

Clasificación de Imágenes de Radiografía de Pulmón utilizando Aprendizaje Bayesiano

Adrián Armando Fazio*, Juliana Gambini* y Juan Miguel Santos*

*CIDIA, Universidad Nacional de Hurlingham

adrian.fazio@estudiantes.unahur.edu.ar; {juliana.gambini,juan.santos}@unahur.edu.ar

RESUMEN

El objetivo del presente trabajo es clasificar radiografías de torax para colaborar en el diagnóstico de neumonía provocada por el virus SAR-COV 2 o por otras causas, así como detectar pulmones sanos. Para llevar a cabo este estudio, se utilizan imágenes de radiografías de tres tipos, pulmones sanos, lesiones pulmonares por neumonía por COVID-19 o neumonía por otras causas. En primer lugar, se realiza un preprocesamiento de estas imágenes, que incluye recortes de la mitad inferior de la imagen (donde se observaban diferencias entre las distintas imágenes) y la división en cuadrados para analizar sus histogramas. Luego, se realizan varias series de experimentos utilizando el método de Bayes para la clasificación y posterior inferencia.

Se utilizan dos clasificadores, uno que separa las imágenes de pulmones sanos, de aquellas que sean de pulmones enfermos, y otro que distingue neumonía por COVID-19 y neumonía por otras causas. Los resultados son prometedores.

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Esta línea de investigación tiene como objetivo el diseño y desarrollo de un sistema que, dada una imagen de radiografía torácica puede decidir si el paciente tiene la enfermedad de COVID-19, tiene neumonía provocada por otras causas o si tiene los pulmones sanos. Esta decisión se toma entrenando un modelo de aprendizaje supervisado basado en inferencia Bayesiana.

La primera parte de este desarrollo consiste en dividir la imagen en rectángulos de píxeles no superpuestos, luego para cada rectángulo se calcula el histograma. Así cada imagen queda caracterizada por un conjunto de histogramas.

es posible calcular la probabilidad a posteriori de que una imagen pertenezca a una de las tres clases y luego elegir la clase más probable. Para esto utilizamos el método de Bayes ingenuo, es decir, realizando la simplificación de que el color de cada píxel dentro de la imagen es independiente del resto de los mismos. Luego calculamos el histograma de cada imagen y la probabilidad de que el ese histograma pertenezca a una clase.

Los objetivos de esta línea de investigación son los siguientes:

- Tomar una base de datos con imágenes etiquetadas en tres clases: COVID-19, Neumonía por otras causas, pulmones sanos.
- Entrenar un modelo basado en el método de Bayes ingenuo para clasificación supervisada, utilizando como características un conjunto de histogramas ordenados.
- Evaluar los resultados utilizando métricas de evaluación estándar, como *Accuracy*, *Precision*, *Recall*, etc.
- Construir un sistema que, dada una imagen de radiografía torácica pueda decidir a qué clase pertenece.

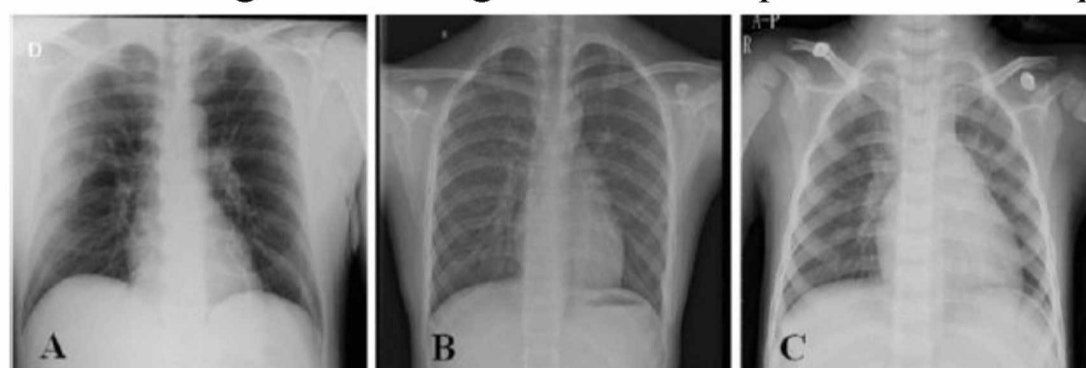


Imagen de muestra de rayos X del conjunto de datos: (A) neumonía por COVID-19, (B) pulmones normales y (C) neumonía por otras causas.

Experimentos y Resultados

Los parámetros utilizados en los experimentos:

Parámetro	Valores
Tamaño del parche	16 × 16; 23 × 23; 32 × 32
Cantidad de bins	8; 16; 32
Repeticiones del experimento	5; 10
Umbral	100; 150; 250
Cantidad de imágenes a procesar	250; 500; 750; 1000; 1250

Resultados de las métricas de evaluación, utilizando histogramas de 32 bins en parches de 32 × 32 y analizando la Clase 0 (Pulmones normales) contra 1 y 2 (Pulmones con COVID-19 y con neumonía por otras enfermedades) con 250 imágenes

Bins	Parche - Tamaño	Clase 0				Cantidad imágenes - CO-c1y2	
		Precision	Acurrancy	Recall	F1Score		
32	32*32	Media	0,8833	0,6375	0,3146	0,4622	250
		Desviación	0,0754	0,0388	0,0604	0,0745	
32	32*32	Media	0,7737	0,6902	0,5379	0,6324	500
		Desviación	0,0558	0,0466	0,0748	0,0653	
32	32*32	Media	0,8129	0,7601	0,6773	0,7371	750
		Desviación	0,0187	0,0201	0,0572	0,0323	
32	32*32	Media	0,8169	0,7482	0,6423	0,7187	1000
		Desviación	0,0512	0,0305	0,0234	0,0310	
32	32*32	Media	0,8052	0,7648	0,7013	0,7488	1250
		Desviación	0,0305	0,0139	0,0293	0,0146	

Resultados utilizando histogramas de 32 bins, en parches de 16 × 16 y analizando la Clase 1 (Pulmones con Covid-19) contra 2 (Pulmones con neumonía por otras enfermedades) con 750 imágenes

Bins	Parche - Tamaño	Clase 1 contra 2				Cantidad imágenes - CO-c1y2	
		Precision	Acurrancy	Recall	F1Score		
32	16*16	Media	0,8777	0,7427	0,5587	0,6772	250
		Desviación	0,0913	0,0801	0,1336	0,1188	
32	16*16	Media	0,8747	0,7680	0,6284	0,7278	500
		Desviación	0,0494	0,0364	0,0790	0,0555	
32	16*16	Media	0,9571	0,8034	0,6343	0,7622	750
		Desviación	0,0272	0,0357	0,0559	0,0470	
32	16*16	Media	0,8638	0,7414	0,5760	0,6883	1000
		Desviación	0,0347	0,0227	0,0622	0,0404	
32	16*16	Media	0,8699	0,7299	0,5440	0,6672	1250
		Desviación	0,0417	0,0150	0,0439	0,0276	

- Para el método de la Clase 0 (pulmones sanos) contra las clases 1 y 2 (pulmones con COVID-19 y pulmones con neumonía por otras enfermedades, respectivamente), con 32 bins, parches de 32 × 32 obtuvimos: Precisión: 88% y Exactitud: 63, 75%
- Para el método de la Clase 1 (pulmones con COVID-19) y 2 (pulmones con neumonía por otras enfermedades) con 32 bins, parches de 16 × 16 obtuvimos: Precisión: 95, 71% y Exactitud: 80, 34%.