

Deteccción del SARS-CoV-2 en radiografías de tórax por medio de descriptores intermedios y técnicas de *machine learning*

Gonzalo Bardález-Trigoso

17100802@ue.edu.pe / Universidad ESAN

Jean Pablo Bazán-Arzapalo

16100402@ue.edu.pe / Universidad ESAN

Junior Fabián

jfabian@esan.edu.pe / Universidad ESAN

Pedro Montenegro-Montori

16100041@ue.edu.pe / Universidad ESAN

Recepción: 31/8/2020 Aceptación: 1/10/2020

RESUMEN. El SARS-CoV-2, que causa la enfermedad del COVID-19, es un virus que se ha expandido rápidamente por el mundo, teniendo como lugar de inicio la ciudad de Wuhan, en China. A la fecha se han detectado más de 36 738 525 casos a nivel mundial. La tasa de infectados aumenta diariamente y la capacidad sanitaria no se da abasto. Por estas razones, se ha venido proponiendo una variedad de métodos para identificar el novel coronavirus con mayor rapidez y a menor costo. Un ejemplo de estos métodos para identificar la enfermedad es el COVID-Net, una red convolucional que identifica el COVID-19, neumonía o pulmones en condición normal. En este trabajo se propone una metodología para identificar y clasificar imágenes de radiografías de tórax que tienen el COVID-19, neumonía o sin condición. Para esto se utilizaron extractores de características intermedias: HOG+PCA, SIFT+K-means y SURF+K-means, combinados con un SVM como clasificador; además, se emplearon tres estructuras CNN: VGG19, Densenet121 y MobilnetV2. Se utilizó la base de datos COVIDx3 que consta de 15 476 imágenes radiográficas de pulmón. Se obtuvieron buenos resultados, y se determinó que la mejor de las combinaciones fue la que utilizó MobilnetV2 con aumento de datos obteniendo una sensibilidad por clase COVID-19 de 0,97 y en promedio una precisión y

sensitividad de 0,92 y 0,91. Debido al contexto de la crisis sanitaria generada por el COVID-19, este trabajo se presenta como un apoyo para la detección de esta enfermedad y como marco de referencia para futuras investigaciones.

PALABRAS CLAVE: COVID-19 / aprendizaje automático / aprendizaje profundo
/ visión computacional / descriptores de nivel intermedio

Detection of SARS-CoV-2 in Chest X-Rays by Means of Mid-Level Image Descriptors and Machine Learning Techniques

ABSTRACT. COVID-19, the disease caused by the SARS-CoV-2 and originated in the Chinese city of Wuhan, has quickly spread around the world. To date, there have been more than 36,738,525 confirmed cases worldwide. Rates of COVID-19 cases increase on a daily basis and access to healthcare is not enough. For these reasons, a series of methods have been proposed to identify the novel coronavirus faster and at lower cost. An example of said methods is COVID-NET, a convolutional neural network that identifies COVID-19, pneumonia or normal lungs. This research proposes a methodology to identify and classify chest X-ray images according to three categories: COVID-19, pneumonia or normal lungs. To that end, mid-level image descriptors were employed: HOG+PCA, SIFT+K-means and SURF+K-means, combined with a SVM classifier. In addition, three CNN structures were used: VGG19, DenseNet121 and MobilNetV2. The COVIDx3 dataset, consisting of 15,746 chest X-rays, was used. Good results were obtained, where MobilnetV2 plus data augmentation showed the best performance, with a recall of 0.97 for the COVID-19 class, and an average precision and recall of 0.92 and 0.91, respectively. Given the current COVID-19 health crisis, this approach may be used for detecting the virus and as a reference for future research.

KEYWORDS: COVID-19 / machine learning / deep learning / computer vision
/ mid-level image descriptors

1. INTRODUCCIÓN

El COVID-19, causado por el SARS-CoV-2, es una enfermedad viral que se ha expandido rápidamente por el mundo, teniendo como lugar de inicio la ciudad de Wuhan. Según la Organización Mundial de la Salud (9 de octubre del 2020), a la fecha se han detectado 36 738 525 casos y 1 066 412 muertes a causa del virus a nivel mundial. Las pruebas serológicas (RT-PCR) pueden detectar el RNA del SARS-CoV-2 en el cuerpo, es el estándar usado para ver si una persona se encuentra infectada (Wang, Xu, Gao, Lu, Han, Wu y Tan, 2020); sin embargo, este proceso es muy trabajoso, toma tiempo, es caro y hay una alta demanda por este tipo de pruebas. Asimismo, existe el creciente problema del colapso sanitario, debido a la carencia de infraestructura para albergar una mayor cantidad de pacientes en cada hospital.

El desarrollo de aplicaciones de aprendizaje profundo ha sido muy popular en cuanto a tareas de detección de objetos o en tareas de clasificación de imágenes médicas; por ello, utilizar los beneficios de la toma de decisiones médicas basadas en sistemas asistidos por computadora se está volviendo cada vez más una realidad (Wang *et al.*, 2020). En este escenario aparece la oportunidad de utilizar estas tecnologías como apoyo a las pruebas estándar para reducir la brecha en infraestructura y aliviar el colapso del sistema sanitario.

Por estas razones, se han venido proponiendo distintas metodologías para identificar el SARS-CoV-2 por medio de técnicas de aprendizaje profundo, con mayor rapidez y a menor costo. Un ejemplo de estos métodos para identificar la enfermedad es el COVID-Net (Wang y Wong, 2020) que es una red convolucional que discrimina en imágenes de radiografías de pulmón entre las clases COVID-19, neumonía y condición normal; se obtuvieron resultados muy satisfactorios, superando a redes como VGG19 y ResNet-50, con una sensibilidad de 93,3 % y una *accuracy* de 93,3 %. Además, esta propuesta ha sido el inicio para todos los trabajos relacionados a identificación del COVID-19 por medio de radiografías de tórax. Estos nuevos métodos no son un reemplazo de las pruebas serológicas, sino son un apoyo y referente para la identificación de esta enfermedad. De tal forma, Oh, Park y Ye (2020) utilizaron *datasets* con data limitada y una red ResNet-18 para extraer las características de las radiografías de tórax computarizadas y clasificarlas, obteniendo una sensibilidad de 94,3 % y precisión de 87,6 %, proponiendo además el uso de aumentación de datos para potenciar la clase COVID-19 que contenía una reducida cantidad de imágenes. Asimismo, Apostolopoulos y Mpesiana (2020) utilizaron las redes VGG19 y MobilnetV2, aplicando *transfer learning*, siendo MobilnetV2 la que mejor se desempeñó, debido a que clasificó mejor los verdaderos positivos, y tuvo como resultado 98,66 % en sensibilidad y 96,46 % en especificidad.

En este trabajo se propone una nueva metodología para identificar y clasificar radiografías de tórax que tienen el COVID-19, neumonía o sin condición. Es importante aclarar que la metodología propuesta se presenta como un marco de apoyo para la identificación de la enfermedad en radiografías, mas no como un reemplazo para las pruebas serológicas. Para esto se utilizaron los siguientes extractores de características intermedias: Histogram of

Oriented Gradients (HOG) (Dalal y Triggs, 2005) + Principal Component Analysis (PCA), Scale-Invariant Feature Transform (SIFT) (Lowe, 2004) + K-means y Speeded Up Robust Features (SURF) (Bay, Tuytelaars y Van Gool, 2006) + K-means, utilizando como clasificador Máquina de Soporte Vectorial (SVM); además, se emplearon tres estructuras de redes neuronales convolucionales (CNN): VGG19 (Simonyan y Zisserman, 2014), MobilnetV2 (Sandler, Howard, Zhu, Zhmoginov y Chen, 2018) y Densenet121 (Gao, Zhuang, Van der Maaten y Weinberger, 2018). El trabajo se encuentra estructurado de la siguiente forma: la sección 2 describe la metodología, la sección 3 los resultados y la sección 4 las conclusiones.

2. METODOLOGÍA

La metodología que se propuso consta de cuatro bloques principales: (I) recolección de datos, (II) preprocesamiento de los datos, (III) procesamiento de los datos y (IV) clasificación. Se puede apreciar el flujo de trabajo en la figura 1.

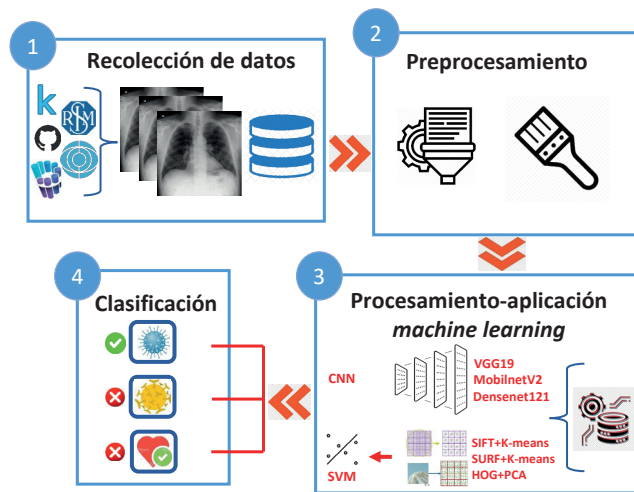


Figura 1. Metodología propuesta en esta investigación

Elaboración propia

2.1 Recolección de datos

En esta investigación se utilizó la base de datos COVIDx3 (Wang *et al.*, 2020) que contiene un total de 15 476 imágenes de radiografías de pulmón. COVIDx3 está conformado por distintas fuentes como la Sociedad de Radiología Italiana, la Sociedad de Radiología Europea, entre otras. Asimismo, se compone de distintos *datasets* de radiografías de tórax de acceso abierto como Covid chestxray-dataset, Figure1-COVID-chestxray-dataset, Actualmed-COVID-chestxray-dataset,

COVID-19 Radiography Database y RSNA Pneumonia Detection Challenge (estas bases de datos se pueden encontrar en el apartado de referencias).

De las 15 476 imágenes, se separaron en muestras para entrenamiento y prueba siguiendo la distribución 80/20. De esta manera se tuvieron 12 380 radiografías para el entrenamiento: 459 que contenían COVID-19, 7080 en condición normal y 4841 con neumonía; y 3096 para la prueba: 114 que contenían COVID-19, 1771 en condición normal y 1211 con neumonía.

2.2 Preprocesamiento

El primer proceso que se realizó fue ordenar los datos de la base de datos COVIDx3, obteniendo una carpeta por cada una de las tres categorías.

Como la categoría del COVID-19 tenía un total de 573 radiografías, estaba claramente desbalanceada con respecto a las otras clases. Esto significaría que a la hora de emplear las CNN, estas podrían generar resultados deficientes. Por esta razón, primero se realizó el entrenamiento con los datos desbalanceados con los descriptores intermedios y las CNN; luego, se hizo otro experimento con las tres estructuras CNN con aumento de datos solo para la clase del COVID-19. Para el proceso de generación de imágenes se implementó una secuencia que generaba datos sintéticos, que consistía en distintas etapas que aplicaban diferentes modificaciones al azar. Las etapas de este proceso consistían en realizar lo siguiente: primero se detalla la cantidad de imágenes a generar por imagen original, en este caso este valor fue de diez; después, la función podría generar una imagen al azar que tuviera al menos una combinación de las siguientes transformaciones: rotaciones entre -10 y 10° , recortes y *padding*, traslaciones en el eje X e Y entre $-0,07$ y $0,07$, aplicación del filtro CLAHE, aplicación de un filtro gaussiano con valores entre 0 y 5, aplicación de filtro de media y mediana con valores de 2 a 5, generación de ruido con Poisson con intensidades de 0 a 5 y generación de ruido gaussiano con intensidades de 0 a 6. De esta manera, se generaron unas 5730 imágenes, teniendo al final un total de 6302 imágenes para la clase del COVID-19.

Asimismo, antes de realizar los procesos de extracción de características, se convirtieron todas las imágenes a escala de grises. Para el caso de HOG se redimensionaron las imágenes al tamaño 1024×1024 , el cual era el tamaño estándar para la mayoría de radiografías.

2.3 Procesamiento

Esta subsección se divide en dos partes, primero se detallará el procesamiento involucrando los descriptores intermedios y después el procesamiento con las redes neuronales convolucionales. Durante este procesamiento, las librerías usadas fueron OpenCV (versión 3.4.2), SKLearn versión 0.22.2.post1 y tensorflow 2.2.0.

En el caso de HOG se utilizó la librería de *skimage* para generar el vector característico. Los parámetros usados para el HOG fueron los siguientes: las imágenes de entrada eran de tamaño 1024×1024 , la cantidad de *bins* para el histograma era de $9,32 \times 32$ píxeles por celda y bloques de celdas de 2×2 . De tal forma que por imagen se generaban vectores de tamaño $34\ 1596 \times 1$; sabiendo que son 15 000 imágenes, se concluyó que utilizar todos los valores como vector característico era inviable debido al costo computacional. Por ello, se aplicaron componentes principales a este conjunto de vectores característicos, se aplicó PCA 50, 100, 200 y 300. Se generaron así distintos vectores de estos tamaños respectivamente; de esta forma se convirtió al descriptor básico HOG en un descriptor intermedio, ya que el PCA logró potenciar al descriptor resumiendo las características y volviéndolas más representativas.

Finalmente, se pasó como entrada esta matriz de valores a un SVM, con la división de muestra y entrenamiento mencionada en la sección de recolección de datos. Se generó una búsqueda de grilla con validación cruzada 5 y los siguientes parámetros: $\text{kernel}='rbf'$, $C=[1,10,100,1000]$, $\text{gamma}=[0.1,0.01,0.001,0.0001]$.

Para obtener los vectores característicos con SIFT y SURF los parámetros que se utilizaron fueron los que vienen predeterminados en la función. A diferencia de HOG, estos descriptores no exigen que todas las imágenes sean del mismo tamaño.

Para SIFT, luego de generar los descriptores de cada imagen, obteniendo vectores de número de KeyPoints $\times 128$, se requería un solo vector característico por cada imagen, por lo que se decidió aplicar el método de *clustering*, en este caso se usó K-means. Cabe resaltar que antes de realizar este paso, se creó una nube de puntos para que sea el *input* del K-means. Como parámetros para el K-means se utilizaron clústeres de 100 y 200. Con el clúster de 100, se obtuvo un vector con 100 columnas, mientras que con 200, un vector de 200. Esto hizo que el descriptor se transforme en un descriptor intermedio.

En el caso de SURF, como la cantidad de descriptores varía dependiendo de la imagen, se obtuvieron distintos números de descriptores para cada imagen; para solo obtener uno por imagen pasamos a generar un Bag of Visual Words, mediante la clusterización de los puntos con K-means. Todos los puntos descriptores de las imágenes se concatenaron de manera vertical obteniendo una nube de puntos con cerca de 4 700 000 puntos. Se sometieron a dos procesos de clusterización, el primero fue con $K=100$ y el segundo $K=200$. Una vez entrenado el modelo de K-means, se usó para predecir en cada descriptor de cada imagen; luego se crearon histogramas de las predicciones en los descriptores de las imágenes; y este vector histograma va a ser nuestro descriptor intermedio.

Finalmente, para someter los descriptores al clasificador, se entrenó un modelo escalador que le quita la media a los datos y lo divide entre su desviación estándar. Una vez escalados, los descriptores intermedios se utilizaron de entrada para una SVM. Para identificar el mejor modelo de SVM se sometió a una búsqueda de grilla de la misma forma que en el caso de HOG.

Al observar los resultados de los descriptores intermedios se determinó que separados tenían una capacidad limitada y que combinándolos se podría mejorar su rendimiento. De esta forma, se concatenaron los vectores característicos de las tres mejores combinaciones de descriptores: HOG+PCA100, SIFT+K-means200 y SURF+K-means200. En este mismo orden se concatenaron, ordenados según sus resultados en la clasificación individual. De este modo, se obtuvo un vector característico de tamaño 500 por imagen que luego se pasó como *input* a un SVM clasificador, generando un *grid search* de validación cruzada 5 con las mismas características explicadas en los puntos anteriores.

En el caso de las CNN, las imágenes de entrada fueron redimensionadas a tensores de (224 224 3) y reescaladas con un factor de 1/255. De tal forma, se usaron modelos preentrenados con los pesos de Imagenet sin considerar la capa de salida para realizar aprendizaje por transferencia y reentrenarlos. Para la estructura VGG19, se le quitó las últimas capas *fully connected* y *softmax*, reemplazándolas con dos *fully connected* de 512 y 1024, dos capas *dropout* de 0,5 para prevenir el sobreajuste, y una capa *softmax* con 3 clases. Esta estructura se determinó después de realizar varios experimentos, dado que con otras configuraciones el nivel predictivo era muy bajo; se probó con una y dos capas *fully connected* de 4096, solo una capa de 1024 sin *dropout* y un *maxpooling3d* antes de la capa FC, todas estas pruebas no brindaron buenos resultados, obteniendo en casos *accuracies* de 0 en dos de las tres clases; además, se probó cambiando los optimizadores entre Adam y Stochastic Gradient Descent (SGD), siendo SGD el mejor en este caso. Por esta razón se llegó a probar con la arquitectura inicialmente mencionada, con la que se obtuvieron mejores resultados que se mostrarán en la siguiente sección. Se entrenó durante 100 épocas, con un tamaño de lote de 32 y una tasa de aprendizaje de 0,001.

Asimismo, para la red MobilNetV2 se retiraron las últimas capas *fully connected* y *softmax*, reemplazando, primero, con un *GlobalAveragePooling2D*, luego, con 3 *fully connected* de 1024, 1024 y 512, cada uno con su función de activación ReLU y con una capa de *softmax* con 3 clases. Esta estructura ha sido la que mejor resultado ha demostrado luego de una serie de intentos. Esta estructura se optimizó haciendo uso del optimizador SGD, se entrenó con 100 épocas con un tamaño de lote de 32 y la tasa de aprendizaje fue de 0,001.

Finalmente, para la estructura Densenet121, se agregó una capa de *AveragePooling2D* de 4×4 ; luego, se volvió vector por medio de *flatten*, y se insertó a una *fully connected* de 1024 con activación ReLU y una capa *softmax* de tres clases. Para los parámetros de entrenamiento se usó el optimizador de SGD, 100 épocas de entrenamiento, tamaño de lote de 32 y tasa de aprendizaje de 0,001.

2.3.1 Métricas

Para medir el nivel predictivo de los métodos empleados se utilizaron dos indicadores: precisión, ecuación 1; y sensibilidad, ecuación 2. Cabe resaltar que para este tipo de problemas de

clasificación de radiografías lo más importante es clasificar correctamente las que sí presentan el COVID-19, por ello la métrica más relevante será la de sensibilidad o *recall*. A continuación, se detallan las fórmulas de cada una de ellas:

$$\text{Precisión} = \frac{TP}{TP + FP} \tag{1}$$

Donde TP se refiere al número de verdaderos positivos y FP al de falsos positivos.

$$\text{Sensibilidad} = \frac{TP}{TP + FN} \tag{2}$$

Similar a la precisión, TP es verdaderos positivos y FN los falsos negativos.

3. RESULTADOS

El entrenamiento fue realizado por medio de una validación cruzada 5 en el caso de los descriptores intermedios. Los resultados de estos descriptores se presentan en las tablas 1, 2 y 3. En las tablas 1 y 2 se presentan la sensibilidad y precisión de cada uno de los experimentos realizados con los descriptores intermedios HOG+PCA, SIFT+K-means y SURF+K-means. Mientras que en la tabla 3 se muestran los resultados promedio de estos métodos.

Tabla 1
Comparando resultados por sensibilidad de los descriptores intermedios

Métodos	Sensibilidad		
	COVID-19	Normal	Neumonía
HOG+PCA50	0,66	0,91	0,78
HOG+PCA100	0,75	0,92	0,80
HOG+PCA200	0,73	0,91	0,82
HOG+PCA300	0,72	0,92	0,82
SIFT+K-means100	0,50	0,88	0,75
SIFT+K-means200	0,51	0,88	0,79
SURF+K-means100	0,41	0,91	0,72
SURF+K-means200	0,47	0,90	0,75
(HOG+PCA100)+(SIFT+K-means200) +(SURF+K-means200)	0,85	0,96	0,91

Elaboración propia

Tabla 2

Comparando resultados según precisión de los descriptores intermedios

Método	Precisión		
	COVID-19	Normal	Neumonía
HOG+PCA50	0,83	0,86	0,78
HOG+PCA100	0,86	0,87	0,80
HOG+PCA200	0,91	0,87	0,82
HOG+PCA300	0,92	0,88	0,82
SIFT+K-means100	0,81	0,83	0,75
SIFT+K-means200	0,78	0,86	0,79
SURF+K-means100	0,83	0,82	0,72
SURF+K-means200	0,84	0,84	0,75
(HOG+PCA100)+(SIFT+K-means200) +(SURF+K-means200)	0,94	0,94	0,93

Elaboración propia

En las tablas 1 y 2 se puede observar que HOG+PCA100, aun siendo un descriptor intermedio, logró predecir con un nivel aceptable las radiografías con el COVID, obteniendo una sensibilidad en el COVID-19 de 0,75 y un promedio de 0,82. Asimismo, cuando se concatenaron los vectores de los descriptores HOG, SIFT y SURF se logró potenciar la capacidad predictiva de estos descriptores logrando obtener 0,94 de precisión en la prueba y una sensibilidad de 0,85 en el COVID-19.

HOG+PCA100 demuestra ser mejor al ver los resultados específicos, debido a que es el que mejor clasifica la clase COVID-19 correctamente entre los descriptores intermedios, es por esto que se concatenó junto con los mejores descriptores de las otras categorías y se obtuvo un descriptor mucho mejor, con una precisión y sensibilidad promedio de 0,94 y 0,91 respectivamente.

Se puede ver que los descriptores intermedios han logrado acercarse a los resultados presentados en el estado del arte del COVID-Net, teniendo nuestro mejor modelo, la combinación de HOG-PCA100+SURF-K-means200+SIFT-K-means200, resultados casi al nivel de esta arquitectura con una sensibilidad en cada clase (COVID-19, condición normal y neumonía) de 0,85, 0,96 y 0,91 respectivamente versus 0,91, 0,95 y 0,94 del COVID-Net (Wang *et al.*, 2020). Cabe recalcar que esta comparación es puramente experimental, debido a que la distribución de nuestros datos es distinta a la usada en el trabajo de Wang.

Tabla 3
Comparando resultados promedio de los descriptores intermedios

Métodos	Precisión	Sensitividad
HOG+PCA50	0,84	0,79
HOG+PCA100	0,86	0,82
HOG+PCA200	0,88	0,82
HOG+PCA300	0,89	0,82
SIFT+K-means100	0,81	0,71
SIFT+K-means200	0,82	0,73
SURF+K-means100	0,82	0,68
SURF+K-means200	0,83	0,71
(HOG+PCA100)+(SIFT+K-means200) +(SURF+K-means200)	0,94	0,91

Elaboración propia

Los resultados que se obtuvieron con las CNN VGG19, MobilnetV2 y Densenet121 con y sin aumento de datos se pueden observar en términos específicos en las tablas 4 y 5, y sus resultados en promedio en la tabla 6.

Tabla 4
Comparando resultados según sensibilidad de las CNN

Métodos	Sensitividad		
	COVID-19	Normal	Neumonía
VGG19	0,46	0,91	0,89
VGG19+DATA AUG	0,96	0,93	0,82
MobilnetV2	0,74	0,93	0,85
MobilnetV2+DATA AUG	0,97	0,95	0,81
Densenet121	0,45	0,92	0,84
Densenet12+DATA AUG	0,94	0,92	0,81

Elaboración propia

En la tabla 4 podemos observar que la CNN que obtuvo mejores resultados en sensibilidad fue la MobilnetV2 con un *recall* en el COVID-19 de 0,74 sin aumento de datos y 0,97 con aumento de datos.

Tabla 5
Comparando resultados según precisión de las CNN

Métodos	Precisión		
	COVID-19	Normal	Neumonía
VGG19	0,90	0,91	0,84
VGG19+DATA AUG	0,96	0,88	0,88
MobilenetV2	0,87	0,90	0,89
MobilenetV2+DATA AUG	0,97	0,88	0,91
Densenet121	0,80	0,89	0,85
Densenet121+DATA AUG	0,95	0,87	0,87

Elaboración propia

Se obtuvo que VGG19 fue la mejor CNN en términos de precisión con un 0,90 en la clase COVID-19 sin aumento de datos.

Tabla 6
Comparando resultados promedio de las CNN

Métodos	Precisión	Sensitividad
VGG19	0,88	0,75
VGG19+DATA AUG	0,91	0,90
MobilenetV2	0,89	0,84
MobilenetV2+DATA AUG	0,92	0,91
Densenet121	0,84	0,74
Densenet121+DATA AUG	0,90	0,89

Elaboración propia

Finalmente, se puede observar que la mejor CNN sería MobilenetV2, con un optimizador SGD con 100 épocas, con un tamaño de lote de 32 y la tasa de aprendizaje de 0,001; además de tener una buena sensibilidad por clase respecto a las demás, presentó un promedio de precisión y sensibilidad de 0,89 y 0,84 respectivamente. Al utilizar los datos generados con aumento de datos, el mejor modelo obtenido fue también MobilenetV2 con una sensibilidad y precisión en la clase COVID-19 de 0,97 y 0,97 respectivamente, como se puede ver en las tablas 4 y 5, además de una precisión y sensibilidad promedio de 0,92 y 0,91 respectivamente, tabla 6. Esto nos demuestra que, como explica el estado del arte, las CNN necesitan datos balanceados para poder aprender las características de las imágenes correctamente. En la figura 2 se presentan las matrices de confusión de los tres mejores modelos: MobilenetV2, MobilenetV2 + aumento de datos y la combinación de los tres mejores descriptores intermedios.

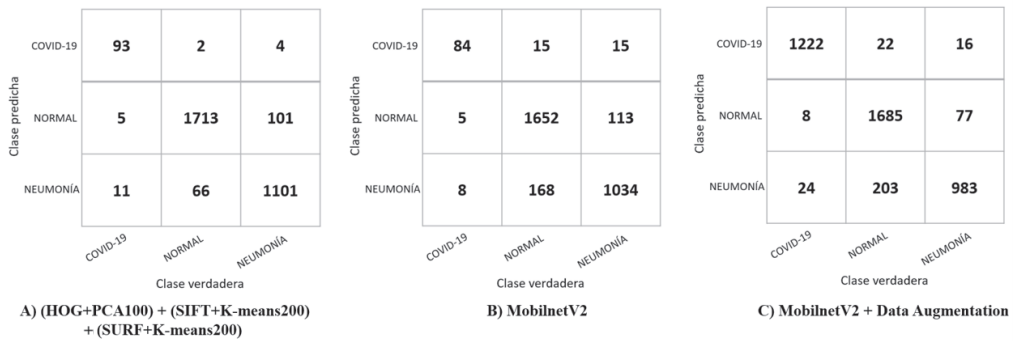


Figura 2. Matrices de confusión de los tres mejores modelos
Elaboración propia

4. CONCLUSIONES

En este trabajo se propuso una metodología para la detección del COVID-19 en radiografías de tórax, clasificando en tres categorías: COVID-19, condición normal y neumonía. Así se logró demostrar la capacidad que tienen los descriptores intermedios y cómo las CNN se pueden potenciar para lograr buenos resultados en este tipo de problemáticas.

En conclusión, se obtuvieron buenos resultados en los casos de radiografías con neumonía o normales con el uso de extractores de vectores característicos intermedios; de esta forma, en el caso del COVID-19, solo con HOG+PCA300 se visualizan resultados satisfactorios con una precisión promedio de 0,89 y sensibilidad de 0,82. La mayor importancia recae en la métrica de sensibilidad por su importancia en el caso de clasificación de radiografías (mencionado anteriormente), debido a que el objetivo principal es clasificar bien qué radiografía de tórax tiene el COVID-19 y cuál no. En este caso se logró demostrar cómo la configuración del HOG propuesta en esta investigación ha logrado muy buenos resultados en comparación al estado del arte al ser combinada con componentes principales. Asimismo, al concatenar los tres mejores descriptores intermedios, se pudo observar que se potenciaron mutuamente y elevaron considerablemente la sensibilidad y precisión para la clase COVID-19, 0,85 y 0,94, logrando muy buenos resultados con una precisión y sensibilidad promedio de 0,94 y 0,91, comparable experimentalmente a los resultados del COVID-Net de Wang *et al.* (2020). Se obtienen, con esta combinación de descriptores intermedios, mejores resultados que con las CNN sin aplicar aumento de datos.

El resultado de MobilnetV2, a pesar de tener una base de datos considerablemente desbalanceada, ha demostrado ser un clasificador que arroja muy buenos resultados, obteniendo una buena sensibilidad en comparación con las otras CNN. De esta forma, observamos que este

resultado mejoró al utilizar aumento de datos, logrando una precisión y sensibilidad promedio de 0,92 y 0,91, superando así a los demás métodos mostrados en la investigación.

Asimismo, la Densenet121 con la configuración de salida utilizada tuvo el peor desempeño en precisión del COVID-19; esto se puede deber a la baja cantidad de imágenes de esta clase y que esta red no ha podido aprender muy bien la información. Al utilizar el aumento de datos, esta red también se potenció mejorando su capacidad predictiva. Algo interesante es que en este escenario HOG+PCA (50, 100, 200, 300) se desarrolló mejor, al igual que la combinación de los descriptores intermedios HOG+SIFT+SURE.

Por otro lado, al entrenar la red VGG19 con aumento de datos se potenció su poder predictivo y se llegó a una sensibilidad en la clase COVID-19 de 0,96; además de lograr, en promedio, una precisión y sensibilidad de 0,91 y 0,90. De esta forma, se demostró que para las CNN es de vital importancia que los datos estén balanceados o, en su defecto, tengan miles de imágenes en cada una de las clases; de lo contrario, como se vio en los experimentos con las otras redes, no se llegará a aprender correctamente las características entre clases.

Finalmente, este trabajo y sus resultados son presentados como un marco de referencia para investigaciones futuras y no como un modelo listo para entrar a proceso de producción.

REFERENCIAS

- Apostolopoulos, I. D., y Mpesiana, T. A. (2020). COVID-19: Automatic Detection from X-Ray Images Utilizing Transfer Learning with Convolutional Neural Networks. *Physical and Engineering Sciences in Medicine*, 43(2), 635-640. <https://doi.org/10.1007/s13246-020-00865-4>
- Bay, H., Tuytelaars, T., y Van Gool L. (2006). SURF: Speeded Up Robust Features. *ECCV 2006*. 3951, 404-417. 10.1007/11744023_32
- Cohen, J. P., Morrison, P., y Dao, L. (2020). COVID-19 Image Data Collection: Prospective Predictions Are the Future. <https://arxiv.org/pdf/2003.11597.pdf>
- Bell *et al.* (s. f.). *COVID-19*. <https://radiopaedia.org/articles/covid-19-4>
- Dalal, N., y Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886-893. 10.1109/CVPR.2005.177
- European Society of Radiology. (2020). *COVID*. <https://www.eurorad.org/advanced-search?search=COVID>
- Gao, H., Zhuang, L., van der Maaten, L., y Weinberger, K. (2018). *Densely Connected Convolutional Networks*. <https://arxiv.org/pdf/1608.06993.pdf>

- Lowe, D. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60, 91-110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- Oh, Y., Park, S., y Ye, J. C. (2020). Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets. *IEEE Transactions on Medical Imaging*, 39(8), 2688-2700. <https://doi.org/10.1109/TMI.2020.2993291>
- Radiological Society of North America. (Octubre del 2018). RSNA Pneumonia Detection Challenge [Dataset]. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>
- Rahman, T., Chowdhury, M., y Khandakar A. (Marzo del 2020). COVID-19 Radiography [Dataset]. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., y Chen L. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4510-4520. 10.1109/CVPR.2018.00474
- Simonyan, K., y Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. <https://arxiv.org/pdf/1409.1556.pdf>
- Società Italiana di Radiologia Medica e Interventistica. (2020). *COVID-19 Database*. <https://www.sirm.org/category/senza-categoria/COVID-19/>
- Wang, L., Wong, A., Qiu Lin, Z., McInnis, P., Chung, A., y Gunraj, H. (2020). Actualmed COVID-19 Chest X-Ray Dataset Initiative [Dataset]. <https://github.com/agchung/Actualmed-COVID-chestxray-dataset>
- Wang, L., Qiu Lin, Z., y Wong, A. (2020). Figure 1 COVID-19 Chest X-Ray Dataset Initiative. <https://github.com/agchung/Figure1-COVID-chestxray-dataset>
- Wang, L., Qiu Lin, Z., y Wong, A. (2020). COVIDx Dataset (version 3) [Dataset]. <https://github.com/lindawang/COVID-Net/blob/master/docs/COVIDx.md>
- Wang, L., y Wong, A. (2020). COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. arXiv preprint arXiv:2003.09871.
- Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., y Tan, W. (2020). Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *Jama*, 323(18), 1843-1844.
- World Health Organization. (9 de octubre del 2020). WHO Coronavirus Disease (COVID-19) *Dashboard*. https://covid19.who.int/?gclid=CjwKCAjw_Y_8BRBiEiwA5MCBjHROPzfziFPFQFT_QltAEQ_F29AGqJd21iBefvHeHeVvQCSxUJJJbhoCowwQAvD_BwE