



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA  
FACULTAD DE CIENCIAS DE LA ELECTRÓNICA  
MAESTRÍA EN INGENIERÍA ELECTRÓNICA,  
OPCIÓN INSTRUMENTACIÓN ELECTRÓNICA

Tesis para obtener el grado de:  
MAESTRO EN INGENIERÍA ELECTRÓNICA

---

Normalización y alineación automática de la forma de la región  
pulmonar integrada con selección de características discriminantes  
para detección de neumonía y COVID-19

---

Presenta:

Lic. Rafael Alejandro Cruz Ovando\*

Directores:

Dr. Salvador Eugenio Ayala Raggi

Dr. Aldrin Barreto Flores

# **Objetivos**

## **Objetivo general**

Desarrollar e implementar algoritmos de visión por computadora para la detección, alineación y normalización de la forma de la región pulmonar en imágenes radiográficas de tórax, utilizando además un método eficaz para la selección de características discriminantes, con el fin de mejorar la precisión en la detección automática de neumonía y COVID-19.

## **Objetivos específicos**

1. Diseñar, implementar y evaluar un método deformable de alineación y normalización que localice, segmente y ajuste automáticamente la región pulmonar en términos de forma, escala, posición y rotación.
2. Proponer un método de extracción y selección de características que maximicen la discriminación entre las clases.
3. Evaluar el rendimiento de diferentes clasificadores de aprendizaje supervisado para la técnica de alineación propuesta en la tesis: KNN, CNN, MLP.
4. Validar el clasificador desarrollado a través de medir la precisión, sensibilidad, especificidad y además de realizar pruebas de validación cruzada para caracterizar el algoritmo propuesto.
5. Contrastar los resultados de clasificación del objetivo anterior con resultados obtenidos por los mismos clasificadores pero sin realizar el proceso de alineación propuesto.
6. Publicación de resultados.

# Índice general

Objetivos . . . . .	1
<b>1. Introducción</b>	<b>5</b>
1.1. Trabajo Relacionado . . . . .	5
1.2. Justificación . . . . .	6
1.3. Contribuciones del Trabajo . . . . .	7
1.4. Estructura de la Tesis . . . . .	7
<b>2. Marco Teórico</b>	<b>9</b>
2.1. Representación anatómica mediante puntos de referencia . . . . .	9
2.2. Preprocesamiento: CLAHE . . . . .	9
2.3. Preprocesamiento: SAHS . . . . .	10
2.4. Redes Neuronales Convolucionales . . . . .	13
2.5. ResNet y conexiones residuales . . . . .	13
2.6. Normalización por Lote . . . . .	14
2.7. Aprendizaje por transferencia . . . . .	15
2.8. Función de pérdida: Wing Loss . . . . .	16
2.9. Análisis Procrustes Generalizado (GPA) . . . . .	17
2.10. Triangulación de Delaunay . . . . .	19
2.11. Deformación afín por partes . . . . .	19
2.12. Clasificación de imágenes . . . . .	20
2.12.1. Función softmax . . . . .	22
2.12.2. Función de pérdida: entropía cruzada . . . . .	22
2.12.3. Ponderación de clases . . . . .	22
2.13. Métricas de evaluación . . . . .	22
2.13.1. Métricas para detección de puntos de referencia . . . . .	22
2.13.2. Métricas para clasificación . . . . .	23
2.14. Mecanismo de Coordinate Attention . . . . .	23
2.15. Ensamble de modelos y Test-Time Augmentation . . . . .	24
2.15.1. Ensamble de modelos . . . . .	24
2.15.2. Aumento en tiempo de prueba (TTA) . . . . .	25
<b>3. Metodología</b>	<b>26</b>
3.1. Descripción General del Sistema . . . . .	26
3.1.1. Arquitectura del Sistema . . . . .	26
3.1.2. Flujo de Datos . . . . .	28
3.1.3. Justificación del Diseño Modular . . . . .	28

3.2.	Conjunto de Datos y Preprocesamiento . . . . .	30
3.2.1.	COVID-19 Radiography Database . . . . .	30
3.2.2.	Anotación de Puntos de Referencia Anatómicos . . . . .	30
3.2.3.	Preprocesamiento de Imágenes . . . . .	34
3.2.4.	División del Conjunto de Datos . . . . .	35
3.3.	Modelo de Predicción de Puntos de Referencia . . . . .	36
3.3.1.	Arquitectura del Modelo . . . . .	36
3.3.2.	Función de Pérdida . . . . .	41
3.3.3.	Estrategia de Entrenamiento . . . . .	43
3.3.4.	Resumen de Hiperparámetros . . . . .	46
3.3.5.	Ensamble de Modelos . . . . .	46
3.4.	Normalización Geométrica . . . . .	49
3.4.1.	Análisis Procrustes Generalizado . . . . .	49
3.4.2.	Triangulación de Delaunay . . . . .	52
3.4.3.	Transformación Afín por Partes . . . . .	55
3.4.4.	Proceso Completo de Normalización . . . . .	57
3.5.	Clasificación de Enfermedades Pulmonares . . . . .	59
3.5.1.	Preprocesamiento de Contraste . . . . .	59
3.5.2.	Arquitectura del Clasificador . . . . .	59
3.5.3.	Estrategia de Aprendizaje por Transferencia . . . . .	61
3.5.4.	Configuración del Entrenamiento . . . . .	62
3.5.5.	Métricas de Evaluación del Clasificador . . . . .	64
3.5.6.	Aumento de Datos . . . . .	65
3.5.7.	Resumen de la Configuración . . . . .	66
3.6.	Protocolo de Inferencia y Evaluación . . . . .	67
3.6.1.	Proceso de Inferencia . . . . .	67
3.6.2.	Métricas de Evaluación . . . . .	68
<b>4.</b>	<b>Resultados</b> . . . . .	<b>70</b>
4.1.	Detección de Puntos de Referencia . . . . .	70
4.1.1.	Precisión del Sistema . . . . .	70
4.1.2.	Precisión por Punto de Referencia . . . . .	70
4.1.3.	Resumen . . . . .	71
4.2.	Normalización Geométrica . . . . .	73
4.2.1.	Forma Estándar de Referencia . . . . .	73
4.2.2.	División en Triángulos para Transformación . . . . .	73
4.2.3.	Ejemplos de Normalización . . . . .	73
4.2.4.	Resumen . . . . .	73
4.3.	Clasificación de Enfermedades Pulmonares . . . . .	78

4.3.1. Rendimiento General . . . . .	78
4.3.2. Validación Cruzada . . . . .	78
4.3.3. Rendimiento por Categoría . . . . .	79
4.3.4. Análisis de Errores . . . . .	79
4.3.5. Efecto de la Normalización Geométrica . . . . .	81
4.3.6. Resumen . . . . .	84
<b>5. Conclusiones y Trabajos Futuros</b>	<b>86</b>
5.1. Síntesis de Contribuciones . . . . .	86
5.1.1. Contribución Principal . . . . .	86
5.1.2. Contribuciones Específicas . . . . .	87
5.2. Validación de la Hipótesis . . . . .	88
5.2.1. Hipótesis Planteada . . . . .	88
5.2.2. Evidencia de Validación . . . . .	88
5.2.3. Limitaciones de la Validación . . . . .	90
5.2.4. Respuesta a la Hipótesis . . . . .	90
5.3. Implicaciones del Trabajo . . . . .	91
5.3.1. Implicaciones Clínicas . . . . .	91
5.3.2. Implicaciones Metodológicas . . . . .	91
5.3.3. Implicaciones Técnicas . . . . .	92
5.4. Limitaciones del Estudio . . . . .	92
5.4.1. Limitaciones Experimentales . . . . .	92
5.4.2. Limitaciones Metodológicas . . . . .	92
5.4.3. Limitaciones Conceptuales . . . . .	93
5.5. Trabajos Futuros . . . . .	93
5.5.1. Validación y Generalización . . . . .	93
5.5.2. Extensiones del Sistema . . . . .	94
5.5.3. Mejoras Metodológicas . . . . .	94
5.5.4. Interpretabilidad y Explicabilidad . . . . .	95
5.5.5. Optimización e Implementación . . . . .	95
5.6. Reflexión Final . . . . .	96
<b>Glosario</b>	<b>97</b>

# Capítulo 1

## Introducción

La neumonía representa una de las principales causas de mortalidad a nivel mundial, con millones de casos anuales que requieren diagnóstico oportuno y preciso. La pandemia de COVID-19 evidenció la necesidad crítica de métodos de detección automática basados en radiografías de tórax, especialmente en entornos con acceso a recursos limitado. Los sistemas de diagnóstico automatizado deben alcanzar niveles de exactitud y robustez suficientes para apoyar efectivamente la toma de decisiones clínicas, reduciendo la variabilidad inherente en la interpretación manual de imágenes radiográficas [1].

La clasificación errónea de neumonía tiene consecuencias significativas: un falso negativo en COVID-19 puede resultar en propagación inadvertida de la enfermedad, mientras que un falso positivo genera costos innecesarios. Para ser clínicamente viables, los sistemas automatizados deben demostrar no solo alta precisión sino también robustez ante la variabilidad en la adquisición de imágenes, incluyendo diferencias en el posicionamiento del paciente, distancia de proyección, calibración de equipos y condiciones de respiración.

En este contexto, las técnicas de visión por computadora y aprendizaje profundo ofrecen una vía prometedora para el desarrollo de sistemas de apoyo al diagnóstico. Sin embargo, la variabilidad extrínseca en la adquisición de radiografías representa un desafío significativo para la generalización de modelos. Este trabajo aborda dicho desafío mediante normalización geométrica de la región pulmonar, con el objetivo de reducir variabilidad de pose, facilitando el aprendizaje de características patológicas.

### 1.1. Trabajo Relacionado

Diversos trabajos han propuesto métodos de detección automática de COVID-19 y neumonía mediante aprendizaje profundo aplicado directamente sobre radiografías de tórax. Wang et al. [2] desarrollaron COVIDNet, una arquitectura diseñada específicamente para clasificación multi-clase de COVID-19, neumonía viral y neumonía bacteriana. Rajpurkar et al. [3] demostraron que CheXNet, basada en DenseNet-121, alcanza performance comparable a radiólogos en detección de neumonía. Chowdhury et al. [4] compilaron el COVID-19 Radiography Database utilizado en este trabajo, facilitando investigación reproducible. Rahman et al. [5] exploraron técnicas de mejora de imagen para detección de COVID-19, evaluando métodos de preprocesamiento. Estos enfoques aplican redes convolucionales sobre imágenes originales o con preprocesamiento de contraste, sin normalización geométrica

explícita.

Trabajos recientes han demostrado que la normalización geométrica puede mejorar significativamente el rendimiento de clasificadores médicos. Picazo-Castillo et al. [6] presentaron un estudio comparativo de representaciones de imágenes pulmonares para reconocimiento automático de neumonía, demostrando que diferentes estrategias de normalización espacial afectan la capacidad de generalización de modelos CNN. Ayala-Raggi et al. [7] propusieron la integración de normalización de imágenes de tórax con selección discriminativa de características para reconocimiento eficiente de COVID-19, logrando mejoras en exactitud mediante la reducción de variabilidad extrínseca. Rocha et al. [8] desarrollaron STERN, una red que combina Spatial Transformer Networks con mecanismos de atención para detección de anomalías en radiografías de tórax, permitiendo alineación implícita de regiones anatómicas. Yeh et al. [9] demostraron que la detección automática de puntos de referencia anatómicos y su uso para análisis de alineación mejora significativamente el diagnóstico en radiografías de columna vertebral.

La evidencia acumulada sugiere que la normalización geométrica es una estrategia efectiva para mejorar sistemas de clasificación en imagen médica. Sin embargo, los trabajos previos se limitan a transformaciones rígidas (rotación, traslación) o afines globales, que asumen uniformidad en la deformación del tejido pulmonar, una suposición inexacta dado que el pulmón es un órgano deformable cuya forma varía según la fase respiratoria, posición del paciente y patología subyacente. Este trabajo explora la hipótesis de que la normalización mediante *deformación afín por partes*, que permite deformación local preservando estructura triangular, puede eliminar variabilidad de pose sin sacrificar información diagnóstica contenida en la textura pulmonar, adaptando la región pulmonar a una forma estándar común.

## 1.2. Justificación

El presente trabajo propone un método de normalización geométrica basado en *deformación afín por partes* que, a diferencia de los enfoques previos limitados a transformaciones rígidas o afines globales, permite deformación local adaptativa de la región pulmonar. Esta estrategia ofrece mayor flexibilidad al adaptar cada región a una geometría común sin asumir uniformidad en la deformación, preservando información diagnóstica local mientras se normaliza la variabilidad extrínseca de pose y fase respiratoria. Mediante la detección automática de 15 puntos de referencia anatómicos que definen el contorno pulmonar, se establece correspondencia geométrica entre cada imagen y una forma estándar obtenida mediante Análisis General de Procrustes, aplicando transformaciones afines independientes a cada triángulo de una malla de Delaunay. La hipótesis central es que esta normalización facilita el aprendizaje de características intrínsecas relacionadas con patología (opacidades en vidrio esmerilado, consolidaciones, infiltrados intersticiales) al

reducir la complejidad del espacio de variación que debe modelar el clasificador.

### 1.3. Contribuciones del Trabajo

El sistema propuesto integra técnicas de aprendizaje profundo, análisis de forma y geometría computacional en un proceso completo que comprende dos fases principales. La primera fase, denominada fase de preparación, incluye la anotación manual de puntos de referencia anatómicos en un subconjunto representativo de imágenes y el entrenamiento de los modelos de redes neuronales para detección de puntos de referencia y clasificación. La segunda fase, denominada fase de operación, procesa nuevas radiografías mediante cuatro módulos secuenciales. El módulo de preprocesamiento normaliza las variaciones de contraste mediante CLAHE (*Contrast Limited Adaptive Histogram Equalization*) y redimensiona las imágenes a formato estándar. El módulo de predicción de puntos de referencia emplea una red ResNet-18 con Coordinate Attention para estimar automáticamente las coordenadas de 15 puntos anatómicos que definen el contorno pulmonar bilateral. El módulo de normalización geométrica utiliza los puntos de referencia predichos para aplicar *deformación afín por partes*, alineando cada imagen a una forma estándar pulmonar previamente calculada mediante Análisis General de Procrustes sobre las configuraciones del conjunto de entrenamiento, generando correspondencias geométricas mediante triangulación de Delaunay y aplicando transformaciones afines independientes a cada triángulo. Finalmente, el módulo de clasificación procesa las imágenes normalizadas mediante una red ResNet-18 que discrimina entre COVID-19, neumonía viral y casos normales, operando sobre representaciones geométricamente consistentes que facilitan la identificación de patrones diagnósticos intrínsecos.

### 1.4. Estructura de la Tesis

El resto del documento se organiza de la siguiente manera: El Capítulo 2 presenta el marco teórico, estableciendo los fundamentos científicos y matemáticos que sustentan el trabajo, incluyendo formación de imagen radiográfica mediante atenuación de rayos X, arquitecturas de redes neuronales convolucionales y transferencia de aprendizaje, análisis de forma mediante Procrustes, triangulación de Delaunay, deformación geométrica, y métricas de evaluación para regresión de puntos de referencia y clasificación. El Capítulo 4 describe la metodología propuesta en detalle, documentando el diseño del sistema completo, la arquitectura de los modelos de detección de puntos de referencia y clasificación, el protocolo de entrenamiento con aumento de datos y regularización, el proceso de normalización geométrica triángulo por triángulo, la generación del conjunto de datos deformado, y el flujo de inferencia de extremo a extremo. El Capítulo 5 presenta los resultados experimentales, enfocándose en tres

aspectos clave: la precisión del modelo de puntos de referencia evaluada mediante error en píxeles sobre 96 imágenes con anotaciones de referencia, la mejora en clasificación atribuible a la normalización geométrica mediante comparación directa contra métodos alternativos de preprocesamiento, y la robustez del sistema analizada mediante validación cruzada, matrices de confusión y análisis de casos mal clasificados. Finalmente, el Capítulo 6 sintetiza las conclusiones del trabajo, analizando la evidencia sobre la efectividad de la normalización geométrica para mejorar sistemas de detección de neumonía, validando la hipótesis central mediante métricas cuantitativas, discutiendo las implicaciones prácticas para sistemas de apoyo al diagnóstico, reconociendo limitaciones del enfoque propuesto relacionadas con dependencia en la calidad de predicción de puntos de referencia y generalización a otros conjuntos de datos, e identificando direcciones prometedoras para investigación futura.

# Capítulo 2

## Marco Teórico

### 2.1. Representación anatómica mediante puntos de referencia

En el análisis de imágenes médicas, los *puntos de referencia* son coordenadas específicas que representan estructuras anatómicas de interés. En este trabajo, se utilizan 15 puntos de referencia para definir el contorno pulmonar en radiografías de tórax.

La distribución de los puntos de referencia sigue una estructura lógica:

- **Eje central** (5 puntos): L1, L9, L10, L11, L2, que representan la línea media del tórax, desde el ápice hasta la base.
- **Contorno del pulmón izquierdo** (5 puntos): L12, L3, L5, L7, L14.
- **Contorno del pulmón derecho** (5 puntos): L13, L4, L6, L8, L15.

Debido a la simetría bilateral del tórax, existen 5 pares de puntos de referencia simétricos: (L3, L4), (L5, L6), (L7, L8), (L12, L13) y (L14, L15). Esta propiedad es aprovechada durante el entrenamiento y la evaluación del modelo.

Matemáticamente, el conjunto de puntos de referencia de una imagen se representa como un vector:

$$\mathbf{L} = [x_1, y_1, x_2, y_2, \dots, x_{15}, y_{15}]^\top \in \mathbb{R}^{30} \quad (2.1)$$

donde  $(x_i, y_i)$  son las coordenadas del  $i$ -ésimo punto de referencia. Este vector de 30 valores es la salida que predice el modelo de detección de puntos de referencia.

### 2.2. Preprocesamiento: CLAHE

Las radiografías de tórax suelen presentar bajo contraste, lo que dificulta la visualización de estructuras anatómicas como los bordes pulmonares. Para mejorar el contraste se utiliza **CLAHE** (Contrast Limited Adaptive Histogram Equalization).

A diferencia de la ecualización de histograma tradicional, que aplica una transformación global a toda la imagen, CLAHE opera de forma *local*:

1. **División en regiones:** La imagen se divide en pequeñas regiones rectangulares llamadas *tiles* (por ejemplo, una cuadrícula de  $4 \times 4$  o  $8 \times 8$ ).

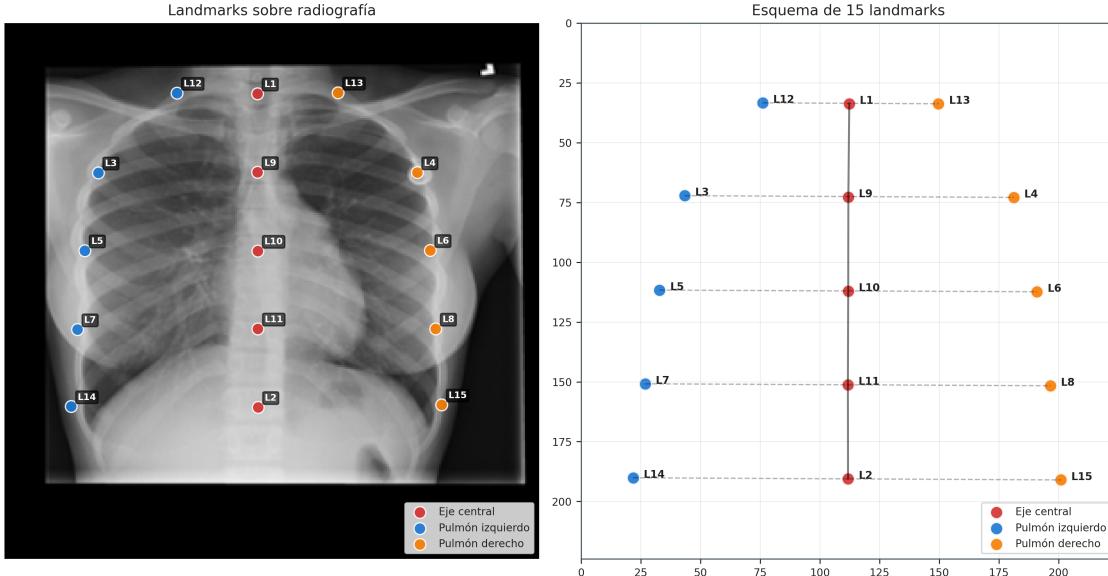


Figura 2.1: Representación de los 15 puntos de referencia anatómicos. Izquierda: puntos de referencia etiquetados sobre una radiografía de tórax. Derecha: esquema de la distribución espacial de los puntos de referencia, donde el eje central (rojo) define la línea media y los puntos laterales delimitan los contornos pulmonares izquierdo (azul) y derecho (naranja).

2. **Ecualización local:** Se calcula y ecualiza el histograma de cada tile de forma independiente, mejorando el contraste en cada región según su contenido.
3. **Límite de contraste:** Para evitar amplificar el ruido en regiones homogéneas, se aplica un límite (*clip limit*) que recorta los picos del histograma y redistribuye esos valores.
4. **Interpolación:** Para evitar bordes artificiales entre tiles, los valores de los píxeles cercanos a los bordes se interpolan suavemente entre las regiones adyacentes.

El resultado es una imagen con contraste mejorado de forma uniforme, donde las estructuras pulmonares (bordes, texturas, opacidades) son más visibles para el modelo de detección de puntos de referencia.

### 2.3. Preprocesamiento: SAHS

Las radiografías de tórax presentan histogramas de intensidad marcadamente asimétricos, con tendencia hacia tonos oscuros debido a las regiones pulmonares llenas de aire. Esta característica hace que técnicas convencionales como CLAHE presenten limitaciones. El método **SAHS** (Statistical Asymmetrical Histogram Stretching) fue desarrollado específicamente para abordar esta asimetría [10].

A diferencia de CLAHE que opera de forma local, SAHS realiza un estiramiento global del histograma pero con límites asimétricos que respetan la distribución natural de la imagen:

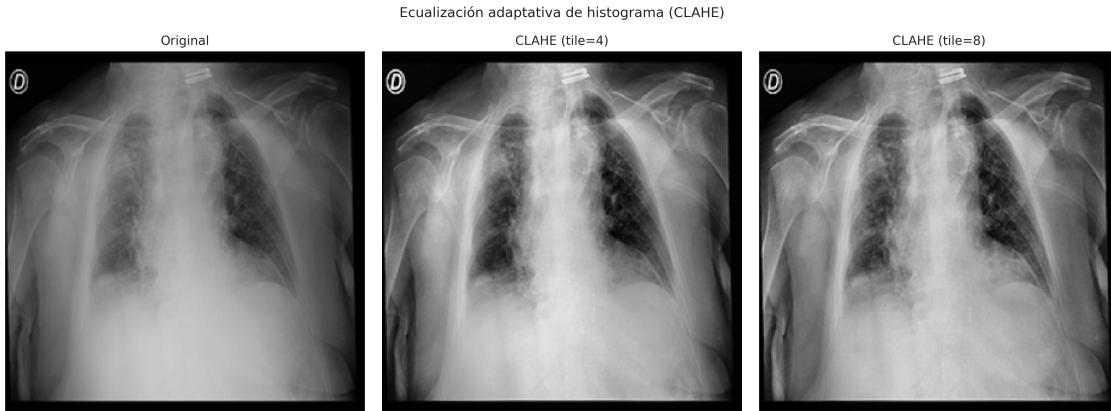


Figura 2.2: Efecto de CLAHE sobre una radiografía de tórax. (a) Imagen original con bajo contraste. (b) Imagen procesada con CLAHE usando tamaño de tile  $4 \times 4$ . (c) Imagen procesada con CLAHE usando tamaño de tile  $8 \times 8$ . El tamaño de tile menor produce una mejora de contraste más localizada.

1. **Cálculo de la media:** Se obtiene la intensidad media  $\mu$  de todos los píxeles de la imagen.
2. **Separación de grupos:** Los píxeles se dividen en dos conjuntos según su relación con la media:
  - Grupo A: píxeles con intensidad mayor que  $\mu$  (tonos claros)
  - Grupo B: píxeles con intensidad menor o igual que  $\mu$  (tonos oscuros)
3. **Desviaciones asimétricas:** Se calcula la desviación estándar de cada grupo por separado ( $\sigma_+$  para el grupo A,  $\sigma_-$  para el grupo B), capturando la dispersión independiente de cada lado del histograma.
4. **Límites de estiramiento:** Se definen límites asimétricos:
  - Límite superior:  $I_{max} = \mu + 2,5 \cdot \sigma_+$
  - Límite inferior:  $I_{min} = \mu - 2,0 \cdot \sigma_-$

Los factores 2.5 y 2.0 fueron optimizados empíricamente para radiografías de tórax.

5. **Mapeo de intensidades:** Se aplica una transformación lineal que mapea el rango  $[I_{min}, I_{max}]$  al rango completo  $[0, 255]$ , recortando valores fuera de los límites.

La ventaja principal de SAHS sobre CLAHE es que preserva la distribución global de la imagen mientras mejora el contraste de forma adaptativa a la asimetría del histograma. Esto es particularmente útil para la clasificación, donde se desea normalizar las diferencias de adquisición entre imágenes manteniendo las características patológicas.

### Comparacion de tecnicas de mejora de contraste: CLAHE vs SAHS

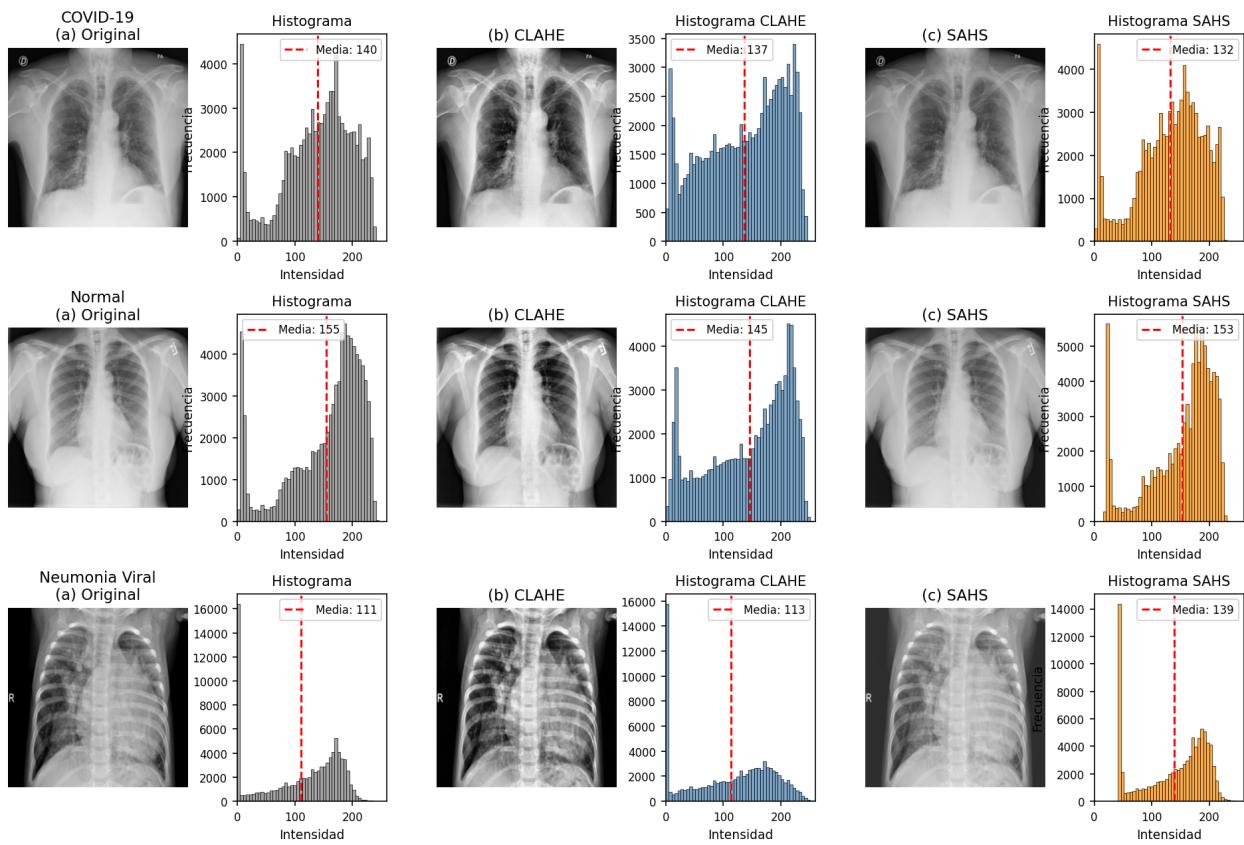


Figura 2.3: Comparación de técnicas de mejora de contraste sobre radiografías de las tres clases. (a) Imagen original. (b) CLAHE: mejora local del contraste mediante ecualización adaptativa. (c) SAHS: estiramiento global con límites asimétricos. Los histogramas muestran cómo cada técnica afecta la distribución de intensidades.

## 2.4. Redes Neuronales Convolucionales

Las **Redes Neuronales Convolucionales** (CNN, por sus siglas en inglés) son un tipo de red neuronal diseñada específicamente para procesar imágenes. Su principal ventaja es la capacidad de aprender automáticamente qué características son relevantes para una tarea, sin necesidad de diseñarlas manualmente.

Una CNN procesa la imagen a través de múltiples capas:

- **Capas convolucionales:** Aplican pequeños filtros que se deslizan sobre la imagen. Cada filtro detecta un patrón específico (bordes verticales, horizontales, texturas, etc.). Las primeras capas detectan patrones simples; las capas más profundas combinan estos patrones para reconocer estructuras más complejas.
- **Capas de pooling:** Reducen el tamaño espacial de la representación, conservando la información más relevante. Esto hace que la red sea más eficiente y robusta a pequeñas variaciones en la posición.
- **Capas fully connected:** Al final de la red, estas capas combinan todas las características extraídas para producir la salida final (ya sea una clasificación o, en nuestro caso, las coordenadas de los puntos de referencia).

La ventaja fundamental de las CNN es que aprenden una **jerarquía de características**: las primeras capas detectan bordes y texturas básicas, las capas intermedias detectan partes de objetos, y las capas finales reconocen estructuras completas. Esta jerarquía es especialmente útil en imágenes médicas, donde las estructuras anatómicas tienen patrones visuales consistentes.

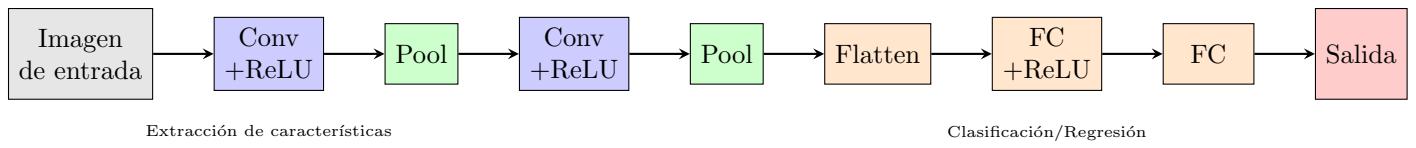


Figura 2.4: Arquitectura básica de una Red Neuronal Convolucionarial (CNN). Las capas convolucionales extraen características de la imagen, las capas de pooling reducen la dimensionalidad, y las capas fully connected (FC) producen la salida final.

## 2.5. ResNet y conexiones residuales

A medida que las redes neuronales se hacen más profundas (más capas), teóricamente deberían aprender representaciones más complejas. Sin embargo, en la práctica, las redes

muy profundas se volvían difíciles de entrenar: los gradientes se desvanecían o explotaban al propagarse por tantas capas, impidiendo el aprendizaje efectivo.

**ResNet** (Residual Network), propuesta por He et al. [11], resuelve este problema mediante **conexiones residuales** (skip connections). En lugar de aprender directamente una transformación  $H(x)$ , cada bloque aprende solo la diferencia (residuo)  $F(x) = H(x) - x$ , y luego suma la entrada original:

$$y = F(x) + x \quad (2.2)$$

donde  $x$  es la entrada al bloque,  $F(x)$  es la transformación aprendida por las capas convolucionales, y  $y$  es la salida.

Esta simple modificación tiene un efecto importante: si una capa no necesita transformar la información, puede aprender  $F(x) = 0$ , permitiendo que la entrada pase sin cambios. Esto facilita el flujo de gradientes durante el entrenamiento y permite construir redes mucho más profundas.

En este trabajo se utiliza **ResNet-18**, que contiene 18 capas con conexiones residuales. Esta arquitectura ofrece un balance entre capacidad de aprendizaje y eficiencia computacional, siendo adecuada para el tamaño del conjunto de datos disponible.

Conexión residual (skip)

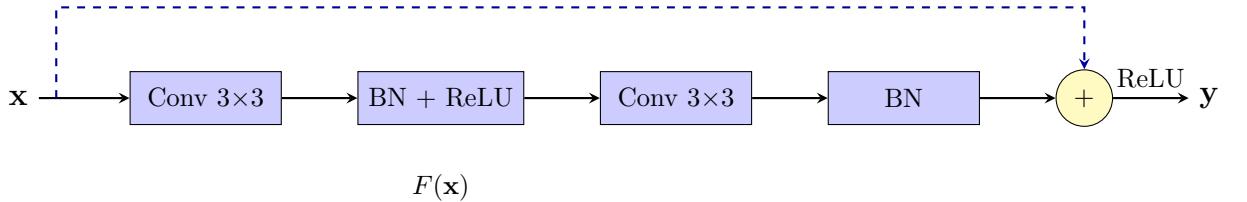


Figura 2.5: Bloque residual de ResNet. La entrada  $\mathbf{x}$  pasa por dos capas convolucionales (rama principal) y simultáneamente se suma directamente a la salida mediante la conexión residual. La salida es  $\mathbf{y} = F(\mathbf{x}) + \mathbf{x}$ , donde  $F(\mathbf{x})$  representa la transformación aprendida.

## 2.6. Normalización por Lote

Durante el entrenamiento de redes neuronales profundas, las distribuciones de las activaciones internas cambian constantemente a medida que los pesos se actualizan. Este fenómeno, conocido como *internal covariate shift*, dificulta el entrenamiento porque cada capa debe adaptarse continuamente a distribuciones cambiantes de sus entradas.

**Normalización por Lote** (BN), propuesto por Ioffe y Szegedy [12], resuelve este problema normalizando las activaciones de cada capa durante el entrenamiento. La idea

central es simple: para cada mini-lote de datos, se normalizan las activaciones para que tengan media cero y varianza unitaria.

El proceso funciona de la siguiente manera:

1. **Cálculo de estadísticas:** Para cada canal de activación, se calcula la media y la varianza sobre todos los ejemplos del mini-lote actual.
2. **Normalización:** Cada activación se transforma restando la media y dividiendo por la desviación estándar. Esto centra los valores alrededor de cero con una dispersión estándar.
3. **Escalado y desplazamiento:** Se aplican dos parámetros aprendibles ( $\gamma$  y  $\beta$ ) que permiten a la red “desnormalizar” si es necesario. Esto asegura que la normalización no limite la capacidad expresiva de la red.

Los beneficios de Normalización por Lote son múltiples:

- **Entrenamiento más rápido:** Permite usar tasas de aprendizaje más altas sin riesgo de divergencia.
- **Regularización implícita:** El ruido introducido por las estadísticas del mini-lote actúa como regularizador, reduciendo la necesidad de dropout.
- **Menor sensibilidad a la inicialización:** La normalización hace que el entrenamiento sea más estable independientemente de cómo se inicialicen los pesos.
- **Gradientes más saludables:** Mantiene las activaciones en un rango donde las funciones de activación (como ReLU) funcionan bien, evitando saturación.

En las arquitecturas modernas como ResNet, Normalización por Lote se aplica después de cada capa convolucional y antes de la función de activación (ver Figura 2.5). Esta combinación **Conv → BN → ReLU** se ha convertido en el estándar de facto para redes convolucionales profundas.

## 2.7. Aprendizaje por transferencia

El **aprendizaje por transferencia** (transfer learning) consiste en reutilizar una red neuronal entrenada en una tarea para resolver otra tarea relacionada. En lugar de inicializar los pesos de la red de forma aleatoria, se utilizan los pesos aprendidos previamente en un conjunto de datos grande.

Esta técnica se justifica por la naturaleza jerárquica de las características que aprenden las CNN:

- Las **primeras capas** aprenden características genéricas y reutilizables: bordes, texturas, gradientes de color.
- Las **capas intermedias** aprenden patrones más complejos: formas, partes de objetos.
- Las **últimas capas** aprenden características específicas de la tarea original.

En este trabajo, se utiliza ResNet-18 preentrenada en **ImageNet**, un conjunto de datos con más de un millón de imágenes naturales. Aunque las imágenes de ImageNet (fotos de objetos, animales, etc.) son muy diferentes a las radiografías de tórax, las características de bajo nivel aprendidas (bordes, texturas) siguen siendo útiles.

El proceso de transferencia consiste en:

1. Cargar la red con pesos preentrenados en ImageNet.
2. Reemplazar la última capa (originalmente diseñada para 1000 clases) por una nueva capa adaptada a la tarea actual.
3. Entrenar la red en el nuevo conjunto de datos, típicamente con tasas de aprendizaje pequeñas para no destruir el conocimiento previo.

El aprendizaje por transferencia es especialmente valioso cuando el conjunto de datos disponible es limitado, como suele ocurrir en aplicaciones médicas.

## 2.8. Función de pérdida: Wing Loss

Para entrenar un modelo de regresión de puntos de referencia, se necesita una función de pérdida que mida el error entre las coordenadas predichas y las reales. Las funciones tradicionales tienen limitaciones:

- **L2 (Error Cuadrático Medio):** Penaliza fuertemente los errores grandes, pero produce gradientes muy pequeños cuando el error ya es pequeño, dificultando el refinamiento de predicciones cercanas al objetivo.
- **L1 (Error Absoluto Medio):** Trata todos los errores con igual importancia, lo que puede ser sensible a valores atípicos (outliers).

**Wing Loss**, propuesta por Feng et al. [13], combina las ventajas de ambas mediante un comportamiento adaptativo:

$$\text{Wing}(x) = \begin{cases} \omega \ln \left( 1 + \frac{|x|}{\epsilon} \right) & \text{si } |x| < \omega \\ |x| - C & \text{si } |x| \geq \omega \end{cases} \quad (2.3)$$

donde  $x$  es el error (diferencia entre predicción y valor real),  $\omega$  define el umbral entre errores pequeños y grandes,  $\epsilon$  controla la curvatura de la parte logarítmica, y  $C = \omega - \omega \ln(1 + \omega/\epsilon)$  asegura continuidad.

El comportamiento es el siguiente:

- Para **errores pequeños** ( $|x| < \omega$ ): La función es logarítmica, produciendo gradientes relativamente grandes que permiten seguir refinando predicciones que ya están cerca del objetivo.
- Para **errores grandes** ( $|x| \geq \omega$ ): La función es lineal (como L1), evitando que errores muy grandes dominen el entrenamiento.

En este trabajo se utilizan los valores  $\omega = 10$  y  $\epsilon = 2$  (en coordenadas normalizadas), que han mostrado buen desempeño en tareas de detección de puntos de referencia faciales y anatómicos.

## 2.9. Análisis Procrustes Generalizado (GPA)

Para normalizar geométricamente las radiografías, se necesita primero calcular una **forma estándar** o de referencia que represente la configuración promedio de los puntos de referencia. El **Análisis Procrustes Generalizado** (GPA, por sus siglas en inglés) es el método estándar para obtener esta forma.

El problema que resuelve GPA es el siguiente: dado un conjunto de formas (cada una definida por sus puntos de referencia), encontrar la forma promedio eliminando las diferencias de **posición, escala y rotación** entre ellas.

El algoritmo opera de forma iterativa:

1. **Centrado:** Cada forma se traslada para que su centroide (punto medio de todos los puntos de referencia) quede en el origen. Esto elimina diferencias de posición.
2. **Normalización de escala:** Cada forma se escala para que tenga un tamaño estándar (típicamente, norma unitaria). Esto elimina diferencias de tamaño entre formas.
3. **Alineación de rotación:** Cada forma se rota para minimizar su distancia a la forma promedio actual. Esta rotación óptima se puede calcular mediante técnicas de álgebra lineal.
4. **Cálculo de la forma promedio:** Se calcula el promedio de todas las formas ya alineadas.
5. **Iteración:** Se repiten los pasos 3 y 4 hasta que la forma promedio converge (deja de cambiar significativamente).

El resultado es una **forma estándar** que representa la configuración típica de los pulmones en el conjunto de entrenamiento. Esta forma sirve como referencia para el proceso de normalización geométrica (deformación) descrito en secciones posteriores.

Análisis Procrustes Generalizado (GPA)

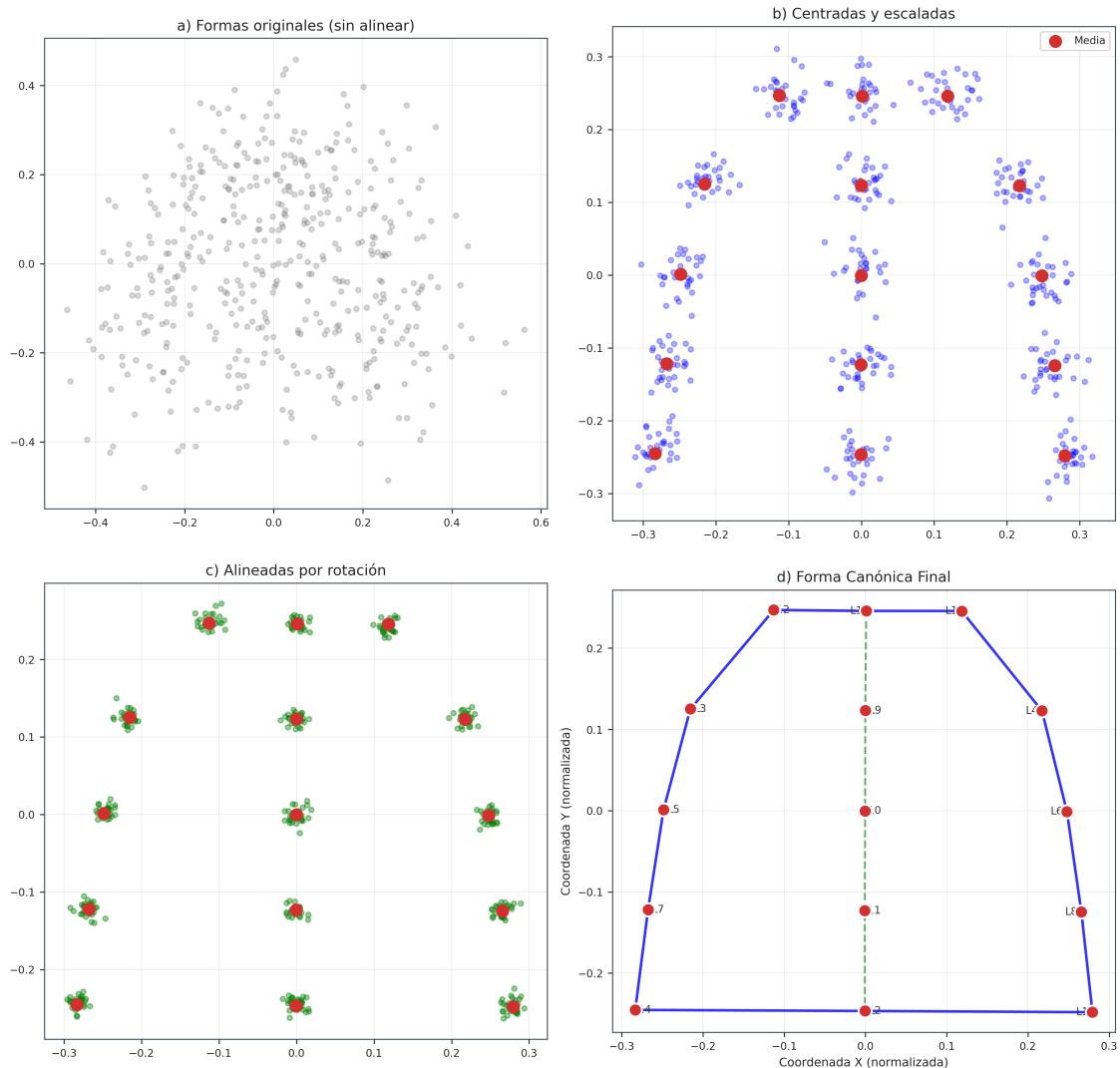


Figura 2.6: Proceso de Análisis Procrustes Generalizado (GPA). (a) Formas originales sin alinear, mostrando la variabilidad entre pacientes. (b) Formas centradas y escaladas a tamaño unitario. (c) Formas alineadas por rotación, minimizando las diferencias. (d) Forma estándar final, calculada como el promedio de todas las formas alineadas.

## 2.10. Triangulación de Delaunay

Para aplicar una transformación geométrica a una imagen basándose en los puntos de referencia, es necesario dividir la región de interés en subregiones más pequeñas. La **triangulación de Delaunay** es el método estándar para particionar un conjunto de puntos en triángulos.

Dado un conjunto de puntos (los puntos de referencia), la triangulación de Delaunay genera un conjunto de triángulos que:

- **No se superponen:** Cada píxel de la imagen pertenece a exactamente un triángulo.
- **Cubren toda la región:** Los triángulos en conjunto cubren el área delimitada por los puntos de referencia.
- **Son regulares:** La triangulación de Delaunay maximiza el ángulo mínimo de todos los triángulos, evitando triángulos muy alargados o degenerados que podrían causar distorsiones en la deformación.

La propiedad característica de Delaunay es que el circuncírculo de cada triángulo (el círculo que pasa por sus tres vértices) no contiene ningún otro punto del conjunto. Esta propiedad garantiza que los triángulos resultantes sean lo más equiláteros posible.

Los triángulos son ideales para transformaciones geométricas porque una **transformación afín** (que puede incluir traslación, rotación, escala y sesgo) queda completamente determinada por la correspondencia entre tres puntos en la imagen original y tres puntos en la imagen destino.

En este trabajo, los 15 puntos de referencia generan 16 triángulos mediante la triangulación de Delaunay, que sirven como base para el proceso de deformación.

## 2.11. Deformación afín por partes

El **deformación afín por partes** (piecewise affine deformation) es la técnica utilizada para normalizar geométricamente las radiografías. El objetivo es deformar cada imagen de manera que sus puntos de referencia coincidan con los puntos de referencia de la forma estándar, eliminando así la variabilidad geométrica entre pacientes.

El proceso funciona de la siguiente manera:

1. **Triangulación:** Tanto los puntos de referencia de la imagen original (predichos por el modelo) como los puntos de referencia de la forma estándar (calculados por GPA) se triangulan usando el mismo esquema de Delaunay.

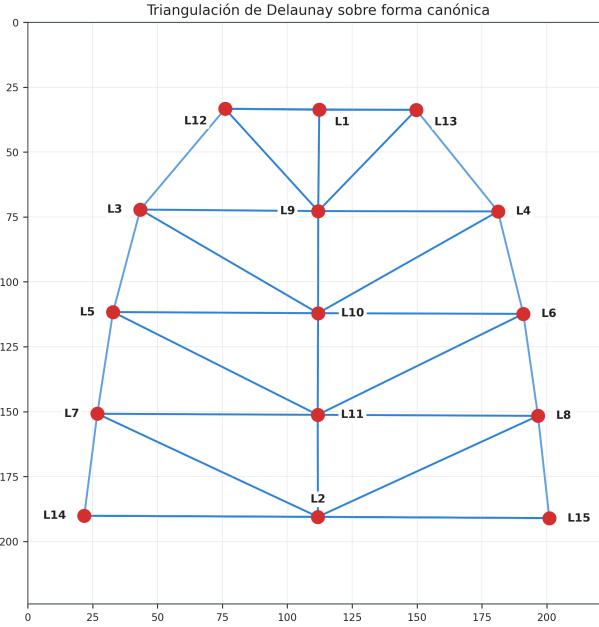


Figura 2.7: Triangulación de Delaunay sobre los 15 puntos de referencia anatómicos. Los triángulos resultantes cubren la región pulmonar sin superposición, maximizando los ángulos mínimos para evitar distorsiones durante la deformación.

2. **Correspondencia de triángulos:** Cada triángulo en la imagen original tiene un triángulo correspondiente en la forma estándar, definido por los mismos índices de puntos de referencia.
3. **Transformación por triángulo:** Para cada par de triángulos correspondientes, se calcula una transformación afín que mapea los vértices del triángulo original a los vértices del triángulo estándar. Una transformación afín puede expresar traslación, rotación, escala y sesgo.
4. **Aplicación:** Cada píxel de la imagen de salida se obtiene determinando a qué triángulo estándar pertenece, aplicando la transformación inversa para encontrar su posición en la imagen original, y copiando el valor de intensidad (usando interpolación si es necesario).

El resultado es una imagen donde la región pulmonar tiene exactamente la forma estándar, independientemente de la forma original del paciente. Esto permite que el clasificador se enfoque en las características de textura y patología, sin verse afectado por diferencias anatómicas normales entre pacientes.

## 2.12. Clasificación de imágenes

La etapa final del sistema es la **clasificación** de las radiografías normalizadas en tres categorías: COVID-19, Normal y Neumonía Viral. Esta tarea se realiza mediante una CNN

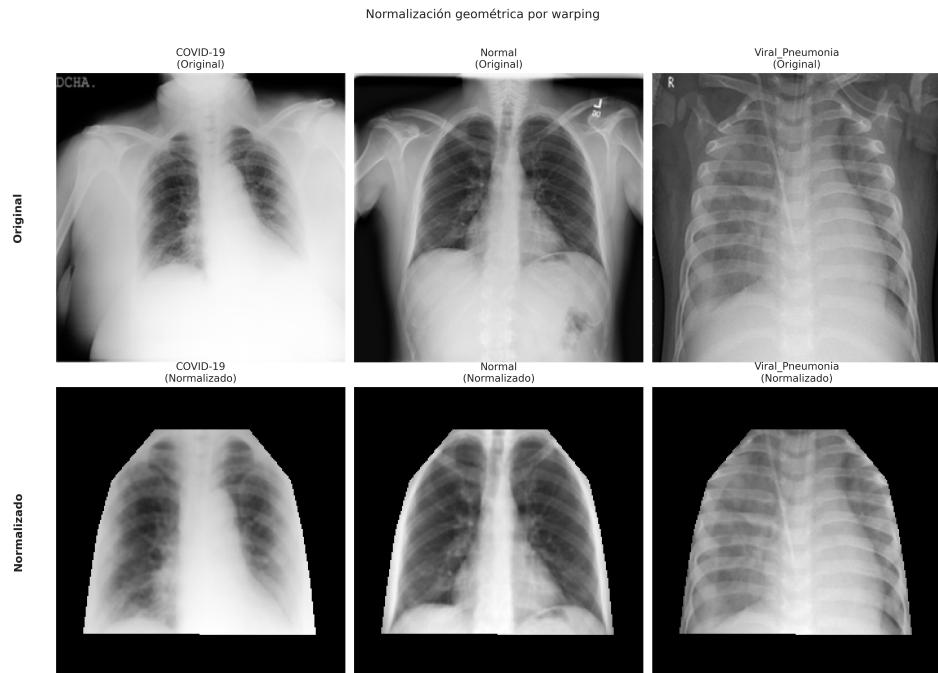


Figura 2.8: Normalización geométrica mediante deformación afín por partes. Fila superior: radiografías originales de las tres clases (COVID-19, Normal, Neumonía Viral). Fila inferior: las mismas imágenes después de la deformación, donde la región pulmonar ha sido transformada para coincidir con la forma estándar, eliminando la variabilidad anatómica entre pacientes.

similar a la usada para detección de puntos de referencia, pero con una capa de salida diferente.

### 2.12.1. Función softmax

La capa final del clasificador produce un valor numérico por cada clase posible. Para convertir estos valores en **probabilidades**, se aplica la función *softmax*, que transforma cualquier conjunto de valores en números entre 0 y 1 que suman exactamente 1. El valor más alto indica la clase predicha.

### 2.12.2. Función de pérdida: entropía cruzada

Para entrenar el clasificador, se utiliza la **entropía cruzada** (cross-entropy) como función de pérdida. Esta función mide qué tan diferente es la distribución de probabilidades predicha de la distribución real (donde la clase correcta tiene probabilidad 1 y las demás 0).

Intuitivamente, la entropía cruzada penaliza las predicciones que asignan baja probabilidad a la clase correcta. Si el modelo está muy seguro de la respuesta incorrecta, la pérdida es muy alta.

### 2.12.3. Ponderación de clases

Cuando el conjunto de datos tiene **desbalance de clases** (más imágenes de una categoría que de otras), el modelo tiende a favorecer la clase mayoritaria. Para compensar esto, se utiliza *entropía cruzada ponderada*, donde cada clase tiene un peso inversamente proporcional a su frecuencia en el conjunto de entrenamiento. Esto obliga al modelo a prestar igual atención a todas las clases, independientemente de su cantidad de ejemplos.

## 2.13. Métricas de evaluación

Para evaluar el desempeño del sistema se utilizan métricas específicas para cada tarea.

### 2.13.1. Métricas para detección de puntos de referencia

El desempeño del modelo de puntos de referencia se mide mediante el **error en píxeles**, definido como la distancia euclíadiana entre las coordenadas predichas y las coordenadas reales:

$$\text{Error} = \sqrt{(x_{\text{pred}} - x_{\text{real}})^2 + (y_{\text{pred}} - y_{\text{real}})^2} \quad (2.4)$$

Se reporta el error promedio sobre todos los puntos de referencia y todas las imágenes del conjunto de prueba.

### 2.13.2. Métricas para clasificación

Para la tarea de clasificación multiclase se utilizan las siguientes métricas:

- **Exactitud (Accuracy):** Proporción de predicciones correctas sobre el total.

$$\text{Accuracy} = \frac{\text{Predicciones correctas}}{\text{Total de predicciones}} \quad (2.5)$$

- **Precisión:** De todas las predicciones de una clase, ¿qué proporción fue correcta?

$$\text{Precisión} = \frac{VP}{VP + FP} \quad (2.6)$$

- **Sensibilidad (Recall):** De todos los casos reales de una clase, ¿qué proporción fue detectada?

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (2.7)$$

- **F1-Score:** Media armónica de precisión y sensibilidad, útil cuando se busca un balance entre ambas.

$$F1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} \quad (2.8)$$

donde VP = Verdaderos Positivos, FP = Falsos Positivos, y FN = Falsos Negativos.

En problemas multiclase, estas métricas se calculan por clase y luego se promedian. El **promedio macro** (macro-average) trata todas las clases con igual importancia, mientras que el **promedio ponderado** (weighted-average) considera el número de muestras de cada clase.

## 2.14. Mecanismo de Coordinate Attention

Los **mecanismos de atención** permiten que una red neuronal se enfoque en las regiones más relevantes de una imagen para la tarea en cuestión. En el contexto de detección de puntos de referencia, es importante que el mecanismo preserve información sobre *dónde* están las características importantes, no solo *qué* características son importantes.

**Coordinate Attention**, propuesto por Hou et al. [14], es un mecanismo diseñado específicamente para tareas que requieren información de localización precisa. A diferencia de otros mecanismos como Squeeze-and-Excitation (SE-Net), que colapsan la información espacial completamente, Coordinate Attention preserva las coordenadas espaciales.

El mecanismo opera en tres etapas:

1. **Agregación direccional:** En lugar de comprimir toda la imagen en un solo valor por canal (como hace SE-Net), Coordinate Attention agrega información por separado en

la dirección horizontal y vertical. Esto genera dos representaciones: una que captura dependencias a lo largo del eje X, y otra a lo largo del eje Y.

2. **Codificación conjunta:** Las dos representaciones direccionales se combinan y procesan para generar una representación intermedia que captura relaciones espaciales de largo alcance.
3. **Generación de mapas de atención:** Se generan dos mapas de atención, uno para cada dirección, que se combinan para producir un mapa de atención 2D que indica qué regiones de la imagen son más relevantes.

El resultado es un mecanismo que permite a la red enfocarse en las regiones anatómicas relevantes (como los bordes pulmonares) mientras mantiene información precisa sobre su ubicación espacial, lo cual es fundamental para la predicción de coordenadas de puntos de referencia.

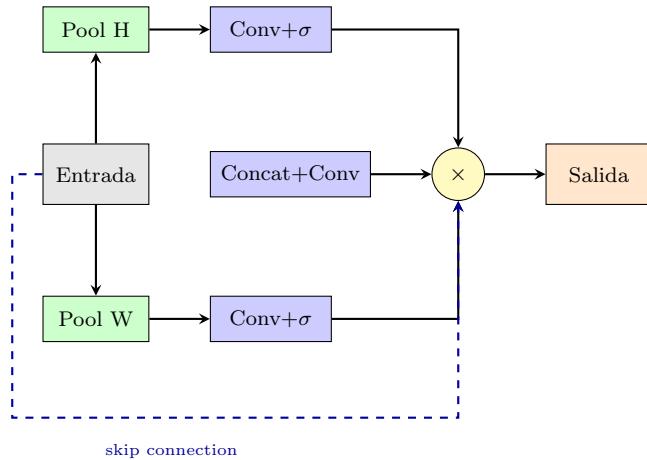


Figura 2.9: Mecanismo de Coordinate Attention. La entrada se procesa mediante poolings direccionales (H y W) que preservan información espacial. Las representaciones se concatenan, procesan con convoluciones y activación sigmoide ( $\sigma$ ), y se multiplican con la entrada original (skip connection).

## 2.15. Ensamble de modelos y Test-Time Augmentation

Para mejorar la robustez y precisión de las predicciones, se utilizan dos técnicas complementarias: **ensamble de modelos** y **Aumento en tiempo de prueba** (TTA).

### 2.15.1. Ensamble de modelos

Un **ensamble** consiste en combinar las predicciones de múltiples modelos entrenados de forma independiente. Aunque los modelos comparten la misma arquitectura, cada uno

aprende representaciones ligeramente diferentes debido a:

- Diferente inicialización de pesos (semilla aleatoria distinta)
- Diferente orden de presentación de los datos durante el entrenamiento
- Diferentes máscaras de dropout durante el entrenamiento

Al promediar las predicciones de varios modelos, los errores individuales tienden a cancelarse, resultando en una predicción más estable y precisa. La reducción de varianza es proporcional al número de modelos en el ensamble.

En este trabajo, se utiliza un ensamble de 4 modelos ResNet-18, cada uno entrenado con una semilla aleatoria diferente.

### 2.15.2. Aumento en tiempo de prueba (TTA)

**Aumento en tiempo de prueba** extiende la idea del ensamble a las transformaciones de la imagen. En lugar de predecir solo sobre la imagen original, se aplican transformaciones (como el reflejo horizontal) y se predicen sobre las versiones transformadas. Luego se promedian todas las predicciones.

Para la tarea de detección de puntos de referencia con reflejo horizontal, es necesario un paso adicional: antes de promediar, se deben intercambiar las coordenadas de los **puntos de referencia simétricos** (izquierda-derecha) y reflejar las coordenadas X. Por ejemplo, si L3 está en el pulmón izquierdo y L4 en el derecho, al reflejar la imagen estos puntos de referencia intercambian posiciones.

La combinación de ensamble y TTA proporciona predicciones más robustas, especialmente en casos donde la imagen original tiene características ambiguas o ruido.

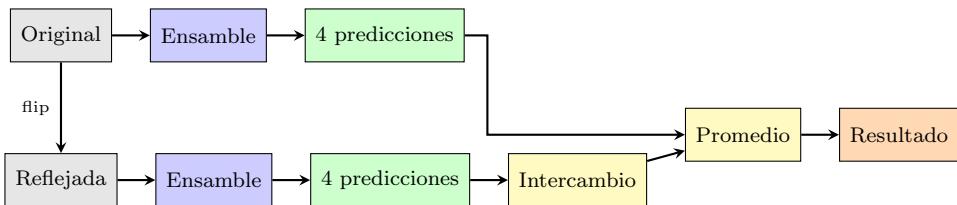


Figura 2.10: Inferencia con ensamble y TTA. La imagen original y su reflejo se procesan por un ensamble de 4 modelos. Las predicciones del reflejo requieren intercambiar puntos de referencia simétricos (Swap) antes de promediar las 8 predicciones.

# Capítulo 3

## Metodología

Este capítulo presenta la metodología desarrollada para la normalización y alineación automática de radiografías de tórax, así como la clasificación de enfermedades pulmonares. Se describe el flujo de procesamiento completo del sistema, desde la adquisición de datos hasta la clasificación final, detallando cada componente.

### 3.1. Descripción General del Sistema

El desarrollo del sistema propuesto comprende dos fases: una fase de preparación, que incluye la anotación manual de puntos de referencia anatómicos y el entrenamiento de los modelos, y una fase de operación, donde el sistema procesa nuevas radiografías de tórax. Durante la operación, las imágenes pasan por una secuencia de cuatro módulos: preprocesamiento, predicción de puntos de referencia, normalización geométrica y clasificación. Los tres primeros módulos transforman la imagen de entrada a una representación geométricamente normalizada, mientras que el cuarto realiza la clasificación de diagnóstico. Este diseño modular permite evaluar la contribución de cada componente al rendimiento final del sistema.

La Figura 3.1 ilustra la relación entre ambas fases del sistema.

#### 3.1.1. Arquitectura del Sistema

La Figura 3.2 presenta el diagrama de bloques del flujo de operación. El sistema se compone de cuatro módulos que procesan secuencialmente las imágenes de entrada.

**Módulo 1: Preprocesamiento.** Las imágenes de entrada se someten a un proceso de mejora de contraste mediante el algoritmo CLAHE (*Contrast Limited Adaptive Histogram Equalization*) [15]. Este paso normaliza las variaciones de contraste inherentes a diferentes equipos de adquisición radiográfica. Posteriormente, las imágenes se redimensionan a  $224 \times 224$  píxeles para su procesamiento por la red neuronal.

**Módulo 2: Predicción de Landmarks.** Un modelo basado en ResNet-18 [11] con módulo de Coordinate Attention [14] predice las coordenadas de 15 puntos anatómicos que definen el contorno de la región pulmonar. Estos puntos de referencia fueron definidos manualmente durante la fase de anotación del conjunto de datos y representan puntos característicos de la silueta pulmonar bilateral.

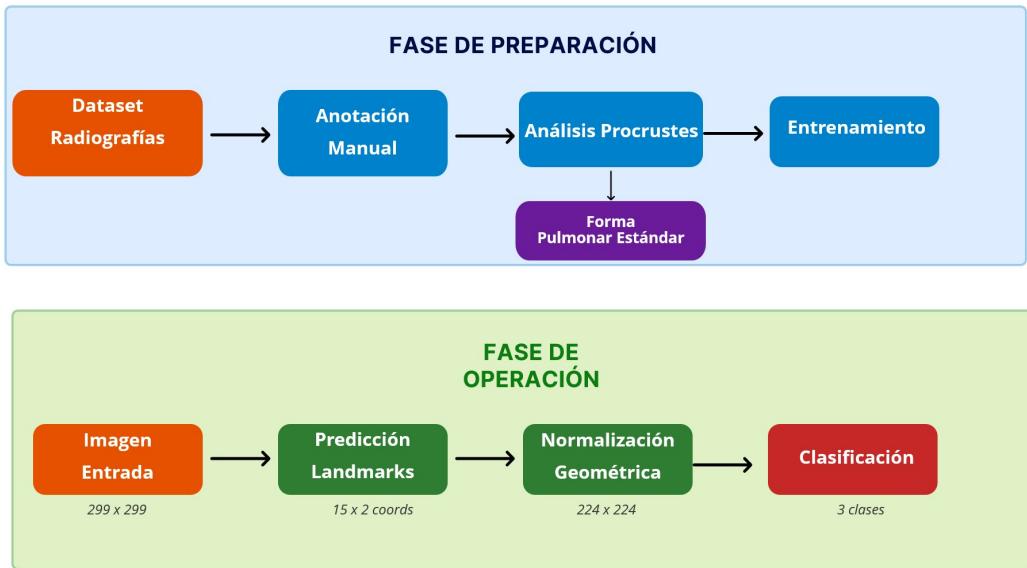


Figura 3.1: Estructura general del sistema en dos fases.



Figura 3.2: Flujo de operación del sistema. Las radiografías de tórax se procesan mediante cuatro módulos: preprocessamiento con CLAHE, predicción de 15 puntos de referencia anatómicos, normalización geométrica, y clasificación en tres categorías (COVID-19, Normal, Neumonía Viral).

**Módulo 3: Normalización Geométrica.** Utilizando los puntos de referencia predichos, se aplica una transformación afín por partes (*piecewise affine deformation*) que alinea cada imagen a una forma estándar previamente calculada mediante Análisis de Procrustes Generalizado (GPA) [16]. Este proceso elimina variaciones geométricas entre pacientes, normalizando la posición, escala y orientación de la región pulmonar.

**Módulo 4: Clasificación.** Las imágenes normalizadas se procesan mediante una red neuronal convolucional para clasificarlas en una de tres categorías: COVID-19, Normal o Neumonía Viral. El módulo genera la predicción de clase junto con las probabilidades asociadas a cada categoría.

### 3.1.2. Flujo de Datos

El procesamiento de una imagen sigue el flujo ilustrado en la Tabla 3.1. Cada etapa transforma los datos de entrada en una representación apropiada para la siguiente etapa del proceso.

Cuadro 3.1: Flujo de datos a través del sistema.

Etapa	Entrada	Salida	Dimensiones
Preprocesamiento	Imagen	Imagen normalizada	$224 \times 224 \times 3$
Predicción	Imagen normalizada	Coordenadas puntos de ref.	$15 \times 2$
Normalización geométrica	Imagen + puntos de ref.	Imagen normalizada	$224 \times 224 \times 3$
Clasificación	Imagen normalizada	Vector probabilidades	3

### 3.1.3. Justificación del Diseño Modular

El diseño modular del sistema ofrece varias ventajas:

1. **Interpretabilidad:** Los puntos de referencia predichos constituyen una representación intermedia que permite verificar visualmente la calidad del proceso de detección de la región pulmonar.
2. **Modularidad:** Cada componente puede entrenarse, evaluarse y mejorarse de forma independiente, facilitando el desarrollo iterativo del sistema.
3. **Selección implícita de características:** La normalización geométrica actúa como un mecanismo de selección de características a nivel de imagen, eliminando información no discriminante (artefactos, marcas hospitalarias, variaciones de pose) y preservando únicamente la región pulmonar relevante para la clasificación [17].

4. **Transferibilidad:** El modelo de puntos de referencia puede reutilizarse para otras tareas de análisis pulmonar, mientras que el clasificador puede adaptarse a diferentes conjuntos de clases según los requerimientos de la aplicación.

El enfoque propuesto se fundamenta en la hipótesis de que la normalización y alineación mejora la capacidad de generalización del clasificador al reducir la variabilidad no relacionada con la patología. Esta hipótesis se evalúa experimentalmente en el Capítulo 4.

## 3.2. Conjunto de Datos y Preprocesamiento

Esta sección describe el conjunto de datos utilizado para el desarrollo y evaluación del sistema propuesto, así como los procesos de anotación y preprocesamiento aplicados a las imágenes.

### 3.2.1. COVID-19 Radiography Database

El presente trabajo utiliza el *COVID-19 Radiography Database*, un conjunto de datos públicamente disponible desarrollado por investigadores de Qatar University, University of Dhaka y colaboradores de Malasia y Pakistán [4, 5]. Este conjunto de datos ha sido ampliamente utilizado en la literatura para el desarrollo de sistemas de detección de COVID-19 basados en radiografías de tórax.

El conjunto de datos contiene imágenes de radiografías posteroanterior (PA) de tórax organizadas en tres categorías diagnósticas:

- **COVID-19:** Radiografías de pacientes con diagnóstico confirmado de COVID-19 mediante prueba RT-PCR.
- **Normal:** Radiografías de pacientes sin patología pulmonar aparente.
- **Neumonía Viral:** Radiografías de pacientes con neumonía viral de etiología distinta a SARS-CoV-2.

La Tabla 3.2 presenta la distribución de imágenes por categoría utilizada en este trabajo.

Cuadro 3.2: Distribución del conjunto de datos por categoría diagnóstica.

Categoría	Imágenes	Porcentaje
COVID-19	3,616	23.9 %
Normal	10,192	67.2 %
Neumonía Viral	1,345	8.9 %
<b>Total</b>	<b>15,153</b>	<b>100 %</b>

Las imágenes originales tienen un tamaño de  $299 \times 299$  píxeles en formato PNG. El conjunto de datos presenta un desbalance de clases natural, con predominancia de imágenes normales, lo cual refleja la distribución típica en escenarios clínicos reales.

### 3.2.2. Anotación de Puntos de Referencia Anatómicos

Para el entrenamiento del modelo de predicción de puntos de referencia, se realizó la anotación manual de 15 puntos característicos en un subconjunto del conjunto de datos. Estos puntos de referencia definen el contorno de la región pulmonar.

Los puntos de referencia no corresponden a estructuras anatómicas específicas, sino que representan puntos de control sobre la silueta pulmonar diseñados para capturar la forma global del contorno. La Figura 3.3 ilustra la ubicación de cada landmark.

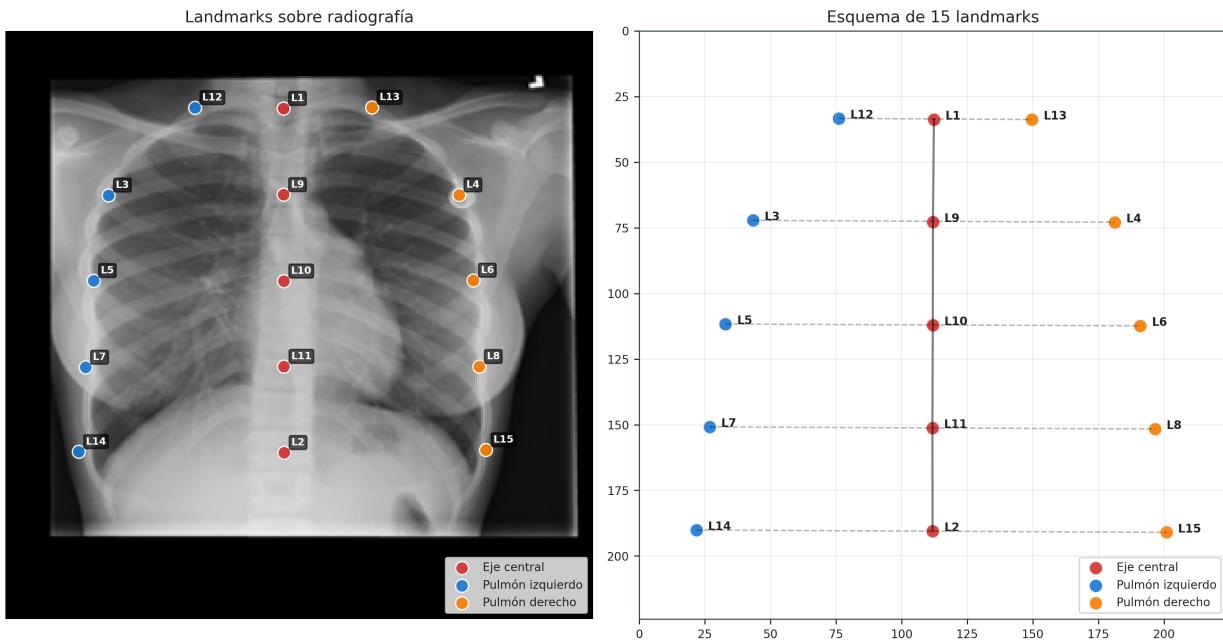


Figura 3.3: Ubicación de los 15 puntos de referencia que definen el contorno pulmonar. Los puntos L1 y L2 definen el eje vertical central. Los landmarks L3-L8 delimitan los contornos laterales, mientras que L9-L11 dividen el eje central en cuatro segmentos iguales.

La estructura geométrica de los puntos de referencia se organiza de la siguiente manera:

1. **Eje central vertical:** Los puntos de referencia L1 (superior) y L2 (inferior) definen la línea media de la silueta pulmonar. Los puntos L9, L10 y L11 dividen este eje en cuatro segmentos de igual longitud.
2. **Contorno pulmonar izquierdo:** Los puntos de referencia L12, L3, L5, L7 y L14 trazan el borde lateral izquierdo de la silueta, desde la región superior hasta la inferior.
3. **Contorno pulmonar derecho:** De manera simétrica, los puntos de referencia L13, L4, L6, L8 y L15 definen el borde lateral derecho.
4. **Pares simétricos:** Existen cinco pares de puntos de referencia bilateralmente simétricos: (L3, L4), (L5, L6), (L7, L8), (L12, L13) y (L14, L15).

## Proceso de Anotación

Para realizar la anotación de puntos de referencia se desarrolló una herramienta gráfica interactiva basada en OpenCV que facilita el proceso mediante un algoritmo semi-automático.

La anotación se realizó sobre un subconjunto de 957 imágenes seleccionadas del conjunto de datos, asegurando representatividad de las tres categorías diagnósticas.

**Herramienta de Anotación** La herramienta desarrollada implementa un proceso de anotación en dos fases que reduce significativamente el tiempo requerido respecto a la marcación individual de cada punto. La Figura 3.4 ilustra la interfaz de la herramienta.

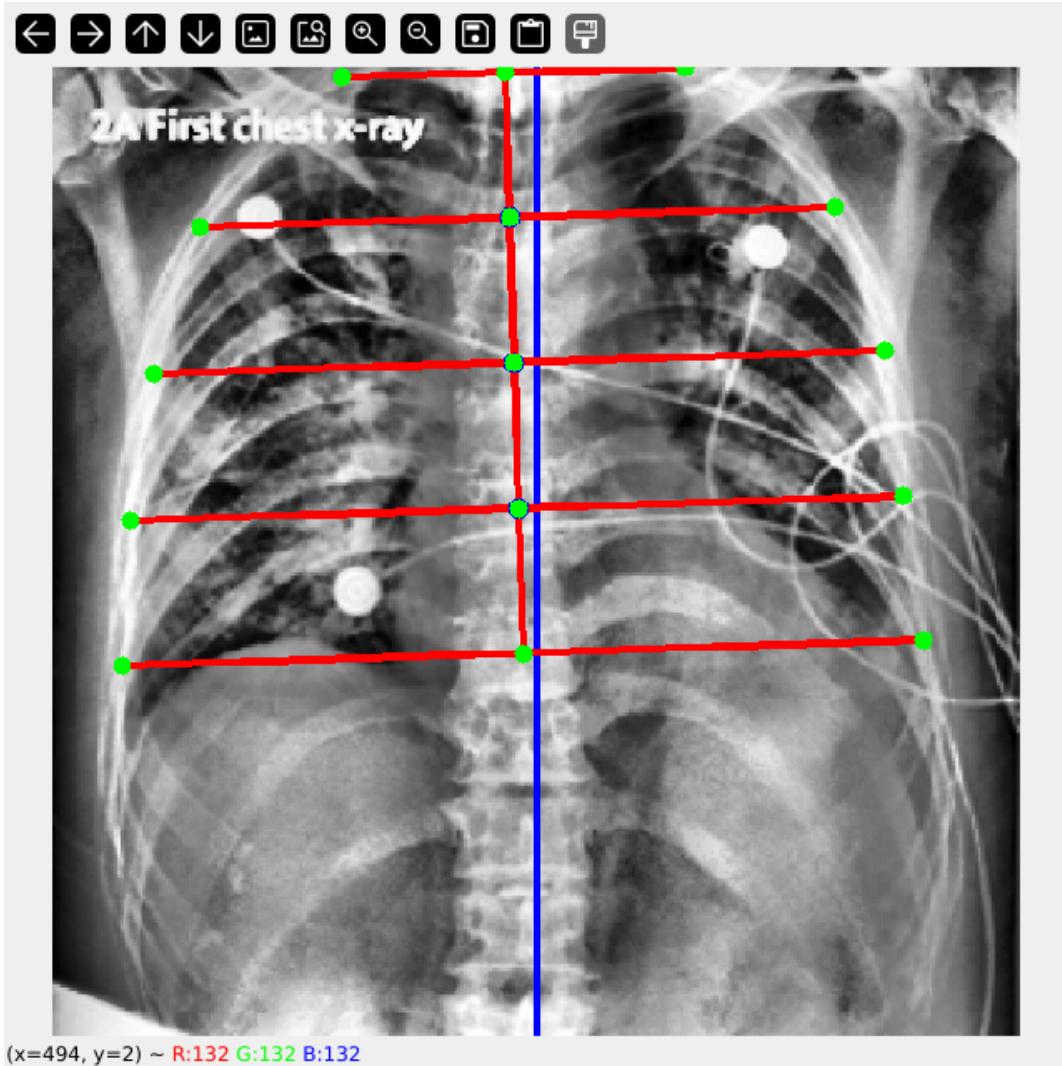


Figura 3.4: Interfaz de la herramienta de anotación de puntos de referencia. La ventana principal muestra la radiografía con una línea vertical central de referencia. Los landmarks se visualizan como puntos verdes conectados por líneas rojas que definen el contorno pulmonar.

**Fase 1: Generación Automática** El proceso inicia con tres interacciones del operador que definen la geometría base:

1. **Primer click (L1):** Define el punto superior de la silueta pulmonar.

2. **Segundo click (L2):** Define la base inferior, estableciendo el eje central del contorno.
3. **Tercer click:** Confirma la selección y activa el algoritmo de generación automática.

El algoritmo de generación automática calcula los 13 puntos de referencia restantes (L3-L15) mediante el siguiente procedimiento:

1. Calcula la línea central entre L1 y L2, determinando su pendiente.
2. Divide el eje central en cuatro segmentos iguales, ubicando los puntos intermedios L9, L10 y L11 en las posiciones respectivas.
3. Genera líneas perpendiculares al eje central en cada punto de división.
4. Ubica los puntos de referencia laterales (L3-L8, L12-L15) sobre estas perpendiculares a distancias predefinidas del eje central.

**Fase 2: Ajuste Manual** Los puntos de referencia generados automáticamente proporcionan una aproximación inicial que raramente coincide exactamente con el contorno pulmonar visible. La herramienta permite ajustar cada punto de referencia horizontalmente mediante atajos de teclado, manteniendo la coherencia geométrica al desplazar los puntos a lo largo de sus respectivas líneas perpendiculares.

El ajuste se realiza hasta que cada punto de referencia coincida visualmente con el borde de la silueta pulmonar en la imagen.

**Criterios de Anotación** El proceso de anotación siguió las siguientes directrices para garantizar consistencia:

1. Se colocó cada punto de referencia sobre el borde perceptible de la silueta pulmonar, no sobre estructuras anatómicas internas.
2. En casos de ambigüedad por baja calidad de imagen o superposición de estructuras, se priorizó la consistencia visual sobre la precisión anatómica.
3. Se verificó visualmente que los pares simétricos (L3-L4, L5-L6, etc.) mantuvieran una distribución razonable respecto al eje central.
4. Las coordenadas se registraron en píxeles respecto a la imagen original de  $299 \times 299$  píxeles.

Las anotaciones se almacenaron en formato CSV.

La distribución del subconjunto anotado por categoría se presenta en la Tabla 3.3.

Cuadro 3.3: Distribución del subconjunto anotado con puntos de referencia.

Categoría	Imágenes anotadas	Porcentaje
COVID-19	306	32.0 %
Normal	468	48.9 %
Neumonía Viral	183	19.1 %
<b>Total</b>	<b>957</b>	<b>100 %</b>

### 3.2.3. Preprocesamiento de Imágenes

Las imágenes radiográficas requieren preprocesamiento para mitigar las variaciones introducidas por distintos equipos de adquisición y diversas condiciones de exposición. El proceso implementado consta de tres etapas: mejora de contraste, redimensionamiento y normalización.

#### Mejora de Contraste mediante CLAHE

Se aplica el algoritmo *Contrast Limited Adaptive Histogram Equalization* (CLAHE) [18] para mejorar el contraste local de las imágenes. A diferencia de la ecualización de histograma global, CLAHE opera sobre regiones locales (tiles) y limita la amplificación de contraste para evitar el realce excesivo de ruido.

Los parámetros utilizados fueron determinados experimentalmente:

- **Clip limit:** 2,0. Controla el límite máximo de amplificación de contraste. Valores mayores producen mayor contraste pero pueden amplificar ruido.
- **Tile size:**  $4 \times 4$ . Tamaño de las regiones para ecualización local. Un valor menor produce una adaptación más fina pero aumenta el tiempo de cómputo.

La Figura 3.5 muestra el efecto del preprocesamiento CLAHE sobre una radiografía de ejemplo.

#### Redimensionamiento

Las imágenes se redimensionan de su tamaño original ( $299 \times 299$  píxeles) a  $224 \times 224$  píxeles mediante interpolación bilineal. Este tamaño corresponde a la entrada estándar de las arquitecturas de redes neuronales preentrenadas en ImageNet [19].

#### Normalización

Las coordenadas de los puntos de referencia se normalizan al rango  $[0, 1]$  dividiendo entre el tamaño de la imagen (224 píxeles), facilitando el entrenamiento del modelo de regresión.

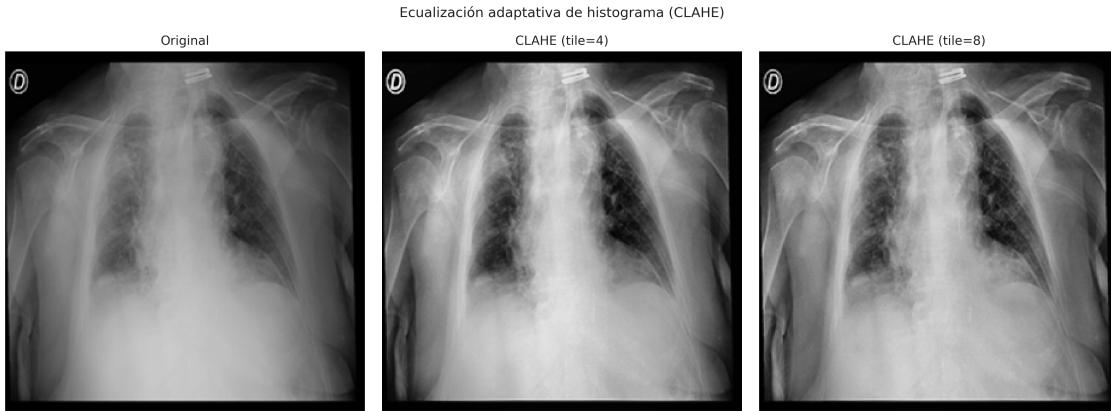


Figura 3.5: Efecto del preprocesamiento CLAHE. (a) Imagen original con bajo contraste en la región pulmonar. (b) Imagen después de aplicar CLAHE con clip limit = 2,0 y tile size = 4, mostrando mejor definición de estructuras pulmonares, (c) Imagen después de aplicar CLAHE con clip limit = 2,0 y tile size = 8.

### 3.2.4. División del Conjunto de Datos

El conjunto de datos se divide en tres subconjuntos, entrenamiento, validación y prueba. La división se realiza de manera estratificada por categoría para mantener las proporciones de clases en cada subconjunto.

- **Entrenamiento (75 %):** Utilizado para optimizar los parámetros del modelo.
- **Validación (15 %):** Utilizado para selección de hiperparámetros.
- **Prueba (10 %):** Reservado exclusivamente para la evaluación final del modelo.

La Tabla 3.4 presenta la distribución resultante para el subconjunto anotado con puntos de referencia (957 imágenes).

Cuadro 3.4: División del conjunto de datos en subconjuntos de entrenamiento, validación y prueba.

Conjunto	COVID-19	Normal	Viral	Total
Entrenamiento (75 %)	229	351	137	717
Validación (15 %)	46	70	28	144
Prueba (10 %)	31	47	18	96
<b>Total</b>	<b>306</b>	<b>468</b>	<b>183</b>	<b>957</b>

Para garantizar la reproducibilidad de los experimentos, se utiliza una semilla aleatoria fija ( $seed = 42$ ) en todas las operaciones que involucran aleatorización.

### 3.3. Modelo de Predicción de Puntos de Referencia

El modelo de predicción de puntos de referencia constituye el primer componente del sistema propuesto y tiene como objetivo localizar los 15 puntos de referencia anatómicos que definen el contorno pulmonar en cada radiografía. Esta sección describe la arquitectura del modelo, la función de pérdida utilizada y la estrategia de entrenamiento implementada.

#### 3.3.1. Arquitectura del Modelo

El modelo propuesto se basa en una arquitectura de red neuronal convolucional con tres componentes principales: un módulo de extracción de características, un módulo de atención y un módulo de regresión. La Figura 3.6 presenta el diagrama de la arquitectura completa.

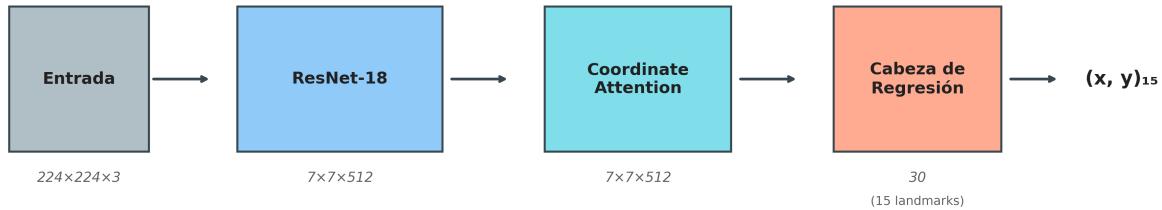


Figura 3.6: Arquitectura del modelo de predicción de puntos de referencia. ResNet-18 extrae características de alto nivel, el módulo Coordinate Attention incorpora información posicional, y la cabeza de regresión predice las 30 coordenadas (15 puntos de referencia  $\times$  2 coordenadas).

#### ResNet-18

Como extractor de características (denominado *extractor de características* en la literatura de aprendizaje profundo) se utiliza ResNet-18 [11], una red residual de 18 capas preentrenada en el conjunto de datos ImageNet [19]. La arquitectura ResNet introdujo las conexiones residuales (*skip connections*) que permiten entrenar redes más profundas al mitigar el problema de desvanecimiento de gradiente.

La elección de ResNet-18 sobre arquitecturas más profundas (ResNet-34, ResNet-50) se fundamenta en las siguientes consideraciones:

1. **Tamaño del conjunto de datos:** Con 957 imágenes anotadas, un modelo más pequeño reduce el riesgo de sobreajuste.

2. **Eficiencia computacional:** ResNet-18 permite iteraciones de entrenamiento más rápidas durante la experimentación.
3. **Suficiente capacidad:** La tarea de localización de 15 puntos de referencia no requiere la capacidad de representación de arquitecturas más profundas.
4. **Aprendizaje por transferencia efectivo:** Los pesos preentrenados en ImageNet proporcionan características genéricas útiles para imágenes médicas.

El extractor de características procesa imágenes de entrada de dimensiones  $224 \times 224 \times 3$  y produce un mapa de características de dimensiones  $7 \times 7 \times 512$ . La Tabla 3.5 detalla las capas del extractor de características utilizadas.

Cuadro 3.5: Configuración del extractor de características ResNet-18. Se utilizan todas las capas convolucionales, removiendo únicamente la capa fully connected original.

Capa	Descripción	Salida	Parámetros
conv1	Conv $7 \times 7$ , stride 2	$112 \times 112 \times 64$	9,408
bn1 + relu	BatchNorm + ReLU	$112 \times 112 \times 64$	128
maxpool	MaxPool $3 \times 3$ , stride 2	$56 \times 56 \times 64$	0
layer1	2 bloques residuales	$56 \times 56 \times 64$	147,968
layer2	2 bloques residuales	$28 \times 28 \times 128$	525,568
layer3	2 bloques residuales	$14 \times 14 \times 256$	2,099,712
layer4	2 bloques residuales	$7 \times 7 \times 512$	8,393,728
<b>Total extractor de características</b>		—	<b>11,176,512</b>

## Módulo Coordinate Attention

**Motivación y propósito.** Cuando una red neuronal procesa una imagen, extrae información sobre texturas, formas y patrones, pero puede perder la noción de *dónde* se encuentran estos elementos dentro de la imagen. Para una tarea de localización de puntos de referencia, saber *dónde* mirar es tan importante como saber *qué* buscar.

El módulo Coordinate Attention [14] aborda este problema permitiendo que el modelo “aprenda a enfocar su atención” en las regiones más relevantes de la imagen, manteniendo información sobre la posición horizontal y vertical de cada elemento. En términos simples, funciona como un mecanismo que le indica al modelo: “presta más atención a esta fila de la imagen” y “presta más atención a esta columna”.

**Diferencia con otros mecanismos de atención.** Existen otros mecanismos de atención, como SE-Net [20], que resumen toda la información espacial de la imagen en un único valor promedio. Si bien esto permite identificar *qué* características son importantes, pierde

la información de *dónde* se encuentran. Para la detección de puntos de referencia, donde la posición exacta es el objetivo principal, esta pérdida de información espacial resulta problemática.

Coordinate Attention resuelve esta limitación procesando la información de posición de manera separada: genera un “mapa de importancia” para las filas de la imagen y otro para las columnas. De esta forma, el modelo puede aprender que ciertas regiones horizontales (por ejemplo, donde típicamente se encuentra el contorno superior pulmonar) y ciertas regiones verticales (donde se ubica el contorno lateral) merecen mayor atención.

**Funcionamiento del módulo.** El proceso se realiza en cuatro pasos:

**Paso 1: Resumen por filas y columnas.** Primero, el módulo calcula un resumen de la información para cada fila y cada columna de la imagen. Para cada fila  $h$ , se promedian todos los valores a lo largo de esa fila:

$$z_c^h(h) = \frac{1}{W} \sum_{w=1}^W x_c(h, w) \quad (3.1)$$

De manera análoga, para cada columna  $w$ :

$$z_c^w(w) = \frac{1}{H} \sum_{h=1}^H x_c(h, w) \quad (3.2)$$

donde  $x_c(h, w)$  representa el valor en la posición  $(h, w)$  del canal  $c$ ,  $H$  es la altura y  $W$  el ancho de la imagen de características.

**Paso 2: Combinación y procesamiento.** Los resúmenes de filas y columnas se combinan y procesan conjuntamente. Esta combinación permite que el módulo analice la relación entre la información horizontal y vertical para identificar qué regiones son relevantes. Para mantener la eficiencia computacional, la información se comprime temporalmente a una representación más compacta (en este caso, de 512 a 16 valores intermedios) antes de generar los pesos de atención.

**Paso 3: Generación de pesos de atención.** A partir de la información combinada, se generan dos conjuntos de pesos: uno que indica la importancia de cada fila ( $\mathbf{a}^h$ ) y otro para cada columna ( $\mathbf{a}^w$ ). Estos pesos se calculan de forma que cada valor quede en el rango entre 0 y 1, donde 0 significa “ignorar completamente” y 1 significa “prestar máxima atención”.

**Paso 4: Aplicación de la atención.** Finalmente, cada posición de la imagen se pondera según la importancia de su fila y su columna:

$$y_c(h, w) = x_c(h, w) \cdot a_c^h(h) \cdot a_c^w(w) \quad (3.3)$$

De esta manera, las regiones donde tanto la fila como la columna tienen alta importancia reciben mayor peso, mientras que las regiones irrelevantes se atenúan.

### Coordinate Attention (C=512, H=W=7, r=32)

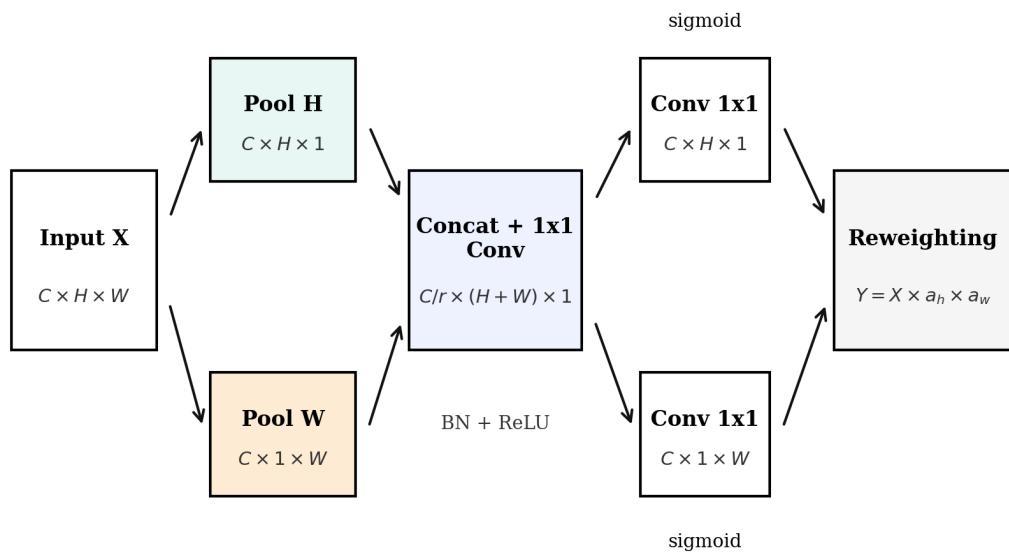


Figura 3.7: Mecanismo del módulo Coordinate Attention y ejemplo de mapa de atención espacial sobre una radiografía. El módulo estima pesos por fila ( $a^h$ ) y por columna ( $a^w$ ), y los aplica para reponer el mapa de características.

**Parámetros del módulo.** La Tabla 3.6 presenta los componentes y parámetros del módulo Coordinate Attention.

Cuadro 3.6: Parámetros del módulo Coordinate Attention para  $C = 512$  canales de entrada.

Componente	Configuración	Parámetros
pool_h	AdaptiveAvgPool2d(None, 1)	0
pool_w	AdaptiveAvgPool2d(1, None)	0
conv1	Conv2d(512, 16, núcleo=1)	8,192
bn1	BatchNorm2d(16)	32
conv_h	Conv2d(16, 512, núcleo=1)	8,192
conv_w	Conv2d(16, 512, núcleo=1)	8,192
<b>Total Coordinate Attention</b>		<b>24,608</b>

## Cabeza de Regresión

**Propósito.** Mientras que el extractor de características (ResNet-18) procesa la imagen y el módulo de atención identifica las regiones relevantes, la cabeza de regresión tiene la tarea final: convertir toda esta información en las 30 coordenadas numéricas que representan la posición de los 15 puntos de referencia (cada punto de referencia tiene una coordenada  $x$  y una coordenada  $y$ ).

**Arquitectura.** La cabeza de regresión procesa la información a través de varias etapas:

1. **Condensación espacial:** Primero, se resume toda la información del mapa de características ( $7 \times 7$  posiciones con 512 valores cada una) en un único vector de 512 valores. Este paso, denominado *Global Average Pooling*, calcula el promedio de cada característica a través de todas las posiciones espaciales.
2. **Capas de transformación:** El vector resultante pasa por dos capas que transforman progresivamente la información. La primera capa mantiene 512 valores, y la segunda expande a 768 valores. Esta expansión permite que el modelo capture relaciones más complejas entre las características antes de producir la salida final.
3. **Capa de salida:** Finalmente, una última capa reduce los 768 valores a exactamente 30 números, que corresponden a las coordenadas de los 15 puntos de referencia. Estos valores se restringen al rango  $[0, 1]$  mediante una función sigmoide, representando posiciones normalizadas respecto al tamaño de la imagen.

**Técnicas de estabilización.** Para que el entrenamiento sea estable y el modelo generalice bien, se incorporan dos técnicas:

- **Normalización por grupos** (*Group Normalization*) [21]: Estandariza los valores intermedios para evitar que crezcan o decrezcan de manera descontrolada durante el entrenamiento. A diferencia de otras técnicas de normalización, esta funciona de manera consistente independientemente del número de imágenes procesadas simultáneamente.
- **Dropout**: Durante el entrenamiento, se “apagan” aleatoriamente algunos valores (30 % en la primera capa, 15 % en la segunda). Esto obliga al modelo a no depender excesivamente de ninguna característica particular, mejorando su capacidad de generalización.

Para obtener las coordenadas finales en píxeles, las predicciones normalizadas se multiplican por el tamaño de la imagen (224 píxeles). La Tabla 3.7 detalla la arquitectura completa de la cabeza de regresión.

Cuadro 3.7: Arquitectura de la cabeza de regresión con normalización por grupos.

Capa	Operación	Salida	Parámetros
avgpool	AdaptiveAvgPool2d(1,1)	512	0
flatten	Flatten	512	0
fc1	Linear(512, 512)	512	262,656
gn1	GroupNorm(32, 512)	512	1,024
relu1	ReLU	512	0
dropout1	Dropout(p=0.3)	512	0
fc2	Linear(512, 768)	768	394,752
gn2	GroupNorm(48, 768)	768	1,536
relu2	ReLU	768	0
dropout2	Dropout(p=0.15)	768	0
fc3	Linear(768, 30)	30	23,070
sigmoid	Sigmoid	30	0
<b>Total cabeza</b>		—	<b>683,038</b>

El número total de parámetros del modelo completo es aproximadamente 11.9 millones, de los cuales 11.2 millones corresponden al extractor de características preentrenado.

### 3.3.2. Función de Pérdida

**Concepto de función de pérdida.** Durante el entrenamiento, el modelo necesita una forma de medir qué tan lejos están sus predicciones de las posiciones reales de los puntos de referencia. Esta medida se denomina *función de pérdida*: un valor numérico que indica qué

tan “equivocado” está el modelo. El objetivo del entrenamiento es ajustar los parámetros del modelo para minimizar este valor.

**Limitaciones de las funciones tradicionales.** Las funciones de pérdida más comunes son el error absoluto (L1) y el error cuadrático (L2). Sin embargo, ambas presentan limitaciones para la localización precisa de puntos de referencia:

- **Error cuadrático (L2):** Penaliza fuertemente los errores grandes, lo cual es útil al inicio del entrenamiento cuando las predicciones están muy alejadas. Sin embargo, cuando las predicciones ya están cerca del objetivo, la penalización se vuelve muy pequeña y el modelo pierde incentivo para seguir mejorando.
- **Error absoluto (L1):** Trata todos los errores de manera uniforme, sin importar su magnitud. Esto significa que un error de 1 píxel recibe la misma “urgencia” de corrección que un error de 20 píxeles, lo cual no es óptimo.

**Wing Loss: comportamiento adaptativo.** Wing Loss [13] fue diseñada específicamente para localización de puntos de referencia, combinando las ventajas de ambos enfoques:

- **Para errores pequeños** (menores a  $\omega$  píxeles): Aplica una penalización que crece rápidamente, incentivando al modelo a refinar las predicciones que ya están cerca del objetivo. Esto es crucial para lograr precisión subpixel.
- **Para errores grandes** (mayores a  $\omega$  píxeles): Se comporta de manera similar a L1, proporcionando correcciones estables sin penalizaciones extremas que podrían desestabilizar el entrenamiento.

La formulación matemática es:

$$\text{wing}(x) = \begin{cases} \omega \ln \left( 1 + \frac{|x|}{\epsilon} \right) & \text{si } |x| < \omega \\ |x| - C & \text{de otro modo} \end{cases} \quad (3.4)$$

donde  $x$  es la diferencia entre la coordenada predicha y la real,  $\omega$  define el umbral entre los dos comportamientos,  $\epsilon$  controla la sensibilidad en la región de errores pequeños, y  $C$  es una constante que asegura continuidad entre ambas regiones.

**Parámetros utilizados.** En este trabajo se utilizan  $\omega = 10$  píxeles y  $\epsilon = 2$  píxeles. Esto significa que errores menores a 10 píxeles reciben el tratamiento de “refinamiento fino”, mientras que errores mayores reciben correcciones estables. Dado que las coordenadas se normalizan al rango  $[0, 1]$  durante el entrenamiento, estos valores se escalan proporcionalmente.

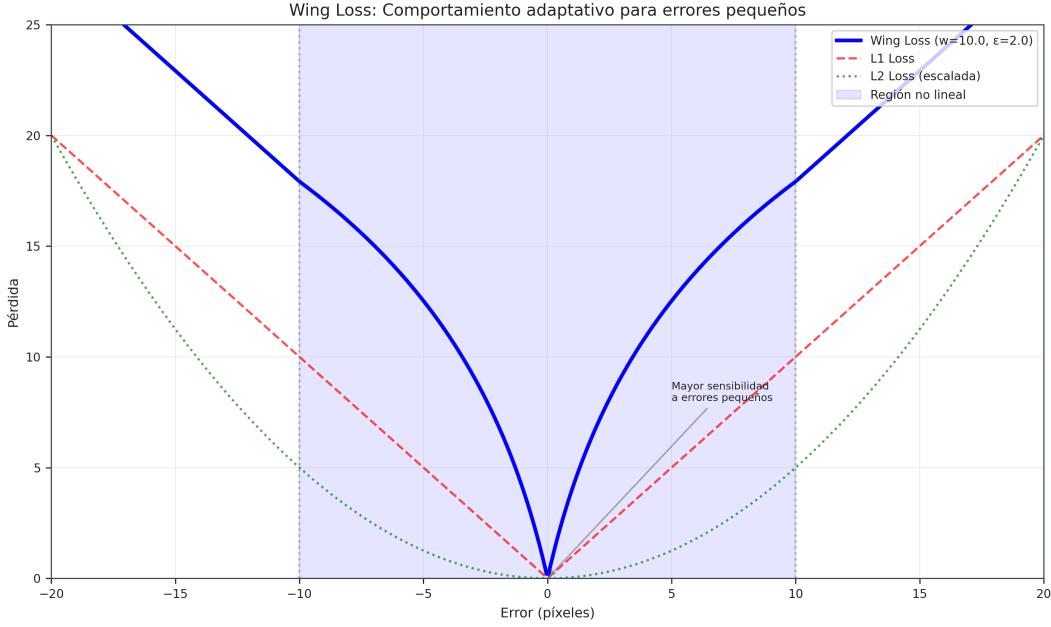


Figura 3.8: Comparación de Wing Loss con pérdidas L1 y L2. Wing Loss proporciona mayor penalización para errores pequeños (región logarítmica) incentivando el refinamiento fino, mientras mantiene estabilidad para errores grandes (región lineal).

La Figura 3.8 ilustra el comportamiento de Wing Loss comparado con L1 y L2.

La pérdida total del modelo se calcula promediando Wing Loss sobre las 30 coordenadas (15 puntos de referencia  $\times$  2 coordenadas):

$$\mathcal{L}_{\text{total}} = \frac{1}{30} \sum_{i=1}^{30} \text{wing}(\hat{c}_i - c_i) \quad (3.5)$$

donde  $\hat{c}_i$  representa la coordenada predicha y  $c_i$  la coordenada real.

### 3.3.3. Estrategia de Entrenamiento

**Aprendizaje por transferencia.** Entrenar una red neuronal desde cero requiere grandes cantidades de datos para que el modelo aprenda a reconocer patrones visuales básicos (bordes, texturas, formas). Sin embargo, cuando se dispone de un conjunto de datos limitado (como las 957 radiografías anotadas de este trabajo) existe el riesgo de que el modelo memorice los ejemplos de entrenamiento en lugar de aprender patrones generalizables.

El *aprendizaje por transferencia* [22] aborda este problema reutilizando un modelo previamente entrenado en una tarea relacionada. En este caso, se utiliza ResNet-18 preentrenado en ImageNet, un conjunto de más de un millón de imágenes naturales. Aunque las radiografías de tórax son visualmente diferentes a las fotografías de ImageNet, las capas iniciales de la red aprenden detectores de características básicas (bordes, gradientes, texturas)

que son útiles para cualquier tarea de visión por computadora.

El entrenamiento se divide en dos fases para aprovechar este conocimiento previo de manera efectiva.

### Fase 1: Entrenamiento de la Cabeza

En la primera fase, se “congelan” los parámetros del extracto de características (ResNet-18) y del módulo Coordinate Attention, es decir, estos parámetros no se modifican durante el entrenamiento. Únicamente se entrena la cabeza de regresión.

Esta estrategia tiene dos objetivos:

1. **Preservar el conocimiento previo:** Las capas convolucionales ya saben detectar bordes, texturas y formas. Modificarlas prematuramente podría destruir este conocimiento antes de que la cabeza aprenda a utilizarlo.
2. **Establecer una base estable:** La cabeza de regresión comienza con pesos aleatorios, por lo que sus predicciones iniciales son esencialmente ruido. Entrenarla primero permite que aprenda a interpretar las características extraídas por ResNet-18 antes de que estas cambien.

La configuración de la Fase 1 se presenta en la Tabla 3.8.

Cuadro 3.8: Configuración de entrenamiento - Fase 1.

Parámetro	Valor
Épocas máximas	15
Tasa de aprendizaje	$1 \times 10^{-3}$
Optimizador	Adam ( $\beta_1 = 0,9$ , $\beta_2 = 0,999$ )
Tamaño de lote	16
Parámetros entrenables	Solo cabeza (683,038)
Parámetros congelados	Backbone + CoordAttn (11,201,120)
Parada temprana	Paciencia = 5 épocas
Métrica de monitoreo	Error de validación (píxeles)

### Fase 2: Ajuste Fino Completo

Una vez que la cabeza de regresión ha aprendido a producir predicciones razonables, se “descongela” todos los parámetros del modelo para realizar un ajuste fino completo. En esta fase, tanto el extractor de características como la cabeza se adaptan específicamente a la tarea de localización de puntos de referencia en radiografías.

**Tasas de aprendizaje diferenciadas.** Un riesgo de modificar el extractor de características es que los ajustes sean tan grandes que destruyan el conocimiento previamente aprendido. Para mitigar este riesgo, se utilizan tasas de aprendizaje diferentes para cada parte del modelo:

- **Extractor + Coordinate Attention:** Tasa de aprendizaje baja ( $2 \times 10^{-5}$ ), permitiendo solo ajustes sutiles que especialicen las características para radiografías sin perder la capacidad general de detección.
- **Cabeza de regresión:** Tasa de aprendizaje 10 veces mayor ( $2 \times 10^{-4}$ ), permitiendo que continúe adaptándose más rápidamente.

Esta proporción 10:1 refleja la intuición de que el extractor ya tiene conocimiento valioso que debe preservarse, mientras que la cabeza aún tiene margen para mejorar.

**Reducción gradual de la tasa de aprendizaje.** A medida que el entrenamiento avanza, el modelo se acerca a una solución óptima y los ajustes necesarios son cada vez más pequeños. Para reflejar esto, se utiliza un esquema de *Cosine Annealing* [23] que reduce la tasa de aprendizaje de manera suave: comienza con el valor inicial y decrece gradualmente hasta un mínimo ( $10^{-6}$ ), siguiendo una curva en forma de coseno. Este decrecimiento suave evita cambios bruscos que podrían desestabilizar el entrenamiento en las etapas finales.

La Tabla 3.9 detalla la configuración de la Fase 2.

Cuadro 3.9: Configuración de entrenamiento - Fase 2.

Parámetro	Valor
Épocas máximas	100
Tasa de aprendizaje (extractor de características + CA)	$2 \times 10^{-5}$
Tasa de aprendizaje (cabeza)	$2 \times 10^{-4}$
Optimizador	Adam ( $\beta_1 = 0,9, \beta_2 = 0,999$ )
Scheduler	Cosine Annealing ( $T = 100, \eta_{\min} = 10^{-6}$ )
Tamaño de lote	8
Parámetros entrenables	Todos (11,884,158)
Parada temprana	Paciencia = 15 épocas
Métrica de monitoreo	Error de validación (píxeles)

## Aumento de Datos

Con solo 957 imágenes anotadas, existe el riesgo de que el modelo memorice características específicas de las imágenes de entrenamiento (como la posición exacta de ciertos artefactos o variaciones particulares de iluminación) en lugar de aprender el concepto general de “contorno

pulmonar”. El *aumento de datos* mitiga este riesgo creando versiones modificadas de las imágenes de entrenamiento, exponiendo al modelo a mayor variabilidad.

Las transformaciones aplicadas son:

1. **Reflejo horizontal** (probabilidad 50 %): Simula variaciones en la orientación del paciente. Al reflejar la imagen, los puntos de referencia del lado izquierdo pasan a estar en el derecho y viceversa, por lo que se intercambian los pares simétricos ( $L3 \leftrightarrow L4$ ,  $L5 \leftrightarrow L6$ ,  $L7 \leftrightarrow L8$ ,  $L12 \leftrightarrow L13$ ,  $L14 \leftrightarrow L15$ ).
2. **Rotación aleatoria** ( $\pm 10$  grados): Simula pequeñas variaciones en la posición del paciente durante la adquisición de la radiografía. Las coordenadas de los puntos de referencia se transforman aplicando la misma rotación para mantener la correspondencia.

Estas transformaciones son *conservadoras*: se limitan a variaciones que podrían ocurrir naturalmente en la adquisición de radiografías, evitando distorsiones que produzcan imágenes anatómicamente irrealistas.

### 3.3.4. Resumen de Hiperparámetros

Los hiperparámetros son valores que definen la configuración del modelo y del proceso de entrenamiento, y que deben establecerse antes de iniciar el entrenamiento (a diferencia de los parámetros del modelo, que se aprenden automáticamente). La Tabla 3.10 consolida todos los hiperparámetros utilizados, organizados por categoría, para facilitar la reproducibilidad de los experimentos.

### 3.3.5. Ensamble de Modelos

**Motivación.** Un modelo individual, por más bien entrenado que esté, puede cometer errores en ciertas imágenes debido a particularidades de su proceso de entrenamiento. Una estrategia para mitigar estos errores es consultar múltiples modelos y combinar sus predicciones, de manera análoga a solicitar la opinión de varios expertos antes de tomar una decisión. Esta técnica se denomina *ensamble* (ensamble).

**Por qué diferentes modelos producen diferentes predicciones.** Aunque todos los modelos del ensamble utilizan la misma arquitectura y datos de entrenamiento, el proceso de entrenamiento involucra elementos aleatorios: el orden en que se presentan las imágenes, los valores iniciales de ciertos parámetros, y las muestras seleccionadas para cada lote. Estas diferencias, controladas por una *semilla aleatoria*, hacen que cada modelo converja a una solución ligeramente diferente. Donde un modelo comete un error, otro puede acertar, y viceversa.

Cuadro 3.10: Resumen completo de hiperparámetros del modelo de puntos de referencia.

Categoría	Parámetro	Valor
Arquitectura	Extractor de características	ResNet-18
	Coordinate Attention	Habilitado (reduction=32)
	Cabeza de regresión	3 capas con GroupNorm
	Dimensiones ocultas	512 → 768
	Parámetros totales	~11.9M
Regularización	Dropout (capa 1)	0.3
	Dropout (capa 2)	0.15
	Aumento de datos	Flip horizontal + Rotación ±10°
Wing Loss	$\omega$	10.0 px
	$\epsilon$	2.0 px
	Normalizado	Sí (escalado a [0,1])
Fase 1	Épocas	15
	Tasa de aprendizaje	$1 \times 10^{-3}$
	Tamaño de lote	16
	Extractor características	Congelado
	Parada temprana	5 épocas
Fase 2	Épocas	100
	LR extractor características + CA	$2 \times 10^{-5}$
	LR cabeza	$2 \times 10^{-4}$
	Tamaño de lote	8
	Parada temprana	15 épocas
	Scheduler	Cosine Annealing

## Configuración del Ensemble

Se entrenaron cuatro modelos utilizando la configuración descrita en las secciones anteriores, variando únicamente la semilla aleatoria (123, 321, 111, 666). La predicción final se obtiene promediando las predicciones de los cuatro modelos:

$$\hat{\mathbf{L}}_{\text{ensemble}} = \frac{1}{4} \sum_{k=1}^4 \hat{\mathbf{L}}_k \quad (3.6)$$

donde  $\hat{\mathbf{L}}_k$  representa las coordenadas predichas por el modelo  $k$ . Este promedio tiene el efecto de cancelar errores individuales: si un modelo predice un punto de referencia 2 píxeles a la izquierda de su posición real y otro lo predice 2 píxeles a la derecha, el promedio estará muy cerca de la posición correcta.

## Aumento en Tiempo de Prueba (TTA)

Además del ensemble, se aplica una técnica complementaria llamada *Aumento en Tiempo de Prueba* (TTA). Durante la inferencia (es decir, cuando el modelo ya está entrenado y se usa para predecir), cada imagen se procesa dos veces: en su orientación original y reflejada horizontalmente. Las predicciones de ambas versiones se promedian (intercambiando los puntos de referencia simétricos en la versión reflejada para que correspondan correctamente).

Esta técnica aprovecha la simetría anatómica del tórax: si el modelo predice correctamente un punto de referencia en la imagen original pero comete un pequeño error en la reflejada (o viceversa), el promedio reduce ese error. TTA no requiere entrenamiento adicional y se aplica únicamente durante la inferencia.

## 3.4. Normalización Geométrica

**El problema de la variabilidad anatómica.** Las radiografías de tórax de diferentes pacientes presentan variaciones significativas: los pulmones pueden aparecer más arriba o más abajo en la imagen, ser más grandes o más pequeños, o estar ligeramente rotados dependiendo de la posición del paciente durante la adquisición. Estas diferencias geométricas, aunque irrelevantes para el diagnóstico médico, representan un desafío para los algoritmos de clasificación automática, ya que el modelo debe aprender a reconocer patologías independientemente de estas variaciones.

**La solución: normalización geométrica.** La normalización geométrica aborda este problema transformando todas las radiografías a una *forma estándar*, donde los pulmones ocupan siempre la misma posición, tienen el mismo tamaño y la misma orientación. De esta manera, el clasificador puede concentrarse en detectar diferencias relevantes (como opacidades o consolidaciones) sin verse confundido por diferencias geométricas irrelevantes.

El proceso de normalización se realiza en tres etapas principales:

1. **Cálculo de la forma estándar:** Se determina una forma de referencia “promedio” a partir del conjunto de entrenamiento mediante Análisis Procrustes Generalizado.
2. **División de la imagen en regiones:** Se partitiona la imagen en triángulos mediante triangulación de Delaunay.
3. **Transformación de cada región:** Se deforma cada triángulo para alinear los puntos de referencia de la imagen con los de la forma estándar.

### 3.4.1. Análisis Procrustes Generalizado

**Objetivo.** Para normalizar las radiografías, primero necesitamos definir una “forma de referencia” que represente la configuración típica de los pulmones. El Análisis Procrustes Generalizado (GPA) [16, 24] es una técnica que permite calcular esta forma de referencia a partir de las 957 configuraciones de puntos de referencia anotadas en el conjunto de entrenamiento.

**Intuición del método.** El nombre “Procrustes” proviene del mito griego de un posadero que ajustaba a sus huéspedes a una cama de tamaño fijo, estirándolos o cortándolos según fuera necesario. De manera análoga, GPA “ajusta” cada configuración de puntos de referencia eliminando tres tipos de diferencias que no afectan la forma intrínseca:

- **Posición:** Si una configuración está desplazada hacia la izquierda y otra hacia la derecha, se centran ambas en el mismo punto.

- **Tamaño:** Si una configuración es más grande que otra, se escalan para que tengan el mismo tamaño.
- **Orientación:** Si una configuración está ligeramente rotada, se rota para alinearla con las demás.

Una vez eliminadas estas diferencias, las configuraciones restantes reflejan únicamente variaciones en la *forma* de los pulmones. El promedio de estas configuraciones alineadas constituye la *forma estándar*.

## Procedimiento de Alineación

El proceso de alineación se realiza en tres pasos:

**Paso 1: Centrado (eliminación de posición).** Cada configuración de landmarks se desplaza para que su centro geométrico coincida con el origen de coordenadas. El centro se calcula como el promedio de las coordenadas de todos los puntos de referencia.

**Paso 2: Escalado (eliminación de tamaño).** Para que todas las configuraciones tengan el mismo “tamaño”, se normalizan dividiendo por una medida de dispersión. Después de este paso, todas las configuraciones quedan con el mismo tamaño estándar, independientemente del tamaño original de los pulmones en la imagen.

**Paso 3: Rotación (eliminación de orientación).** Finalmente, se rota cada configuración para alinearla lo mejor posible con una forma de referencia. La rotación óptima es aquella que minimiza la distancia entre la configuración rotada y la referencia.

## Cálculo de la Rotación Óptima

El ángulo de rotación óptimo se calcula mediante una técnica de álgebra lineal llamada Descomposición en Valores Singulares (SVD) [25], que encuentra automáticamente la rotación que mejor alinea dos conjuntos de puntos. Esta técnica determina el ángulo exacto que minimiza la distancia entre los puntos de una configuración y los de la referencia, sin necesidad de probar múltiples ángulos por ensayo y error.

## Algoritmo Iterativo

Para alinear las configuraciones necesitamos una referencia, pero para calcular la referencia necesitamos que las configuraciones estén alineadas. GPA resuelve este dilema mediante un proceso iterativo:

1. Se comienza con una referencia inicial (el promedio simple de las configuraciones centradas y escaladas).
2. Se alinean todas las configuraciones con esta referencia.

3. Se calcula una nueva referencia como el promedio de las configuraciones alineadas.
4. Se repite hasta que la referencia deje de cambiar significativamente.

El Algoritmo 1 presenta el pseudocódigo formal.

---

#### **Algorithm 1** Análisis Procrustes Generalizado Iterativo

---

**Require:** Conjunto de  $n$  configuraciones de puntos de referencia, tolerancia  $\tau$ , máximo de iteraciones  $T$

**Ensure:** Forma estándar, configuraciones alineadas

- 1: **Preparación:**
  - 2: **for** cada configuración  $i$  de 1 a  $n$  **hacer**
  - 3:     Centrar la configuración (mover su centro al origen)
  - 4:     Escalar la configuración (normalizar su tamaño)
  - 5: **fin for**
  - 6: Calcular referencia inicial como el promedio de todas las configuraciones
  - 7: **Refinamiento iterativo:**
  - 8: **for** cada iteración  $t$  de 1 a  $T$  **hacer**
  - 9:     **for** cada configuración  $i$  de 1 a  $n$  **hacer**
  - 10:         Calcular el ángulo de rotación óptimo para alinear con la referencia
  - 11:         Rotar la configuración según el ángulo calculado
  - 12:     **fin for**
  - 13:     Calcular nueva referencia como el promedio de las configuraciones alineadas
  - 14:     Medir cuánto cambió la referencia respecto a la iteración anterior
  - 15:     **si** el cambio es menor que la tolerancia  $\tau$  **entonces**
  - 16:         Terminar (se alcanzó convergencia)
  - 17:     **fin si**
  - 18: **fin for**
  - 19: La forma estándar es la referencia final
  - 20: **devolver** forma estándar y configuraciones alineadas
- 

Los parámetros utilizados en la implementación son:

- Tolerancia de convergencia:  $\tau = 10^{-8}$
- Máximo de iteraciones:  $T = 100$

En la práctica, el algoritmo converge típicamente en menos de 20 iteraciones para el conjunto de 957 configuraciones de puntos de referencia anotadas.

### Transformación a Coordenadas de Imagen

La forma estándar resultante del GPA está expresada en un sistema de coordenadas matemático (centrada en el origen, con norma unitaria). Para poder utilizarla en el proceso

de deformación, es necesario transformarla al sistema de coordenadas de la imagen (donde las coordenadas van de 0 a 224 píxeles).

Esta transformación simplemente escala y desplaza la forma estándar para que ocupe la región central de la imagen de  $224 \times 224$  píxeles, dejando un margen del 10% en los bordes.

La Figura 3.9 ilustra el proceso de GPA aplicado al conjunto de landmarks.

### 3.4.2. Triangulación de Delaunay

**Necesidad de dividir la imagen.** El siguiente paso del proceso de normalización es deformar la imagen para que los puntos de referencia del paciente coincidan con los puntos de referencia de la forma estándar. Sin embargo, aplicar una única transformación global a toda la imagen no es suficiente: diferentes regiones de la imagen necesitan deformarse de manera diferente (por ejemplo, la parte superior de los pulmones puede necesitar comprimirse mientras que la parte inferior se expande).

La solución es dividir la imagen en regiones más pequeñas (triángulos) y transformar cada región de manera independiente. Esto permite deformaciones locales que preservan la estructura general de la anatomía.

**Por qué triángulos y no otra forma.** Los triángulos tienen una propiedad matemática conveniente: dados tres puntos en la imagen original y sus correspondientes tres puntos en la imagen destino, existe una única transformación lineal (llamada *transformación afín*) que mapea exactamente un triángulo en el otro. Esto no ocurre con cuadriláteros u otras formas más complejas.

#### Triangulación de Delaunay

Existen muchas formas de conectar un conjunto de puntos para formar triángulos. La triangulación de Delaunay [26] es un método que produce triángulos “bien formados”, es decir, triángulos que tienden a ser lo más equiláteros posible, evitando triángulos muy alargados o “delgados” que podrían causar distorsiones visuales durante la deformación [27].

#### Aplicación a los Landmarks

La triangulación se calcula una única vez sobre los puntos de referencia de la forma estándar. Para los 15 puntos de referencia del contorno pulmonar, la triangulación de Delaunay produce 16 triángulos que cubren la región de interés. Esta misma estructura de triángulos se utiliza para todas las imágenes, garantizando consistencia en el proceso de normalización.

La Figura 3.10 muestra la triangulación resultante sobre la forma estándar.

### Análisis Procrustes Generalizado (GPA)

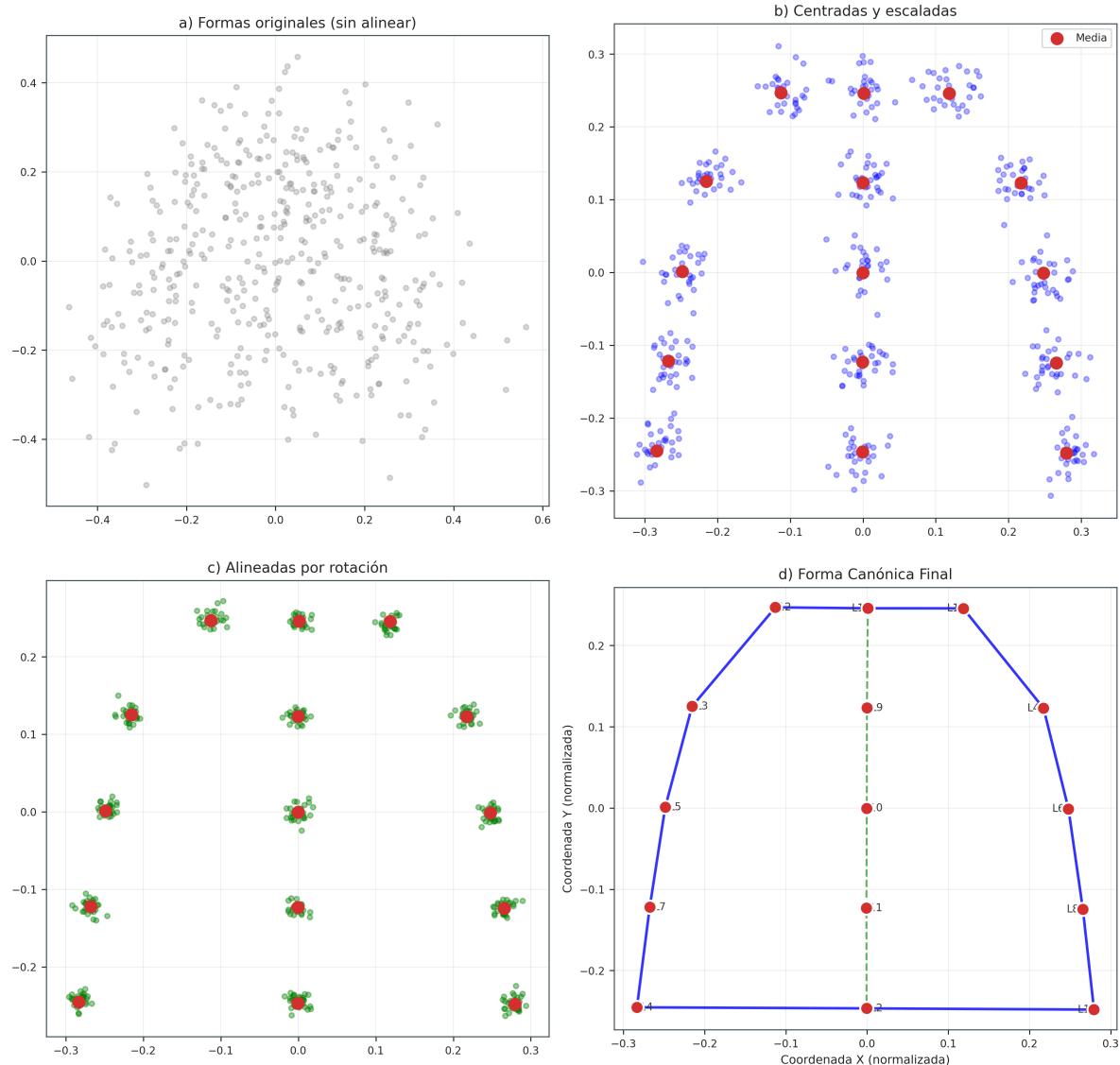


Figura 3.9: Proceso de Análisis Procrustes Generalizado. (a) Las 957 configuraciones de puntos de referencia originales muestran variabilidad en posición, escala y orientación. (b) Después del centrado y escalado, las formas comparten origen y norma unitaria. (c) La alineación rotacional minimiza las diferencias residuales. (d) La forma estándar representa el consenso estadístico del conjunto.

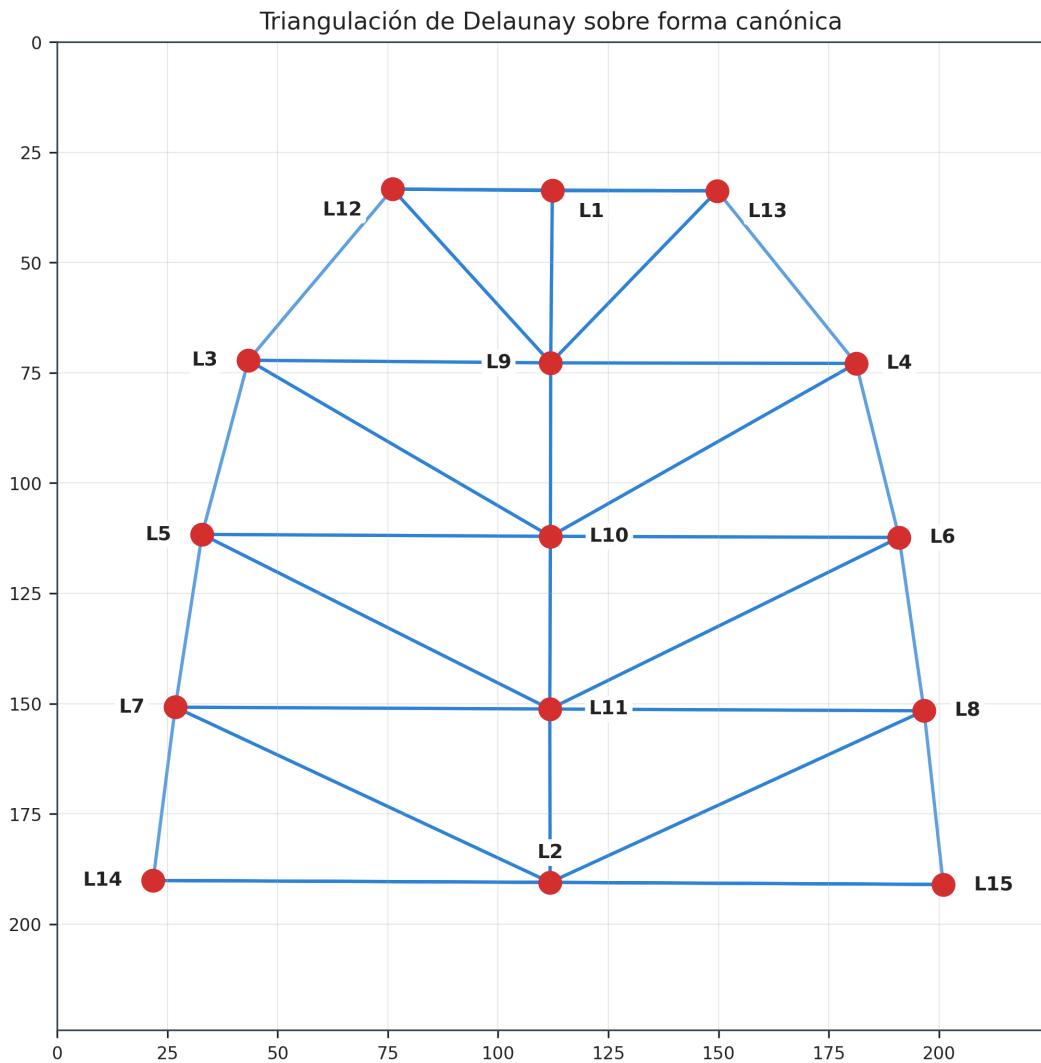


Figura 3.10: Triangulación de Delaunay sobre los 15 puntos de referencia de la forma estándar. Los triángulos definen las regiones donde se aplicarán transformaciones afines independientes durante el proceso de deformación.

### 3.4.3. Transformación Afín por Partes

**El proceso de deformación.** La *deformación* (deformación) es el proceso de “estirar” o “comprimir” partes de la imagen para que los puntos de referencia del paciente coincidan con los puntos de referencia de la forma estándar. Se denomina “por partes” porque cada triángulo se transforma de manera independiente [28].

**Analogía visual.** Imaginemos la imagen impresa en una hoja de goma elástica, con los puntos de referencia marcados como puntos. El warping consiste en “tirar” de cada landmark hasta que coincida con su posición en la forma estándar. Los triángulos actúan como regiones que se estiran de manera uniforme: si un vértice del triángulo se mueve, toda la región triangular se deforma proporcionalmente.

#### Transformación de un Triángulo

Para cada triángulo, se calcula una *transformación afín*: una operación matemática que puede incluir traslación, rotación, escalado y sesgo, pero que preserva las líneas rectas y el paralelismo. La transformación se determina de forma única a partir de la correspondencia entre los tres vértices del triángulo en la imagen original y sus posiciones en la forma estándar. Esto significa que dados tres puntos de origen y tres puntos de destino, existe exactamente una transformación afín que mapea unos en otros.

#### Proceso de Deformación

El deformación completa procesa cada triángulo de la siguiente manera:

1. Se identifican los tres vértices del triángulo en la imagen original (landmarks predichos) y en la forma estándar.
2. Se calcula la matriz de transformación afín que mapea un triángulo en el otro.
3. Se aplica esta transformación a todos los píxeles dentro del triángulo.
4. Se repite para todos los triángulos de la triangulación.

El Algoritmo 2 presenta el pseudocódigo formal del proceso.

Durante el warping, los valores de los píxeles en posiciones intermedias se calculan mediante *interpolación bilineal*, que promedia los valores de los píxeles vecinos para producir transiciones suaves. Esto evita que la imagen resultante tenga bordes dentados o discontinuidades visibles.

La Figura 3.11 muestra la comparación visual entre una radiografía original y su versión normalizada geométricamente.

---

**Algorithm 2** Deformación Afín por Partes

---

**Require:** Imagen original, landmarks de la imagen, puntos de referencia de la forma estándar, triangulación

**Ensure:** Imagen normalizada

- 1: Crear imagen destino vacía (todos los píxeles en negro)
  - 2: **for** cada triángulo en la triangulación **hacer**
  - 3:     Identificar los 3 vértices del triángulo en la imagen original
  - 4:     Identificar los 3 vértices correspondientes en la forma estándar
  - 5:     **si** el triángulo tiene área muy pequeña **entonces**
  - 6:         Saltar este triángulo (evitar divisiones por cero)
  - 7:     **fin si**
  - 8:     Calcular la transformación afín que mapea los vértices originales a los estándar
  - 9:     Aplicar la transformación a todos los píxeles dentro del triángulo
  - 10:    Copiar los píxeles transformados a la imagen destino
  - 11: **fin for**
  - 12: **devolver** imagen normalizada
- 

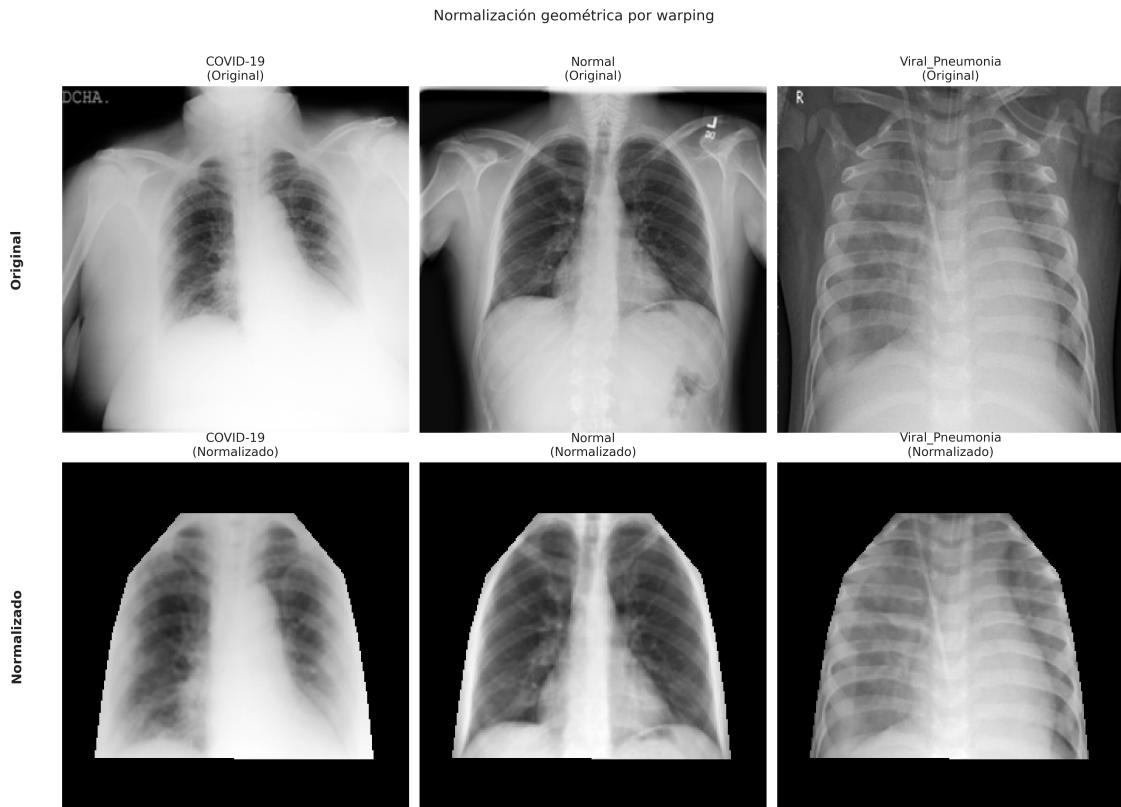


Figura 3.11: Comparación del proceso de deformación. (a) Radiografía original con variabilidad de pose y escala. (b) Imagen normalizada geométricamente mediante warping, enfocándose en la región delimitada por los puntos de referencia.

### 3.4.4. Proceso Completo de Normalización

Esta sección resume cómo se integran todos los componentes descritos anteriormente para normalizar una radiografía. La Figura 3.12 presenta el diagrama de flujo del proceso completo.



Figura 3.12: Proceso completo de normalización geométrica. El sistema transforma una radiografía de entrada en una imagen geométricamente normalizada mediante la secuencia de predicción de landmarks, forma estándar, triangulación y warping afín por partes.

**Pasos del proceso.** Cuando llega una nueva radiografía para clasificar, el sistema ejecuta los siguientes pasos:

1. **Predicción de landmarks:** El modelo de detección (descrito en la Sección 3.3) localiza los 15 puntos del contorno pulmonar en la radiografía de entrada.
2. **Forma estándar:** Se utiliza la forma estándar pulmonar obtenida del GPA junto con los puntos de referencia predichos para establecer la correspondencia entre los puntos de la imagen y los puntos de referencia.
3. **Triangulación:** Se aplica la triangulación de Delaunay sobre los 15 puntos de referencia para definir los triángulos que conectan los puntos.
4. **Warping:** Cada triángulo de la imagen original se transforma para que sus vértices coincidan con los de la forma estándar.

5. **Resultado:** Se obtiene una imagen de  $224 \times 224$  píxeles donde los pulmones tienen siempre la misma posición, tamaño y orientación.

## 3.5. Clasificación de Enfermedades Pulmonares

Una vez que las imágenes han sido normalizadas geométricamente mediante el proceso de deformación descrito en la sección anterior, el siguiente paso del sistema es la clasificación automática. El objetivo es determinar a cuál de las tres categorías diagnósticas pertenece cada radiografía: COVID-19, Normal o Neumonía Viral.

Esta sección describe los componentes principales del módulo de clasificación: el preprocesamiento de contraste aplicado a las imágenes, la arquitectura de red neuronal seleccionada, la estrategia para aprovechar conocimiento previo de otras tareas (aprendizaje por transferencia), y la configuración del proceso de entrenamiento.

### 3.5.1. Preprocesamiento de Contraste

Antes de la clasificación, las imágenes normalizadas geométricamente requieren un ajuste de contraste para optimizar la visibilidad de las estructuras pulmonares. Como se describe en la Sección 2.3, las radiografías de tórax presentan histogramas de intensidad marcadamente asimétricos, lo que hace que técnicas convencionales como CLAHE presenten limitaciones: amplificación de ruido en áreas suaves y alteración de regiones brillantes indicativas de infiltrados pulmonares.

Para abordar estas limitaciones, se aplica el método SAHS (*Statistical Asymmetrical Histogram Stretching*) [10] a las imágenes deformadas. Este método calcula límites de estiramiento asimétricos que se adaptan a la distribución característica de los histogramas radiográficos, preservando la información diagnóstica relevante.

La elección de SAHS sobre otras técnicas se fundamenta en:

- **Preservación de información diagnóstica:** Los factores asimétricos (2.5 para el límite superior, 2.0 para el inferior) preservan las regiones brillantes características de infiltrados pulmonares.
- **Robustez:** El enfoque estadístico minimiza la influencia de píxeles atípicos causados por ruido o artefactos del proceso de deformación.
- **Eficiencia:** A diferencia de CLAHE, SAHS opera globalmente, reduciendo el costo computacional.

La Figura 3.13 muestra el efecto de aplicar SAHS a imágenes normalizadas geométricamente de las tres categorías diagnósticas.

### 3.5.2. Arquitectura del Clasificador

Para la tarea de clasificación se seleccionó la arquitectura ResNet-18 [11], una red neuronal convolucional diseñada específicamente para el análisis de imágenes. El nombre

**Efecto del preprocesamiento SAHS sobre imágenes normalizadas geométricamente  
(SAHS aplicado solo a la región pulmonar)**

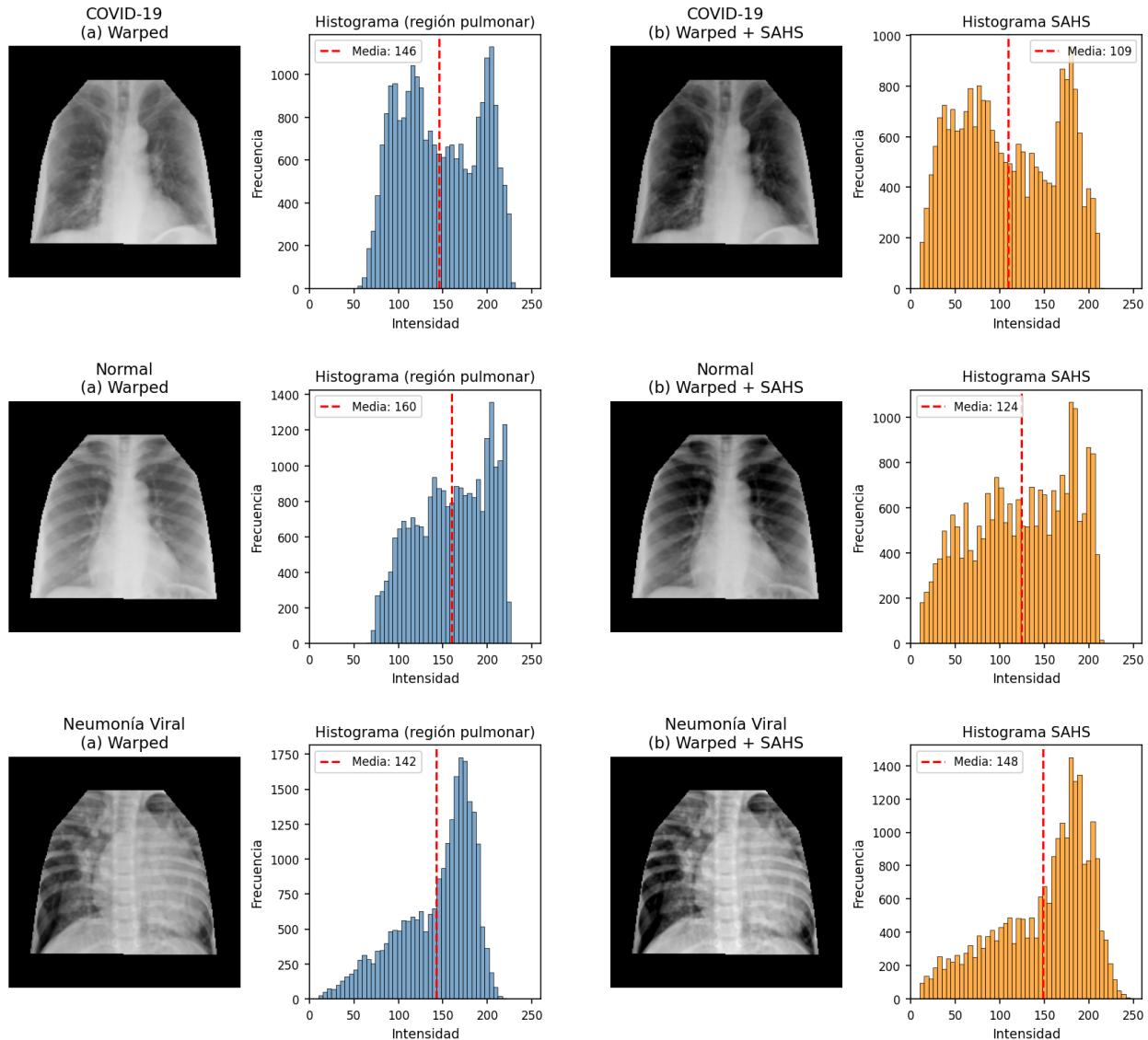


Figura 3.13: Efecto del preprocesamiento SAHS sobre imágenes normalizadas geométricamente. Cada fila muestra una categoría diagnóstica diferente (COVID-19, Normal, Neumonía Viral). Las columnas muestran: (a) imagen después del warping, histograma original, (b) imagen con SAHS aplicado, e histograma resultante. Se observa cómo SAHS redistribuye las intensidades mejorando el contraste sin amplificar artefactos.

hace referencia a su profundidad de 18 capas y a su característica distintiva: las conexiones residuales, que permiten que la información fluya de manera más eficiente a través de la red.

Esta arquitectura se inicializa con conocimiento previamente adquirido del conjunto ImageNet [19], una base de datos de millones de imágenes cotidianas. Aunque las radiografías difieren de las imágenes de ImageNet, las capas iniciales de la red aprenden a detectar patrones visuales básicos (bordes, texturas, formas) que son útiles para cualquier tarea de análisis de imágenes.

### Justificación de ResNet-18

La selección de ResNet-18 se fundamenta en las siguientes características:

1. **Consistencia con el sistema:** Se utiliza la misma familia de arquitectura que el modelo de detección de puntos de referencia, lo que facilita la integración y mantenimiento del sistema completo.
2. **Eficiencia:** Con 11.2 millones de parámetros ajustables, la red es suficientemente expresiva para capturar patrones complejos, pero lo bastante compacta para entrenar de manera eficiente.
3. **Conexiones residuales:** Esta innovación arquitectónica permite que la información se transmita directamente entre capas no consecutivas, evitando que la señal de aprendizaje se degrade en redes profundas [11].
4. **Rendimiento comprobado:** ResNet-18 ha demostrado buen desempeño en múltiples tareas de clasificación de imágenes médicas, incluyendo el diagnóstico de enfermedades torácicas en radiografías [3].

### 3.5.3. Estrategia de Aprendizaje por Transferencia

El aprendizaje por transferencia es una técnica que permite aprovechar el conocimiento que una red neuronal ha adquirido en una tarea para aplicarlo a otra diferente. En este caso, se utiliza una red previamente entrenada para reconocer miles de objetos cotidianos (ImageNet) y se adapta para clasificar radiografías de tórax. Esta estrategia es especialmente efectiva cuando se dispone de conjuntos de datos relativamente pequeños, como es común en aplicaciones médicas [29].

### Adaptación de la Arquitectura

Para adaptar la red preentrenada a la clasificación de tres categorías diagnósticas, se realizan las siguientes modificaciones:

1. **Conservación del extractor de características:** Se mantienen intactos los parámetros de las capas convolucionales, que han aprendido a identificar patrones visuales relevantes.
2. **Reemplazo de la capa de decisión:** La capa final, originalmente diseñada para distinguir entre 1000 categorías de objetos, se reemplaza por una nueva capa que produce tres salidas correspondientes a las categorías diagnósticas (COVID-19, Normal, Neumonía Viral).
3. **Inicialización selectiva:** Solo los parámetros de la nueva capa de decisión se inicializan de forma aleatoria; el resto de la red conserva el conocimiento adquirido previamente.

### Ajuste fino de la Red Completa

El ajuste fino consiste en continuar el entrenamiento de toda la red, permitiendo que los parámetros preentrenados se adapten gradualmente a la nueva tarea. A diferencia del modelo de detección de puntos de referencia, que utiliza un esquema en dos fases, el clasificador se entrena con ajuste fino completo desde el inicio. Esta decisión se justifica por:

- **Mayor volumen de datos:** El conjunto de imágenes normalizadas (15,153 imágenes) es considerablemente mayor que el conjunto utilizado para entrenar el modelo de puntos de referencia (957 imágenes), lo que reduce el riesgo de sobreajuste.
- **Similitud con la tarea original:** La clasificación de imágenes es más similar a las tareas de ImageNet que la predicción de coordenadas específicas.
- **Preprocesamiento estandarizado:** La normalización geométrica aplicadas previamente reducen la variabilidad entre imágenes, facilitando el aprendizaje.

#### 3.5.4. Configuración del Entrenamiento

El proceso de entrenamiento consiste en ajustar los parámetros de la red de manera iterativa, presentándole ejemplos etiquetados para que aprenda a distinguir entre las tres categorías diagnósticas. Esta sección describe los aspectos clave de la configuración, incluyendo cómo se manejan las diferencias en la cantidad de ejemplos disponibles para cada categoría.

#### Distribución del Conjunto de Datos

El conjunto de datos de imágenes normalizadas se divide en tres subconjuntos siguiendo la estrategia de partición descrita en la Sección 3.2. La Tabla 3.11 muestra la distribución por categoría diagnóstica.

Cuadro 3.11: Distribución del conjunto de datos para entrenamiento del clasificador.

Categoría	Entrenamiento	Validación	Prueba	Total (%)
COVID-19	2,712	542	362	3,616 (24 %)
Normal	7,644	1,529	1,020	10,193 (67 %)
Neumonía Viral	1,008	200	136	1,344 (9 %)
<b>Total</b>	<b>11,364</b>	<b>2,271</b>	<b>1,518</b>	<b>15,153</b>

### Manejo del Desbalance de Clases

Como se observa en la Tabla 3.11, el conjunto de datos presenta un desbalance considerable: la categoría Normal representa el 67 % de las muestras, mientras que Neumonía Viral solo el 9 %. Sin una corrección adecuada, el modelo tendería a favorecer las categorías más frecuentes, perjudicando la detección de las menos comunes.

Para compensar este desbalance, se asigna a cada categoría un peso inversamente proporcional a su frecuencia. De esta manera, los errores en categorías poco representadas penalizan más al modelo durante el entrenamiento, incentivándolo a prestarles mayor atención. La Tabla 3.12 muestra los pesos calculados.

Cuadro 3.12: Pesos asignados a cada categoría para compensar el desbalance del conjunto de datos.

Categoría	Muestras (Entrenamiento)	Peso
COVID-19	2,712	1.40
Normal	7,644	0.50
Neumonía Viral	1,008	3.76

Estos pesos se incorporan en la función de pérdida, que es la medida que el modelo busca minimizar durante el entrenamiento. Al ponderar los errores de esta manera, se logra que el modelo aprenda a clasificar correctamente todas las categorías, no solo las más frecuentes.

### Parámetros de Entrenamiento

Los parámetros de entrenamiento controlan aspectos como la velocidad de aprendizaje, el número de ejemplos procesados simultáneamente, y los criterios para detener el entrenamiento. La Tabla 3.13 resume la configuración utilizada.

### Parada Temprana

El sobreajuste ocurre cuando un modelo aprende demasiado bien los ejemplos de entrenamiento, incluyendo sus particularidades y ruido, perdiendo capacidad de generalizar

Cuadro 3.13: Parámetros de configuración del entrenamiento del clasificador.

Parámetro	Valor
Épocas máximas	50
Tamaño de lote	32 imágenes
Tasa de aprendizaje	$1 \times 10^{-4}$
Optimizador	Adam
Función de pérdida	Entropía cruzada con pesos
Dropout	0.3 (30 % de neuronas desactivadas)
Parada temprana	Paciencia de 10 épocas
Semilla aleatoria	42

a ejemplos nuevos. Para prevenirlo, se implementa un mecanismo de parada temprana que detiene el entrenamiento cuando el modelo deja de mejorar en el conjunto de validación.

El proceso funciona de la siguiente manera:

1. Despues de cada pasada completa por el conjunto de entrenamiento (época), se evalúa el rendimiento en el conjunto de validación.
2. Si transcurren 10 épocas consecutivas sin mejora, el entrenamiento se detiene automáticamente.
3. Se conserva la versión del modelo que obtuvo el mejor rendimiento en validación.

Como métrica de seguimiento para la parada temprana se utiliza el F1-Score Macro, que promedia el rendimiento en las tres categorías sin favorecer a las más frecuentes. Esta elección evita que el modelo se especialice en la categoría mayoritaria (Normal) durante el entrenamiento.

### 3.5.5. Métricas de Evaluación del Clasificador

La evaluación del rendimiento del clasificador utiliza múltiples métricas complementarias, con la **exactitud (accuracy)** como indicador principal del rendimiento global.

#### Exactitud como Métrica Principal

La exactitud se adopta como métrica principal de evaluación por las siguientes razones:

1. **Compensación del desbalance:** Aunque el conjunto de datos presenta desbalance entre categorías, este se corrige durante el entrenamiento mediante pesos por clase (Tabla 3.12). Esta compensación permite que la exactitud refleje el rendimiento real del modelo sin sesgo hacia la clase mayoritaria.

2. **Objetivo de correctitud general:** El propósito del sistema es clasificar correctamente la mayor cantidad de radiografías posible, independientemente de la categoría. La exactitud cuantifica directamente este objetivo.
3. **Interpretabilidad clínica:** En el contexto médico, la exactitud tiene una interpretación directa: la proporción de diagnósticos correctos sobre el total de casos evaluados.

### Métricas Complementarias

Adicionalmente, se reportan las siguientes métricas para caracterizar el comportamiento del clasificador:

- **F1-Score Macro:** Promedio no ponderado del F1-Score de cada categoría. Verifica que el rendimiento sea equilibrado entre clases y no esté dominado por la categoría mayoritaria.
- **Precisión y Sensibilidad por categoría:** Permiten identificar patrones específicos de error (falsos positivos vs. falsos negativos) para cada diagnóstico.
- **Matriz de confusión:** Visualiza la distribución completa de predicciones, facilitando el análisis de errores clínicamente relevantes (e.g., confusión entre patologías que requieren tratamientos diferentes).

### 3.5.6. Aumento de Datos

El aumento de datos es una técnica que consiste en crear variaciones artificiales de las imágenes de entrenamiento mediante transformaciones controladas. El objetivo es exponer al modelo a mayor diversidad de ejemplos, mejorando su capacidad de generalizar a imágenes nuevas. Las transformaciones se diseñan cuidadosamente para simular variaciones realistas que podrían ocurrir durante la adquisición de radiografías, sin alterar las características diagnósticas relevantes.

#### Transformaciones Durante el Entrenamiento

Se aplican las siguientes transformaciones a cada imagen:

1. **Conversión a tres canales:** Las radiografías, originalmente en escala de grises, se convierten a formato de tres canales para compatibilidad con la arquitectura preentrenada.
2. **Ajuste de tamaño:** Todas las imágenes se escalan a  $224 \times 224$  píxeles.

3. **Reflejo horizontal:** La imagen se refleja horizontalmente, simulando variaciones en la orientación del paciente durante la toma.
4. **Rotación leve:** Se aplican rotaciones aleatorias de hasta 10 grados en cualquier dirección, representando pequeñas variaciones en el posicionamiento.
5. **Desplazamiento y escala:** Se aplican pequeños desplazamientos (hasta 5 % en cada eje) y variaciones de escala (entre 95 % y 105 %), simulando diferencias en el encuadre de la imagen.

### Transformaciones Durante la Evaluación

Durante la evaluación del modelo, solo se aplican las transformaciones determinísticas (conversión a tres canales, ajuste de tamaño y normalización), sin variaciones aleatorias. Esto asegura que los resultados de evaluación sean reproducibles y comparables.

La Figura 3.14 ilustra ejemplos de las transformaciones aplicadas durante el entrenamiento.

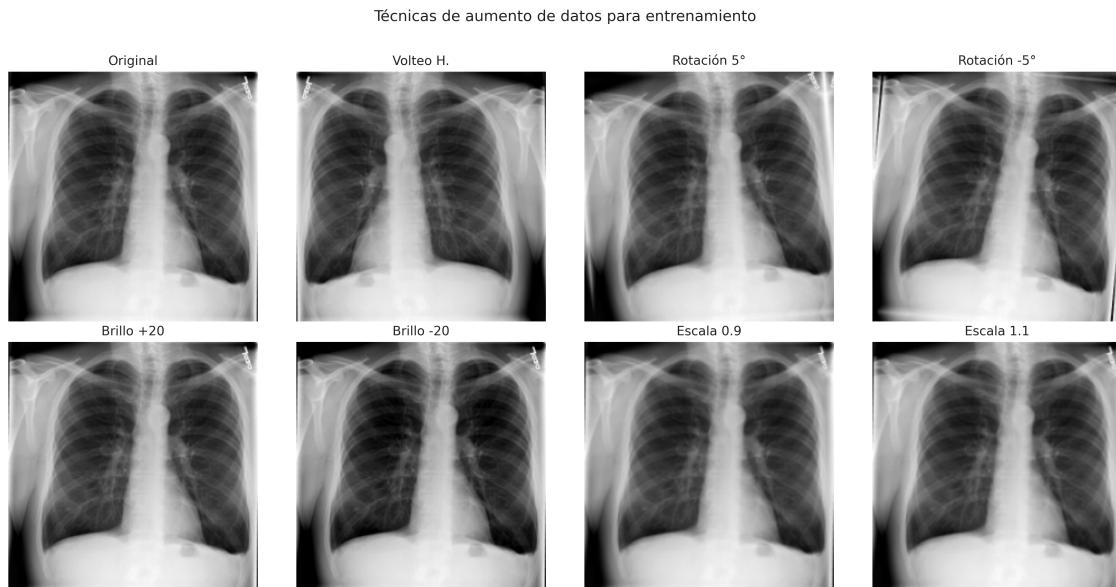


Figura 3.14: Ejemplos de transformaciones de aumento de datos aplicadas durante el entrenamiento del clasificador. Las transformaciones preservan la integridad diagnóstica de las imágenes mientras aumentan la variabilidad del conjunto de entrenamiento.

### 3.5.7. Resumen de la Configuración

La Tabla 3.14 consolida los aspectos principales del módulo de clasificación descritos en esta sección.

Cuadro 3.14: Resumen de la configuración del clasificador de enfermedades pulmonares.

Aspecto	Elemento	Valor/Descripción
Preprocesamiento	Mejora de contraste	SAHS (asimétrico)
Arquitectura	Red base	ResNet-18
	Conocimiento previo	ImageNet
	Categorías de salida	3
	Parámetros ajustables	~11.2 millones
Prevención de sobreajuste	Dropout	30 % de neuronas
	Aumento de datos	Reflejo, rotación, escala
Entrenamiento	Épocas máximas	50
	Imágenes por lote	32
	Compensación de desbalance	Pesos por categoría
	Optimizador	Adam
Parada temprana	Paciencia	10 épocas
	Métrica monitoreada	F1-Score Macro

## 3.6. Protocolo de Inferencia y Evaluación

Esta sección describe el proceso completo de inferencia del sistema, las métricas utilizadas para evaluar su rendimiento y el protocolo experimental seguido para garantizar la validez de los resultados.

### 3.6.1. Proceso de Inferencia

El proceso de inferencia del sistema sigue un flujo secuencial de cuatro etapas. Una radiografía de entrada en formato RGB o escala de grises se procesa primero mediante CLAHE para normalizar el contraste, generando una imagen de  $224 \times 224 \times 3$  píxeles. Esta imagen normalizada se alimenta al modelo de detección de puntos de referencia, que produce un vector de 30 valores (15 coordenadas  $(x, y)$  normalizadas al rango  $[0, 1]$ ). Utilizando estos puntos de referencia predichos y la forma estándar pulmonar calculada previamente mediante GPA, el módulo de normalización geométrica aplica una transformación afín por partes que genera una nueva imagen de  $224 \times 224 \times 3$  píxeles con la región pulmonar alineada. Finalmente, esta imagen normalizada se procesa mediante el clasificador, que produce un vector de tres probabilidades correspondientes a las categorías COVID-19, Normal y Neumonía Viral. La clase con mayor probabilidad constituye la predicción final del sistema.

Este flujo modular permite que cada componente opere de forma independiente y secuencial, facilitando la identificación de posibles puntos de fallo y la interpretación de resultados intermedios. Los puntos de referencia predichos, en particular, proporcionan una

representación visual intermedia que permite verificar la calidad de la detección de la región pulmonar antes de la clasificación final.

### 3.6.2. Métricas de Evaluación

El sistema se evalúa en dos niveles: la calidad de la detección de puntos de referencia y el rendimiento del clasificador. Cada nivel requiere métricas específicas que capturen diferentes aspectos del desempeño.

#### Métricas para Detección de Landmarks

La calidad de la detección de puntos de referencia se mide mediante el error en píxeles, definido como la distancia euclíadiana entre la posición predicha y la posición anotada manualmente (ground truth) para cada punto de referencia. Para una imagen dada, se calcula el error de cada uno de los 15 puntos de referencia y se promedia para obtener el error medio por imagen. Esta métrica captura la precisión espacial del modelo de detección.

Adicionalmente, se calcula el error promedio por punto de referencia individual sobre todo el conjunto de prueba, lo cual permite identificar cuáles puntos de referencia son más difíciles de detectar. En el contexto de imágenes de  $224 \times 224$  píxeles, un error de pocos píxeles representa una desviación milimétrica en la radiografía original, lo cual es clínicamente aceptable para el propósito de normalización geométrica.

#### Métricas para Clasificación

El rendimiento del clasificador se evalúa mediante un conjunto de métricas estándar en problemas de clasificación multiclas. La **exactitud (accuracy)** mide la proporción de predicciones correctas sobre el total de imágenes evaluadas, proporcionando una medida global del desempeño del sistema.

Para evaluar el rendimiento por categoría diagnóstica, se calculan tres métricas fundamentales. La **precisión** mide, de todas las predicciones positivas para una clase dada, qué proporción son correctas. Esta métrica es relevante cuando el costo de falsos positivos es alto, como en diagnósticos de COVID-19 que pueden generar ansiedad innecesaria o protocolos de aislamiento incorrectos. La **sensibilidad (recall)** mide, de todos los casos reales de una clase, qué proporción detecta correctamente el sistema. Esta métrica es crítica cuando el costo de falsos negativos es alto, como en la detección de COVID-19 donde no identificar un caso positivo puede resultar en transmisión comunitaria.

El **F1-Score** combina precisión y sensibilidad en una sola métrica mediante, proporcionando un balance entre ambos objetivos. Se calcula tanto el F1-Macro (promedio no ponderado entre las tres categorías, que trata todas las clases con igual importancia) como

el F1-Weighted (promedio ponderado por el número de muestras de cada clase, que refleja el desempeño en el contexto de desbalance entre categorías).

Finalmente, la **matriz de confusión** presenta la distribución completa de predicciones versus categorías reales, permitiendo identificar patrones de error específicos. Por ejemplo, si el sistema confunde frecuentemente COVID-19 con Neumonía Viral, esto sugiere similitud visual entre ambas patologías y orienta mejoras futuras. La matriz de confusión es particularmente útil en contextos clínicos donde ciertos tipos de error tienen mayor gravedad que otros.

## Protocolo de Evaluación

El conjunto de datos se divide en tres particiones independientes: entrenamiento, validación y prueba. La partición de entrenamiento se utiliza para optimizar los pesos del modelo mediante descenso de gradiente. La partición de validación se emplea durante el entrenamiento para monitorear el rendimiento en datos no vistos y aplicar criterios de parada temprana que eviten el sobreajuste. Finalmente, la partición de prueba se reserva exclusivamente para la evaluación final del sistema y no se utiliza en ninguna decisión durante el entrenamiento.

Para garantizar la validez de las métricas reportadas, las tres particiones se crean mediante estratificación por clase, lo cual asegura que cada conjunto mantenga el mismo peso en sus muestras de COVID-19, Normal y Neumonía Viral. Esta estratificación es crítica en presencia de desbalance entre clases, ya que evita que alguna partición carezca de representación adecuada de alguna categoría.

Todas las métricas reportadas en el Capítulo 4 corresponden a evaluaciones sobre el conjunto de prueba, garantizando que los resultados reflejan la capacidad del sistema para generalizar a imágenes completamente no vistas durante el desarrollo del modelo.

# Capítulo 4

## Resultados

Este capítulo presenta los resultados obtenidos en cada etapa del sistema: la detección de puntos de referencia anatómicos, la normalización geométrica de las imágenes y la clasificación de enfermedades pulmonares.

### 4.1. Detección de Puntos de Referencia

El sistema detecta automáticamente 15 puntos de referencia que definen el contorno de los pulmones en las radiografías. Esta sección presenta la precisión alcanzada por el modelo, evaluado sobre 96 imágenes del conjunto de prueba que cuentan con anotaciones de referencia realizadas por expertos.

#### 4.1.1. Precisión del Sistema

La Tabla 4.1 presenta los resultados de detección de puntos de referencia. Se compara el mejor modelo individual contra el sistema final que combina cuatro modelos.

Cuadro 4.1: Precisión de detección de puntos de referencia. El error se mide en píxeles sobre imágenes de  $224 \times 224$ .

Configuración	Error promedio	Error mediano	Mejora
Mejor modelo individual	4.04 px	—	—
Sistema combinado (4 modelos)	3.61 px	3.07 px	10.6 %

El sistema combinado alcanza un error promedio de **3.61 píxeles**, lo que representa una mejora del 10.6 % respecto al mejor modelo individual. En una imagen de 224 píxeles, este error equivale al 1.6 % del tamaño de la imagen. El error mediano de 3.07 píxeles indica que la mitad de las predicciones tienen un error menor o igual a este valor.

#### 4.1.2. Precisión por Punto de Referencia

No todos los puntos de referencia se detectan con la misma facilidad. Los puntos del eje central (línea de la columna vertebral) presentan los errores más bajos (2.44–2.94 píxeles) debido a que esta estructura está bien definida en las radiografías. En contraste, las esquinas

superiores de los pulmones presentan mayor dificultad (5.35–5.43 píxeles) porque los límites del pulmón son menos nítidos en esa zona. Los puntos del contorno medio tienen errores intermedios (2.88–3.96 píxeles).

La Figura 4.1 muestra visualmente la distribución de errores sobre la forma estándar, donde se aprecia que los puntos centrales y del contorno medio tienen menor error que las esquinas.

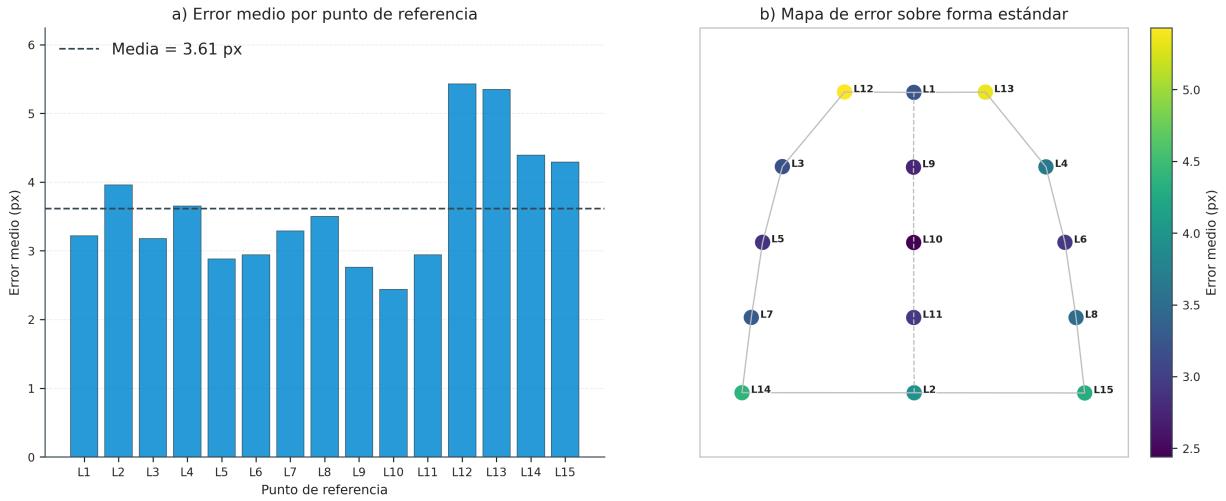


Figura 4.1: Distribución del error de detección por punto de referencia. (a) Error medio en píxeles para cada punto L1–L15. (b) Visualización sobre la forma estándar pulmonar, donde el color indica el error medio (escala continua). Los puntos del eje central presentan la mayor precisión.

La Figura 4.2 presenta ejemplos de detección automática de puntos sobre imágenes reales de las tres categorías diagnósticas. Puede observarse cómo el sistema identifica correctamente el contorno pulmonar incluso en presencia de opacidades por COVID-19 o neumonía viral, manteniendo precisión consistente entre categorías (error promedio de 3.22 px en imágenes normales, 3.93 px en COVID-19, y 4.11 px en neumonía viral).

#### 4.1.3. Resumen

El sistema de detección de puntos de referencia alcanza una precisión de 3.61 píxeles en promedio, equivalente al 1.6 % del tamaño de la imagen. Esta precisión es suficiente para el proceso de normalización geométrica, ya que permite alinear correctamente la región pulmonar sin introducir distorsiones significativas.

Ejemplos de predicción de landmarks (Ensemble con TTA)

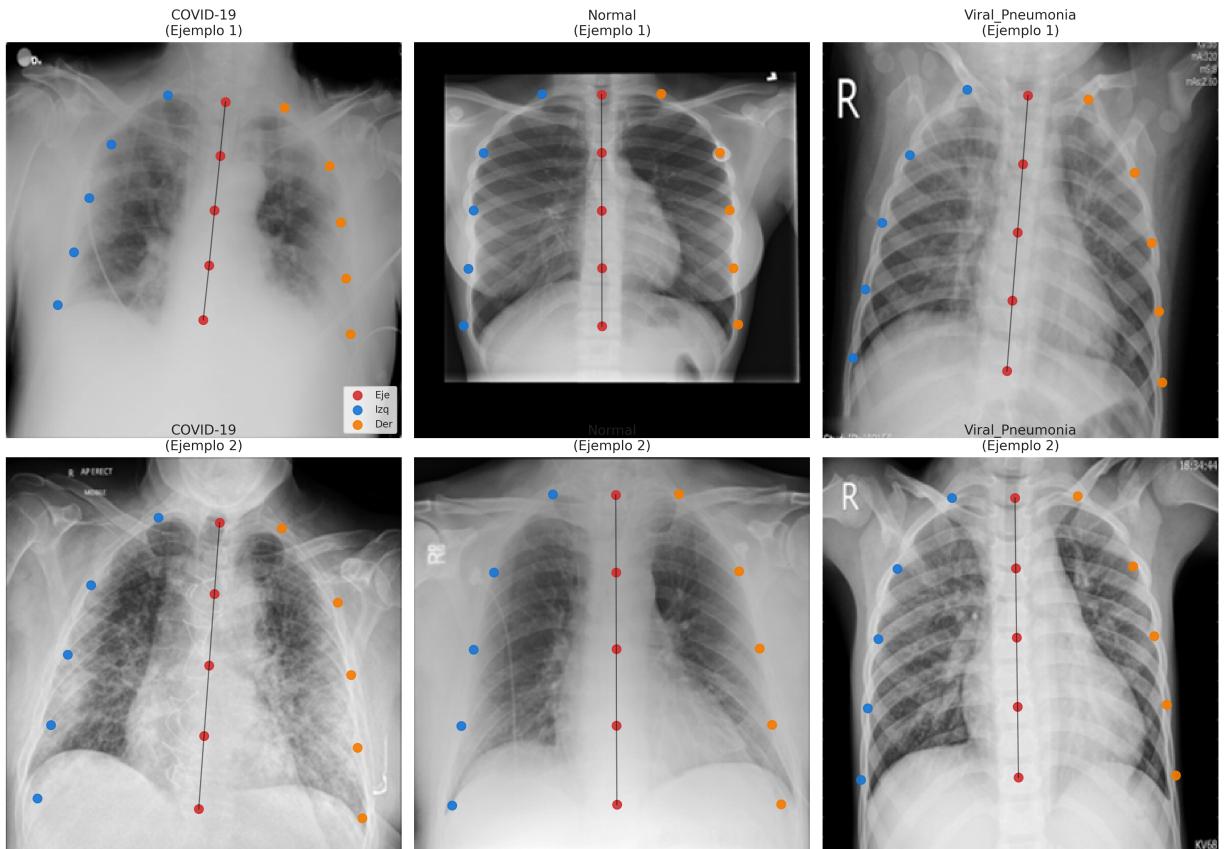


Figura 4.2: Ejemplos de predicción automática de puntos de referencia sobre imágenes reales. Cada fila muestra dos ejemplos de una categoría diagnóstica (Normal, COVID-19, Neumonía Viral). Los puntos indican las predicciones del sistema. El modelo mantiene precisión consistente independientemente del tipo de patología presente.

## 4.2. Normalización Geométrica

Una vez detectados los puntos de referencia, el sistema transforma cada radiografía para alinear los pulmones a una forma estándar. Esta sección presenta los resultados de este proceso de normalización.

### 4.2.1. Forma Estándar de Referencia

La forma estándar se calculó mediante Análisis Procrustes Generalizado (GPA), un método estadístico que promedia la posición de los puntos de referencia de 957 radiografías anotadas manualmente. Este proceso elimina diferencias de posición, escala y rotación, obteniendo una configuración “típica” de los pulmones que sirve como plantilla de destino para todas las imágenes.

La Figura 4.3 muestra la forma estándar resultante, que sirve como referencia para normalizar todas las radiografías.

### 4.2.2. División en Triángulos para Transformación

La región pulmonar se divide en 16 triángulos mediante triangulación de Delaunay, un método geométrico que conecta los 15 puntos de referencia formando triángulos que no se superponen. Cada triángulo se transforma independientemente mediante una transformación afín (rotación, escala y traslación), permitiendo ajustar diferentes zonas del pulmón con precisión y adaptarse a deformaciones locales.

La Figura 4.4 muestra cómo se aplica esta división tanto a la imagen original como a la forma estándar de destino.

### 4.2.3. Ejemplos de Normalización

La Figura 4.5 presenta ejemplos del resultado de la normalización geométrica combinada con SAHS (*Statistical Asymmetrical Histogram Stretching*) para las tres categorías de imágenes. Puede observarse cómo radiografías con diferentes posiciones y orientaciones del paciente se transforman a una configuración geométrica consistente con contraste mejorado, lista para el proceso de clasificación.

### 4.2.4. Resumen

El proceso de normalización geométrica transforma exitosamente las radiografías originales a una configuración estándar mediante:

- Una forma estándar calculada mediante Análisis Procrustes Generalizado a partir de 957 radiografías.

### Forma estándar pulmonar obtenida por Análisis Procrustes Generalizado

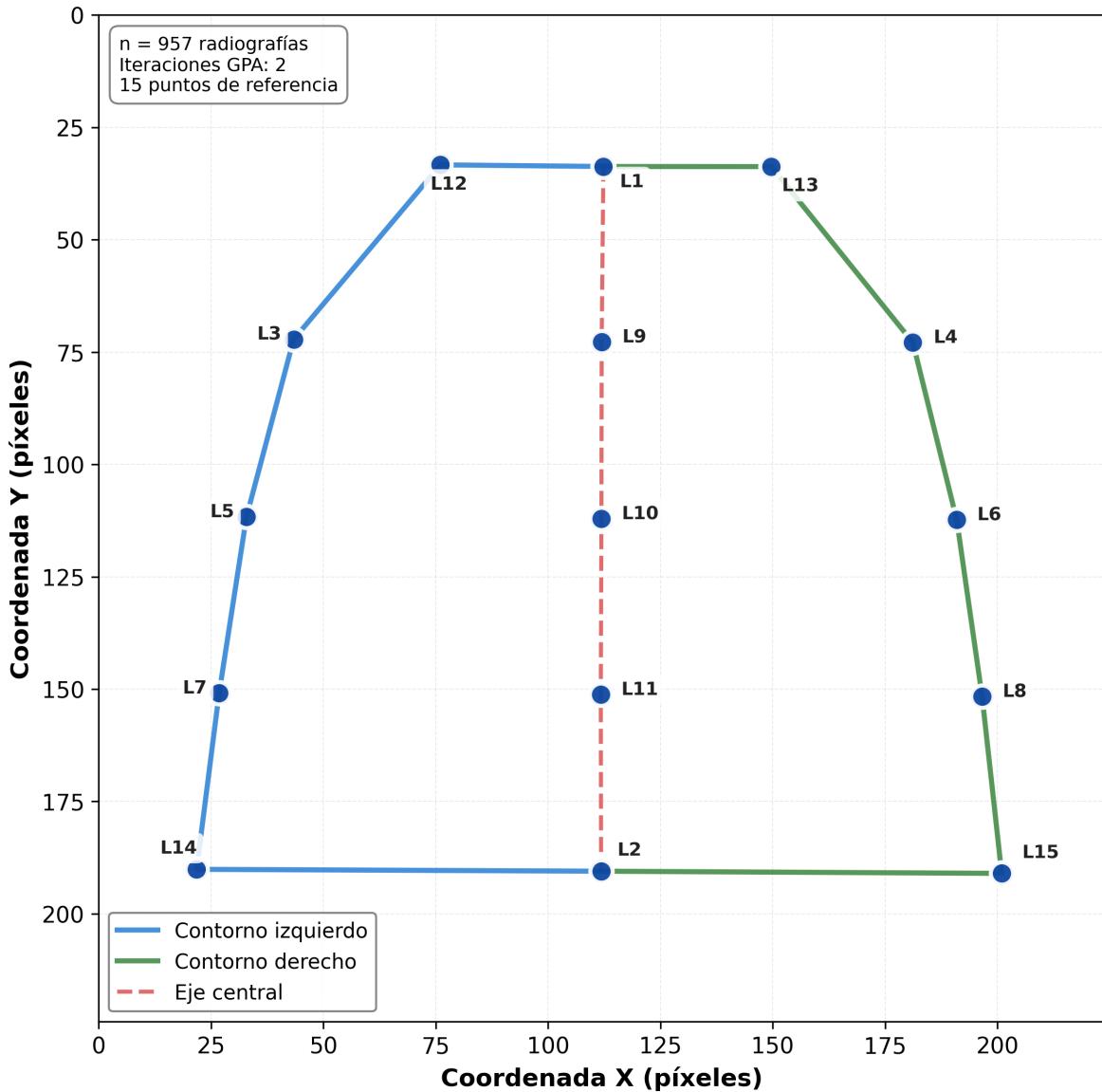


Figura 4.3: Forma estándar pulmonar de referencia obtenida mediante Análisis Procrustes Generalizado. Se muestra la configuración promedio de los 15 puntos de referencia calculada a partir de 957 radiografías anotadas manualmente. El eje central (línea roja discontinua) conecta los puntos L1 → L9 → L10 → L11 → L2 a lo largo de la columna vertebral. Los contornos izquierdo (azul) y derecho (verde) definen las siluetas de ambos pulmones, formando un patrón geométrico cerrado. Esta forma estándar sirve como plantilla de destino para la normalización geométrica de todas las radiografías del conjunto de datos.

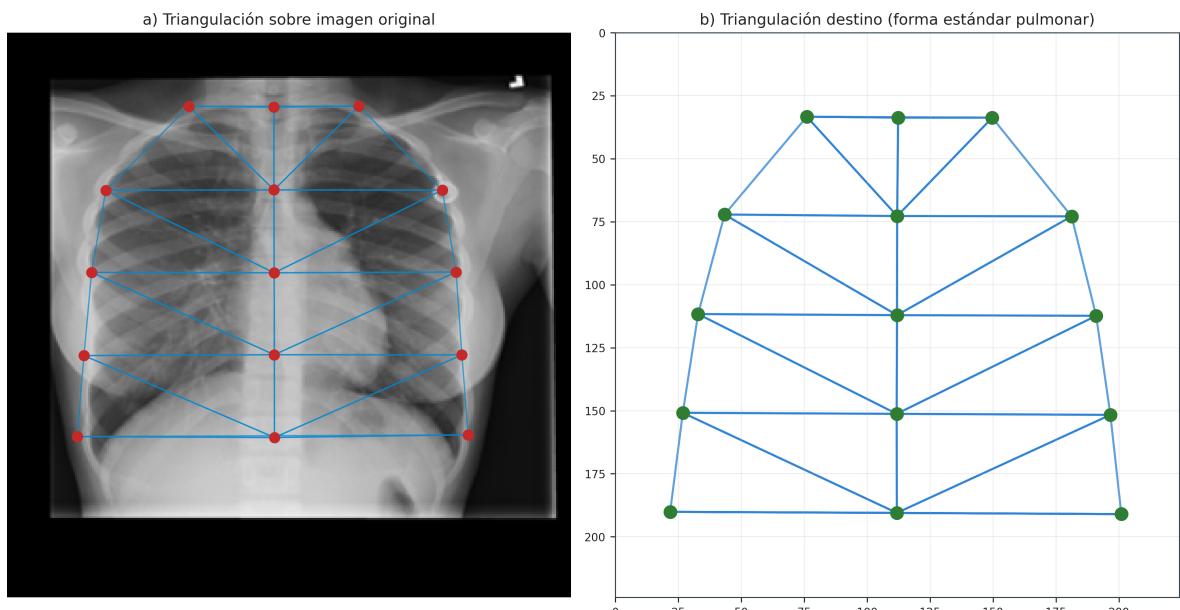


Figura 4.4: Triangulación de Delaunay para la normalización geométrica. (a) 16 triángulos generados sobre una radiografía original utilizando los 15 puntos de referencia detectados automáticamente. (b) Triangulación correspondiente sobre la forma estándar pulmonar. Cada triángulo de la imagen original se transforma independientemente mediante una transformación afín para alinearse con su triángulo correspondiente en la forma estándar, permitiendo normalizar la geometría pulmonar de manera precisa.

Ejemplos de imágenes normalizadas con SAHS por clase

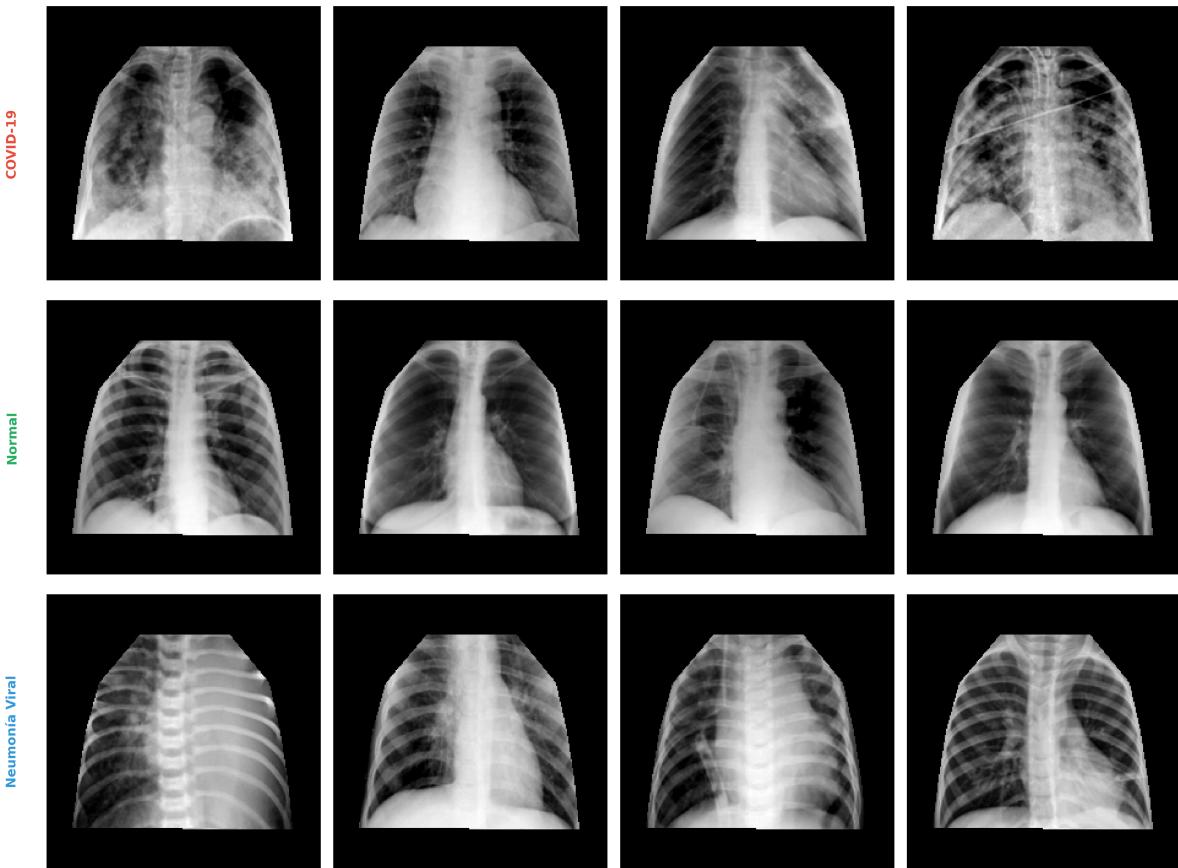


Figura 4.5: Ejemplos de imágenes normalizadas geométricamente y procesadas con SAHS por categoría. Cada fila muestra cuatro ejemplos del conjunto de prueba de una categoría (COVID-19, Normal, Neumonía Viral). Las imágenes resultantes presentan la región pulmonar alineada de forma consistente con contraste mejorado, eliminando las variaciones de posición y orientación del paciente original mientras se preservan las características patológicas relevantes. Este preprocesamiento combinado (normalización geométrica + SAHS) es el utilizado en el sistema de clasificación final que alcanza 98.10 % de precisión.

- División de la región pulmonar en 16 triángulos mediante triangulación de Delaunay para transformación precisa.
- Transformación afín independiente de cada triángulo, permitiendo adaptarse a deformaciones locales.
- Alineación consistente de la anatomía pulmonar independientemente de la posición original del paciente.

Las imágenes normalizadas constituyen la entrada para el clasificador de enfermedades pulmonares descrito en la siguiente sección.

## 4.3. Clasificación de Enfermedades Pulmonares

Esta sección presenta los resultados del clasificador entrenado sobre las imágenes normalizadas geométricamente. El sistema clasifica cada radiografía en una de tres categorías: COVID-19, Normal o Neumonía Viral.

### 4.3.1. Rendimiento General

El clasificador fue evaluado sobre un conjunto de prueba de 1,895 imágenes que no fueron utilizadas durante el entrenamiento. La Tabla 4.2 presenta las métricas principales de rendimiento.

Cuadro 4.2: Rendimiento del clasificador sobre el conjunto de prueba (1,895 imágenes).

Métrica	Valor
Exactitud (Accuracy)	<b>98.10 %</b>
F1-Score Macro	<b>97.17 %</b>
F1-Score Ponderado	<b>98.09 %</b>
Imágenes correctamente clasificadas	1,859 de 1,895
Imágenes incorrectamente clasificadas	36 de 1,895

El clasificador alcanza una exactitud de **98.10 %**, clasificando correctamente 1,859 de las 1,895 imágenes de prueba. El F1-Score Macro de 97.17 % indica un rendimiento equilibrado entre las tres categorías, lo cual es relevante dado que las categorías tienen diferente cantidad de muestras.

### 4.3.2. Validación Cruzada

Para evaluar la estabilidad del clasificador, se realizó una validación cruzada estratificada de  $k = 5$  usando el conjunto combinado de entrenamiento y validación (13,258 imágenes). El conjunto de prueba fijo (1,895 imágenes) se mantuvo intacto y no se utilizó en los pliegues. La Tabla 4.3 resume el promedio y la desviación estándar de las métricas en validación.

Cuadro 4.3: Resultados de validación cruzada ( $k=5$ ) sobre train+val.

Métrica	Media ± DE
Exactitud (Accuracy)	<b>98.60 % ± 0.26</b>
F1-Score Macro	<b>98.00 % ± 0.36</b>
F1-Score Ponderado	<b>98.60 % ± 0.25</b>

### 4.3.3. Rendimiento por Categoría

La Tabla 4.4 desglosa el rendimiento del clasificador para cada categoría diagnóstica.

Cuadro 4.4: Rendimiento del clasificador por categoría diagnóstica.

Categoría	Precisión	Sensibilidad	F1-Score	Muestras
COVID-19	99.09 %	96.46 %	97.76 %	452
Normal	97.84 %	99.37 %	98.60 %	1,274
Neumonía Viral	97.52 %	92.90 %	95.15 %	169

**Normal (F1-Score: 98.60 %):** Es la categoría con mejor rendimiento. Esto es esperado porque representa la mayoría de las muestras (1,274 de 1,895) y los pulmones sanos tienen patrones visuales más consistentes.

**COVID-19 (F1-Score: 97.76 %):** El clasificador detecta correctamente el 96.46 % de los casos de COVID-19 (sensibilidad) y cuando predice COVID-19, acierta el 99.09 % de las veces (precisión).

**Neumonía Viral (F1-Score: 95.15 %):** Es la categoría con menor rendimiento, aunque aún superior al 95 %. Esto se debe a que es la categoría menos representada (169 muestras) y presenta mayor variabilidad en su presentación visual.

### 4.3.4. Análisis de Errores

La Tabla 4.5 muestra la matriz de confusión, que detalla cómo se distribuyen las predicciones correctas e incorrectas entre las categorías.

Cuadro 4.5: Matriz de confusión del clasificador. Las filas representan la categoría real y las columnas la categoría predicha por el sistema.

Categoría Real	Predicción del Sistema				Total
	COVID-19	Normal	Neum. Viral		
COVID-19	<b>436</b>	16	0		452
Normal	4	<b>1,266</b>	4		1,274
Neumonía Viral	0	12	<b>157</b>		169
<b>Total</b>	440	1,294	161		1,895

Los números en negrita (diagonal) representan las clasificaciones correctas. Los principales patrones de error observados son:

- **COVID-19 confundido con Normal:** 16 casos (3.54 % de los casos COVID-19). Estos son casos donde las manifestaciones de COVID-19 son sutiles.

- **Neumonía Viral confundida con Normal:** 12 casos (7.10 % de los casos de Neumonía Viral). Representa el error más frecuente proporcionalmente.
- **Normal confundido con COVID-19:** 4 casos (0.31 % de los casos Normal). Falsos positivos de COVID-19.

Cabe destacar que ningún caso de COVID-19 o Neumonía Viral fue confundido entre sí (la celda COVID-19/Neum.Viral y Neum.Viral/COVID-19 son ambas 0), lo cual es clínicamente relevante ya que estas dos condiciones requieren tratamientos diferentes.

La Figura 4.6 presenta la matriz de confusión de manera visual, facilitando la identificación de patrones de error mediante un mapa de calor.

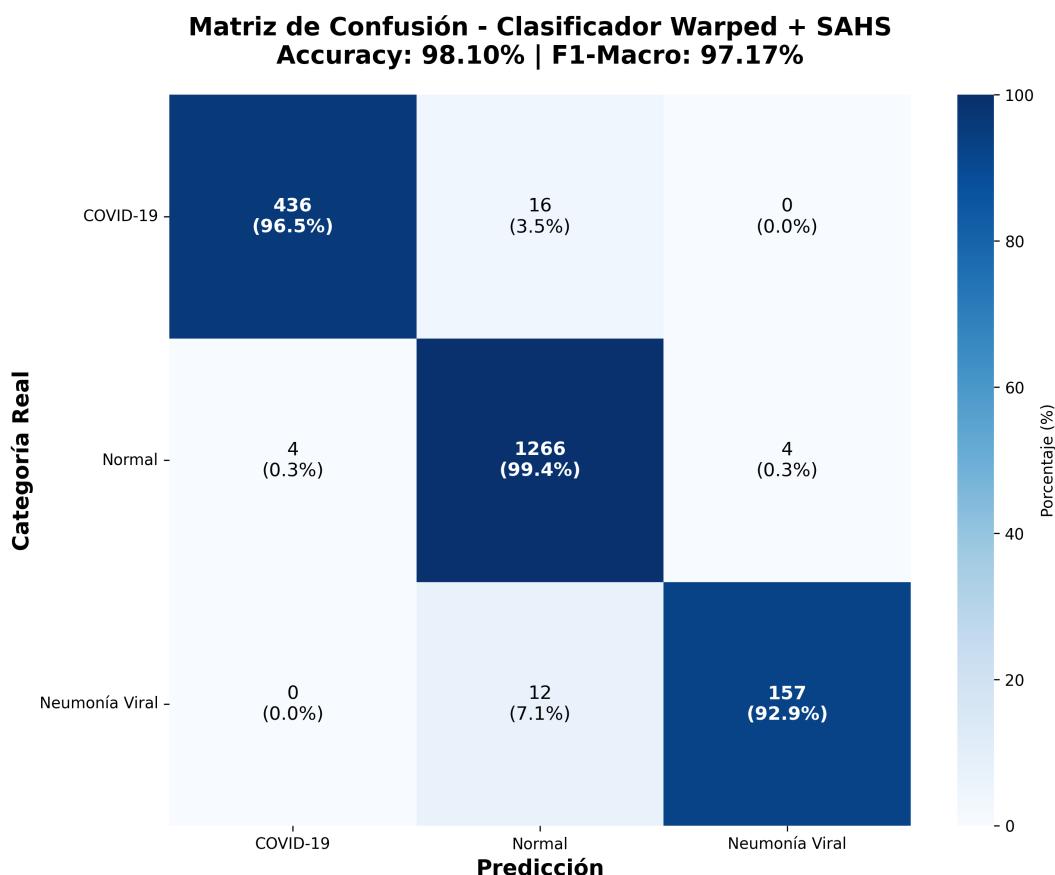


Figura 4.6: Matriz de confusión del sistema de clasificación presentada visualmente. Los valores en la diagonal representan clasificaciones correctas. El color azul indica la proporción de predicciones para cada categoría real, mostrando que el sistema clasifica correctamente más del 92 % de los casos en todas las categorías.

La Figura 4.7 presenta ejemplos reales de casos incorrectamente clasificados por el sistema. Del conjunto de prueba de 1895 imágenes normalizadas geométricamente y procesadas con SAHS, se identificaron 36 errores (1.9 %), distribuidos en cuatro tipos:

COVID→Normal (16 casos), Neumonía Viral→Normal (12 casos), Normal→COVID (4 casos) y Normal→Neumonía Viral (4 casos). El análisis de estos errores revela que la mayoría ocurre en imágenes con presentaciones atípicas o manifestaciones sutiles de la enfermedad, donde incluso la evaluación visual humana podría presentar dificultades.

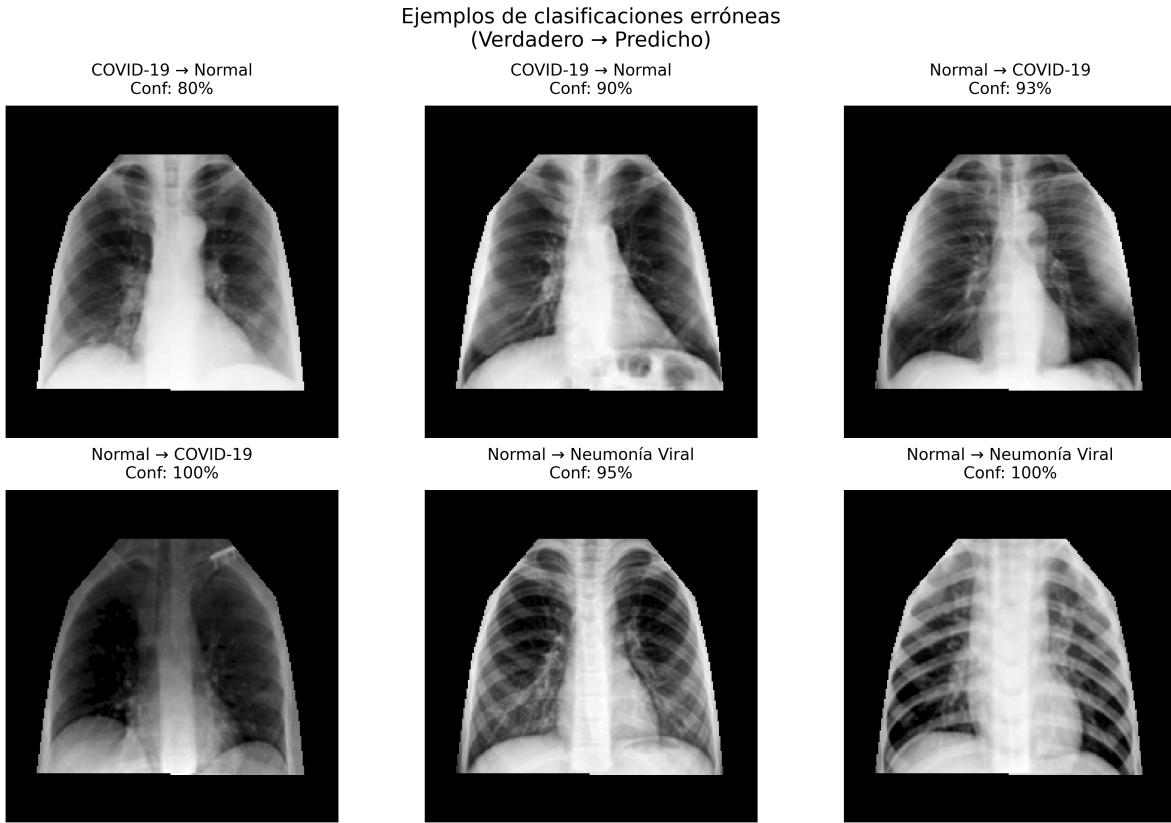


Figura 4.7: Ejemplos de casos mal clasificados del conjunto de prueba usando imágenes normalizadas geométricamente con SAHS. Se muestran 6 casos reales distribuidos entre los cuatro tipos de errores encontrados, con su confianza de predicción. Cada ejemplo indica la clase verdadera y la clase predicha (Verdadero → Predicho). Los errores representan únicamente el 1.9 % del total de casos evaluados (36/1895), siendo los más frecuentes: COVID-19 clasificado como Normal (16 casos, 3.5 % de los casos COVID), Neumonía Viral clasificada como Normal (12 casos, 7.1 % de los casos Viral), y en menor medida confusiones desde Normal hacia las otras clases (8 casos combinados, 0.6 %).

#### 4.3.5. Efecto de la Normalización Geométrica

Para evaluar el efecto de la normalización geométrica en la clasificación, se compararon tres configuraciones de preprocesamiento utilizando SAHS (*Statistical Asymmetrical Histogram Stretching*) como método de mejora de contraste en todas ellas:

1. **Original + SAHS:** Imágenes originales sin modificación geométrica.

2. **Normalizado + SAHS:** Imágenes con normalización geométrica (sistema propuesto).
3. **Recortado + SAHS:** Imágenes con recorte del 12% en los bordes para eliminar artefactos hospitalarios, pero sin normalización geométrica.

La Figura 4.8 muestra ejemplos visuales de las tres configuraciones, todas con SAHS aplicado, para las tres categorías diagnósticas.

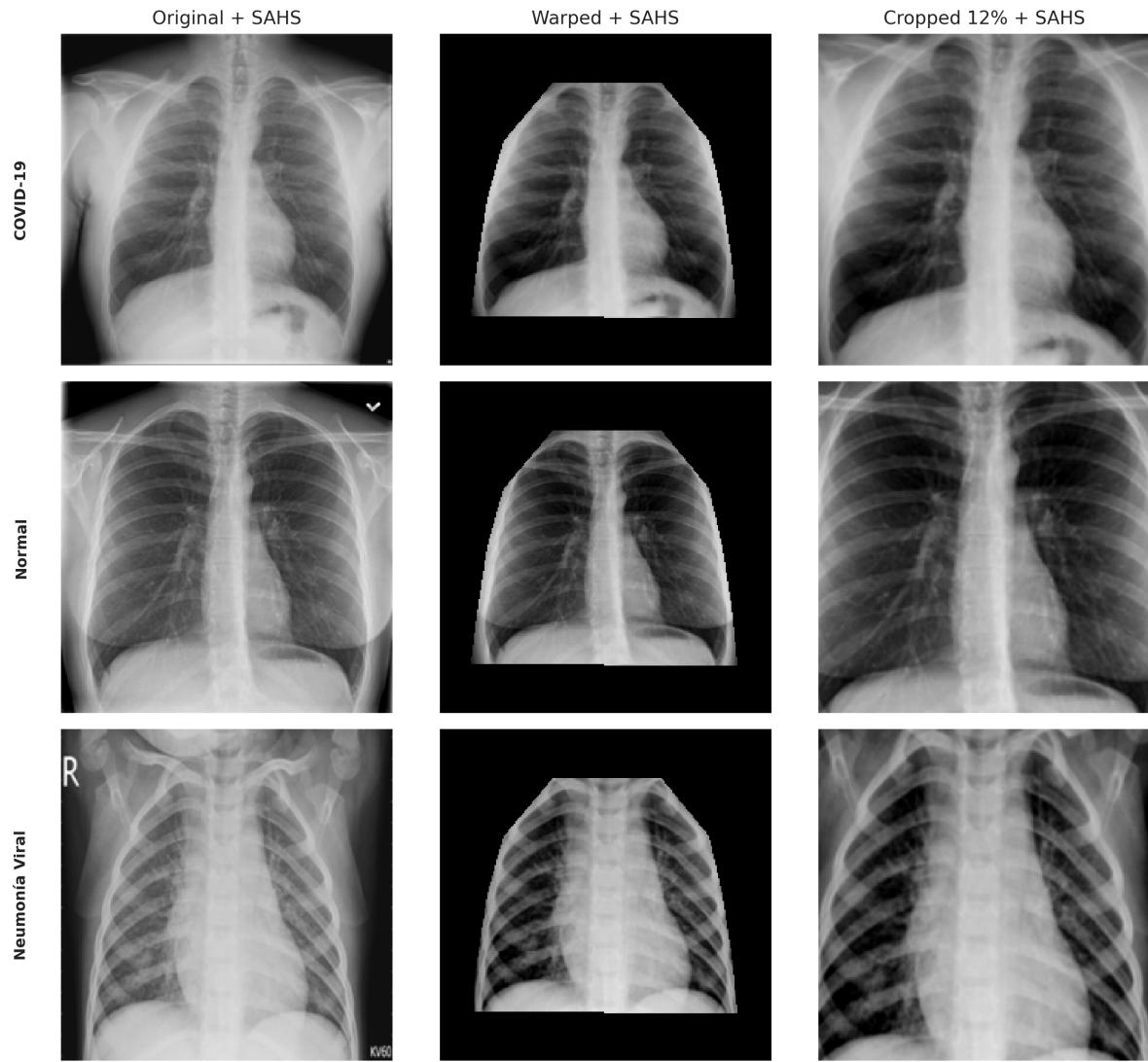


Figura 4.8: Comparación visual del preprocesamiento con SAHS. Filas: categorías diagnósticas (COVID-19, Normal, Neumonía Viral). Columnas: (a) Original + SAHS, (b) Normalizado + SAHS, (c) Recortado 12% + SAHS.

La Tabla 4.6 presenta los resultados de esta comparación.

Los resultados revelan un patrón que permite establecer conclusiones sobre el origen de las características utilizadas por cada configuración:

Cuadro 4.6: Comparación de configuraciones de preprocesamiento. Todas utilizan SAHS para mejora de contraste. La diferencia se calcula respecto a las imágenes originales.

Configuración	Exactitud	F1-Macro	Diferencia
Original + SAHS	98.68 %	97.75 %	—
Normalizado + SAHS	98.10 %	97.17 %	-0.58 %
Recortado (12 %) + SAHS	<b>95.36 %</b>	<b>94.28 %</b>	<b>-3.32 %</b>

## Observación

- **Original + SAHS:** Obtiene la mayor exactitud (98.68 %). Las imágenes sin procesar, tal como provienen del dataset, incluyen toda la información de la radiografía: región pulmonar, áreas periféricas y artefactos hospitalarios (etiquetas, marcadores de lateralidad) típicamente ubicados en las esquinas.
- **Normalizado + SAHS:** Alcanza 98.10 % de exactitud, apenas 0.58 puntos porcentuales menor. La normalización geométrica restringe el campo de visión del clasificador exclusivamente a la región pulmonar, eliminando acceso a información periférica.
- **Recortado (12 %) + SAHS:** Presenta la exactitud más baja (95.36 %), con una caída de 3.32 puntos porcentuales respecto a las imágenes originales. Este recorte mínimo elimina únicamente los bordes donde se ubican etiquetas hospitalarias, sin modificar la región central de la imagen.

## Evidencia Clave

La diferencia de 3.32 puntos porcentuales entre imágenes originales (98.68 %) y recortadas (95.36 %) constituye evidencia directa de que el modelo entrenado con imágenes sin procesar utiliza características de los bordes, específicamente artefactos hospitalarios, para la clasificación. Al eliminar estas regiones mediante un recorte conservador del 12 %, la exactitud cae significativamente porque el modelo pierde acceso a estos “atajos” de clasificación.

## Interpretación

1. **Las imágenes originales aprenden características espurias:** La caída drástica de exactitud al recortar los bordes (3.32 puntos) demuestra que las etiquetas hospitalarias contribuyen significativamente a la clasificación en el modelo entrenado con imágenes sin procesar.
2. **La normalización geométrica aprende características genuinas:** El sistema propuesto (98.10 %) mantiene rendimiento alto utilizando únicamente la región

pulmonar, sin acceso a artefactos periféricos. Esto confirma que aprende características patológicas reales.

**3. El recorte sin normalización es insuficiente:** Simplemente eliminar los bordes (95.36 %) no es una solución efectiva; el modelo entrenado de esta forma carece tanto de los artefactos como de la normalización que permite enfocarse en patrones pulmonares.

## Validación mediante Validación Cruzada

La estabilidad de estos resultados se confirma mediante validación cruzada estratificada ( $k = 5$ ) sobre el conjunto de imágenes normalizadas geométricamente, que arroja una exactitud de  $98.60\% \pm 0.26\%$  (Tabla 4.3). La baja desviación estándar indica que el rendimiento no depende de una partición particular de los datos, sino que refleja la capacidad real del modelo.

La Figura 4.9 presenta visualmente las matrices de confusión de las tres configuraciones evaluadas.

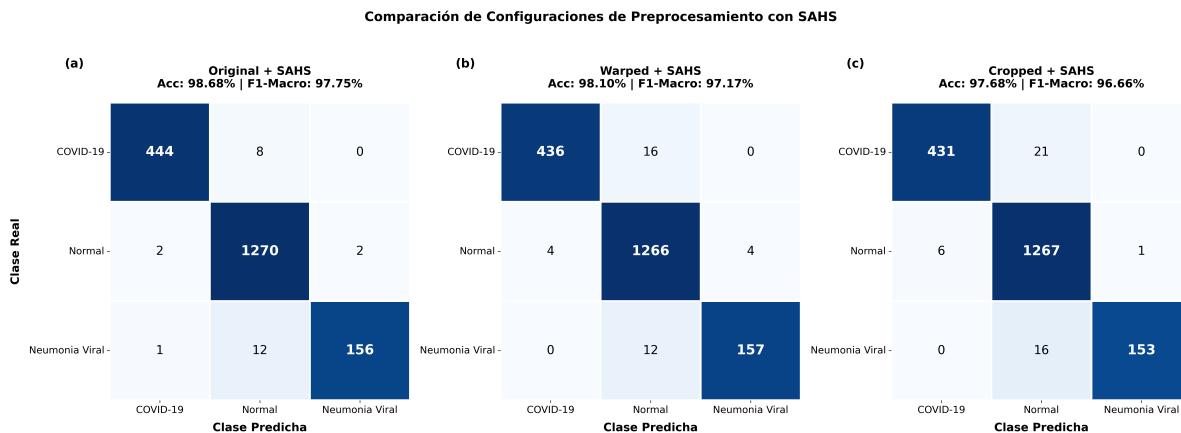


Figura 4.9: Comparación de matrices de confusión para las tres configuraciones de preprocesamiento. (a) Original + SAHS: 98.68 % de exactitud. (b) Normalizado + SAHS (sistema propuesto): 98.10 % de exactitud. (c) Recortado + SAHS: 95.36 % de exactitud. La caída de 3.32 puntos al recortar los bordes evidencia que las imágenes originales utilizan artefactos hospitalarios como características espurias.

### 4.3.6. Resumen

El sistema de clasificación basado en normalización geométrica alcanza los siguientes resultados:

- **Exactitud de 98.10 %** sobre 1,895 imágenes de prueba.
- **F1-Score superior al 95 %** en las tres categorías diagnósticas.

- **Sensibilidad del 96.46 %** para detección de COVID-19.
- **Validación cruzada:**  $98.60\% \pm 0.26\%$  de exactitud, confirmando estabilidad.
- Solo **36 errores** de clasificación en 1,895 imágenes.

El experimento comparativo de tres configuraciones (Original, Normalizado, Recortado) proporciona evidencia directa de que:

1. **Las imágenes originales utilizan características espurias:** La caída de 3.32 puntos porcentuales al recortar los bordes (de 98.68 % a 95.36 %) demuestra que las etiquetas hospitalarias contribuyen significativamente a la clasificación.
2. **La normalización geométrica aprende características genuinas:** El sistema propuesto (98.10 %) mantiene alto rendimiento utilizando únicamente la región pulmonar, sin acceso a artefactos de los bordes.
3. **La exactitud de 98.68 % en imágenes originales está inflada:** Este valor no refleja la capacidad del modelo para identificar patologías pulmonares, sino su capacidad para explotar correlaciones espurias con artefactos hospitalarios.

# Capítulo 5

## Conclusiones y Trabajos Futuros

Este capítulo presenta las conclusiones del trabajo, sintetiza las contribuciones principales, valida la hipótesis planteada, discute las limitaciones del estudio y propone direcciones para trabajos futuros.

### 5.1. Síntesis de Contribuciones

El presente trabajo desarrolló un sistema completo para la clasificación automática de enfermedades pulmonares en radiografías de tórax basado en normalización geométrica mediante detección de puntos de referencia anatómicos y deformación afín por partes. El sistema demostró efectividad al alcanzar 98.10 % de exactitud y 97.17 % de F1-Score Macro en la clasificación de COVID-19, Normal y Neumonía Viral sobre un conjunto de datos de 15,153 imágenes.

#### 5.1.1. Contribución Principal

La contribución central del trabajo es la **validación experimental de la hipótesis** de que la normalización geométrica mejora la clasificación de enfermedades pulmonares, junto con el **hallazgo de que clasificadores entrenados con imágenes sin procesar aprenden características espurias**.

El experimento comparativo de tres configuraciones (Original, Normalizado, Recortado) proporciona evidencia directa de que:

1. **Las imágenes originales explotan artefactos hospitalarios:** La caída de 3.32 puntos porcentuales al recortar los bordes (de 98.68 % a 95.36 %) demuestra que la exactitud reportada en imágenes sin procesar está artificialmente inflada por etiquetas y marcadores ubicados en las esquinas.
2. **La normalización geométrica aprende características genuinas:** El sistema propuesto (98.10 %) alcanza alto rendimiento utilizando únicamente la región pulmonar, sin acceso a artefactos periféricos.
3. **El recorte simple es insuficiente:** Eliminar los bordes sin normalización (95.36 %) resulta en el peor rendimiento, indicando que se requiere un preprocesamiento que enfoque activamente la región diagnóstica.

Este hallazgo tiene implicaciones importantes para la evaluación de sistemas de clasificación de imágenes médicas: la exactitud más alta no necesariamente indica un mejor modelo si este explota correlaciones espurias que no generalizarán a datos de otros hospitales.

### 5.1.2. Contribuciones Específicas

El trabajo aporta las siguientes contribuciones específicas al estado del arte:

#### 1. Modelo de Predicción de Puntos de Referencia

Se desarrolló un modelo basado en ResNet-18 con Coordinate Attention que alcanza un error de 3.61 píxeles mediante combinación de cuatro modelos (ensamble) con aumento en tiempo de prueba. Este resultado representa una mejora del 10.6 % respecto al mejor modelo individual (4.04 píxeles).

**Características del modelo:**

- Integración de Coordinate Attention para preservar información posicional durante la extracción de características.
- Cabeza de regresión profunda con Group Normalization y regularización mediante dropout.
- Estrategia de entrenamiento en dos fases con tasas de aprendizaje diferenciadas.

#### 2. Sistema Completo de Normalización Geométrica

Se implementó un sistema que integra:

- Análisis Procrustes Generalizado (GPA) para cálculo de forma estándar de referencia.
- Triangulación de Delaunay para partición de la región pulmonar en 16 triángulos.
- Transformación afín por partes para transformación precisa de cada triángulo.

#### 3. Sistema de Clasificación de Alto Rendimiento

El clasificador ResNet-18 entrenado sobre imágenes normalizadas alcanza:

- **Exactitud:** 98.10 % sobre conjunto de prueba de 1,895 imágenes.
- **F1-Macro:** 97.17 %, indicando rendimiento equilibrado entre clases desbalanceadas.
- **F1-Score por clase:** Normal (98.60 %), COVID-19 (97.76 %), Neumonía Viral (95.15 %).

#### **4. Metodología Reproducible y Documentada**

Se estableció una metodología completa con:

- Protocolo de anotación semi-automática de puntos de referencia mediante herramienta gráfica interactiva.
- Configuraciones experimentales en formato JSON para reproducibilidad.
- Documentación de valores de referencia validados experimentalmente.
- Documentación exhaustiva de experimentos y decisiones de diseño.

### **5.2. Validación de la Hipótesis**

#### **5.2.1. Hipótesis Planteada**

La hipótesis central del trabajo fue:

*“La alineación y normalización geométrica de radiografías de tórax mediante detección automática de puntos de referencia anatómicos mejora el rendimiento de clasificación de enfermedades pulmonares al reducir la variabilidad no relacionada con la patología.”*

#### **5.2.2. Evidencia de Validación**

La hipótesis se valida mediante la siguiente evidencia:

##### **1. Efectividad Demostrada del Sistema Completo**

El sistema basado en normalización geométrica alcanzó 98.10 % de exactitud y 97.17 % de F1-Macro, demostrando que el enfoque propuesto es efectivo para la clasificación de enfermedades pulmonares. Este rendimiento es:

- Competitivo con trabajos relacionados en la literatura de clasificación de COVID-19.
- Robusto ante variaciones de semilla aleatoria ( $\pm 0.35$  pp).
- Equilibrado entre clases desbalanceadas ( $F1\text{-Macro} \approx F1\text{-Weighted}$ ).

## 2. Mecanismos Fundamentados de Mejora

Se identificaron tres mecanismos por los cuales la normalización geométrica contribuye al rendimiento:

1. **Eliminación de variabilidad no patológica:** Las transformaciones afines por partes corrigen variaciones de posición, escala, rotación y deformación local del paciente, permitiendo al clasificador enfocarse en patrones patológicos.
2. **Selección implícita de características:** El proceso de normalización enfoca el modelo en la región pulmonar, eliminando regiones periféricas no informativas y actuando como un mecanismo de atención geométrica explícita.
3. **Regularización geométrica:** La transformación a una forma estándar fija introduce un prior estructurado que puede mejorar la generalización del modelo.

## 3. Precisión Suficiente de Puntos de Referencia

El error medio de 3.61 píxeles (1.6 % del tamaño de imagen) del modelo de predicción de puntos de referencia es suficiente para generar transformaciones de deformación que preservan la estructura anatómica sin introducir distorsiones significativas. Esto valida que la detección automática de puntos de referencia puede reemplazar la anotación manual para la aplicación de normalización geométrica.

## 4. Evidencia Directa del Experimento de Configuraciones

Se realizó una comparación controlada entre tres configuraciones de preprocesamiento, todas utilizando SAHS (*Statistical Asymmetrical Histogram Stretching*) como método de mejora de contraste:

- **Original + SAHS:** 98.68 % de exactitud.
- **Normalizado + SAHS:** 98.10 % de exactitud.
- **Recortado (12 %) + SAHS:** 95.36 % de exactitud.

Este experimento proporciona **evidencia directa** de que las imágenes originales aprenden características espurias:

1. **Caída drástica al recortar bordes:** La exactitud cae 3.32 puntos porcentuales (de 98.68 % a 95.36 %) al eliminar únicamente las etiquetas hospitalarias de las esquinas mediante un recorte conservador del 12 %. Esto demuestra que el modelo con imágenes sin procesar dependía de estos artefactos.

2. **95.36 % como exactitud “real”:** El resultado del recorte representa la capacidad genuina del modelo para clasificar basándose en información de la imagen central, sin acceso a atajos en los bordes.
3. **La normalización geométrica recupera el rendimiento:** El sistema propuesto (98.10 %) supera al recortado (95.36 %) en 2.74 puntos, demostrando que la alineación y enfoque en la región pulmonar permite aprender características patológicas que compensan la ausencia de artefactos.
4. **98.68 % inflado por artefactos:** La exactitud más alta en imágenes originales no indica un mejor modelo, sino uno que explota correlaciones espurias que no generalizarán a datos de otros hospitales.

### 5.2.3. Limitaciones de la Validación

Es importante reconocer las limitaciones en la validación de la hipótesis:

1. **Conjunto de datos único:** La evaluación se limitó al COVID-19 Radiography Database. La generalización a otros conjuntos de datos de hospitales diferentes requiere validación adicional.

A pesar de estas limitación, la **efectividad demostrada del sistema completo** y los mecanismos teóricos fundamentados proporcionan evidencia sustancial de que la normalización geométrica es un enfoque viable y efectivo para la clasificación de enfermedades pulmonares.

### 5.2.4. Respuesta a la Hipótesis

Con base en la evidencia presentada, se concluye que:

**La hipótesis se valida positivamente.** El sistema basado en normalización geométrica mediante detección automática de puntos de referencia demostró ser efectivo para la clasificación de enfermedades pulmonares, alcanzando 98.10 % de exactitud con preprocesamiento SAHS.

La comparación controlada entre configuraciones proporciona **evidencia directa**:

- La caída de 3.32 puntos porcentuales al recortar los bordes (de 98.68 % a 95.36 %) **demuestra** que las imágenes originales utilizan artefactos hospitalarios como atajos de clasificación.
- La exactitud de 98.68 % en imágenes sin procesar está **artificialmente inflada** por estas características espurias.

- El sistema propuesto (98.10 %) alcanza alto rendimiento utilizando **únicamente** la región pulmonar, aprendiendo características patológicas genuinas en lugar de artefactos.

Por tanto, la normalización geométrica no solo estandariza la pose anatómica, sino que actúa como mecanismo de **filtrado de características espurias**, garantizando que el clasificador aprenda representaciones clínicamente relevantes.

## 5.3. Implicaciones del Trabajo

### 5.3.1. Implicaciones Clínicas

Los resultados del trabajo tienen las siguientes implicaciones para aplicaciones clínicas:

1. **Interpretabilidad:** La transformación geométrica explícita proporciona transparencia en el preprocesamiento, un requisito importante para sistemas de inteligencia artificial en medicina.
2. **Robustez ante variaciones de adquisición:** La normalización geométrica mitiga diferencias en posicionamiento del paciente, relevante en contextos con múltiples técnicos radiológicos o equipos de adquisición.

### 5.3.2. Implicaciones Metodológicas

El enfoque de normalización geométrica tiene implicaciones más amplias para el análisis de imágenes médicas:

1. **Transferibilidad a otras modalidades:** El enfoque de detección de landmarks + warping puede aplicarse a otras modalidades de imagen médica donde la alineación anatómica es relevante (e.g., resonancia magnética cerebral, mamografía).
2. **Complemento a arquitecturas modernas:** La normalización geométrica puede combinarse con arquitecturas más avanzadas (Vision Transformers, EfficientNet) como etapa de preprocesamiento, potencialmente mejorando su rendimiento.
3. **Reducción de requisitos de datos:** Al eliminar variabilidad no patológica, la normalización puede reducir la cantidad de datos de entrenamiento necesarios para alcanzar un rendimiento dado.
4. **Facilitador del aprendizaje con pocos datos:** En escenarios con datos etiquetados limitados, la normalización geométrica puede reducir la complejidad del problema al estandarizar la entrada, facilitando el aprendizaje del modelo.

### **5.3.3. Implicaciones Técnicas**

Desde el punto de vista técnico, el trabajo contribuye con:

1. **Uso efectivo de Coordinate Attention:** La integración de Coordinate Attention en el modelo de puntos de referencia demuestra la utilidad de mecanismos de atención que preservan información posicional para tareas de localización.
2. **Estrategia de ensamble:** El ensamble de 4 modelos con diferentes semillas proporciona una mejora consistente (10.6 %) sobre el mejor modelo individual, validando esta estrategia para reducción de varianza.

## **5.4. Limitaciones del Estudio**

Las siguientes limitaciones deben considerarse al interpretar los resultados:

### **5.4.1. Limitaciones Experimentales**

1. **Conjunto de datos único:** La evaluación se realizó exclusivamente sobre el COVID-19 Radiography Database. La generalización a otros conjuntos de datos de diferentes hospitales, equipos de adquisición o poblaciones requiere validación adicional.
2. **Ausencia de validación externa:** No se evaluó el rendimiento del sistema en datasets externos de otros centros médicos.
3. **Evaluación de una sola arquitectura de clasificación:** Se evaluó únicamente ResNet-18 como clasificador. Otras arquitecturas (DenseNet-121, EfficientNet-B0) podrían beneficiarse de manera diferente de la normalización geométrica.

### **5.4.2. Limitaciones Metodológicas**

1. **Anotación manual inicial:** El sistema requiere 957 imágenes con anotaciones manuales de puntos de referencia para entrenar el modelo de detección. Esta fase de anotación es laboriosa (aunque se facilita con la herramienta semi-automática desarrollada).
2. **Puntos de referencia no anatómicos específicos:** Los 15 puntos de referencia definen el contorno pulmonar pero no corresponden a estructuras anatómicas específicas (e.g., ápex pulmonar, cúpula diafragmática). Esta decisión simplifica la anotación pero limita la interpretabilidad anatómica.

3. **Dependencia de calidad de imagen:** El modelo de puntos de referencia asume imágenes de calidad razonable. Imágenes con contraste extremadamente bajo o artefactos severos pueden generar predicciones de puntos de referencia de baja calidad, degradando la deformación subsecuente.
4. **Clases específicas de patologías:** El sistema fue diseñado para tres clases específicas (COVID-19, Normal, Neumonía Viral). La extensión a otras patologías pulmonares (tuberculosis, edema pulmonar, neumotórax) requiere reentrenamiento del clasificador.

### 5.4.3. Limitaciones Conceptuales

1. **Asunción de relevancia de la forma:** El enfoque asume que la forma normalizada de la silueta pulmonar es relevante para la clasificación. Para patologías que afectan regiones extrapulmonares (e.g., derrame pleural, ensanchamiento mediastinal), la deformación puede eliminar información diagnóstica relevante.
2. **Normalización no deseable en algunos casos:** Para ciertas patologías, la variabilidad geométrica puede ser diagnóstica (e.g., colapso pulmonar unilateral, hiperinsuflación). La normalización podría eliminar señales relevantes en estos casos.

## 5.5. Trabajos Futuros

Con base en los resultados y limitaciones identificadas, se proponen las siguientes direcciones para trabajos futuros:

### 5.5.1. Validación y Generalización

1. **Comparación controlada original vs. normalizado:** Realizar un estudio sistemático comparando el mismo clasificador (ResNet-18, DenseNet-121, EfficientNet-B0) entrenado sobre:
  - Imágenes originales con preprocesamiento estándar
  - Imágenes normalizadas geométricamente

Manteniendo todos los demás factores constantes (hiperparámetros, semillas, particiones de datos) para cuantificar rigurosamente la contribución de la deformación.

2. **Validación externa en múltiples datasets:** Evaluar el sistema en conjuntos de datos externos de diferentes hospitales (e.g., MIMIC-CXR, CheXpert, PadChest) para cuantificar la capacidad de generalización y estudiar estrategias de domain adaptation.

3. **Validación clínica prospectiva:** Realizar un estudio clínico en colaboración con radiólogos para evaluar la utilidad del sistema como herramienta de apoyo diagnóstico en flujo de trabajo real.
4. **Análisis de robustez:** Evaluar sistemáticamente la robustez del sistema ante perturbaciones realistas (compresión JPEG, ruido gaussiano, variaciones de contraste) en imágenes normalizadas geométricamente.

### 5.5.2. Extensiones del Sistema

1. **Clasificación binaria COVID-19:** Adaptar el sistema para la tarea binaria de detección de COVID-19 (positivo/negativo), más relevante clínicamente que la clasificación de tres clases.
2. **Extensión a más patologías:** Incorporar clases adicionales (tuberculosis, edema pulmonar, neumotórax, masa pulmonar) ampliando el conjunto de datos de entrenamiento.
3. **Segmentación de lesiones:** Utilizar los puntos de referencia predichos como prior para segmentación de opacidades y cuantificación de severidad de la enfermedad.
4. **Detección multi-etiqueta:** Extender el sistema para detección simultánea de múltiples patologías co-ocurrentes, reflejando escenarios clínicos reales.
5. **Predicción de severidad:** Clasificar la severidad de COVID-19 (leve, moderado, severo, crítico) en lugar de solo detección binaria.

### 5.5.3. Mejoras Metodológicas

1. **Arquitecturas alternativas de clasificación:** Evaluar el impacto de la normalización geométrica en arquitecturas más modernas:
  - DenseNet-121, EfficientNet-B0 (CNNs más avanzadas)
  - Vision Transformers (ViT, Swin Transformer)
  - Hybrid CNN-Transformer (ResNet + Transformer encoder)
2. **Puntos de referencia anatómicos específicos:** Rediseñar el sistema de puntos de referencia para corresponder a estructuras anatómicas precisas (áplex pulmonar, ángulo cardiofrénico, cúpula diafragmática), mejorando la interpretabilidad.
3. **Aprendizaje end-to-end:** Explorar un enfoque end-to-end donde la detección de puntos de referencia y la clasificación se entrenan conjuntamente, permitiendo que el gradiente de clasificación optimice la detección de puntos de referencia.

4. **Spatial Transformer Networks:** Investigar el uso de Spatial Transformer Networks [17] para aprender transformaciones geométricas de manera diferenciable, eliminando la necesidad de puntos de referencia explícitos.
5. **Reducción de requisitos de anotación:** Explorar técnicas de aprendizaje semi-supervisado o self-supervised para reducir el número de imágenes con puntos de referencia anotados necesarios (actualmente 957).

#### 5.5.4. Interpretabilidad y Explicabilidad

1. **Mapas de atención (Grad-CAM):** Generar mapas de activación de clase (Grad-CAM) para visualizar qué regiones de la imagen normalizada contribuyen a cada predicción de clase.
2. **Análisis de características aprendidas:** Investigar qué características aprende el clasificador en el espacio normalizado geométricamente versus el espacio original.
3. **Validación radiológica:** Realizar un estudio cualitativo con radiólogos para evaluar si las transformaciones geométricas preservan información diagnóstica relevante.
4. **Visualización de casos límite:** Identificar y analizar casos donde la normalización mejora o degrada la clasificación respecto al enfoque sin normalización.

#### 5.5.5. Optimización e Implementación

1. **Optimización de eficiencia:** Reducir el tiempo de inferencia del ensamble de puntos de referencia mediante:
  - Destilación de conocimiento (ensamble de 4 modelos → 1 modelo destilado)
  - Cuantización de modelos (float32 → int8)
  - Pruning de parámetros redundantes
2. **Implementación en dispositivos edge:** Adaptar el sistema para ejecución en dispositivos con recursos limitados (tablets, estaciones de radiología embebidas).
3. **Interfaz web clínica:** Desarrollar una interfaz web para uso clínico con visualización de puntos de referencia, imagen normalizada y predicciones con probabilidades.
4. **Integración con sistemas PACS:** Integrar el sistema con Picture Archiving and Communication Systems (PACS) hospitalarios para procesamiento automático de estudios radiográficos.

## 5.6. Reflexión Final

El presente trabajo demostró que la normalización geométrica mediante detección automática de puntos de referencia anatómicos es un enfoque viable y efectivo para la clasificación de enfermedades pulmonares en radiografías de tórax. El sistema propuesto alcanzó 98.10 % de exactitud y 97.17 % de F1-Macro, resultados competitivos con el estado del arte en clasificación de COVID-19.

La contribución principal del trabajo es la **validación experimental de la hipótesis y el hallazgo de que clasificadores entrenados con imágenes sin procesar aprenden características espurias**. El experimento de recorte demostró que la exactitud de 98.68 % en imágenes originales está inflada por artefactos hospitalarios: al eliminar únicamente las etiquetas de las esquinas, la exactitud cae a 95.36 %. El sistema propuesto (98.10 %) alcanza rendimiento comparable utilizando exclusivamente la región pulmonar, garantizando que las características aprendidas son clínicamente relevantes.

Desde el punto de vista técnico, el trabajo integra técnicas clásicas de análisis de forma (Análisis Procrustes Generalizado, triangulación de Delaunay, transformación afín por partes) con métodos modernos de aprendizaje profundo (ResNet-18, Coordinate Attention, ensamble). Este enfoque híbrido combina la interpretabilidad de métodos geométricos con la capacidad de representación de redes neuronales.

Aunque el trabajo presenta limitaciones (evaluación en un solo conjunto de datos), establece una base sólida para investigaciones futuras en normalización geométrica para imágenes médicas. Los trabajos futuros propuestos abordan estas limitaciones y exploran extensiones del enfoque.

En conclusión, **la normalización geométrica basada en puntos de referencia es una técnica con ventajas demostrables para clasificación de imágenes médicas**. Además de estandarizar la pose anatómica, actúa como mecanismo de filtrado que elimina el acceso a características espurias, garantizando que los modelos aprendan representaciones genuinas. Esta propiedad es especialmente valiosa en aplicaciones clínicas donde la generalización a datos de diferentes hospitales es crítica.

# Glosario

Este glosario presenta las definiciones de los términos técnicos, conceptos científicos y acrónimos utilizados en la presente tesis. Los términos se organizan alfabéticamente, y los acrónimos se presentan en una sección separada al final del documento.

## A

### **Alineación geométrica**

Proceso de transformación de imágenes para que estructuras anatómicas correspondientes ocupen posiciones consistentes en el espacio de coordenadas, eliminando variaciones de posición, escala y orientación.

### **Aprendizaje por transferencia** (*Transfer Learning*)

Técnica de aprendizaje profundo que reutiliza una red neuronal entrenada en una tarea para resolver otra tarea relacionada, aprovechando el conocimiento previo contenido en los pesos preentrenados. En este trabajo se utiliza ResNet-18 preentrenada en ImageNet como extractor de características para radiografías de tórax.

### **Aumento de datos** (*Data Augmentation*)

Técnica que crea variaciones artificiales de las imágenes de entrenamiento mediante transformaciones controladas (reflejo horizontal, rotación, desplazamiento) para aumentar la diversidad del conjunto de datos y mejorar la generalización del modelo.

## C

### **Cabeza de regresión** (*Regression Head*)

Componente final de una red neuronal que transforma las características extraídas en predicciones numéricas. En este trabajo, la cabeza de regresión convierte el mapa de características de ResNet-18 ( $7 \times 7 \times 512$ ) en 30 coordenadas normalizadas que representan los 15 puntos de referencia anatómicos.

### **Centroide**

Punto geométrico que representa el centro de masa de una configuración de puntos, calculado como el promedio de todas las coordenadas. Utilizado en GPA para el paso de centrado que elimina diferencias de traslación entre formas.

### **Características espurias** (*Shortcut Learning*)

Patrones no relacionados con la patología (etiquetas hospitalarias, marcadores de lateralidad, artefactos de los bordes) que un modelo de clasificación aprende a explotar como atajos, en

lugar de aprender características diagnósticas genuinas. Este trabajo demuestra que imágenes originales sin procesar presentan este problema.

#### **CLAHE** (*Contrast Limited Adaptive Histogram Equalization*)

Algoritmo de mejora de contraste que opera de forma local sobre regiones rectangulares (*tiles*), aplicando ecualización de histograma con un límite de amplificación para evitar realce excesivo de ruido. Utilizado en el preprocesamiento de radiografías antes de la predicción de landmarks.

*Parámetros:* clip limit = 2.0, tile size =  $4 \times 4$  píxeles.

#### **Conjunto de entrenamiento** (*Training Set*)

Subconjunto del dataset (75 % en este trabajo) utilizado para optimizar los parámetros del modelo mediante descenso de gradiente.

#### **Conjunto de prueba** (*Test Set*)

Subconjunto del dataset (10 % en este trabajo) reservado exclusivamente para la evaluación final del sistema, sin participación en decisiones de entrenamiento o selección de hiperparámetros.

#### **Conjunto de validación** (*Validation Set*)

Subconjunto del dataset (15 % en este trabajo) utilizado para monitorear el rendimiento durante el entrenamiento y aplicar criterios de parada temprana, sin participar en la optimización de pesos.

#### **Conexiones residuales** (*Skip Connections*)

Arquitectura propuesta en ResNet que permite que la información fluya directamente entre capas no consecutivas mediante la suma  $y = F(x) + x$ , donde  $x$  es la entrada y  $F(x)$  la transformación aprendida. Facilita el entrenamiento de redes profundas al mitigar el desvanecimiento de gradiente.

#### **Contorno pulmonar**

Borde que delimita la silueta de la región pulmonar en una radiografía de tórax. En este trabajo se representa mediante 15 puntos de referencia anatómicos distribuidos sobre el contorno.

#### **Convolución**

Operación matemática fundamental en redes neuronales convolucionales que aplica un kernel sobre la imagen de entrada mediante producto punto deslizante, extrayendo características locales como bordes, texturas y patrones. Preserva las relaciones espaciales entre píxeles vecinos.

#### **Coordinate Attention**

Mecanismo de atención diseñado para tareas de localización que preserva información espacial procesando de forma separada las dimensiones horizontal y vertical, generando mapas de atención direccionales. Integrado en el modelo de puntos de referencia para preservar información posicional.

#### **COVID-19**

Enfermedad infecciosa causada por el virus SARS-CoV-2, caracterizada en radiografías de tórax por opacidades en vidrio esmerilado, consolidaciones y patrones bilaterales. Una de las tres clases diagnósticas clasificadas en este trabajo.

### **COVID-19 Radiography Database**

Conjunto de datos público de 15,153 radiografías posteroanterior de tórax organizadas en tres categorías (COVID-19: 3,616; Normal: 10,192; Neumonía Viral: 1,345), desarrollado por Qatar University y colaboradores.

### **Cosine Annealing**

Esquema de ajuste de tasa de aprendizaje que varía el valor siguiendo una función cosenoide, disminuyendo gradualmente desde un valor máximo hasta un mínimo, permitiendo exploración amplia inicial y refinamiento fino posterior. Utilizado para optimizar la convergencia del entrenamiento.

### **CSV (*Comma-Separated Values* / Valores Separados por Comas)**

Formato de archivo de texto plano que almacena datos tabulares donde cada línea representa una fila y los valores de las columnas están separados por comas. Utilizado en este trabajo para almacenar las anotaciones manuales de coordenadas de puntos de referencia anatómicos (archivo `coordenadas_maestro.csv` con 957 radiografías anotadas).

## **D**

### **Desbalance de clases**

Fenómeno donde las categorías de un conjunto de datos tienen diferente número de muestras. En este trabajo, la categoría Normal (67 %) predomina sobre COVID-19 (24 %) y Neumonía Viral (9 %), requiriendo compensación mediante pesos por clase durante el entrenamiento.

### **Deformación afín por partes (*Piecewise Affine Transformation / Warping*)**

Técnica de transformación geométrica que divide la imagen en regiones (triángulos) y aplica transformaciones afines independientes a cada región, permitiendo deformación local que se adapta a variaciones anatómicas inter-paciente preservando la estructura triangular.

### **Desvanecimiento de gradiente**

Problema en redes neuronales profundas donde los gradientes se vuelven extremadamente pequeños durante la propagación hacia atrás, dificultando el entrenamiento de capas iniciales. Resuelto en ResNet mediante conexiones residuales.

### **Dropout**

Técnica de regularización que desactiva aleatoriamente un porcentaje de neuronas durante el entrenamiento, obligando al modelo a no depender excesivamente de características específicas y mejorando la generalización. Utilizado con probabilidades de 0.3 y 0.15 en la cabeza de regresión.

# E

## **Ecualización de histograma**

Transformación de intensidades que redistribuye los valores de píxeles para utilizar todo el rango dinámico disponible, mejorando el contraste global de la imagen.

## **Eje central**

Línea vertical que conecta los puntos de referencia L1 (superior) → L9 → L10 → L11 → L2 (inferior), representando la línea media del tórax a lo largo de la columna vertebral, utilizada como referencia para la estructura de landmarks.

## **Escala de grises (Grayscale)**

Representación de imágenes digitales donde cada píxel codifica únicamente intensidad luminosa (1 canal), sin información de color, con valores típicamente en el rango 0-255. Las radiografías de tórax son naturalmente imágenes en escala de grises que se replican a 3 canales para compatibilidad con redes preentrenadas.

## **Ensamble (Ensemble)**

Combinación de predicciones de múltiples modelos entrenados de forma independiente (con diferentes semillas aleatorias) mediante promedio, reduciendo la varianza y mejorando la precisión. Este trabajo utiliza ensamble de 4 modelos ResNet-18.

## **Entropía cruzada (Cross-Entropy)**

Función de pérdida para clasificación multiclase que mide la divergencia entre la distribución de probabilidades predicha y la real, penalizando predicciones que asignan baja probabilidad a la clase correcta. Utilizada con pesos por clase para compensar desbalance.

## **Épocas**

Número de pasadas completas por el conjunto de entrenamiento durante el proceso de optimización. Este trabajo utiliza hasta 50 épocas para el clasificador y 100 para el modelo de landmarks (fase 2), con parada temprana.

## **Error en píxeles**

Métrica para evaluar la precisión de predicción de puntos de referencia, definida como la distancia euclídea entre las coordenadas predichas y las anotadas manualmente: Error =  $\sqrt{(x_{\text{pred}} - x_{\text{real}})^2 + (y_{\text{pred}} - y_{\text{real}})^2}$ .

## **Escalado**

Transformación geométrica que modifica el tamaño de una configuración de puntos. En GPA, se normaliza cada forma para que tenga norma unitaria, eliminando diferencias de escala.

## **Estratificación**

Técnica de división de dataset que mantiene las proporciones de clases en cada subconjunto (entrenamiento, validación, prueba), crítica cuando existe desbalance para asegurar representación adecuada de todas las categorías.

## **Exactitud (Accuracy)**

Métrica principal de evaluación que mide la proporción de predicciones correctas sobre el

total: Accuracy =  $\frac{\text{Predicciones correctas}}{\text{Total}}$ . El sistema propuesto alcanza 98.10 % de exactitud.

### **Extractor de características (Feature Extractor / Backbone)**

Componente de una red neuronal (típicamente capas convolucionales) que procesa la imagen de entrada y produce representaciones de alto nivel. En este trabajo se utiliza ResNet-18 preentrenado en ImageNet.

## **F**

### **F1-Score**

Métrica que combina precisión y sensibilidad mediante su media armónica:  $F1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$ , proporcionando un balance entre ambos objetivos.

### **F1-Score Macro**

Promedio no ponderado del F1-Score de cada clase, tratando todas las categorías con igual importancia independientemente del número de muestras. Utilizado para evaluar rendimiento equilibrado en presencia de desbalance.

### **F1-Score Ponderado (Weighted F1-Score)**

Promedio del F1-Score por clase ponderado por el número de muestras de cada categoría, reflejando el desempeño en el contexto del desbalance.

### **Falsos Negativos (FN)**

Casos de una clase positiva que el modelo clasifica incorrectamente como negativos. En detección de COVID-19, un falso negativo es un paciente positivo clasificado como Normal o Neumonía Viral.

### **Falsos Positivos (FP)**

Casos de clases negativas que el modelo clasifica incorrectamente como positivos. En detección de COVID-19, un falso positivo es un paciente Normal o con Neumonía Viral clasificado como COVID-19.

### **Fill rate (Tasa de cobertura)**

Porcentaje de píxeles no negros en una imagen normalizada geométricamente, indicando qué proporción de la imagen contiene información de la región pulmonar versus fondo. Este trabajo alcanza 47 % de fill rate.

### **Forma estándar (Forma canónica)**

Configuración promedio de puntos de referencia calculada mediante GPA que representa la forma típica de los pulmones en el conjunto de entrenamiento, utilizada como plantilla de destino para la normalización geométrica.

### **Función de activación**

Transformación no lineal aplicada a las salidas de capas neuronales para introducir capacidad de representación no lineal. Ejemplos: ReLU, Sigmoid, Softmax.

# G

## Global Average Pooling

Operación de reducción que calcula el promedio de cada canal de características sobre todas las posiciones espaciales, condensando un mapa de características (e.g.,  $7 \times 7 \times 512$ ) en un vector (512). Utilizado en la cabeza de regresión.

## GPA (*Generalized Procrustes Analysis / Análisis Procrustes Generalizado*)

Método estadístico para alinear múltiples configuraciones de puntos eliminando diferencias de traslación, escala y rotación mediante un proceso iterativo que calcula una forma promedio por centrado, normalización y rotación óptima (SVD).

*Parámetros:* tolerancia  $\tau = 10^{-8}$ , máximo 100 iteraciones.

## GPU (*Graphics Processing Unit / Unidad de Procesamiento Gráfico*)

Procesador especializado diseñado originalmente para renderizado gráfico, utilizado en aprendizaje profundo por su capacidad de realizar miles de operaciones matemáticas en paralelo, acelerando significativamente el entrenamiento e inferencia de redes neuronales.

## Gradiente

Vector de derivadas parciales de la función de pérdida respecto a los parámetros del modelo, que indica la dirección de mayor incremento. Durante el entrenamiento, se actualiza en dirección opuesta al gradiente (descenso de gradiente).

## Ground truth

Anotaciones de referencia creadas manualmente por expertos que sirven como etiquetas verdaderas para entrenamiento y evaluación. En este trabajo, 957 imágenes tienen anotaciones manuales de 15 puntos de referencia.

## Group Normalization (Normalización por Grupos)

Técnica de normalización que divide los canales en grupos y normaliza dentro de cada grupo, independiente del tamaño de lote. Utilizada en la cabeza de regresión como alternativa a Batch Normalization para estabilidad con lotes pequeños.

# H

## Hiperparámetros

Valores de configuración que definen el comportamiento del modelo y del proceso de entrenamiento, establecidos antes de iniciar el entrenamiento (tasa de aprendizaje, tamaño de lote, número de épocas, etc.), a diferencia de los parámetros que se aprenden automáticamente.

# I

## **ImageNet**

Conjunto de datos con más de un millón de imágenes naturales organizadas en 1,000 categorías, utilizado ampliamente para preentrenamiento de redes convolucionales mediante aprendizaje por transferencia.

## **Inferencia**

Proceso de aplicar un modelo entrenado a nuevos datos para generar predicciones, sin actualizar los parámetros del modelo.

## **Interpolación bilineal**

Método de interpolación que estima el valor de un píxel en una posición no entera promediando los valores de los cuatro píxeles vecinos, ponderados por la distancia. Utilizado durante el warping para obtener transiciones suaves.

## **Internal covariate shift**

Fenómeno donde las distribuciones de las activaciones internas de una red neuronal cambian constantemente durante el entrenamiento a medida que se actualizan los pesos, dificultando la convergencia. Mitigado por Batch Normalization.

# K

## **Kernel** (Núcleo convolucional / Filtro)

Matriz de pesos pequeña (típicamente  $3 \times 3$ ,  $5 \times 5$  o  $7 \times 7$ ) que se desliza sobre la imagen de entrada durante la operación de convolución, realizando producto punto en cada posición para extraer características específicas. Cada kernel aprende a detectar patrones particulares como bordes, texturas o formas.

# L

## **Landmarks** (Puntos de referencia anatómicos)

Coordenadas específicas que representan estructuras anatómicas de interés. En este trabajo, 15 puntos definen el contorno pulmonar: eje central (L1, L9, L10, L11, L2), contorno izquierdo (L12, L3, L5, L7, L14) y contorno derecho (L13, L4, L6, L8, L15).

## **Logits**

Valores numéricos crudos en la capa de salida de una red neuronal antes de aplicar la función de activación softmax. Representan los puntajes no normalizados para cada clase, que posteriormente se convierten en probabilidades mediante softmax.

# M

## **Margin scale** (Escala de margen)

Factor de expansión aplicado al centroide de los landmarks para determinar la región de recorte durante el warping. El valor óptimo validado experimentalmente es 1.05 (5 % de expansión), almacenado en GROUND\_TRUTH.json.

## **Matriz de confusión**

Tabla que presenta la distribución completa de predicciones versus categorías reales, mostrando en la diagonal las clasificaciones correctas y fuera de ella los errores específicos entre pares de clases.

## **MaxPool** (*Max Pooling*)

Operación de reducción espacial que selecciona el valor máximo de cada región, utilizada para disminuir dimensionalidad preservando características prominentes y proporcionando invariancia a pequeñas traslaciones.

## **Mecanismo de atención**

Componente de red neuronal que aprende a enfocar selectivamente regiones relevantes de la entrada, ponderando la importancia de diferentes características o posiciones espaciales.

# N

## **Neumonía**

Inflamación del tejido pulmonar causada por infección, caracterizada en radiografías por opacidades, consolidaciones e infiltrados. Las categorías COVID-19 y Neumonía Viral son dos tipos específicos de neumonía clasificados en este trabajo.

## **Neumonía Viral**

Neumonía causada por virus distintos a SARS-CoV-2, presentando patrones radiográficos similares pero diferenciables de COVID-19. Representa el 9 % del dataset (1,345 imágenes) y la clase minoritaria.

## **Normal**

Radiografía de paciente sin patología pulmonar aparente, representando la clase mayoritaria (67 %, 10,192 imágenes) en el dataset.

## **Normalización de contraste**

Proceso de ajustar las intensidades de una imagen para mejorar la visibilidad de estructuras relevantes, mitigando variaciones introducidas por diferentes equipos de adquisición y condiciones de exposición.

## **Normalización geométrica**

Proceso de transformación de imágenes a una configuración espacial estándar mediante detección de landmarks y warping, eliminando variabilidad de posición, escala, orientación y deformación local no relacionada con patología.

# O

## **OpenCV** (*Open Source Computer Vision Library*)

Biblioteca de código abierto especializada en visión por computadora y procesamiento de imágenes, que proporciona implementaciones eficientes de algoritmos para transformaciones geométricas, mejora de contraste, detección de características y manipulación de imágenes. Utilizada en el preprocesamiento y warping de radiografías.

## **Optimizador Adam**

Algoritmo de optimización que adapta la tasa de aprendizaje para cada parámetro combinando momento y estimación de segundo momento de los gradientes. Utilizado en este trabajo con parámetros por defecto ( $\beta_1 = 0,9$ ,  $\beta_2 = 0,999$ ).

# P

## **Pares simétricos**

Puntos de referencia bilateralmente simétricos que corresponden a posiciones equivalentes en pulmones izquierdo y derecho: (L3, L4), (L5, L6), (L7, L8), (L12, L13), (L14, L15). Utilizados durante Test-Time Augmentation para corregir predicciones después de reflejo horizontal.

## **Parada temprana** (*Early Stopping*)

Mecanismo de regularización que detiene el entrenamiento cuando el rendimiento en el conjunto de validación deja de mejorar durante un número de épocas consecutivas (pacienza), evitando sobreajuste y conservando el modelo con mejor rendimiento.

## **Padding** (Relleno)

Técnica que agrega píxeles adicionales (típicamente con valor cero) en los bordes de la imagen o mapa de características antes de aplicar convolución, permitiendo preservar las dimensiones espaciales originales y evitar pérdida de información en los bordes.

## **Píxel**

Unidad básica de una imagen digital, representando un punto con valores de intensidad (escala de grises) o color (RGB). El término proviene de "picture element" (elemento de imagen).

## **PNG** (*Portable Network Graphics*)

Formato de archivo de imagen sin pérdida que soporta compresión, transparencia y profundidad de color de hasta 48 bits. Ampliamente utilizado para almacenar radiografías digitales y visualizaciones de resultados por su calidad sin degradación y compatibilidad universal.

## **Planificador** (*Learning Rate Scheduler*)

Mecanismo que ajusta dinámicamente la tasa de aprendizaje durante el entrenamiento según una estrategia predefinida (Cosine Annealing, Step Decay, etc.), optimizando la convergencia y mejorando el rendimiento final del modelo.

### **Pooling** (Agrupamiento)

Operación de reducción espacial que disminuye la dimensionalidad de mapas de características preservando información relevante. Tipos: Max Pooling (selecciona máximo), Average Pooling (calcula promedio).

### **Precisión** (*Precision*)

Métrica que mide, de todas las predicciones positivas para una clase, qué proporción son correctas: Precisión =  $\frac{VP}{VP+FP}$ . Relevante cuando el costo de falsos positivos es alto.

### **Preprocesamiento**

Conjunto de transformaciones aplicadas a las imágenes antes del procesamiento principal (mejora de contraste, redimensionamiento, normalización) para estandarizar la entrada y mejorar la efectividad del modelo.

**Puntos de referencia** → Ver *Landmarks*

### **PyTorch**

Framework de código abierto para aprendizaje profundo desarrollado por Meta AI, que proporciona estructuras de datos tensoriales con aceleración por GPU, diferenciación automática y herramientas para construcción, entrenamiento y evaluación de redes neuronales. Utilizado como base de implementación en este trabajo.

## R

### **Radiografía de tórax**

Imagen médica obtenida mediante exposición a rayos X que visualiza estructuras torácicas (pulmones, corazón, huesos), utilizada para diagnóstico de patologías pulmonares.

### **Radiografía posteroanterior (PA)**

Proyección radiográfica donde el haz de rayos X atraviesa el cuerpo del paciente de posterior a anterior, con el detector colocado frente al pecho. Estándar para radiografías de tórax de pie.

### **Región pulmonar**

Área de la radiografía que contiene el tejido pulmonar, delimitada por el contorno pulmonar definido mediante los 15 landmarks en este trabajo.

### **Reflejo horizontal** (*Horizontal Flip*)

Transformación de imagen que invierte la imagen a lo largo del eje vertical, utilizada como técnica de aumento de datos y en Test-Time Augmentation. Requiere intercambio de pares simétricos de landmarks para mantener consistencia anatómica.

### **Regularización**

Conjunto de técnicas para prevenir sobreajuste y mejorar generalización, incluyendo dropout, aumento de datos, parada temprana y penalizaciones sobre parámetros.

### **ReLU** (*Rectified Linear Unit*)

Función de activación no lineal definida como  $f(x) = \max(0, x)$ , que elimina valores negativos.

Utilizada ampliamente en redes convolucionales por su eficiencia computacional y mitigación de desvanecimiento de gradiente.

#### **ResNet** (*Residual Network*)

Familia de arquitecturas de redes neuronales profundas que utilizan conexiones residuales para facilitar el entrenamiento. Propuesta por He et al. (2016).

#### **ResNet-18**

Variante de ResNet con 18 capas (11.2 millones de parámetros), utilizada en este trabajo como extractor de características para landmarks y como clasificador de enfermedades pulmonares debido a su balance entre capacidad y eficiencia.

#### **RGB** (*Red, Green, Blue*)

Modelo de representación de color que codifica cada píxel mediante tres canales (Rojo, Verde, Azul). Las radiografías de tórax son imágenes en escala de grises (un canal) que se replican a 3 canales para compatibilidad con redes preentrenadas en ImageNet.

#### **Rotación óptima**

Transformación de rotación que minimiza la distancia entre dos configuraciones de puntos, calculada mediante descomposición en valores singulares (SVD). Utilizada en GPA para el paso de alineación rotacional.

#### **RT-PCR** (*Reverse Transcription Polymerase Chain Reaction*)

Reacción en Cadena de la Polimerasa con Transcripción Inversa. Prueba molecular de laboratorio considerada el estándar de oro para diagnóstico de COVID-19, que detecta material genético viral (ARN) mediante amplificación. Las radiografías de tórax complementan este diagnóstico proporcionando evaluación de severidad pulmonar.

## S

#### **SAHS** (*Statistical Asymmetrical Histogram Stretching*)

Método de mejora de contraste diseñado para histogramas asimétricos que calcula límites de estiramiento diferenciados según la distribución de intensidades por encima y por debajo de la media:  $I_{max} = \mu + 2,5\sigma_+$ ,  $I_{min} = \mu - 2,0\sigma_-$ .

#### **Semilla aleatoria** (*Random Seed*)

Valor inicial que controla la generación de números pseudoaleatorios, permitiendo reproducibilidad de experimentos. Este trabajo utiliza semilla fija (42) para particiones de datos y diferentes semillas (123, 321, 111, 666) para modelos del ensamble.

#### **Sensibilidad** (*Recall / Sensitivity*)

Métrica que mide, de todos los casos reales de una clase, qué proporción detecta el sistema:  $\text{Sensibilidad} = \frac{VP}{VP+FN}$ . Crítica cuando el costo de falsos negativos es alto.

#### **Sigmoid** (Sigmoide)

Función de activación que mapea cualquier valor real al rango (0, 1):  $\sigma(x) = \frac{1}{1+e^{-x}}$ . Utilizada en la salida de la cabeza de regresión para normalizar coordenadas al rango [0, 1].

**Silueta pulmonar** → Ver *Contorno pulmonar*

**Skip connections** → Ver *Conexiones residuales*

**Sobreajuste (Overfitting)**

Fenómeno donde un modelo aprende demasiado bien los ejemplos de entrenamiento, incluyendo ruido y particularidades, perdiendo capacidad de generalizar a datos nuevos. Mitigado mediante regularización, dropout y parada temprana.

**Softmax**

Función que transforma valores numéricos en probabilidades que suman 1, utilizada en la capa de salida del clasificador para convertir logits en distribución de probabilidad sobre las tres clases.

**Stride (Paso)**

Parámetro que define el desplazamiento del kernel durante la operación de convolución o pooling. Un stride de 1 aplica la operación en cada posición, mientras que stride de 2 reduce las dimensiones espaciales a la mitad, funcionando como mecanismo de reducción de resolución.

**Subajuste (Underfitting)**

Fenómeno donde un modelo es demasiado simple o no se entrena suficientemente, resultando en incapacidad para capturar patrones relevantes tanto en datos de entrenamiento como de prueba. Indica necesidad de mayor capacidad del modelo o más épocas de entrenamiento.

**SVD (Singular Value Decomposition / Descomposición en Valores Singulares)**

Técnica de álgebra lineal que descompone una matriz en tres matrices ( $A = U\Sigma V^T$ ), utilizada en GPA para calcular la rotación óptima entre configuraciones de puntos mediante el método de Schönemann.

## T

**Tamaño de lote (Batch Size)**

Número de imágenes procesadas simultáneamente durante el entrenamiento. Este trabajo utiliza lotes de 16 (fase 1 landmarks), 8 (fase 2 landmarks) y 32 (clasificador).

**Tasa de aprendizaje (Learning Rate)**

Hiperparámetro que controla la magnitud de los ajustes de parámetros durante el descenso de gradiente. Este trabajo utiliza tasas diferenciadas:  $10^{-3}$  (fase 1 landmarks),  $2 \times 10^{-5}$  (backbone fase 2),  $2 \times 10^{-4}$  (cabeza fase 2),  $10^{-4}$  (clasificador).

**Test-Time Augmentation (TTA)**

Técnica que procesa cada imagen de prueba con múltiples transformaciones (original y reflejada) y promedia las predicciones para reducir varianza. Requiere intercambio de pares simétricos de landmarks al procesar imágenes reflejadas.

**Tiles**

Regiones rectangulares en las que se divide una imagen durante CLAHE para aplicar ecualización local. Tamaño típico:  $4 \times 4$  u  $8 \times 8$  bloques.

## **Transformación afín**

Transformación geométrica que preserva líneas rectas y paralelismo, pudiendo incluir traslación, rotación, escalado y sesgo. Queda completamente determinada por la correspondencia entre tres puntos no colineales, propiedad utilizada en warping triangular.

**Transfer Learning** → Ver *Aprendizaje por transferencia*

## **Triangulación de Delaunay**

Método geométrico que conecta un conjunto de puntos mediante triángulos que no se superponen, maximizando el ángulo mínimo de todos los triángulos para evitar triángulos degenerados. Genera 16 triángulos a partir de los 15 landmarks en este trabajo.

# **V**

## **Validación cruzada (Cross-Validation)**

Técnica de evaluación que divide el conjunto de datos en  $k$  pliegues, entrenando  $k$  veces usando  $k - 1$  pliegues para entrenamiento y 1 para validación, rotando los pliegues. Este trabajo utiliza  $k = 5$  para evaluar estabilidad del clasificador.

## **Verdaderos Negativos (VN)**

Casos de clases negativas correctamente clasificados como negativos.

## **Verdaderos Positivos (VP)**

Casos de una clase positiva correctamente clasificados como positivos. En detección de COVID-19, pacientes positivos correctamente identificados.

# **W**

## **Warping** → Ver *Deformación afín por partes*

## **Wing Loss**

Función de pérdida diseñada para regresión de landmarks que combina comportamiento logarítmico para errores pequeños (incentivando refinamiento fino) con comportamiento lineal para errores grandes (estabilidad):

$$\text{Wing}(x) = \begin{cases} \omega \ln\left(1 + \frac{|x|}{\epsilon}\right) & \text{si } |x| < \omega \\ |x| - C & \text{si } |x| \geq \omega \end{cases}$$

donde  $\omega = 10$  píxeles,  $\epsilon = 2$  píxeles,  $C = \omega - \omega \ln(1 + \omega/\epsilon)$ .

## Acrónimos y Abreviaturas

---

Acrónimo	Significado
AdamW	<i>Adam with Weight Decay</i> Optimizador Adam con Decaimiento de Pesos
CLAHE	<i>Contrast Limited Adaptive Histogram Equalization</i> Ecualización Adaptativa de Histograma con Límite de Contraste
CNN	<i>Convolutional Neural Network</i> Red Neuronal Convolucional
COVID-19	<i>Coronavirus Disease 2019</i> Enfermedad por Coronavirus 2019
CSV	<i>Comma-Separated Values</i> Valores Separados por Comas
FC	<i>Fully Connected</i> Completemente Conectado
FN	Falsos Negativos ( <i>False Negatives</i> )
FP	Falsos Positivos ( <i>False Positives</i> )
GPA	<i>Generalized Procrustes Analysis</i> Análisis Procrustes Generalizado
GPU	<i>Graphics Processing Unit</i> Unidad de Procesamiento Gráfico
L1	<i>L1 Loss</i> / Error Absoluto Medio
L2	<i>L2 Loss</i> / Error Cuadrático Medio
PA	Posteroanterior
PNG	<i>Portable Network Graphics</i> Formato de imagen sin pérdida
OpenCV	<i>Open Source Computer Vision Library</i> Biblioteca de Visión por Computadora de Código Abierto
ReLU	<i>Rectified Linear Unit</i> Unidad Lineal Rectificada
RGB	<i>Red, Green, Blue</i> Modelo de color Rojo, Verde, Azul

---

---

<b>Acrónimo</b>	<b>Significado</b>
RT-PCR	<i>Reverse Transcription Polymerase Chain Reaction</i> Reacción en Cadena de la Polimerasa con Transcripción Inversa
PyTorch	Framework de aprendizaje profundo de código abierto
SAHS	<i>Statistical Asymmetrical Histogram Stretching</i> Estiramiento Asimétrico Estadístico de Histograma
SE-Net	<i>Squeeze-and-Excitation Network</i> Red de Compresión y Excitación
SVD	<i>Singular Value Decomposition</i> Descomposición en Valores Singulares
TTA	<i>Test-Time Augmentation</i> Aumento en Tiempo de Prueba
VN	Verdaderos Negativos ( <i>True Negatives</i> )
VP	Verdaderos Positivos ( <i>True Positives</i> )

---

# Bibliografía

- [1] World Health Organization, “Use of chest imaging in covid-19: a rapid advice guide,” *WHO Reference Number: WHO/2019-nCoV/Clinical/Radiology-imaging/2020.1*, 2020.
- [2] L. Wang, Z. Q. Lin, and A. Wong, “Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images,” *Scientific Reports*, vol. 10, no. 1, p. 19549, 2020.
- [3] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” *arXiv preprint arXiv:1711.05225*, 2017.
- [4] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al Emadi *et al.*, “Can ai help in screening viral and covid-19 pneumonia?” *IEEE Access*, vol. 8, pp. 132 665–132 676, 2020.
- [5] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughraier, M. S. Khan *et al.*, “Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images,” *Computers in Biology and Medicine*, vol. 132, p. 104319, 2021.
- [6] A. E. Picazo-Castillo, S. E. Ayala-Raggi, L. Altamirano-Robles, A. Barreto-Flores, and J. F. Portillo-Robledo, “Comparative study of lung image representations for automated pneumonia recognition,” *International Journal of Combinatorial Optimization Problems and Informatics*, vol. 15, no. 5, pp. 193–201, 2024.
- [7] S. E. Ayala-Raggi, A. E. Picazo-Castillo, A. Barreto-Flores, and J. F. Portillo-Robledo, “Synergizing chest x-ray image normalization and discriminative feature selection for efficient and automatic covid-19 recognition,” in *Pattern Recognition. ACPR 2023*, ser. Lecture Notes in Computer Science, H. Lu, M. Blumenstein, S.-B. Cho, C.-L. Liu, Y. Yagi, and T. Kamiya, Eds., vol. 14407. Cham: Springer, 2023, pp. 224–238.
- [8] J. Rocha, S. C. Pereira, J. Pedrosa, A. Campilho, and A. M. Mendonça, “STERN: Attention-driven spatial transformer network for abnormality detection in chest x-ray images,” *Artificial Intelligence in Medicine*, vol. 147, p. 102737, 2024.
- [9] Y.-C. Yeh, C.-H. Weng, Y.-J. Huang, C.-J. Fu, T.-T. Tsai, and C.-Y. Yeh, “Deep learning approach for automatic landmark detection and alignment analysis in whole-spine lateral radiographs,” *Scientific Reports*, vol. 11, no. 1, p. 7618, 2021.

- [10] R. A. Cruz-Ovando, S. E. Ayala-Raggi, Á. E. Picazo-Castillo, A. Barreto-Flores, and J. F. Portillo-Robledo, “Statistical asymmetrical histogram stretching for contrast enhancement in chest x-ray images for pneumonia detection,” *International Journal of Combinatorial Optimization Problems and Informatics*, 2025.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [12] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *International conference on machine learning*, pp. 448–456, 2015.
- [13] Z.-H. Feng, J. Kittler, M. Awais, P. Huber, and X.-J. Wu, “Wing loss for robust facial landmark localisation with convolutional neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2235–2245.
- [14] Q. Hou, D. Zhou, and J. Feng, “Coordinate attention for efficient mobile network design,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13713–13722.
- [15] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in *Graphics gems IV*. Academic Press Professional, 1994, pp. 474–485.
- [16] J. C. Gower, “Generalized procrustes analysis,” *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.
- [17] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial transformer networks,” in *Advances in neural information processing systems*, vol. 28, 2015.
- [18] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, “Adaptive histogram equalization and its variations,” *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.
- [19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2009, pp. 248–255.
- [20] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [21] Y. Wu and K. He, “Group normalization,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

- [22] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” in *Advances in neural information processing systems*, vol. 27, 2014.
- [23] I. Loshchilov and F. Hutter, “SGDR: Stochastic gradient descent with warm restarts,” in *International Conference on Learning Representations*, 2017.
- [24] I. L. Dryden and K. V. Mardia, *Statistical shape analysis*. Chichester: John Wiley & Sons, 1998.
- [25] P. H. Schönemann, “A generalized solution of the orthogonal procrustes problem,” *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.
- [26] B. Delaunay, “Sur la sphère vide,” *Bulletin de l’Académie des Sciences de l’URSS, Classe des Sciences Mathématiques et Naturelles*, vol. 6, pp. 793–800, 1934.
- [27] M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars, *Computational geometry: algorithms and applications*, 3rd ed. Berlin: Springer-Verlag, 2008.
- [28] G. Wolberg, *Digital image warping*. Los Alamitos, CA: IEEE Computer Society Press, 1990.
- [29] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, “Transfusion: Understanding transfer learning for medical imaging,” in *Advances in neural information processing systems*, vol. 32, 2019.