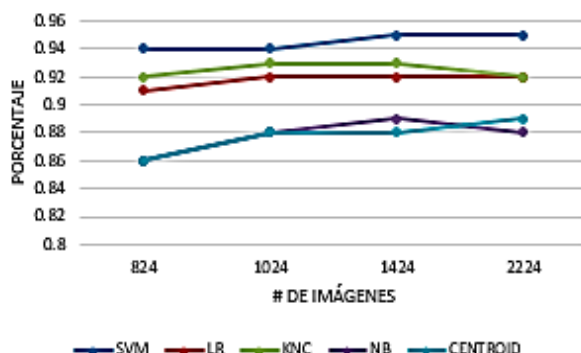
(a)



(b)



(c)



(d)

Fig. 5. (a) y (b) son las métricas de evaluación para el modelo de conjunto de entrenamiento y prueba, exactitud y precisión, respectivamente. (c) y (d) métricas de evaluación para el modelo de validación cruzada. El rango de las ordenadas va de 0.5 a 0.9, que representan el 50 % y 90 % en escala de porcentaje.

TABLA VII. COMPARACIÓN ENTRE EL CONJUNTO DE ENTRENAMIENTO Y PRUEBA (CEYP) Y VALIDACIÓN CRUZADA (VC) PARA LA EXACTITUD.

Número de imágenes	Clasificador	CEYP	VC
824	SVM	0.79	0.94
1,024	SVM	0.83	0.94
1,424	SVM	0.81	0.95
2,224	SVM	0.91	0.95

TABLA VIII. COMPARACIÓN ENTRE EL CONJUNTO DE ENTRENAMIENTO Y PRUEBA (CEYP) Y VALIDACIÓN CRUZADA (VC) PARA LA PRECISIÓN.

Número de imágenes	Clasificador	CEYP	VC
824	SVM	0.81	0.94
1,024	SVM	0.86	0.94
1,424	SVM	0.86	0.95
2,224	SVM	0.92	0.95

IV. CONCLUSIÓN

Los resultados muestran que en el modelo de validación cruzada existe un alto índice de confiabilidad al clasificar las radiografías de tórax, para predecir la neumonía. Sin embargo, el conjunto de entrenamiento y prueba, a medida que aumenta las instancias en el conjunto de entrenamiento, este va adquiriendo valores competitivos, con los reportados en la literatura, superiores al 90 %.

Se ha reportado en el trabajo presente, una alternativa para clasificar radiografías de este tipo, diferentes a los enfoques tradicionales de CNN, donde utilizan otros tipos de redes neuronales convolucionales. Para este trabajo, se reportó los valores obtenidos a partir de la red ResNet50, dicha red, nos ha proporcionado valores que pueden compararse con resultados de artículos ya publicados. La extracción de las características y el procesamiento de imágenes, sin ningún refinamiento a las instancias, son diferentes a cualquier otro modelo previamente entrenado, este trabajo reportado, representa una alternativa para clasificar imágenes de radiografías de tórax, previamente etiquetadas manualmente.

Cuando se ocupa de un diagnóstico médico, un falso positivo (es decir, predecirla enfermedad cuando el paciente está sano) es menos crítico que un falso negativo (predecir la salud cuando el paciente está enfermo). El número de falsos negativos obtenidos con la CNN, que aquí se ha presentado es bajo, alrededor del 5%, comparado con trabajos publicados [28, 29]. Lo que posiciona a los modelos presentados como una herramienta auxiliar y confiable para la detección de neumonía.

Cabe mencionar, que el enfoque mostrado se realizó con 2,224 observaciones en total, podría tomarse en cuenta que, si se aumenta el conjunto de datos, se podrían obtener

mejores resultados y los falsos negativos serían mínimos. Esto podría tomarse como una sugerencia de mejora adicional al presente artículo. El modelo se entrenó varias veces, aumentando de 200, 400 y así sucesivamente duplicado el número de imágenes en el conjunto de entrenamiento, por lo que se mostró resultados favorables con las métricas reportadas: exactitud y precisión.

Las imágenes utilizadas en el experimento fueron imágenes de entrada de un canal, es decir, en escala de grises. Podría considerarse realizar filtros en las imágenes, y poder comparar las métricas con las imágenes originales (escala de grises) y los filtros aplicados, y así verificar qué filtro es adecuado para clasificar, o si no hay necesidad de aplicar filtro para procesar las imágenes.

REFERENCIAS

- [1] Sattar, S.B.A. and S. Sharma, *Bacterial pneumonia*. 2018.
- [2] Metlay, J.P., et al., *Influence of age on symptoms at presentation in patients with community-acquired pneumonia*. Archives of internal medicine, 1997. **157**(13): p. 1453-1459.
- [3] Hoare, Z. and W.S. Lim, *Pneumonia: update on diagnosis and management*. Bmj, 2006. **332**(7549): p. 1077-1079.
- [4] Flanders, S.A., et al., *Performance of a bedside C-reactive protein test in the diagnosis of community-acquired pneumonia in adults with acute cough*. The American journal of medicine, 2004. **116**(8): p. 529-535.
- [5] Parveen, N. and M.M. Sathik, *Detection of pneumonia in chest X-ray images*. Journal of X-ray Science and Technology, 2011. **19**(4): p. 423-428.
- [6] Ayan, E. and H.M. Ünver, *Diagnosis of pneumonia from chest X-ray images using deep learning*. in *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*. 2019. Ieee.
- [7] Abbas, A., M.M. Abdelsamea, and M.M. Gaber, *Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network*. Applied Intelligence, 2021. **51**(2): p. 854-864.
- [8] Self, W.H., et al., *High discordance of chest x-ray and computed tomography for detection of pulmonary opacities in ED patients: implications for diagnosing pneumonia*. The American journal of emergency medicine, 2013. **31**(2): p. 401-405.
- [9] Varshni, D., et al., *Pneumonia detection using CNN based feature extraction*. in *2019 IEEE international conference on electrical, computer and communication technologies (ICECCT)*. 2019. IEEE.
- [10] Olaode, A., G. Naghdy, and C. Todd, *Unsupervised classification of images: a review*. International Journal of Image Processing, 2014. **8**(5): p. 325-342.
- [11] Anthony, G., H. Greg, and M. Tshildizi, *Classification of images using support vector machines*. arXiv preprint arXiv:0709.3967, 2007.
- [12] Ponraj, D.N., et al., *A survey on the preprocessing techniques of mammogram for the detection of breast cancer*. Journal of Emerging Trends in Computing and Information Sciences, 2011. **2**(12): p. 656-664.
- [13] Vasuki, P., J. Kanimozhi, and M.B. Devi, *A survey on image preprocessing techniques for diverse fields of medical imagery*. in *2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE)*. 2017. IEEE.
- [14] Albawi, S., T.A. Mohammed, and S. Al-Zawi, *Understanding of a convolutional neural network*. in *2017 International Conference on Engineering and Technology (ICET)*. 2017. Ieee.
- [15] Hensman, P. and D. Masko, *The impact of imbalanced training data for convolutional neural networks*. Degree Project in Computer Science, KTH Royal Institute of Technology, 2015.
- [16] Refaeilzadeh, P., L. Tang, and H. Liu, *Cross-validation*. Encyclopedia of database systems, 2009. **5**: p. 532-538.
- [17] Storkey, A., *When training and test sets are different: characterizing learning transfer*. Dataset shift in machine learning, 2009. **30**: p. 3-28.

- [18] Browne, M.W., *Cross-validation methods*. Journal of mathematical psychology, 2000. **44**(1): p. 108-132.
- [19] Dalianis, H., *Evaluation Metrics and Evaluation*, in *Clinical Text Mining: Secondary Use of Electronic Patient Records*. 2018, Springer International Publishing: Cham. p. 45-53.
- [20] Hossin, M. and M.N. Sulaiman, *A review on evaluation metrics for data classification evaluations*. International journal of data mining & knowledge management process, 2015. **5**(2): p. 1.
- [21] Visa, S., et al., *Confusion matrix-based feature selection*. MAICS, 2011. **710**: p. 120-127.
- [22] Handelman, G.S., et al., *Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods*. American Journal of Roentgenology, 2019. **212**(1): p. 38-43.
- [23] Flores, É.C. and J.A.H. Soriano, *Análisis de concordancia mediante coeficiente de Kappa de Cohen para la elaboración de un instrumento de categorización de entrevistas biográfico-narrativas*. Educación: aportaciones metodológicas: p. 50.
- [24] Mukti, I.Z. and D. Biswas, *Transfer learning based plant diseases detection using ResNet50*. in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*. 2019. IEEE.
- [25] Baltruschat, I.M., et al., *Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification*. Scientific Reports, 2019. **9**(1): p. 6381.
- [26] Hira, S., A. Bai, and S. Hira, *An automatic approach based on CNN architecture to detect Covid-19 disease from chest X-ray images*. Applied Intelligence, 2021. **51**(5): p. 2864-2889.

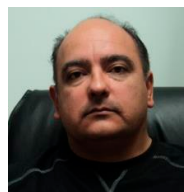
BIOGRAFÍAS



ARÓN HERNÁNDEZ TRINIDAD Estudiante del programa de Doctorado en Física, de la División de Ciencias e Ingenierías del Campus León de la Universidad de Guanajuato. Maestro en Ciencias Aplicadas por la Universidad de Guanajuato. Licenciado en Física por la Universidad Juárez Autónoma de Tabasco.



TEODORO CÓRDOVA-FRAGA Profesor Titular en el Departamento de Ingeniería Física de la División de Ciencias e Ingenierías, Universidad de Guanajuato campus León. Es Doctor en Física (2003) por la Universidad de Guanajuato, en modalidad mixto con la Universidad de Sao Paulo en Brasil. Estancia postdoctoral en Vanderbilt University (2004 - 2005) as Research Fellow. Perteneció al SNI Nivel 2. Integrante del Cuerpo Académico de Física Médica e Instrumentación Biomédica



RAFAEL GUZMÁN-CABRERA Profesor Titular del departamento de Ingeniería Eléctrica de la División de Ingenierías del Campus Irapuato-Salamanca de la Universidad de Guanajuato desde hace 21 años. Doctor en Reconocimiento de Formas e Inteligencias Artificial por la Universidad Politécnica de Valencia, España. Miembro de la Academia Mexicana de Ciencias, SIN 1. Miembro del Cuerpo Académico de Física Aplicada y Tecnologías Avanzadas.