

# Tarea 3 – Reconocimiento de Patrones

Rodrigo Nazar Meier, *Pontificia Universidad Católica de Chile*

**Abstract**—Hoy día estamos viviendo una de las pandemias más grandes que ha tenido que enfrentar la sociedad moderna, la del COVID-19. En la actualidad la enfermedad presenta elevadísimas tasas de contagio, limitados métodos de tratamiento y lentas formas de diagnosticar. Es por esto que en el siguiente informe se presenta una posible forma de diagnosticar la enfermedad a partir de radiografías de tórax.

**Index Terms**—Reconocimiento de Patrones, Radiografía, COVID-19.

## 1 SOLUCIÓN PROPUESTA

PARA este problema, conté con extractos de radiografías de tres distintas clases: pacientes sanos, con neumonía y COVID-19. Como estas imágenes están en escala de grises, me enfoqué en obtener características que describen las texturas de las imágenes.

Las características elegidas para esta solución, son las siguientes:

- **LBP de 59 elementos**
- **Haralick:** Se calcularon las matrices de co-ocurrencia en 8 direcciones y se les midió el contraste, la similitud, la homogeneidad, el ASM, la energía y finalmente la correlación. Obteniendo así  $8 \cdot 6 = 48$  características.
- **Descriptores de Gabor:** Se obtuvo el resultado de la convolución de cada imagen con cuatro kernels. A cada uno de estos, se le calculó la media, la varianza y la suma de sus componentes. Obteniendo así  $4 \cdot 3 = 12$  características.

Luego, se implementaron una serie de estrategias distintas para manipular los datos. Las que se muestran en la figura 1.

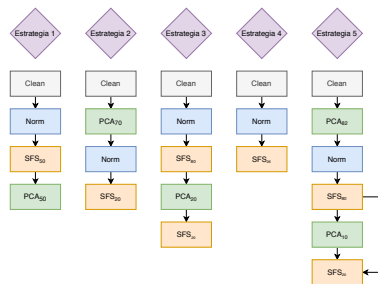


Fig. 1: Estrategias implementadas en el programa.

Se diseñaron dos tipos de clasificadores: el primero que clasifica cada extracto de radiografía de forma individual y el segundo considera los 10 extractos de cada paciente para tomar una decisión de su diagnóstico. Ambos clasificadores se probaron con las estrategias de la figura 1, sólo que el segundo clasificador tomó en cuenta la mayoría de votos de los diagnósticos de cada uno de los pacientes. Ambos clasificadores se testearon

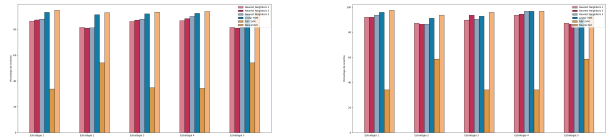


Fig. 2: Porcentaje de aciertos por de cada algoritmo de clasificación para el clasificador 1 y 2 respectivamente.

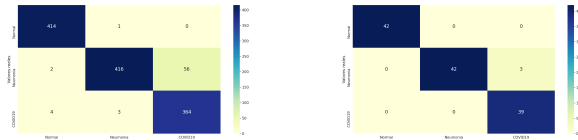


Fig. 3: Matriz de Confusión para el clasificador 1 y 2 respectivamente.

con: KNN de 1, 3 y 5 vecinos, SVM lineal y RBF y finalmente con redes neuronales.

## 2 EXPERIMENTOS REALIZADOS

Se dispuso de 1680 fotografías por clase para el entrenamiento del algoritmo, junto de 420 por clase para el testeo. En la figura 2 se presentan los porcentajes de aciertos obtenidos por cada estrategia (conjunto de columnas) con cada uno de los 5 algoritmos de clasificación empleados (barras de colores), para el clasificador 1 y 2 respectivamente.

El mejor resultado se obtuvo en ambos casos con la estrategia número 1 y un clasificador de redes neuronales, con un 94.7% y 97.6% para los clasificadores 1 y 2 respectivamente. En la figura 3 se presentan las matrices de confusión de ambos casos.

**Importante:** revisar el archivo *README.md* de la tarea para ver detalles de la implementación.

## 3 CONCLUSIONES

Son impresionantes los altos porcentajes de acierto de las técnicas mostradas. Esto podría indicar que esta es una viable ruta para el diagnóstico de la enfermedad, pero para poder concluir eso es necesario complementar estudios anexos al área de la salud que respalden la eficacia del método.