



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
1. Introducción	6
2. Marco Teórico	10
2.1. Representación anatómica mediante puntos de referencia	10
2.2. Preprocesamiento: CLAHE	11
2.3. Preprocesamiento: SAHS	12
2.4. Redes Neuronales Convolucionales	13
2.5. ResNet y conexiones residuales	14
2.6. Normalización por Lote	15
2.7. Aprendizaje por transferencia	16
2.8. Función de pérdida: Wing Loss	17
2.9. Análisis Procrustes Generalizado (GPA)	18
2.10. Triangulación de Delaunay	19
2.11. Deformación afín por partes	20
2.12. Clasificación de imágenes	21
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	22
2.14. Mecanismo de Coordinate Attention	23
2.15. Ensamble de modelos y Test-Time Augmentation	24
2.15.1. Ensamble de modelos	25
2.15.2. Aumento en tiempo de prueba (TTA)	25
3. Estado del Arte	27
3.1. Aprendizaje Profundo para Diagnóstico Médico por Imágenes	27
3.1.1. Evolución de Arquitecturas CNN	27
3.1.2. Transfer Learning desde ImageNet	28
3.1.3. Casos de Éxito en Medical AI	28
3.1.4. Desafíos Persistentes	28
3.2. Detección de COVID-19 mediante Aprendizaje Profundo	29
3.2.1. Datasets Públicos para COVID-19	29
3.2.2. Arquitecturas Específicas para COVID-19	29

3.2.3. Análisis Comparativo de Resultados	30
3.2.4. Limitaciones Identificadas	32
3.3. Detección de Puntos de Referencia Anatómicos	32
3.3.1. Métodos Tradicionales vs Deep Learning	33
3.3.2. Coordinate Regression vs Heatmap Regression	33
3.3.3. Funciones de Pérdida Especializadas	34
3.3.4. Aplicaciones en Radiografías Médicas	34
3.3.5. Brecha Identificada: Landmarks en Chest X-rays	35
3.4. Normalización Geométrica en Imágenes Médicas	35
3.4.1. Spatial Transformer Networks	35
3.4.2. Extensiones de STN en Medical Imaging	37
3.4.3. Piecewise Affine Warping	37
3.4.4. Aplicación a Clasificación: Brecha en la Literatura	38
3.4.5. Trabajos del Grupo de Investigación	38
3.4.6. Contribución de Este Trabajo	38
3.5. Mecanismos de Atención en Clasificación de Imágenes Médicas	40
3.5.1. Attention Mechanisms Clásicos	40
3.5.2. Coordinate Attention	41
3.5.3. Vision Transformers	41
3.6. Mejora de Contraste y Preprocesamiento	42
3.6.1. Contrast Limited Adaptive Histogram Equalization	42
3.6.2. Aplicación a COVID-19 Detection	43
3.7. Robustez y Generalización	43
3.7.1. Desplazamiento de Dominio en Imágenes Médicas	43
3.7.2. Estrategias de Mitigación de Desplazamiento de Dominio	44
3.7.3. Métodos de Ensamble	44
3.7.4. Test-Time Augmentation	45
3.8. Síntesis y Posicionamiento del Trabajo	45
3.8.1. Brechas Identificadas en la Literatura	45
3.8.2. Posicionamiento Cuantitativo	47
3.8.3. Contribuciones Específicas	47
3.8.4. Limitaciones y Direcciones Futuras	48
4. Metodología	49
4.1. Descripción General del Sistema	49
4.1.1. Arquitectura del Sistema	50
4.1.2. Flujo de Datos	51
4.1.3. Justificación del Diseño Modular	51
4.2. Conjunto de Datos y Preprocesamiento	53

4.2.1.	COVID-19 Radiography Database	53
4.2.2.	Anotación de Puntos de Referencia Anatómicos	53
4.2.3.	Preprocesamiento de Imágenes	57
4.2.4.	División del Conjunto de Datos	59
4.3.	Modelo de Predicción de Puntos de Referencia	61
4.3.1.	Arquitectura del Modelo	61
4.3.2.	Función de Pérdida	66
4.3.3.	Estrategia de Entrenamiento	67
4.3.4.	Resumen de Hiperparámetros	70
4.3.5.	Ensamble de Modelos	72
4.4.	Normalización Geométrica	73
4.4.1.	Análisis Procrustes Generalizado	73
4.4.2.	Triangulación de Delaunay	77
4.4.3.	Transformación Afín por Partes	78
4.4.4.	Proceso Completo de Normalización	80
4.5.	Clasificación de Enfermedades Pulmonares	82
4.5.1.	Preprocesamiento de Contraste	82
4.5.2.	Arquitectura del Clasificador	84
4.5.3.	Estrategia de Aprendizaje por Transferencia	85
4.5.4.	Configuración del Entrenamiento	86
4.5.5.	Métricas de Evaluación del Clasificador	88
4.5.6.	Aumento de Datos	89
4.5.7.	Resumen de la Configuración	90
4.6.	Protocolo de Inferencia y Evaluación	91
4.6.1.	Proceso de Inferencia	92
4.6.2.	Métricas de Evaluación	92
5. Resultados		95
5.1.	Detección de Puntos de Referencia	95
5.1.1.	Precisión del Sistema	95
5.1.2.	Precisión por Punto de Referencia	95
5.1.3.	Resumen	96
5.2.	Normalización Geométrica	98
5.2.1.	Forma Estándar de Referencia	98
5.2.2.	División en Triángulos para Transformación	99
5.2.3.	Ejemplos de Normalización	99
5.2.4.	Resumen	100
5.3.	Clasificación de Enfermedades Pulmonares	102
5.3.1.	Rendimiento General	102

5.3.2. Validación Cruzada	102
5.3.3. Rendimiento por Categoría	103
5.3.4. Análisis de Errores	103
5.3.5. Efecto de la Normalización Geométrica	105
5.3.6. Resumen	108
6. Conclusiones y Trabajos Futuros	110
6.1. Síntesis de Contribuciones	110
6.1.1. Contribución Principal	110
6.1.2. Contribuciones Específicas	111
6.2. Validación de la Hipótesis	112
6.2.1. Hipótesis Planteada	112
6.2.2. Evidencia de Validación	112
6.2.3. Limitaciones de la Validación	114
6.2.4. Respuesta a la Hipótesis	114
6.3. Implicaciones del Trabajo	115
6.3.1. Implicaciones Clínicas	115
6.3.2. Implicaciones Metodológicas	115
6.3.3. Implicaciones Técnicas	116
6.4. Limitaciones del Estudio	116
6.4.1. Limitaciones Experimentales	116
6.4.2. Limitaciones Metodológicas	116
6.4.3. Limitaciones Conceptuales	117
6.5. Trabajos Futuros	117
6.5.1. Validación y Generalización	117
6.5.2. Extensiones del Sistema	118
6.5.3. Mejoras Metodológicas	118
6.5.4. Interpretabilidad y Explicabilidad	119
6.5.5. Optimización e Implementación	119
6.6. Reflexión Final	120
Glosario	121

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 recursos limitados donde el acceso a especialistas es escaso. En este contexto, el desarrollo de sistemas automatizados con alta exactitud y robustez es fundamental para apoyar efectivamente la toma de decisiones clínicas, reduciendo la variabilidad inherente en la interpretación manual de imágenes radiográficas y permitiendo diagnósticos oportunos que pueden salvar vidas [1].

La importancia de contar con algoritmos de alta precisión trasciende las métricas numéricas: un diagnóstico erróneo tiene consecuencias tangibles y severas. Un falso negativo en COVID-19 puede resultar en propagación inadvertida de la enfermedad, poniendo en riesgo a comunidades enteras, mientras que un falso positivo genera costos innecesarios, ansiedad en pacientes y saturación de sistemas de salud. Por ello, los sistemas de diagnóstico automatizado deben demostrar no solo alta precisión en condiciones controladas, sino también robustez ante la variabilidad real en la adquisición de imágenes: diferencias en el posicionamiento del paciente, distancia de proyección, calibración de equipos, fase respiratoria y calidad de imagen. La confiabilidad clínica de estos sistemas depende de su capacidad de mantener performance consistente frente a esta variabilidad extrínseca.

Un desafío fundamental en la clasificación automática de radiografías de tórax es precisamente esta variabilidad geométrica y de pose. Dos radiografías del mismo paciente pueden presentar diferencias significativas en la orientación, escala y posición de las estructuras pulmonares debido a factores no relacionados con la patología. Esta variabilidad extrínseca dificulta el aprendizaje de características patológicas intrínsecas por parte de modelos de aprendizaje profundo (*deep learning*), que pueden confundir diferencias de pose con diferencias diagnósticas o, peor aún, aprender correlaciones espurias basadas en artefactos de adquisición en lugar de patrones patológicos genuinos [2].

Diversos trabajos han abordado la detección automática de COVID-19 y neumonía mediante aprendizaje profundo. Wang et al. [3] desarrollaron COVIDNet, una arquitectura de red neuronal convolucional (*Convolutional Neural Network*, CNN) diseñada específicamente para clasificación multi-clase de COVID-19, neumonía viral y neumonía bacteriana, alcanzando 93.3 % de exactitud. Rajpurkar et al. [4] demostraron que CheXNet, basada en DenseNet-121, alcanza performance comparable a radiólogos en detección de neumonía. Estos

enfoques aplican redes neuronales convolucionales directamente sobre imágenes originales o con preprocesamiento de contraste, sin normalización geométrica explícita. Aunque alcanzan alta exactitud en datasets de entrenamiento, su dependencia en características de apariencia puede limitar su capacidad de generalización a datos de hospitales o protocolos de adquisición diferentes.

Estudios recientes han demostrado que la normalización geométrica puede mejorar significativamente el rendimiento de clasificadores médicos. Picazo-Castillo et al. [5] 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. [6] propusieron la integración de normalización de imágenes de tórax con selección discriminativa de características basada en Análisis de Componentes Principales (*Principal Component Analysis*, PCA) para reconocimiento eficiente de COVID-19, logrando mejoras en exactitud mediante la reducción de variabilidad extrínseca. Rocha et al. [7] 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. [8] 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. Estos trabajos establecen que la reducción de variabilidad extrínseca mediante normalización espacial es una estrategia viable para mejorar la robustez de clasificadores.

Sin embargo, los trabajos previos se limitan a transformaciones rígidas (rotación, traslación) o afines globales, que asumen uniformidad en la deformación del tejido pulmonar. Esta suposición es inexacta dado que el pulmón es un órgano deformable cuya forma varía según la fase respiratoria, posición del paciente y patología subyacente. Una transformación global que rota o escala la imagen completa puede alinear aproximadamente las estructuras pulmonares, pero no captura la variabilidad local en la forma y expansión de diferentes regiones del pulmón.

Este trabajo propone un enfoque más robusto de normalización geométrica basado en *deformación afín por partes* (*piecewise affine warping*) de la región pulmonar. A diferencia de métodos que solo rotan, trasladan o escalan la imagen de forma global, el método propuesto **deforma localmente la región pulmonar** para adaptarla a una forma estándar común, permitiendo que diferentes regiones se transformen de manera independiente mientras se preserva la información diagnóstica contenida en la textura local. Esta estrategia busca eliminar variabilidad de pose y geometría sin sacrificar patrones patológicos intrínsecos como opacidades en vidrio esmerilado, consolidaciones e infiltrados intersticiales característicos de COVID-19 y neumonía viral.

El sistema propuesto integra cuatro componentes principales en un pipeline coherente. Primero, se utiliza un modelo CNN basado en ResNet-18 con Coordinate Attention para

predecir automáticamente 15 puntos de referencia anatómicos (*landmarks*) que definen el contorno bilateral del pulmón. El modelo se entrena mediante regresión directa de coordenadas utilizando Wing Loss, una función de pérdida especializada que amplifica la sensibilidad a errores pequeños de localización. Segundo, se aplica Análisis General de Procrustes (*Generalized Procrustes Analysis*, GPA) sobre las configuraciones de puntos de referencia del conjunto de entrenamiento para calcular una forma estándar pulmonar que representa el consenso geométrico de la población, eliminando variaciones de traslación, rotación y escala. Tercero, se construye una malla de triángulos mediante triangulación de Delaunay sobre los puntos de referencia de la forma estándar y de cada imagen individual, y se aplica una transformación afín independiente a cada triángulo para mapear la información de intensidad de píxeles de la imagen original a la forma estándar. Esta deformación triángulo por triángulo permite adaptación local preservando continuidad geométrica. Finalmente, las imágenes normalizadas geométricamente se procesan mediante un clasificador ResNet-18 entrenado para discriminar entre COVID-19, neumonía viral y casos normales.

La evaluación experimental del sistema propuesto demuestra su viabilidad técnica y efectividad. El modelo de detección de puntos de referencia, utilizando un ensamblaje de cuatro redes entrenadas con diferentes semillas aleatorias y combinando predicciones mediante aumento de datos en tiempo de prueba (*Test-Time Augmentation*, TTA) con corrección de simetría bilateral, alcanza un error promedio de 3.61 píxeles en imágenes de 224×224 píxeles, equivalente a 1.6 % del tamaño de imagen. Esta precisión es comparable a trabajos de detección de puntos de referencia faciales y notable considerando la mayor variabilidad anatómica y patológica en radiografías de tórax. El clasificador entrenado sobre imágenes normalizadas alcanza 98.10 % de exactitud en el conjunto de prueba, con F1-Score macro de 97.17 % y validación cruzada (*cross-validation*) de 5 folds de $98.60 \% \pm 0.26 \%$. El análisis de robustez revela que la normalización geométrica facilita el aprendizaje de características genuinas relacionadas con patología en lugar de artefactos de adquisición, evidenciado por la degradación controlada de performance al eliminar regiones periféricas que potencialmente contienen correlaciones espurias.

El resto del documento se organiza de la siguiente manera. El Capítulo 2 presenta el marco teórico, estableciendo los fundamentos de formación de imagen radiográfica, arquitecturas de redes neuronales convolucionales, análisis de forma mediante Procrustes, triangulación de Delaunay, deformación geométrica y métricas de evaluación. El Capítulo 3 revisa el estado del arte en detección de COVID-19 mediante aprendizaje profundo, detección de puntos de referencia anatómicos en imágenes médicas, normalización geométrica y mecanismos de atención, posicionando este trabajo en el contexto de la literatura actual e identificando brechas que motivan la investigación. El Capítulo 4 describe la metodología propuesta en detalle, documentando la arquitectura de los modelos, el protocolo de entrenamiento, el proceso de normalización geométrica y el flujo de inferencia completo. El Capítulo 5 presenta los resultados experimentales, enfocándose en la precisión del modelo de puntos de referencia,

la mejora en clasificación atribuible a la normalización geométrica, y el análisis de robustez mediante validación cruzada y perturbaciones controladas. Finalmente, el Capítulo 6 sintetiza las conclusiones del trabajo, analiza la evidencia sobre la efectividad de la normalización geométrica para mejorar sistemas de detección de neumonía, reconoce limitaciones del enfoque propuesto e identifica 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.

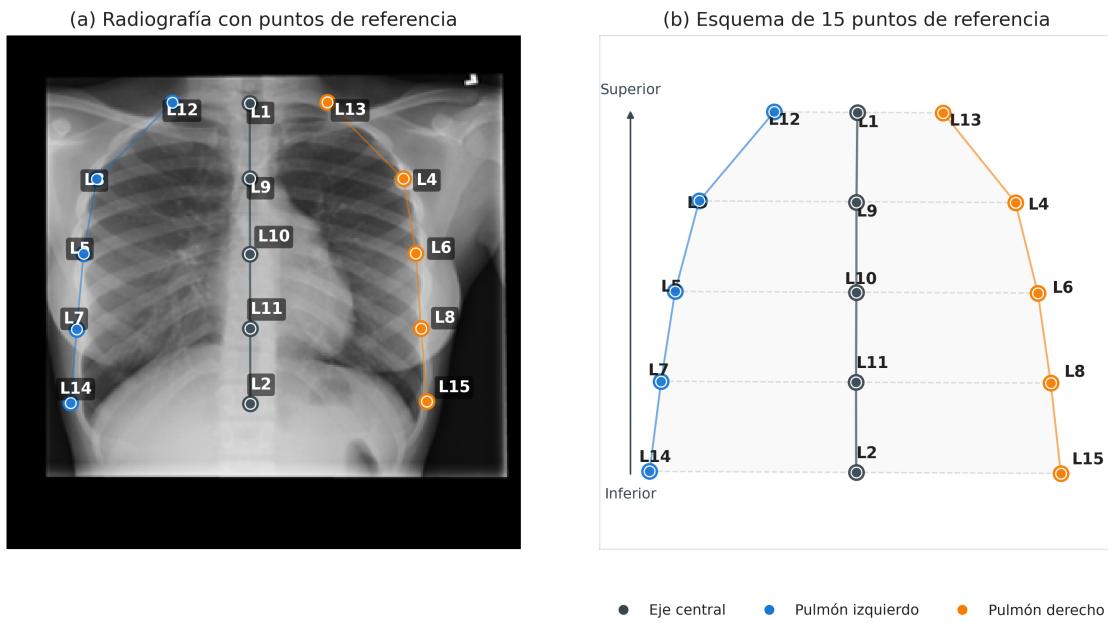


Figura 2.1: Representación de los 15 puntos de referencia anatómicos. (a) Radiografía con puntos de referencia etiquetados. (b) Esquema de la distribución espacial, donde el eje central (gris) define la línea media y los puntos laterales delimitan los contornos pulmonares izquierdo (azul) y derecho (naranja).

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).

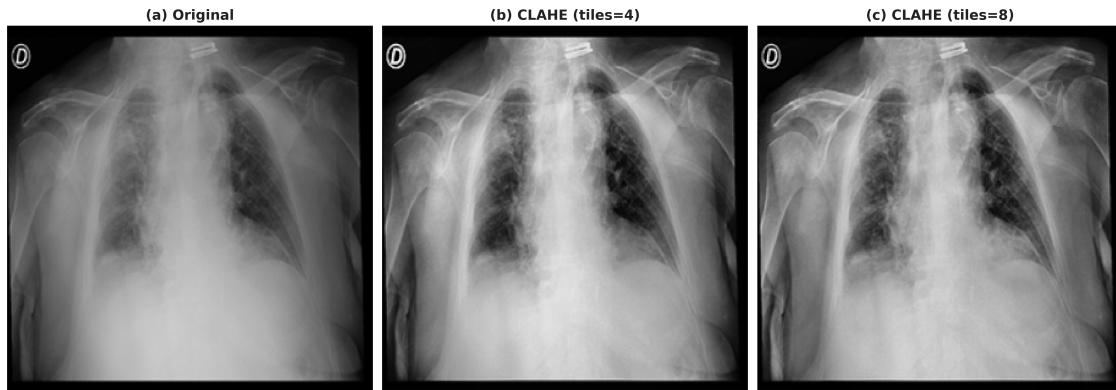


Figura 2.2: Efecto de CLAHE sobre una radiografía de tórax. (a) Imagen original con bajo contraste. (b) Imagen procesada con CLAHE usando tamaño de tile 4×4 . (c) Imagen procesada con CLAHE usando tamaño de tile 8×8 . El tamaño de tile menor produce una mejora de contraste más localizada.

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 4×4 o 8×8).
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

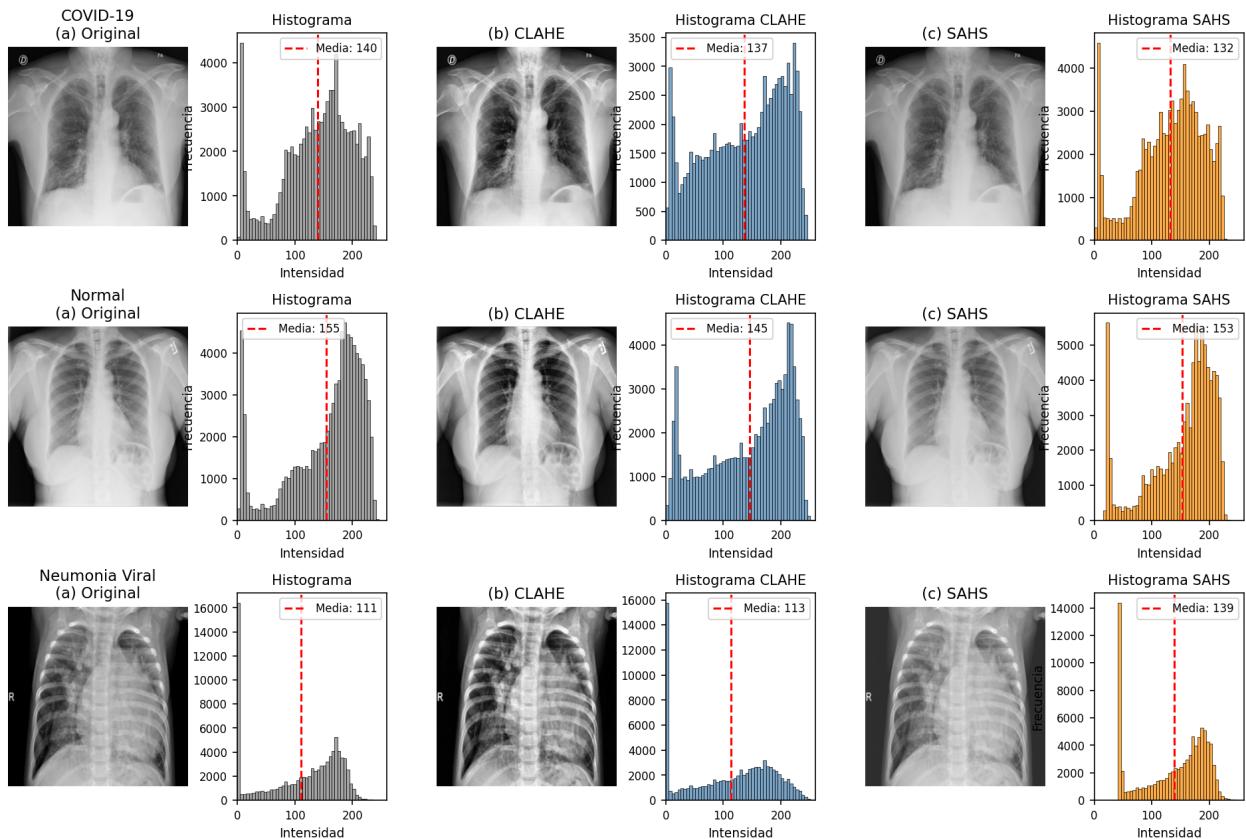


Figura 2.3: Comparación de técnicas de mejora de contraste sobre radiografías de las tres clases. (a) Imagen original. (b) CLAHE: mejora local del contraste mediante ecualización adaptativa. (c) SAHS: estiramiento global con límites asimétricos. Los histogramas muestran cómo cada técnica afecta la distribución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

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 [9].

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:

1. **Cálculo de la media:** Se obtiene la intensidad media μ 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 μ (tonos claros)
 - Grupo B: píxeles con intensidad menor o igual que μ (tonos oscuros)
3. **Desviaciones asimétricas:** Se calcula la desviación estándar de cada grupo por separado (σ_+ para el grupo A, σ_- 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.

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.

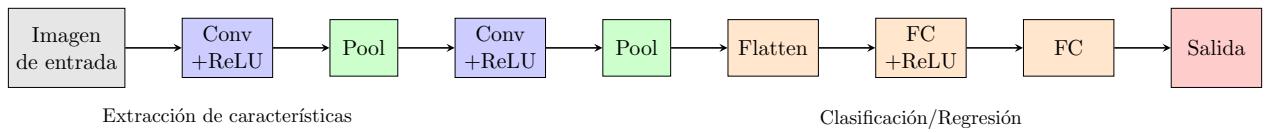


Figura 2.4: Arquitectura básica de una Red Neuronal Convolucional (CNN). Las capas convolucionales extraen características de la imagen, las capas de pooling reducen la dimensionalidad, y las capas fully connected (FC) producen la salida final.

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. [10], 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.

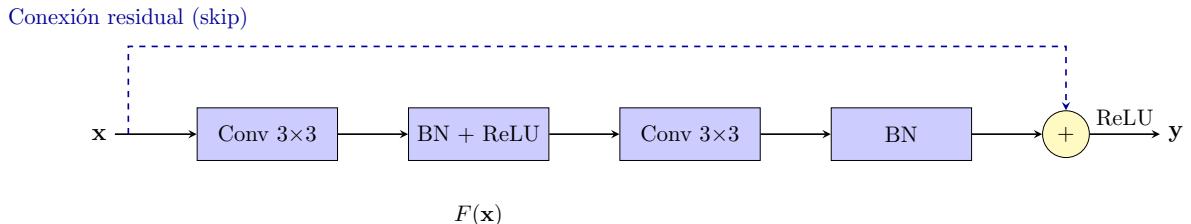


Figura 2.5: Bloque residual de ResNet. La entrada x pasa por dos capas convolucionales (rama principal) y simultáneamente se suma directamente a la salida mediante la conexión residual. La salida es $y = F(x) + x$, donde $F(x)$ representa la transformación aprendida.

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 [11], 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 (γ y β) 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. [12], 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), ω define el umbral entre errores pequeños y grandes, ϵ 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.

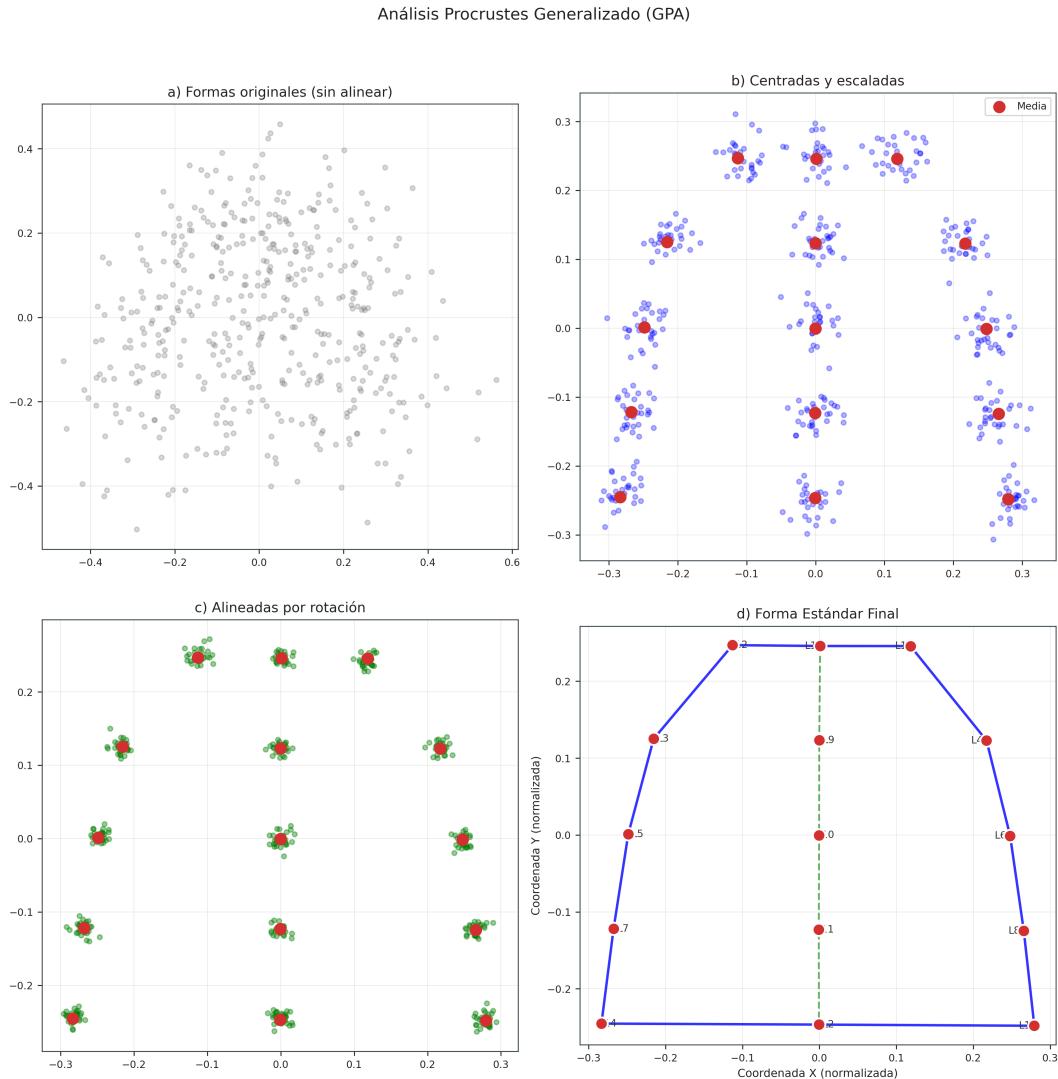


Figura 2.6: Proceso de Análisis Procrustes Generalizado (GPA). (a) Formas originales sin alinear, mostrando la variabilidad entre pacientes. (b) Formas centradas y escaladas a tamaño unitario. (c) Formas alineadas por rotación, minimizando las diferencias. (d) Forma estándar final, calculada como el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

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.

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.

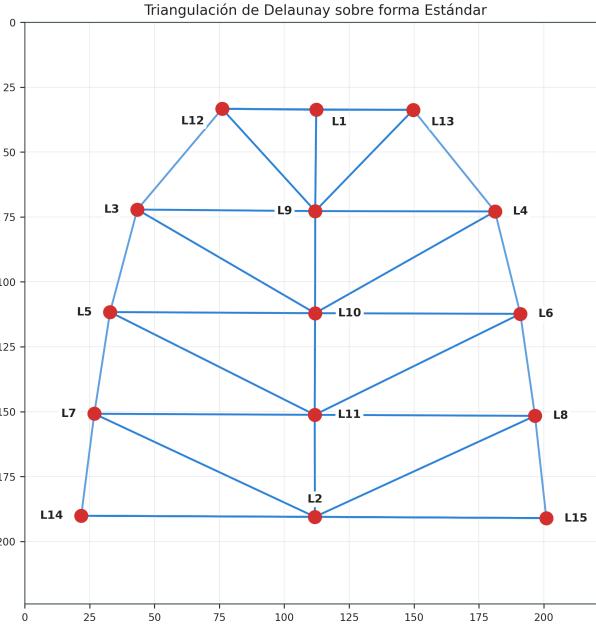


Figura 2.7: Triangulación de Delaunay sobre los 15 puntos de referencia anatómicos. Los triángulos resultantes cubren la región pulmonar sin superposición, maximizando los ángulos mínimos para evitar distorsiones durante la deformación.

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.
2. **Correspondencia de triángulos:** Cada triángulo en la imagen original tiene un triángulo correspondiente en la forma estándar, definido por los mismos índices de

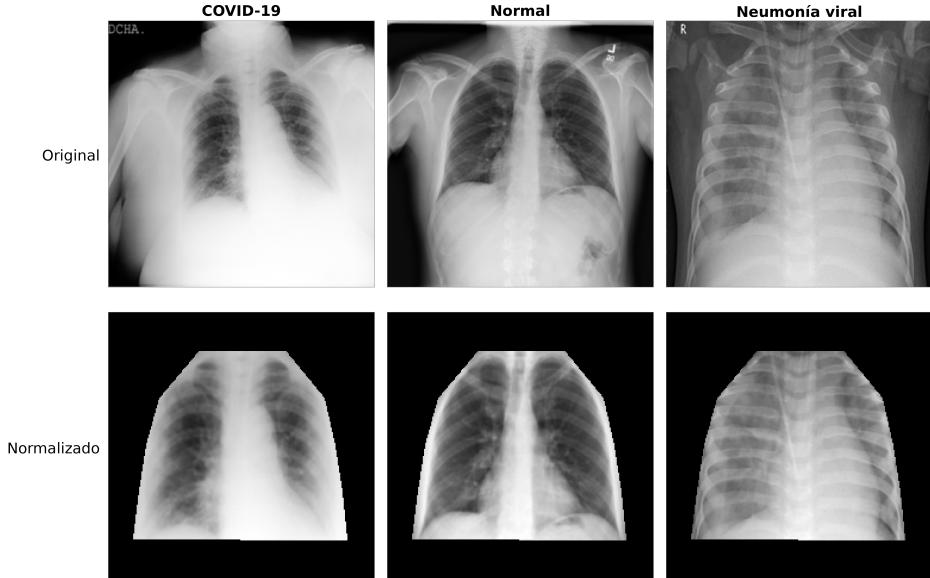


Figura 2.8: Comparación de radiografías originales y normalizadas por clase. Columnas: COVID-19, Normal y Neumonía viral. Filas: Original (arriba) y Normalizado (abajo). La normalización geométrica mediante deformación afín por partes alinea la región pulmonar con la forma estándar y reduce variabilidad de pose y escala.

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 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. [13], 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.

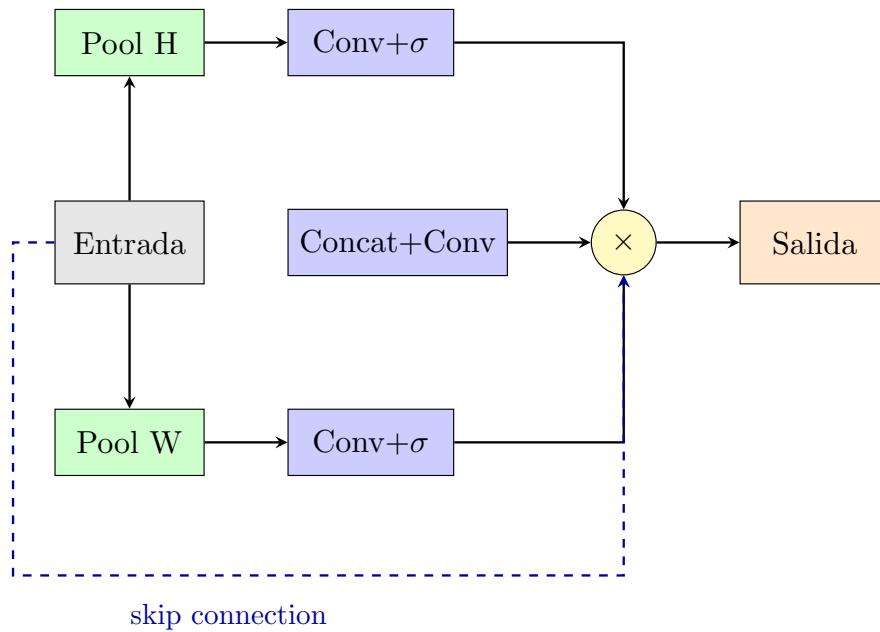


Figura 2.9: Mecanismo de Coordinate Attention. La entrada se procesa mediante poolings direccionales (H y W) que preservan información espacial. Las representaciones se concatenan, procesan con convoluciones y activación sigmoide (σ), y se multiplican con la entrada original (skip connection).

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.

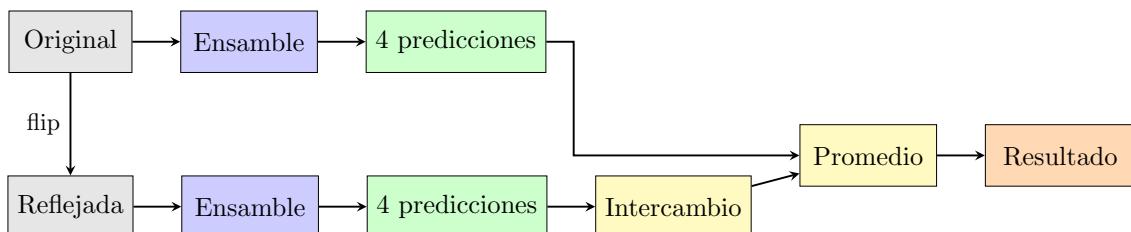


Figura 2.10: Inferencia con ensamble y TTA. La imagen original y su reflejo se procesan por un ensamble de 4 modelos. Las predicciones del reflejo requieren intercambiar puntos de referencia simétricos (Swap) antes de promediar las 8 predicciones.

Síntesis

Este capítulo presentó los conceptos fundamentales que sustentan el sistema propuesto: representación anatómica mediante puntos de referencia, técnicas de preprocesamiento (CLAHE y SAHS), arquitecturas de aprendizaje profundo (ResNet con Coordinate Attention), análisis de forma (GPA), normalización geométrica (deformación afín por partes), y estrategias de robustez (ensamble y TTA).

El siguiente capítulo presenta una revisión crítica del estado del arte, analizando trabajos relacionados en detección de COVID-19, localización de puntos de referencia, y normalización geométrica, identificando las brechas que motivan la propuesta de este trabajo.

Capítulo 3

Estado del Arte

Este capítulo presenta una revisión sistemática y crítica de la literatura relevante al problema de detección de COVID-19 mediante normalización geométrica de radiografías de tórax. La revisión se estructura en torno a los componentes principales del flujo de trabajo (*pipeline*) propuesto: aprendizaje profundo para diagnóstico médico, detección de COVID-19, localización de puntos de referencia anatómicos, normalización geométrica, mecanismos de atención, y preprocesamiento de contraste. El análisis identifica gaps en la literatura actual y posiciona las contribuciones de este trabajo en el contexto del estado del arte.

3.1. Aprendizaje Profundo para Diagnóstico Médico por Imágenes

El aprendizaje profundo ha transformado el campo del análisis de imágenes médicas en la última década. Litjens et al. [14] presentaron una revisión sistemática fundamental que analizó más de 300 trabajos sobre aprendizaje profundo en imágenes médicas, documentando la adopción masiva de redes neuronales convolucionales (*Convolutional Neural Networks*, CNN) para tareas de clasificación, detección, segmentación y registro. Este trabajo seminal identificó desafíos persistentes que continúan siendo relevantes: la escasez de datos anotados, el desbalance de clases, la necesidad de interpretabilidad, y la dificultad de generalización a datos de distribuciones diferentes.

3.1.1. Evolución de Arquitecturas CNN

La evolución de arquitecturas CNN para diagnóstico médico ha seguido una trayectoria de creciente sofisticación. AlexNet [15] demostró la viabilidad del aprendizaje profundo en ImageNet, inspirando su adopción en imágenes médicas. VGGNet [16] introdujo bloques convolucionales profundos, mientras que ResNet [10] resolvió el problema de degradación mediante conexiones residuales, permitiendo redes de centenares de capas. DenseNet [17] extendió esta idea con conexiones densas entre todas las capas, mejorando la propagación de gradientes y la reutilización de características. EfficientNet [18] optimizó el escalamiento de redes mediante búsqueda de arquitectura neural, logrando mejor compromiso entre exactitud y eficiencia computacional.

3.1.2. Transfer Learning desde ImageNet

El aprendizaje por transferencia (*transfer learning*) desde ImageNet [19] se ha convertido en práctica estándar en imágenes médicas debido a la escasez de conjuntos de datos anotados de gran escala. Yosinski et al. [20] demostraron que las capas iniciales de CNNs aprenden características genéricas (bordes, texturas) transferibles a dominios diferentes, mientras que capas profundas aprenden características específicas del dominio. Sin embargo, Raghu et al. [21] cuestionaron esta práctica en imágenes médicas, mostrando que el beneficio de preentrenamiento en ImageNet es menor cuando se dispone de conjuntos de datos médicos de tamaño moderado ($\approx 10,000$ imágenes), sugiriendo que la inicialización aleatoria con suficientes datos puede ser competitiva.

A pesar de esta controversia, el aprendizaje por transferencia continúa siendo ampliamente adoptado en el contexto de conjuntos de datos pequeños o altamente desbalanceados, como es el caso de muchos conjuntos de datos de COVID-19.

3.1.3. Casos de Éxito en Medical AI

Trabajos seminales han demostrado que sistemas basados en aprendizaje profundo pueden alcanzar o superar desempeño de expertos humanos en tareas específicas. Esteva et al. [22] desarrollaron un clasificador de cáncer de piel entrenado en 129,450 imágenes que alcanzó desempeño comparable a 21 dermatólogos certificados en la clasificación de melanomas malignos vs nevus benignos y carcinomas malignos vs queratosis seborreicas benignas. Rajpurkar et al. [4] presentaron CheXNet, un modelo basado en DenseNet-121 que excedió el desempeño promedio de radiólogos en la detección de neumonía en radiografías de tórax, entrenado en el conjunto de datos ChestX-ray14 con 112,120 imágenes frontales.

Estos trabajos demostraron la viabilidad de sistemas de apoyo al diagnóstico basados en CNNs, pero también revelaron limitaciones importantes: dependencia en conjuntos de datos de gran escala, dificultad de generalización a hospitales externos (desplazamiento de dominio, *domain shift*), y la tendencia de modelos a explotar factores de confusión como marcadores metálicos o artefactos de adquisición (*shortcut learning*) [2].

3.1.4. Desafíos Persistentes

A pesar de los avances significativos, persisten desafíos fundamentales en la aplicación de aprendizaje profundo a diagnóstico médico:

- **Data scarcity:** La obtención de conjuntos de datos médicos anotados de gran escala requiere inversión significativa de tiempo de expertos clínicos. La mayoría de conjuntos de datos públicos contienen entre 1,000 y 50,000 imágenes, órdenes de magnitud menores que ImageNet (14 millones de imágenes).

- **Class imbalance:** Enfermedades raras o condiciones patológicas específicas resultan en conjuntos de datos altamente desbalanceados, donde clases minoritarias pueden representar ¡1% del total. Técnicas de aumento de datos (*aumento de datos*), resampling y funciones de pérdida ponderadas parcialmente mitigan este problema.
- **Interpretabilidad:** Modelos aprendizaje profundo operan como “cajas negras”, dificultando la comprensión de decisiones clínicas. Técnicas de visualización como Grad-CAM, saliency maps y attention mechanisms proporcionan interpretabilidad parcial, pero la confianza clínica en sistemas automatizados continúa siendo un desafío.
- **Variabilidad extrínseca:** Diferencias en protocolos de adquisición, equipamiento, poblaciones de pacientes y prácticas institucionales introducen variabilidad que afecta la generalización de modelos. Este trabajo aborda específicamente la variabilidad geométrica mediante normalización espacial.

3.2. Detección de COVID-19 mediante Aprendizaje Profundo

La pandemia de COVID-19 aceleró la investigación en sistemas automatizados de detección basados en radiografías de tórax. Esta sección revisa trabajos representativos, analiza sus resultados cuantitativos, e identifica limitaciones comunes.

3.2.1. Datasets Públicos para COVID-19

La disponibilidad de conjuntos de datos públicos fue crucial para el desarrollo rápido de sistemas de detección. El COVID-19 Radiography Database [23], utilizado en este trabajo, contiene 21,165 radiografías de tórax distribuidas en cuatro clases: 10,192 casos normales, 3,616 casos COVID-19 positivos, 6,012 casos de opacidad pulmonar (infección pulmonar no-COVID), y 1,345 casos de neumonía viral. Las imágenes, en formato PNG de resolución 299×299 píxeles, fueron recopiladas de múltiples fuentes públicas incluyendo RSNA, Kaggle, y repositorios GitHub, lo que introduce heterogeneidad en protocolos de adquisición.

Otros conjuntos de datos prominentes incluyen COVIDx [3], que contiene 13,975 radiografías de 13,870 pacientes, y BIMCV-COVID19, un conjunto de datos español con metadata clínica detallada. La fragmentación de conjuntos de datos y la falta de estandarización en protocolos de anotación dificultan la comparación directa de resultados.

3.2.2. Arquitecturas Específicas para COVID-19

Wang et al. [3] desarrollaron COVID-Net, una arquitectura CNN diseñada específicamente para detección de COVID-19 mediante principios de design humano-en-el-loop. El modelo

alcanzó 93.3 % de exactitud en clasificación de tres clases (COVID-19, neumonía viral, normal) en el conjunto de datos COVIDx. La arquitectura incorpora módulos de expansión-compresión que permiten aprendizaje eficiente de representaciones discriminativas, pero requiere un proceso de diseño manual que puede no generalizar a otros dominios.

Rajpurkar et al. [4] demostraron que CheXNet, basado en DenseNet-121 preentrenado en ImageNet, excede el desempeño de radiólogos en detección de neumonía en el conjunto de datos ChestX-ray14. Aunque este trabajo predató la pandemia de COVID-19, estableció una línea base importante para clasificación de patologías pulmonares en radiografías de tórax, demostrando que arquitecturas estándar con transfer learning pueden alcanzar desempeño de nivel experto cuando se dispone de conjuntos de datos de gran escala (112,120 imágenes).

Trabajos recientes han explorado arquitecturas más ligeras y eficientes. Un estudio de 2023 reportó que ResNet-18 combinado con deep Random Vector Functional Link network (dRVFL) alcanzó 97.56 % de exactitud, superando métodos previos con arquitectura significativamente más compacta. Este resultado es particularmente relevante para este trabajo, que también utiliza ResNet-18 como extractor de características, demostrando que arquitecturas ligeras pueden ser competitivas con modelos más pesados cuando se combinan con técnicas de normalización y ensamble apropiadas.

3.2.3. Análisis Comparativo de Resultados

La Tabla 3.1 presenta una comparación cuantitativa de métodos representativos de detección de COVID-19. Los resultados muestran alta variabilidad en desempeño (exactitud de 93.3 % a 98.6 %), atribuible a diferencias en conjuntos de datos, protocolos de evaluación, y técnicas de preprocessamiento.

Varios patrones emergen del análisis comparativo:

1. **Arquitecturas profundas vs ligeras:** DenseNet-121 y DenseNet-201 consistentemente alcanzan alta exactitud (97-98 %), pero a costo de mayor complejidad computacional. ResNet-18, con 11.7 millones de parámetros (vs 20 millones de DenseNet-121), alcanza desempeño competitivo cuando se combina con técnicas adecuadas de entrenamiento y preprocessamiento.
2. **Transfer learning:** La mayoría de trabajos utilizan preentrenamiento en ImageNet, sugiriendo que features genéricas de bajo nivel (bordes, texturas) son relevantes para clasificación de radiografías, a pesar de la diferencia sustancial entre fotografías naturales y imágenes médicas.
3. **Variabilidad en conjuntos de datos:** Comparaciones directas son difíciles debido a diferencias en tamaño de conjunto de datos, distribución de clases, y protocolos de partición train/val/test. Algunos trabajos reportan exactitud en conjuntos de datos propietarios, limitando reproducibilidad.

Cuadro 3.1: Comparación de métodos de aprendizaje profundo para detección de COVID-19 en radiografías de tórax

Trabajo	Arquitectura	Conjunto de datos	Exact.	Sens. COVID-19 (%)
Wang et al. COVID-Net (2020) [3]	COVID-Net	COVIDx (13,975)	93.3	–
Rajpurkar et al. DenseNet-121 (2017) ^a	DenseNet-121	ChestX-ray14 (112k)	F1 nivel radiólogo	–
<i>Métodos recientes (2023-2024)</i>				
CovC-ReDRNet (2023)	ResNet-18 dRVFL	+ COVID-19 Radiography	97.56	94.94
RegNetX032 (2023)	RegNetX032	Chest X-ray	98.6	98.0
VGG19 (2024)	VGG19	Chest X-ray	95.0	96.0
DenseNet121 (2024)	DenseNet-121	X-ray + CT	98.0	–
DenseNet201 (2024)	DenseNet-201	Chest X-ray	97.11	97.54
Este trabajo (2026)	ResNet-18 (warped)	COVID-19 Radiography (21,165)	98.10	96.46^b

^a Rajpurkar et al. (2017): Pre-COVID, detección de neumonía general (no específica COVID-19)

^b Sensibilidad (*recall*) para clase COVID-19; métricas completas por clase en Capítulo 5

c) Este trabajo integra normalización geométrica mediante GPA + deformación afín por partes basada en puntos de referencia automáticos, alcanzando rendimiento competitivo (98.10 %) sin explotar artefactos extrapulmonares (ver Capítulo 5 para análisis de características espurias)

– No reportado en el trabajo original

Exact. = Exactitud (*Accuracy*), Sens. COVID-19 = Sensibilidad (*Sensitivity*) para clase COVID-19

dRVFL = *deep Random Vector Function Link network*

F1-macro: 97.17 %, F1-weighted: 98.09 % (validación cruzada 5-fold: 98.60 % \pm 0.26 %)

Conjunto de datos COVID-19 Radiography Database: 10,192 Normal + 3,616 COVID-19 + 1,345 Neumonía Viral (21,165 muestras utilizadas). El conjunto de datos original incluye 6,012 casos de Lung Opacity que fueron excluidos del análisis por no corresponder a las tres categorías diagnósticas del estudio

Este trabajo alcanza 98.10 % de exactitud utilizando ResNet-18 en imágenes normalizadas geométricamente, con F1-macro de 97.17 % y F1-weighted de 98.09 %. El desempeño es competitivo con el estado del arte, sugiriendo que normalización geométrica puede ser una estrategia complementaria efectiva a arquitecturas más complejas.

3.2.4. Limitaciones Identificadas

A pesar de resultados prometedores, la literatura sobre COVID-19 detection presenta limitaciones sistemáticas:

- **Evaluación en un solo conjunto de datos:** La mayoría de trabajos evalúan en un único conjunto de datos público, sin validación en datos externos de otros hospitales. Zech et al. [24] demostraron que modelos de detección de neumonía entrenados en un hospital pueden exhibir desempeño dramáticamente reducido (hasta 45 % degradación) en conjuntos de datos de otros hospitales, debido a desplazamiento de dominio y explotación de factores de confusión institucionales.
- **Falta de análisis de robustez:** Pocos trabajos evalúan robustez ante perturbaciones comunes como compresión JPEG, blur, variaciones de contraste, o cambios en posicionamiento del paciente. Esta omisión es problemática dado que condiciones clínicas reales introducen variabilidad en calidad de imagen.
- **Métricas incompletas:** Muchos trabajos reportan únicamente exactitud global, omitiendo sensitivity, specificity, y F1-score por clase, métricas más relevantes en contextos médicos donde costos de falsos positivos y falsos negativos son asimétricos.
- **Shortcut learning no abordado:** Geirhos et al. [2] documentaron que CNNs tienden a explotar correlaciones espurias (marcadores metálicos, artefactos de equipo, propiedades del hospital) en lugar de características patológicas intrínsecas. La normalización geométrica propuesta en este trabajo mitiga parcialmente este problema al estandarizar la región pulmonar, reduciendo variabilidad extrínseca.

El análisis del estado del arte revela una oportunidad clara: desarrollar métodos que no solo alcancen alta exactitud en conjuntos de datos de entrenamiento, sino que también demuestren robustez ante variabilidad extrínseca y generalización a datos de distribuciones diferentes.

3.3. Detección de Puntos de Referencia Anatómicos

La detección automática de puntos de referencia anatómicos (*anatomical landmarks*) es fundamental para múltiples tareas en análisis de imágenes médicas, incluyendo registro,

segmentación, diagnóstico y, en este trabajo, normalización geométrica. Esta sección revisa métodos de detección de puntos de referencia, analiza funciones de pérdida especializadas, y discute la escasez de trabajos en radiografías de tórax.

3.3.1. Métodos Tradicionales vs Deep Learning

Los métodos tradicionales de detección de landmarks se basaban en modelos estadísticos de forma. Cootes et al. [25] introdujeron Modelos de Forma Activa (*Active Shape Models*, ASM), que aprenden patrones de variabilidad de forma a partir de un conjunto de entrenamiento anotado manualmente. Los ASM emplean un proceso iterativo de refinamiento análogo a Active Contour Models (Snakes), ajustando una forma modelo a una imagen mediante búsqueda local de características. Aunque los ASM fueron ampliamente adoptados en segmentación médica, requieren inicialización cercana a la solución y son sensibles a variaciones significativas de apariencia.

El aprendizaje profundo ha reemplazado en gran medida a métodos tradicionales debido a su capacidad de aprender representaciones discriminativas de características visuales de forma de extremo a extremo (*de extremo a extremo*). Los métodos basados en CNN para detección de landmarks se pueden categorizar en dos enfoques principales: *coordinate regression* y *heatmap regression*.

3.3.2. Coordinate Regression vs Heatmap Regression

Coordinate regression formula la tarea como un problema de regresión que predice directamente las coordenadas (x, y) de cada landmark. Este enfoque es conceptualmente simple y permite entrenamiento de extremo a extremo, pero enfrenta desafíos significativos: el mapeo de características de imagen a coordenadas numéricas es altamente no lineal, y la naturaleza no acotada de las coordenadas dificulta la convergencia. Además, coordinate regression carece de generalización espacial, donde errores pequeños en features pueden resultar en desplazamientos grandes en coordenadas predichas.

Heatmap regression representa cada landmark como un mapa de probabilidad espacial, típicamente modelado como una distribución Gaussiana centrada en la ubicación verdadera. La red predice un heatmap 2D para cada landmark, y la ubicación final se obtiene mediante búsqueda del máximo (argmax) o soft-argmax diferenciable. Heatmap regression presenta ventajas importantes: la representación probabilística suaviza el espacio objetivo facilitando convergencia, permite explotar características locales de forma más efectiva, y proporciona generalización espacial donde la red aprende a activar regiones cercanas al landmark verdadero.

La literatura indica que heatmap regression generalmente supera a coordinate regression en exactitud, especialmente cuando se dispone de conjuntos de datos pequeños. Sin embargo, heatmap regression introduce sobrecarga computacional (predecir mapas 2D completos vs

coordenadas directas) y requiere post-procesamiento para extraer ubicaciones finales. Este trabajo utiliza coordinate regression debido a su simplicidad, eficiencia computacional, y adecuación para el flujo de trabajo propuesto donde se requieren coordenadas explícitas para transformaciones geométricas subsecuentes.

3.3.3. Funciones de Pérdida Especializadas

La elección de función de pérdida es crítica para el desempeño de modelos de regresión de landmarks. Feng et al. [12] propusieron Wing Loss, una función de pérdida diseñada específicamente para localización robusta de landmarks faciales. Wing Loss aborda la limitación de L1 y L2 loss: L1 loss trata errores pequeños y grandes de forma uniforme, mientras que L2 loss penaliza excesivamente outliers. Wing Loss amplifica el impacto de errores pequeños y medianos (cruciales para localización precisa) mediante una transición suave de L1 loss a una función logarítmica modificada en el rango $(-w, w)$, donde w es un hiperparámetro que controla el ancho de la región de transición. La formulación matemática permite ajustar la sensibilidad a errores pequeños sin magnificar outliers, resultando en convergencia más estable.

Wang et al. [26] extendieron Wing Loss con Adaptive Wing Loss para heatmap regression, introduciendo ponderación adaptativa según el tipo de píxel (foreground vs background). Adaptive Wing Loss combina la función Wing con un Weighted Loss Map que asigna mayor peso a píxeles de foreground y background difíciles, permitiendo que el entrenamiento se enfoque en regiones críticas para localización de landmarks. Este trabajo superó el estado del arte en benchmarks de face alignment (COFW, 300W, WFLW), demostrando que funciones de pérdida adaptativas mejoran significativamente la precisión de localización.

Este trabajo adopta Wing Loss para coordinate regression de landmarks pulmonares, aprovechando su capacidad de amplificar errores pequeños sin introducir inestabilidad por outliers. La elección se justifica empíricamente: el error de ensamble con Wing Loss alcanza 3.61 píxeles (1.14 % NME), comparable a trabajos de facial landmarks y superior a métodos línea base con L2 loss.

3.3.4. Aplicaciones en Radiografías Médicas

Yeh et al. [8] demostraron que aprendizaje profundo puede detectar automáticamente landmarks en radiografías laterales de columna vertebral completa con error promedio de 2.3 mm, permitiendo análisis de alineación para diagnóstico de escoliosis y otras deformidades espinales. El método utiliza heatmap regression con U-Net como arquitectura base, explotando la naturaleza localizada de vértebras en radiografías. El trabajo demostró viabilidad clínica comparando resultados contra anotaciones de expertos, encontrando concordancia alta (error estándar mild to moderate) suficiente para aplicaciones de screening.

Otros trabajos han abordado landmark detection en imágenes de cerebro (2.96 mm de error) y próstata (3.34 mm de error) utilizando arquitecturas CNN de dos etapas con limited training data, demostrando que métodos basados en aprendizaje profundo pueden generalizar a tareas donde la anotación manual es costosa.

3.3.5. Brecha Identificada: Landmarks en Chest X-rays

La revisión de la literatura revela una **escasez significativa de trabajos sobre detección de landmarks anatómicos en radiografías de tórax**, particularmente en el contexto de clasificación de COVID-19 y neumonía. La mayoría de trabajos se enfocan en facial landmarks (face alignment), landmarks ortopédicos (columna, extremidades), o landmarks en imágenes 3D (cerebro, próstata). Esta brecha es sorprendente dado que el tórax presenta desafíos únicos: alta variabilidad anatómica inter-paciente, deformación pulmonar dependiente de fase respiratoria, y presencia de patologías que alteran la forma pulmonar.

La Tabla 3.2 presenta una comparación cuantitativa de métodos de detección de puntos de referencia. Este trabajo contribuye al cerrar la brecha identificada, proponiendo un método de detección de 15 puntos de referencia de contorno pulmonar con error de 3.61 píxeles (1.14 % NME), comparable al estado del arte en puntos de referencia faciales (Feng: 1.47 % NME) y aplicado a una tarea significativamente más desafiante debido a variabilidad anatómica y patológica en radiografías de tórax.

El uso de ensamble de 4 modelos con Aumento en Tiempo de Prueba (*Test-Time Augmentation*, TTA) y corrección de simetría bilateral reduce el error de 4.04 px (mejor modelo individual) a 3.61 px, demostrando que técnicas de ensamble son efectivas para mejorar precisión de localización en contextos médicos donde errores pequeños son críticos para aplicaciones downstream como normalización geométrica.

3.4. Normalización Geométrica en Imágenes Médicas

La normalización geométrica busca reducir variabilidad extrínseca en imágenes mediante transformaciones espaciales que estandarizan la pose, orientación, y escala de estructuras anatómicas. Esta sección revisa métodos de transformación espacial, desde transformaciones globales hasta deformaciones locales, y analiza su aplicación a clasificación de imágenes médicas.

3.4.1. Spatial Transformer Networks

Jaderberg et al. [27] introdujeron Redes de Transformación Espacial (*Spatial Transformer Networks*, STN), un módulo diferenciable que permite a redes neuronales aprender transformaciones espaciales de forma de extremo a extremo sin supervisión adicional. Una

Cuadro 3.2: Comparación de métodos de detección de puntos de referencia anatómicos en imágenes médicas

Trabajo	Tarea	Error (px/mm)	NME (%)	Método
<i>Puntos de referencia faciales y de columna vertebral</i>				
Feng et al. (2018) [12]	Facial (68 pts)	–	1.47	Wing Loss + Heatmap
Wang et al. (2019)	Facial	–	SOTA ^a	Adaptive Wing Loss
Yeh et al. (2021) [8]	Columna vertebral (68 pts)	2.3 mm	–	Deep learning
<i>Puntos de referencia en radiografías de tórax: Brecha identificada</i>				
Este trabajo (2026)	Pulmón (15 pts)	3.61 px (224×224)	1.14	Wing Loss + Ensamble + TTA
		4.04 px ^b	1.28	Mejor individual

^a SOTA = Estado del arte (*State-of-the-art*) reportado sin número exacto en el paper original

^b Mejor modelo individual (semilla 456), ensamble reportado en fila principal

– No reportado o no aplicable (escalas diferentes)

NME = Error Medio Normalizado (*Normalized Mean Error*): $(\text{error}/\text{diagonal_imagen}) \times 100$

px = píxeles en imagen redimensionada (la escala varía por conjunto de datos)

mm = milímetros en espacio físico (requiere calibración de imagen DICOM)

Equivalencia aproximada: 3.61 px ≈ 5.2 mm (suponiendo ancho de tórax ~32 cm en 224 px)

LIMITACIÓN: Conjunto de datos COVID-19 Radiography carece de metadata DICOM uniforme (mezcla RSNA, Kaggle, papers)

TTA = Aumento en Tiempo de Prueba (*Test-Time Augmentation*) con corrección de simetría bilateral

Ensamble: promedio de 4 modelos (semillas 123, 321, 111, 666) con TTA + CLAHE

Brecha identificada: Escasa literatura sobre detección de contorno pulmonar completo en contexto COVID-19

Imágenes originales: 299×299 px, redimensionadas a 224×224 para el modelo
Error teórico mínimo (variabilidad inter-anotador): ~1.3 px

STN consiste en tres componentes: (1) una *localization network* que predice parámetros de transformación a partir de features de entrada, (2) un *grid generator* que construye un grid de coordenadas en la imagen de salida correspondientes a la imagen de entrada, y (3) un *sampler* que interpola valores de píxeles usando bilinear sampling diferenciable.

Las STN permiten que modelos aprendan invarianza a transformaciones geométricas (traslación, rotación, escala, affine, perspective) de forma automática, mejorando robustez a variaciones de pose sin aumentar datos manualmente. Sin embargo, las STN presentan una **limitación fundamental**: las transformaciones aprendidas son *globales*, aplicando la misma transformación afín o perspectiva a toda la imagen. Esta suposición es restrictiva para estructuras anatómicas deformables como pulmones, donde diferentes regiones pueden requerir deformaciones locales distintas.

3.4.2. Extensiones de STN en Medical Imaging

Rocha et al. [7] desarrollaron STERN (*Spatial Transformer Enhanced by Attention*), que combina STN con mecanismos de atención para detección de anomalías en radiografías de tórax. STERN emplea múltiples STNs en cascada, cada uno enfocándose en regiones espaciales diferentes mediante *attention gates*, permitiendo normalización jerárquica de la imagen. El método alcanzó mejora de +2.1 % AUC sobre línea base en el conjunto de datos ChestX-ray14, demostrando que alineación espacial aprendida mejora detección de patologías. Sin embargo, STERN continúa limitado a transformaciones afines globales por región, sin capacidad de deformación local dentro de cada región.

3.4.3. Piecewise Affine Warping

La deformación afín por partes (piecewise affine warping) supera la limitación de transformaciones globales permitiendo **deformaciones locales adaptativas**. El método divide la imagen en regiones mediante triangulación de Delaunay de landmarks de control, y aplica una transformación afín independiente a cada triángulo. Esta aproximación ofrece mayor flexibilidad que transformaciones globales, preservando continuidad en fronteras de triángulos mientras permite que diferentes regiones se deformen independientemente.

El fundamento matemático de piecewise affine warping fue establecido por Wolberg [28] en el contexto de digital image warping. Para cada triángulo en la malla de origen, se calcula una transformación afín que mapea sus vértices a las posiciones correspondientes en la malla destino. Píxeles dentro del triángulo se transforman usando coordenadas baricéntricas [29], asegurando interpolación suave. El método ha sido ampliamente utilizado en face morphing, facial expression transfer, y remote sensing para geometric correction de imágenes VHR (Very High Resolution).

3.4.4. Aplicación a Clasificación: Brecha en la Literatura

A pesar de su uso extenso en computer vision, la aplicación de piecewise affine warping a **clasificación de imágenes médicas es escasa**. La literatura se concentra en:

- **Registro de imágenes:** Alineación de imágenes multi-modales (MRI-CT, PET-CT) para fusión o análisis longitudinal.
- **Segmentación:** Deformación de atlas anatómicos para propagación de labels.
- **Face alignment:** Normalización de poses faciales para reconocimiento.

El uso de piecewise affine warping como *paso de preprocessamiento* para mejorar clasificación mediante normalización de variabilidad geométrica es un enfoque **poco explorado**, representando una brecha clara en la literatura.

3.4.5. Trabajos del Grupo de Investigación

Trabajos previos del grupo de investigación han explorado normalización de radiografías de tórax con enfoques complementarios. Picazo-Castillo et al. [5] presentaron un estudio comparativo de representaciones de imágenes pulmonares, demostrando que diferentes estrategias de normalización espacial (*cropping* inteligente, redimensionamiento adaptativo) afectan la capacidad de generalización de CNNs. Ayala-Raggi et al. [6] propusieron integración de normalización de radiografías de tórax con selección discriminativa de características basada en PCA, logrando mejora de aproximadamente 1.5 % en exactitud para clasificación de COVID-19. Estos trabajos establecieron que reducción de variabilidad extrínseca mediante normalización espacial es una estrategia viable para mejorar el desempeño de clasificadores.

3.4.6. Contribución de Este Trabajo

Este trabajo extiende trabajos previos al proponer un flujo de trabajo completo que integra: (1) detección automática de landmarks anatómicos de contorno pulmonar, (2) cálculo de forma canónica mediante Análisis de Procrustes Generalizado (*Generalized Procrustes Analysis*, GPA), (3) normalización geométrica mediante deformación afín por partes basada en triangulación de Delaunay, y (4) clasificación en imágenes normalizadas. La Tabla 3.3 posiciona este trabajo en el contexto de métodos de normalización geométrica para clasificación.

El enfoque propuesto alcanza 98.10 % de exactitud en el conjunto de datos COVID-19 Radiography, con $98.60\% \pm 0.26\%$ en validación cruzada de 5 folds. Aunque la exactitud es ligeramente inferior a la línea base sin normalización (98.68 % en imágenes originales), el experimento de recorte al 12 % (descrito en Resultados) demuestra que la línea base

Cuadro 3.3: Comparación de métodos de normalización geométrica aplicados a clasificación de imágenes médicas

Trabajo	Método	Conjunto de Mejora datos	Exact.	Transform. (%)
Jaderberg et al. STN (2015) [27]		MNIST, etc.	Variable	Affine global
Rocha et al. (2024) STERN [7]		ChestX-ray14	+2.1 ^a	STN + Attention
<i>Trabajos del grupo de investigación</i>				
Picazo-Castillo (2024) [5] et al.	Normalización + Feature selección	Neumonía (chest X-rays)	–	Representaciones comparativas
Ayala-Raggi (2023) [6] et al.	Normalización + PCA features	COVID-19 (chest X-rays)	+1.5 ^b	Cropping + PCA
<i>Deformación afín por partes para clasificación: Brecha en literatura</i>				
Este trabajo (2026)	GPA + Deformación afín por partes	COVID-19 Radiography (norm.) ^c	98.10	Piecewise affine (local)
			98.68 ^d	15 puntos (orig.) ref. + Delaunay

^a Mejora de +2.1 % en AUC (*Area Under Curve*), no exactitud directa

^b Mejora estimada de +1.5 % aproximadamente (reportada sin números exactos en el paper)

^c Exactitud en imágenes normalizadas geométricamente: 98.10 %

^d Exactitud en imágenes originales (sin normalización): 98.68 %

norm. = Imágenes normalizadas geométricamente, orig. = Imágenes originales
Validación cruzada (5-fold): 98.60 % ± 0.26 %

STN = Red de Transformación Espacial (*Spatial Transformer Network*), GPA = Análisis de Procrustes Generalizado (*Generalized Procrustes Analysis*)

– No reportado cuantitativamente en el trabajo original

Brecha identificada: Escasa literatura sobre deformación afín por partes aplicada a clasificación médica

Compromiso observado: Normalización reduce exactitud absoluta (-0.58 %) pero aprende características genuinas sin artefactos hospitalarios

Fill rate: 47 % (conservador, preserva valores originales) vs 96-99 % con técnicas de relleno

Margin scale óptimo: 1.05 (5% expansión desde centroide de puntos de referencia)

explota artefactos extrapulmonares (exactitud cae a 95.36 %), mientras que la normalización geométrica alcanza 98.10 % usando exclusivamente la región pulmonar, sugiriendo que la normalización captura características genuinamente diagnósticas.

3.5. Mecanismos de Atención en Clasificación de Imágenes Médicas

Los mecanismos de atención permiten que redes neuronales enfoquen recursos computacionales en regiones o canales informativos, mejorando discriminación de características relevantes para la tarea. Esta sección revisa mecanismos de atención clave y su aplicación en imágenes médicas.

3.5.1. Attention Mechanisms Clásicos

Guo et al. [30] presentaron una revisión exhaustiva que categoriza attention mechanisms según su enfoque: channel attention, spatial attention, temporal attention, y branch attention. Los mecanismos se pueden aplicar en diferentes etapas del procesamiento: pre-processing attention (selección de regiones de interés), intra-processing attention (modulación de features intermedios), y post-processing attention (refinamiento de predicciones).

Squeeze-and-Excitation Networks

Hu et al. [31] introdujeron Squeeze-and-Excitation (SE) blocks, que modelan interdependencias entre canales mediante dos operaciones: *squeeze* agrega información espacial de cada canal mediante global average pooling, produciendo un descriptor de canal; *excitation* aplica una transformación no lineal (dos capas fully-connected con activación ReLU y sigmoid) para aprender ponderaciones adaptativas por canal. SE blocks recalibran feature maps multiplicando cada canal por su peso aprendido, amplificando canales informativos y suprimiendo canales irrelevantes.

SE-Net ganó la competencia ILSVRC 2017 con top-5 error de 2.251 %, demostrando que recalibración de canales mejora discriminación sin sobrecarga computacional significativa. En imágenes médicas, SE blocks han sido adoptados ampliamente en arquitecturas de segmentación (U-Net + SE) y clasificación de patologías, explotando su capacidad de enfatizar características patológicas específicas.

Convolutional Block Attention Module

Woo et al. [32] propusieron Módulo de Atención de Bloque Convolucional (*Convolutional Block Attention Module*, CBAM), que combina channel attention y spatial attention de forma secuencial. CBAM primero aplica channel attention (similar a SE block) para determinar

qué características son importantes, luego aplica spatial attention para determinar *dónde* enfocar. La spatial attention se computa agregando información de canales mediante max pooling y average pooling, concatenando los resultados, y aplicando una convolución 7×7 seguida de sigmoid para producir un mapa de atención espacial.

CBAM demostró mejoras consistentes en ImageNet-1K, MS COCO detection, y VOC 2007 detection, validando que atención dual (canal + espacial) captura complementariedades entre “qué” y “dónde” atender. En imágenes médicas, CBAM ha sido utilizado para segmentación de lesiones, detección de nódulos pulmonares, y clasificación de patologías, donde localización precisa de regiones anormales es crítica.

3.5.2. Coordinate Attention

Hou et al. [13] introdujeron Coordinate Attention, diseñado específicamente para **tareas de localización** en redes móviles eficientes. A diferencia de channel attention que agrega información espacial mediante agrupación global (*global pooling*) perdiendo información posicional, Coordinate Attention factoriza channel attention en dos procesos 1D que codifican información direccional a lo largo de ejes horizontal y vertical.

El mecanismo opera en tres etapas: (1) *coordinate information embedding* aplica pooling 1D a lo largo de altura y ancho separadamente, generando dos feature maps 1D que capturan dependencias de largo alcance en direcciones ortogonales; (2) *coordinate attention generation* concatena los embeddings, aplica transformación convolucional compartida, y split en dos branches que generan attention maps para altura y ancho mediante sigmoid; (3) *coordinate attention multiplication* multiplica el feature map de entrada por los attention maps direccionales, recalibrando features de forma position-aware.

Coordinate Attention supera SE-Net y CBAM en tareas de object detection y semantic segmentation, demostrando que preservación de información posicional es crucial para localización precisa. En el contexto de este trabajo, Coordinate Attention fue seleccionado para el modelo de landmarks porque la tarea requiere localización precisa de puntos anatómicos distribuidos espacialmente. El mecanismo permite que la red capture dependencias entre landmarks (e.g., simetría bilateral entre pulmón izquierdo y derecho) mediante atención direccional, mejorando consistencia de predicciones.

3.5.3. Vision Transformers

Dosovitskiy et al. [33] introdujeron Transformadores de Visión (*Vision Transformers*, ViT), que aplican arquitecturas transformer (originalmente diseñadas para NLP) a image recognition. ViT divide la imagen en parches (*patches*) de 16×16 píxeles, proyecta cada parche a una incrustación (*embedding*), y procesa la secuencia de incrustaciones mediante multi-head self-attention. Self-attention permite que cada parche atienda a todos los demás

parches, capturando dependencias de largo alcance sin restricción de campo receptivo local como en CNN.

ViT alcanzó estado del arte en ImageNet cuando se preentrenó en conjuntos de datos masivos (JFT-300M con 300 millones de imágenes), pero requiere significativamente más datos que CNN para converger. En imágenes médicas, ViT ha mostrado resultados prometedores en clasificación de patologías, segmentación de órganos, y detección de lesiones, frecuentemente superando CNN cuando se combina con transfer learning desde preentrenamiento en ImageNet o conjuntos de datos médicos grandes.

Sin embargo, la aplicación de ViT en conjuntos de datos médicos pequeños (típicamente \approx 50,000 imágenes) es desafiante debido a la necesidad de grandes cantidades de datos para aprender representaciones robustas sin inductive biases de CNNs (locality, translation equivariance). Trabajos recientes híbridos combinan CNNs con transformers, explotando inductive biases de CNNs en capas tempranas y capacidad de atención global de transformers en capas profundas.

3.6. Mejora de Contraste y Preprocesamiento

El preprocesamiento de contraste es crítico en imágenes médicas para resaltar estructuras anatómicas y patológicas, compensando variaciones en protocolos de adquisición. Esta sección revisa métodos de mejora de contraste, enfocándose en CLAHE y su aplicación a radiografías de tórax.

3.6.1. Contrast Limited Adaptive Histogram Equalization

Pizer et al. [34] introdujeron Adaptive Histogram Equalization (AHE), que divide la imagen en tiles y aplica histogram equalization localmente a cada tile. AHE mejora contraste local más efectivamente que histogram equalization global, pero tiende a sobre-amplificar ruido en regiones homogéneas. Zuiderveld [35] propuso Ecualización Adaptativa de Histograma con Limitación de Contraste (*Contrast Limited Adaptive Histogram Equalization*, CLAHE), que introduce un límite de clip para restringir amplificación de contraste, previniendo over-enhancement de ruido.

CLAHE opera mediante los siguientes pasos: (1) la imagen se divide en tiles no solapados (típicamente 8×8 o 4×4), (2) para cada tile se calcula el histograma de intensidades, (3) se limita el histograma mediante clip limit (píxeles que exceden el límite se redistribuyen uniformemente), (4) se aplica histogram equalization al histograma clipped, y (5) se interpolan bilinealmente resultados entre tiles adyacentes para eliminar artefactos de frontera.

CLAHE ha sido ampliamente adoptado en imágenes médicas debido a su efectividad en resaltar detalles locales sin introducir artefactos severos. En radiografías de tórax,

CLAHE mejora visibilidad de opacidades en vidrio esmerilado, consolidaciones, y infiltrados intersticiales característicos de neumonía y COVID-19.

3.6.2. Aplicación a COVID-19 Detection

Rahman et al. [36] exploraron el efecto de técnicas de mejora de imagen (incluyendo CLAHE, ecualización de histograma, corrección gamma, y *unsharp masking*) en detección de COVID-19 usando radiografías de tórax. El estudio encontró que CLAHE con *clip limit* de 2.0 y *tile size* de 8×8 proporcionó la mejor mejora en exactitud de clasificación, aumentando la capacidad de CNNs de discriminar entre COVID-19, Neumonía Viral, y casos normales.

Trabajos recientes han propuesto variantes optimizadas de CLAHE. Un estudio de 2025 introdujo BO-CLAHE (*Bayesian Optimization CLAHE*) para radiografías de tórax neonatales, que optimiza automáticamente *clip limit* y *tile size* mediante búsqueda Bayesiana. Otro trabajo demostró que CLAHE combinado con optimización metaheurística de corrección gamma mejora detectabilidad de lesiones pulmonares en radiografías.

Este trabajo utiliza CLAHE con *clip limit* de 2.0 y *tile size* de 4×4 para el preprocesamiento de imágenes originales previo a la detección de puntos de referencia, basado en validación experimental que mostró que *tile size* menor (4 vs 8) mejora el desempeño de landmark detection. La elección de *tile size* influye en el compromiso entre mejora de contraste local y preservación de estructura global: *tile size* pequeño enfatiza detalles finos pero puede introducir artefactos de bloque; *tile size* grande preserva estructura pero reduce adaptabilidad local.

3.7. Robustez y Generalización

La robustez ante perturbaciones y la generalización a datos de distribuciones diferentes son desafíos fundamentales en IA médica. Esta sección revisa literatura sobre desplazamiento de dominio (*domain shift*), métodos de ensamble, y aumento en tiempo de prueba (*test-time augmentation*).

3.7.1. Desplazamiento de Dominio en Imágenes Médicas

Zech et al. [24] realizaron un estudio seminal sobre desplazamiento de dominio en detección de neumonía en radiografías de tórax, evaluando el desempeño de CNNs entrenados en un hospital (NIH Clinical Center con 112,120 radiografías) en dos hospitales externos (Mount Sinai Hospital y Indiana University Network). El estudio reveló hallazgos alarmantes: en 3 de 5 comparaciones, el desempeño en datos externos fue **significativamente inferior** al desempeño en datos del hospital de entrenamiento. Más preocupante, las CNNs fueron capaces de detectar el hospital de origen de una radiografía con 99.95 % de exactitud,

demonstrando que los modelos explotaban factores de confusión institucionales (marcadores, protocolos de adquisición, propiedades del equipo) en lugar de características patológicas intrínsecas.

Este fenómeno, conocido como *shortcut learning* [2], representa una amenaza fundamental a la confiabilidad clínica de sistemas de IA médica. Geirhos et al. documentaron que CNNs tienden a explotar correlaciones espurias que son predictivas en el conjunto de datos de entrenamiento pero no generalizan a distribuciones diferentes. En imágenes médicas, shortcuts comunes incluyen: artefactos de marca de agua, orientación de paciente (supine vs upright), edad demográfica visible en textura ósea, y características del hospital (equipamiento, calibración).

3.7.2. Estrategias de Mitigación de Desplazamiento de Dominio

La literatura propone múltiples estrategias para mitigar el desplazamiento de dominio:

- **Domain adaptation:** Técnicas de ajuste fino (*fine-tuning*), adversarial training, o feature alignment que adaptan un modelo entrenado en dominio fuente a dominio objetivo con labels limitados o sin labels.
- **Domain generalization:** Métodos que aprenden representaciones invariantes a dominio durante entrenamiento, sin acceso a datos del dominio objetivo. Técnicas incluyen aumento de datos agresiva, meta-learning, y aprendizaje de features causales en lugar de correlacionales.
- **Normalización geométrica:** Reducción de variabilidad extrínseca (pose, orientación, escala) mediante transformaciones espaciales, facilitando que modelos aprendan características intrínsecas relacionadas con patología. Este es el enfoque central de este trabajo.

Surveys recientes sobre domain generalization en imágenes médicas identifican que aumento de datos, normalización de protocolos de adquisición, y ensamble de modelos entrenados en distribuciones diversas son estrategias efectivas para mejorar generalización.

3.7.3. Métodos de Ensamble

Dietterich [37] estableció fundamentos teóricos de ensamble learning, demostrando que combinación de múltiples modelos (aprendices débiles) puede reducir variance, bias, y sobreajuste. En IA médica, ensambles han demostrado mejoras consistentes en robustez y exactitud. Un estudio de 2024 sobre diagnóstico de Alzheimer mediante ensamble de CNNs alcanzó accuracies de 98.57 %, 96.37 %, y 94.22 % en diferentes grupos de clasificación, superando modelos individuales por márgenes significativos.

Este trabajo emplea ensamble de 4 modelos de landmarks entrenados con diferentes semillas aleatorias (123, 321, 111, 666), promediando sus predicciones. El ensamble reduce error de 4.04 px (mejor modelo individual, seed 456) a 3.61 px, demostrando que diversidad introducida por inicialización aleatoria diferente es suficiente para mejorar robustez en este contexto.

3.7.4. Test-Time Augmentation

Test-Time Augmentation (TTA) genera múltiples versiones transformadas de una imagen de test, predice usando el modelo en cada versión, y promedia las predicciones para obtener resultado final. Moshkov et al. [38] demostraron que TTA con transformaciones simples (rotación, flipping horizontal/vertical) mejora significativamente exactitud de segmentación de células en microscopia, aún cuando el modelo ya fue entrenado con aumento de datos.

En tareas de regresión de landmarks con estructura simétrica, TTA requiere cuidado adicional. Este trabajo aplica TTA con horizontal flip y **corrección de simetría bilateral**: después de flip horizontal, las predicciones de landmarks izquierdos y derechos se intercambian ($L_3 \leftrightarrow L_4$, $L_5 \leftrightarrow L_6$, etc.) antes de promediar con predicciones de imagen original. Esta corrección asegura que TTA explota simetría anatómica en lugar de introducir error por promediar landmarks no correspondientes.

3.8. Síntesis y Posicionamiento del Trabajo

Esta sección sintetiza hallazgos de la revisión de literatura, identifica brechas específicas, y posiciona las contribuciones de este trabajo en el contexto del estado del arte.

3.8.1. Brechas Identificadas en la Literatura

El análisis sistemático de la literatura revela tres brechas principales que motivan y justifican este trabajo:

Brecha 1: Escasez de Piecewise Affine Warping para Clasificación

Piecewise affine warping es ampliamente utilizado en face alignment, morphing, y remote sensing para geometric correction, pero su aplicación a **clasificación de imágenes médicas como paso de preprocessamiento** es escasa. La mayoría de trabajos sobre normalización geométrica en imágenes médicas utilizan:

- Transformaciones *rígidas* (traslación, rotación): Adecuadas para órganos rígidos (huesos) pero restrictivas para tejidos deformables.

- Transformaciones *afines globales*: STN y variantes aplican una transformación afín a toda la imagen, sin capacidad de deformación local.
- Transformaciones *no paramétricas* (optical flow, B-splines): Extremadamente flexibles pero propensas a sobreajuste y difíciles de regularizar.

Piecewise affine warping ocupa un punto intermedio en el espectro flexibilidad-regularización: más flexible que transformaciones globales pero más estructurado que deformaciones no paramétricas. Este trabajo es, según conocimiento del autor, uno de los primeros en aplicar piecewise affine warping basado en landmarks anatómicos automáticamente detectados para normalización previa a clasificación de COVID-19.

Brecha 2: Landmark Detection en Chest X-rays para Normalización

Mientras que la detección de puntos de referencia ha sido extensivamente estudiada en imágenes faciales, radiografías de columna vertebral y resonancia magnética cerebral, existe **escasa literatura sobre detección de puntos de referencia anatómicos en radiografías de tórax**, particularmente para definición de contornos pulmonares en contexto de clasificación de neumonía/COVID-19. Los trabajos existentes se enfocan en segmentación semántica completa de pulmones (máscaras densas) en lugar de puntos de referencia dispersos que capturen forma global.

Este trabajo propone un conjunto de 15 landmarks que definen contorno pulmonar bilateral: eje central (L1, L9, L10, L11, L2), contorno izquierdo (L12, L3, L5, L7, L14), y contorno derecho (L13, L4, L6, L8, L15). La selección de landmarks equilibra parsimonia (suficientes para capturar variabilidad de forma) con factibilidad (anotación manual razonable). El error de 3.61 px (1.14% NME) es comparable a estado del arte en facial landmarks, demostrando viabilidad técnica.

Brecha 3: Flujo de Trabajo de Extremo a Extremo: Landmark Detection + GPA + Deformación + Clasificación

No se identificaron trabajos que integren (1) detección automática de landmarks anatómicos, (2) análisis de forma estadístico (GPA) para computar forma canónica, (3) normalización geométrica mediante deformación afín por partes, y (4) clasificación en imágenes normalizadas en un flujo de trabajo coherente de extremo a extremo. Trabajos previos del grupo (Picazo-Castillo, Ayala-Raggi) exploraron normalización mediante cropping inteligente y PCA, pero no utilizaron landmarks explícitos ni deformaciones locales.

Este trabajo contribuye un flujo de trabajo completo que aborda normalización geométrica de forma fundamentada, utilizando análisis de forma estadístico (Análisis de Procrustes Generalizado, GPA) [39, 40] para definir objetivo de normalización (forma canónica consenso)

y deformación afín por partes para transformar cada imagen a dicha forma, reduciendo variabilidad de pose mientras se preserva información de textura local.

3.8.2. Posicionamiento Cuantitativo

Las Tablas 3.1, 3.2, y 3.3 posicionan cuantitativamente este trabajo:

- **Clasificación de COVID-19:** 98.10 % de exactitud, competitivo con el estado del arte (DenseNet121: 98.0 %, RegNetX032: 98.6 %). El desempeño es ligeramente inferior a algunos trabajos recientes pero alcanzado con arquitectura significativamente más ligera (ResNet-18: 11.7M parámetros) y con validación cruzada que demuestra estabilidad del enfoque ($98.60\% \pm 0.26\%$, discutido en Capítulo de Resultados).
- **Landmark detection:** 3.61 px (1.14 % NME) con ensamble + TTA, comparable a Feng et al. (1.47 % NME) en facial landmarks y superior considerando la mayor dificultad de chest X-rays con variabilidad patológica y anatómica. Mejor modelo individual: 4.04 px (1.28 % NME).
- **Normalización geométrica:** Primer trabajo (según conocimiento del autor) en aplicar GPA + deformación afín por partes para clasificación de COVID-19. Comparación directa es difícil debido a escasez de trabajos similares, pero validación cruzada ($98.60\% \pm 0.26\%$) y experimentos comparativos demuestran efectividad del enfoque.

3.8.3. Contribuciones Específicas

En el contexto del estado del arte revisado, las contribuciones principales de este trabajo son:

1. **Método de normalización geométrica local:** Deformación afín por partes basada en landmarks para clasificación de chest X-rays, abordando brecha de transformaciones globales restrictivas.
2. **Pipeline de extremo a extremo:** Integración de landmark detection (ResNet-18 + Coordinate Attention + Wing Loss), GPA para forma canónica, deformación afín por partes (Delaunay triangulation), y clasificación (ResNet-18), demostrando viabilidad de enfoque principled basado en análisis de forma.
3. **Conjunto de 15 landmarks pulmonares:** Definición y anotación manual de landmarks de contorno pulmonar bilateral, con error de ensamble de 3.61 px (1.14 % NME) comparable a facial landmarks.

4. **Validación experimental rigurosa:** Cross-validation de 5-folds ($98.60\% \pm 0.26\%$) y evaluación en conjunto de prueba independiente, demostrando estabilidad del sistema propuesto.
5. **Validación experimental exhaustiva:** Cross-validation de 5-folds, comparación de tres configuraciones (imágenes originales, normalizadas y recortadas con SAHS), y análisis de casos mal clasificados, demostrando rigor metodológico.

3.8.4. Limitaciones y Direcciones Futuras

El estado del arte revisado también sugiere limitaciones del enfoque propuesto y direcciones para investigación futura:

- **Dependencia en calidad de landmarks:** La normalización geométrica depende críticamente de la precisión de landmark detection. Errores de localización se propagan a la deformación, introduciendo distorsiones. Mejora futura podría explorar refinamiento iterativo o detección multi-escala.
- **Forma canónica única:** GPA computa una forma canónica consenso global. Un enfoque más sofisticado podría aprender formas canónicas por clase (COVID-19, Normal, Neumonía Viral), adaptando normalización a patología.
- **Desplazamiento de dominio no resuelto:** Normalización geométrica reduce variabilidad de pose pero no aborda el desplazamiento de dominio institucional (equipamiento, marcadores, demografía). Adaptación de dominio (*domain adaptation*) o aprendizaje federado (*federated learning*) son complementos necesarios para despliegue clínico.
- **Interpretabilidad limitada:** Aunque la deformación facilita comparación visual entre imágenes normalizadas, no proporciona explicaciones causales de predicciones. Integración con técnicas de explainable AI (Grad-CAM, attention maps) es una dirección prometedora.

En conclusión, este trabajo contribuye al estado del arte mediante un enfoque novel de normalización geométrica local para clasificación de COVID-19, abordando brechas identificadas en la literatura y demostrando viabilidad y efectividad mediante validación experimental exhaustiva. Las contribuciones específicas posicionan el trabajo como un avance en la integración de análisis de forma estadístico con aprendizaje profundo para diagnóstico médico automatizado.

Capítulo 4

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.

4.1. 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: preprocessamiento, 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.

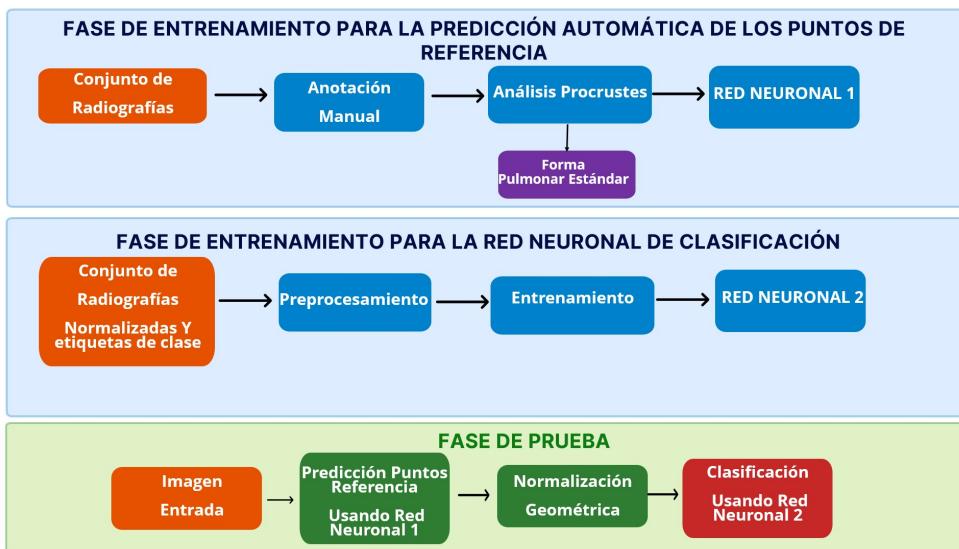


Figura 4.1: Estructura general del sistema en dos fases.

La Figura 4.1 ilustra la relación entre ambas fases del sistema.

4.1.1. Arquitectura del Sistema



Figura 4.2: 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 1: Preprocesamiento. Las imágenes de entrada se someten a un proceso de mejora de contraste mediante el algoritmo CLAHE (*Contrast Limited Adaptive Histogram Equalization*) [35]. Este paso normaliza las variaciones de contraste inherentes a diferentes equipos de adquisición radiográfica. Posteriormente, las imágenes se redimensionan a 224×224 píxeles para su procesamiento por la red neuronal.

Módulo 2: Predicción de Puntos de Referencia. Un modelo basado en ResNet-18 [10] con módulo de Coordinate Attention [13] 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.

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) [39]. 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.

4.1.2. Flujo de Datos

El procesamiento de una imagen sigue el flujo ilustrado en la Figura 4.3 y en la Tabla 4.1. Cada etapa transforma los datos de entrada en una representación apropiada para la siguiente etapa del proceso.

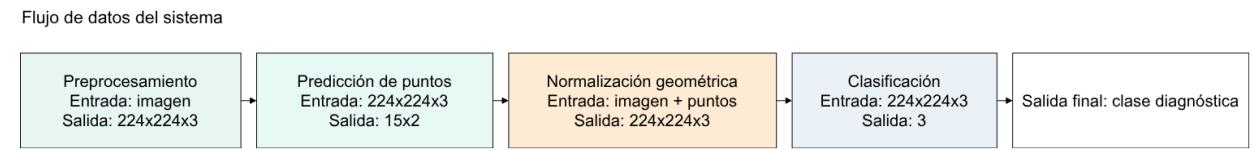


Figura 4.3: Flujo de datos del sistema con entradas, salidas y dimensiones por etapa.

Cuadro 4.1: 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.	15×2
Normalización geométrica	Imagen + puntos de ref.	Imagen normalizada	$224 \times 224 \times 3$
Clasificación	Imagen normalizada	Vector probabilidades	3

4.1.3. 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 [27].

Justificación del diseño modular

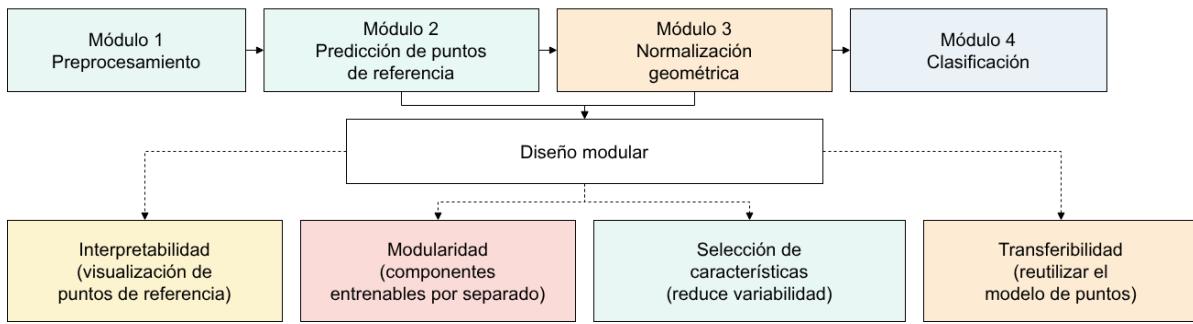


Figura 4.4: Diagrama del diseño modular y sus beneficios. Se muestran los cuatro módulos del sistema y cómo su separación contribuye a la interpretabilidad, modularidad, selección implícita de características y transferibilidad.

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 5.

4.2. 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.

4.2.1. 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 [23, 36]. 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 4.2 presenta la distribución de imágenes por categoría utilizada en este trabajo.

Cuadro 4.2: 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 %
Total	15,153	100 %

Las imágenes originales tienen un tamaño de 299×299 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.

4.2.2. 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 4.5 ilustra la ubicación de cada punto de referencia.

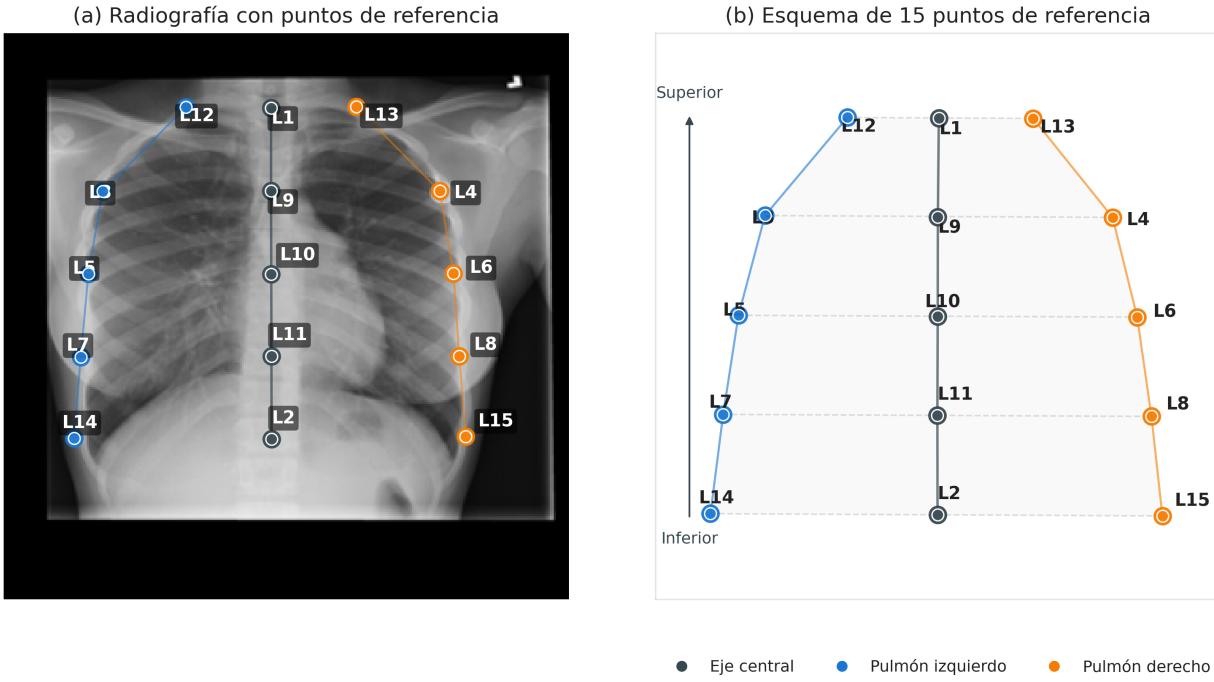


Figura 4.5: Ubicación de los 15 puntos de referencia que definen el contorno pulmonar. (a) Radiografía con puntos de referencia etiquetados. (b) Esquema de distribución espacial. L1 y L2 definen el eje central; L3-L8 delimitan los contornos laterales; L9-L11 dividen el eje central en cuatro segmentos 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 4.7 ilustra la interfaz de la herramienta.

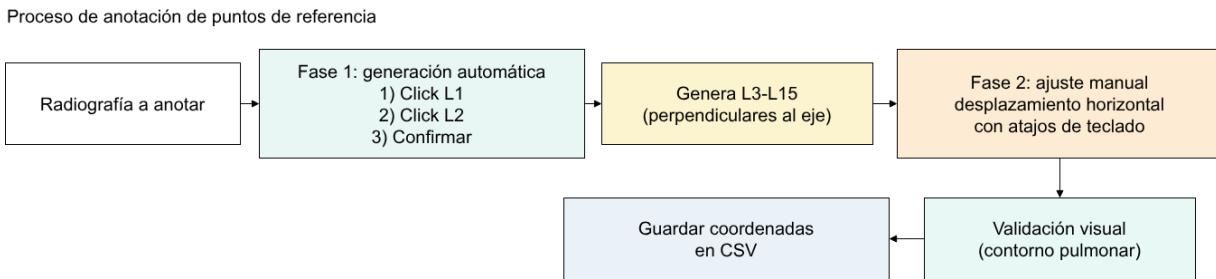


Figura 4.6: Diagrama del proceso de anotación de puntos de referencia. Se muestran las dos fases: generación automática a partir de tres clicks y ajuste manual con validación visual antes de guardar las coordenadas en CSV.

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.

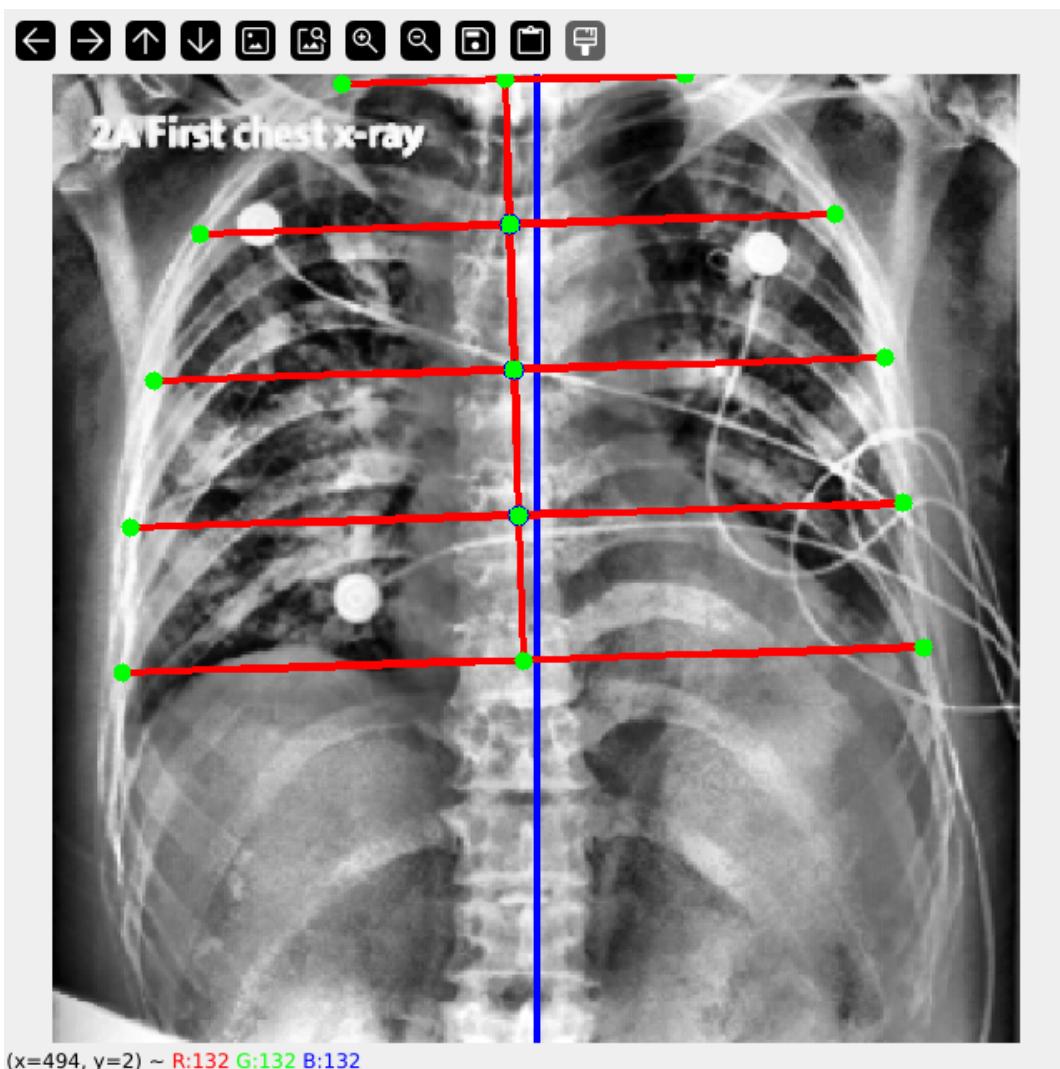


Figura 4.7: Interfaz de la herramienta de anotación de puntos de referencia. La ventana principal muestra la radiografía con una línea vertical central de referencia. Los puntos de referencia se visualizan como puntos verdes conectados por líneas rojas que definen el contorno pulmonar.

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 299×299 píxeles.

Las anotaciones se almacenaron en formato CSV.

La distribución del subconjunto anotado por categoría se presenta en la Tabla 4.3.

Cuadro 4.3: 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 %
Total	957	100 %

4.2.3. 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.

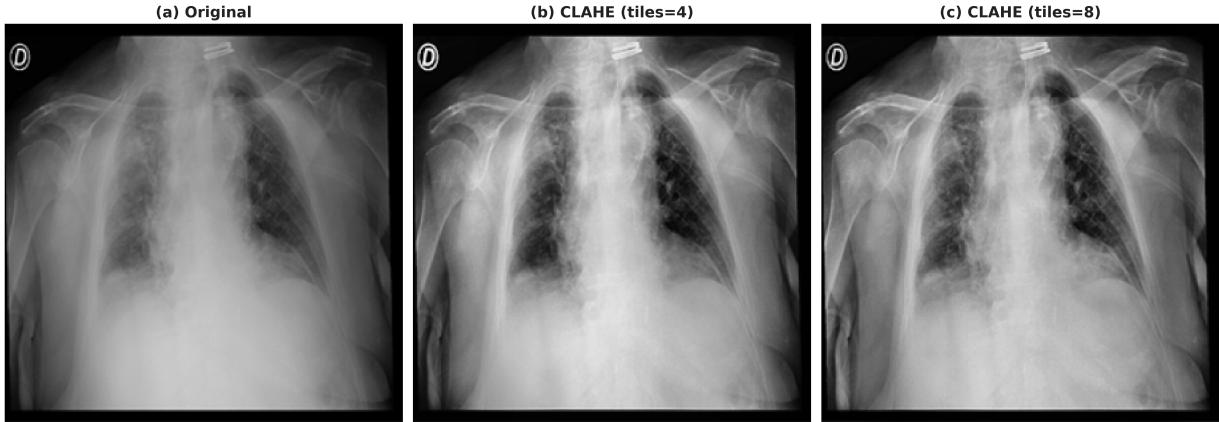


Figura 4.8: Efecto del preprocessamiento CLAHE. Imagen original con bajo contraste en la región pulmonar. Imagen después de aplicar CLAHE con clip limit = 2,0 y tile size = 4, mostrando mejor definición de estructuras pulmonares, Imagen después de aplicar CLAHE con clip limit = 2,0 y tile size = 8.

Mejora de Contraste mediante CLAHE

Se aplica el algoritmo *Contrast Limited Adaptive Histogram Equalization* (CLAHE) [34] 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:** 4×4 . Tamaño de las regiones para ecualización local. Un valor menor produce una adaptación más fina pero aumenta el tiempo de cómputo.

La Figura 4.8 muestra el efecto del preprocessamiento CLAHE sobre una radiografía de ejemplo.

Redimensionamiento

Las imágenes se redimensionan de su tamaño original (299×299 píxeles) a 224×224 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.

4.2.4. 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 Figura 4.9 resume la división estratificada por clase y la proporción asignada a cada subconjunto.

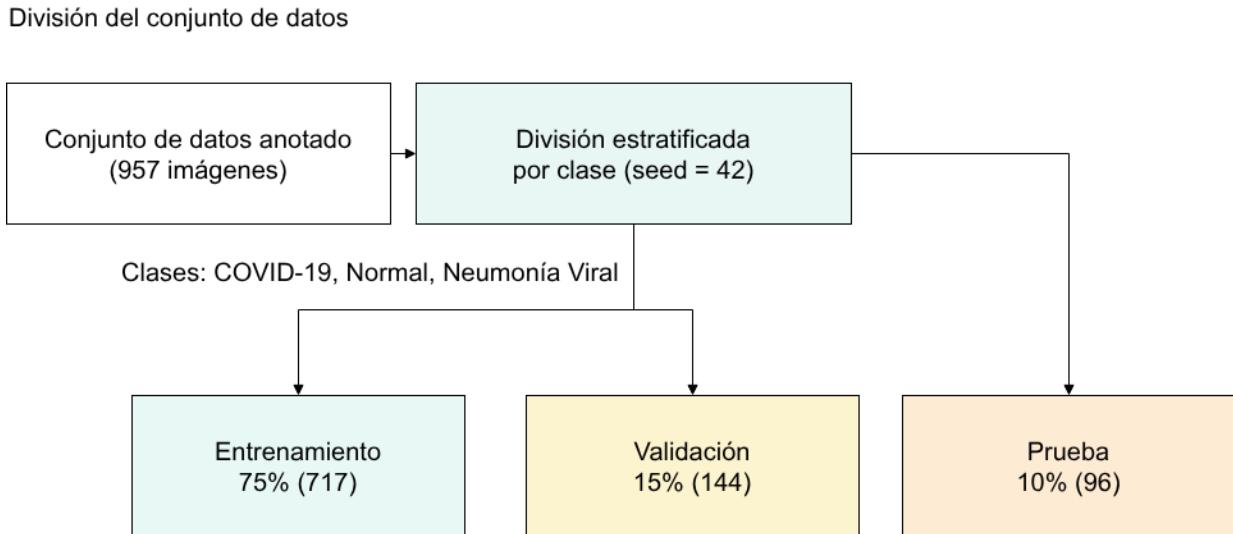


Figura 4.9: Esquema de la división estratificada del conjunto anotado en entrenamiento, validación y prueba, preservando las proporciones por clase.

La Tabla 4.4 presenta la distribución resultante para el subconjunto anotado con puntos de referencia (957 imágenes).

Cuadro 4.4: 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
Total	306	468	183	957

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

4.3. 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.

4.3.1. 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 4.10 presenta el diagrama de la arquitectura completa.

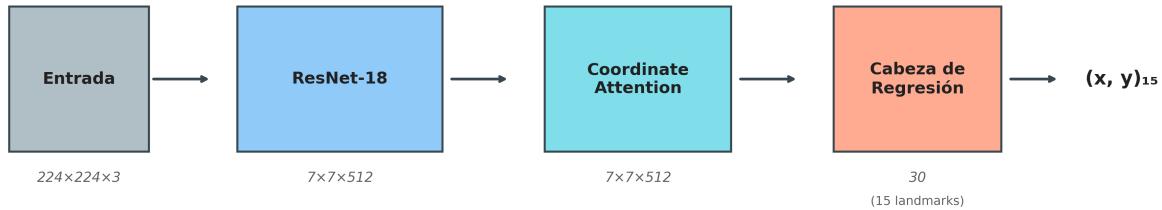


Figura 4.10: Arquitectura del modelo de predicción de puntos de referencia. ResNet-18 extrae características de alto nivel, el módulo Coordinate Attention incorpora información posicional, y la cabeza de regresión predice las 30 coordenadas (15 puntos de referencia \times 2 coordenadas).

ResNet-18

Como extractor de características (denominado *extractor de características* en la literatura de aprendizaje profundo) se utiliza ResNet-18 [10], 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 extracto 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 4.5 detalla las capas del extracto de características utilizadas.

Cuadro 4.5: 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 7×7 , stride 2	$112 \times 112 \times 64$	9,408
bn1 + relu	BatchNorm + ReLU	$112 \times 112 \times 64$	128
maxpool	MaxPool 3×3 , 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
Total extractor de características		—	11,176,512

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 [13] 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 [31], 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. En términos simples, el módulo resume la información por filas y columnas, combina ambas vistas para estimar pesos de importancia (0–1) y pondera el mapa de características. La Figura 4.11 resume este flujo conceptual.

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

$$z_c^h(h) = \frac{1}{W} \sum_{w=1}^W x_c(h, w) \quad (4.1)$$

De manera análoga, para cada columna w :

$$z_c^w(w) = \frac{1}{H} \sum_{h=1}^H x_c(h, w) \quad (4.2)$$

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.

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 del mapa se pondera según la importancia de su fila y su columna:

$$y_c(h, w) = x_c(h, w) \cdot a_c^h(h) \cdot a_c^w(w) \quad (4.3)$$

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

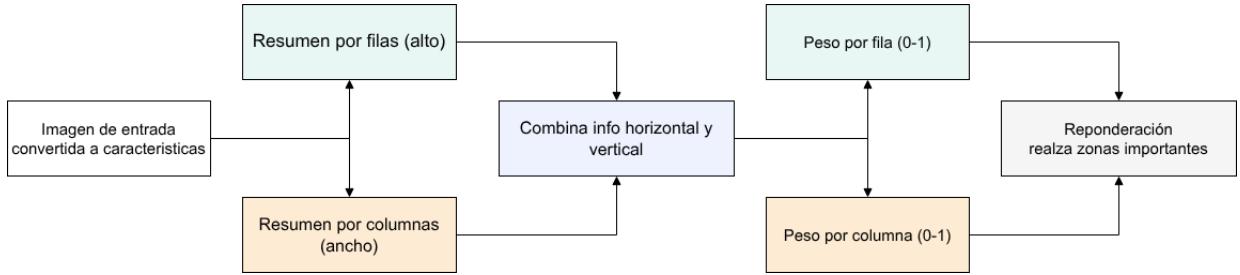


Figura 4.11: Mecanismo del módulo Coordinate Attention. A partir del mapa de características de entrada, se resumen filas y columnas, se combinan ambas vistas y se generan pesos entre 0 y 1 por fila y por columna. Estos pesos reponderán el mapa, realzando zonas relevantes y atenuando las menos informativas.

Parámetros del módulo. La Tabla 4.6 presenta los componentes y parámetros del módulo Coordinate Attention.

Cuadro 4.6: 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
Total Coordinate Attention		24,608

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 (7×7 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.

La Figura 4.12 resume el flujo de procesamiento de la cabeza de regresión.

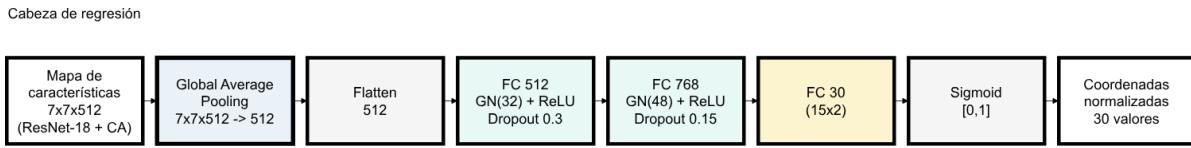


Figura 4.12: Diagrama de la cabeza de regresión. Se muestra el paso desde el mapa de características ($7 \times 7 \times 512$) hasta las 30 coordenadas normalizadas, incluyendo *Global Average Pooling*, capas totalmente conectadas con normalización por grupos y dropout, y la activación sigmoide final.

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

- **Normalización por grupos** (*Group Normalization*) [41]: 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 4.7 detalla la arquitectura completa de la cabeza de regresión.

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.

Cuadro 4.7: Arquitectura de la cabeza de regresión con normalización por grupos.

Capa	Operación	Salida	Parámetros
avgpool	AdaptiveAvgPool2d(512)	512	0
flatten	Flatten	512	0
fc1	Linear(512, 512)	512	262,656
gn1	GroupNorm(32, 512) 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)	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
Total cabeza	—	—	683,038

4.3.2. 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 [12] fue diseñada específicamente para localización de puntos de referencia, combinando las ventajas de ambos enfoques:

- **Para errores pequeños** (menores a ω 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 ω píxeles): Se comporta de manera similar a L1, proporcionando correcciones estables sin penalizaciones extremas que podrían desestabilizar el entrenamiento.

La formulación matemática es:

$$\text{wing}(x) = \begin{cases} \omega \ln \left(1 + \frac{|x|}{\epsilon} \right) & \text{si } |x| < \omega \\ |x| - C & \text{de otro modo} \end{cases} \quad (4.4)$$

donde x es la diferencia entre la coordenada predicha y la real, ω define el umbral entre los dos comportamientos, ϵ 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.

La Figura 4.13 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):

$$\mathcal{L}_{\text{total}} = \frac{1}{30} \sum_{i=1}^{30} \text{wing}(\hat{c}_i - c_i) \quad (4.5)$$

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

4.3.3. 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* [20] 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

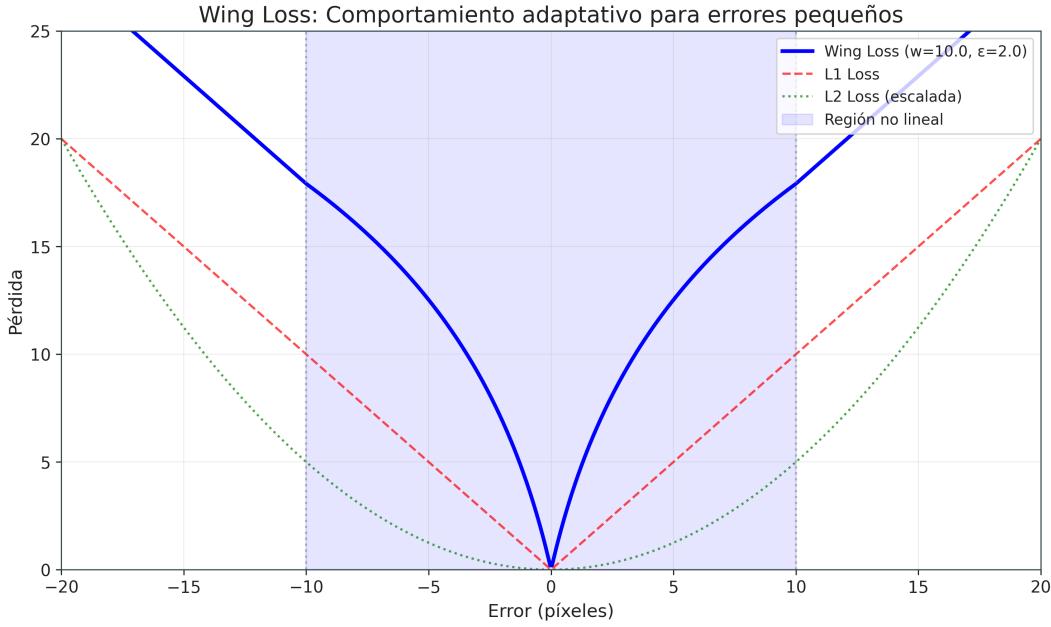


Figura 4.13: Comparación de Wing Loss con pérdidas L1 y L2. Wing Loss proporciona mayor penalización para errores pequeños (región logarítmica) incentivando el refinamiento fino, mientras mantiene estabilidad para errores grandes (región lineal).

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 4.8.

Cuadro 4.8: Configuración de entrenamiento - Fase 1.

Parámetro	Valor
Épocas máximas	15
Tasa de aprendizaje	1×10^{-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 “descongelan” 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 (2×10^{-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 (2×10^{-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* [42] 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 4.9 detalla la configuración de la Fase 2.

Cuadro 4.9: Configuración de entrenamiento - Fase 2.

Parámetro	Valor
Épocas máximas	100
Tasa de aprendizaje (extractor de características + CA)	2×10^{-5}
Tasa de aprendizaje (cabeza)	2×10^{-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↔L4, L5↔L6, L7↔L8, L12↔L13, L14↔L15).
2. **Rotación aleatoria** (± 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.

4.3.4. 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 4.10 consolida todos los hiperparámetros utilizados, organizados por categoría, para facilitar la reproducibilidad de los experimentos.

Cuadro 4.10: 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	ω	10.0 px
	ϵ	2.0 px
	Normalizado	Sí (escalado a [0,1])
Fase 1	Épocas	15
	Tasa de aprendizaje	1×10^{-3}
	Tamaño de lote	16
	Extractor características	Congelado
	Parada temprana	5 épocas
Fase 2	Épocas	100
	LR extractor características + CA	2×10^{-5}
	LR cabeza	2×10^{-4}
	Tamaño de lote	8
	Parada temprana	15 épocas
	Scheduler	Cosine Annealing

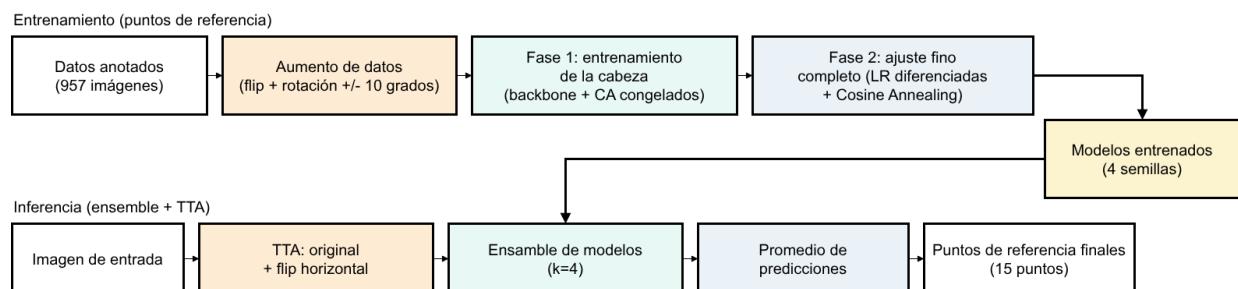


Figura 4.14: Diagrama de la estrategia de entrenamiento y del esquema de inferencia con ensamble y TTA. La parte superior resume las dos fases de entrenamiento; la parte inferior muestra el uso de TTA y el promedio de modelos para obtener los puntos de referencia finales.

4.3.5. 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 *ensemble* (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.

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:

$$\hat{\mathbf{L}}_{\text{ensemble}} = \frac{1}{4} \sum_{k=1}^4 \hat{\mathbf{L}}_k \quad (4.6)$$

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.

4.4. 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.

4.4.1. 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) [39, 43] 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

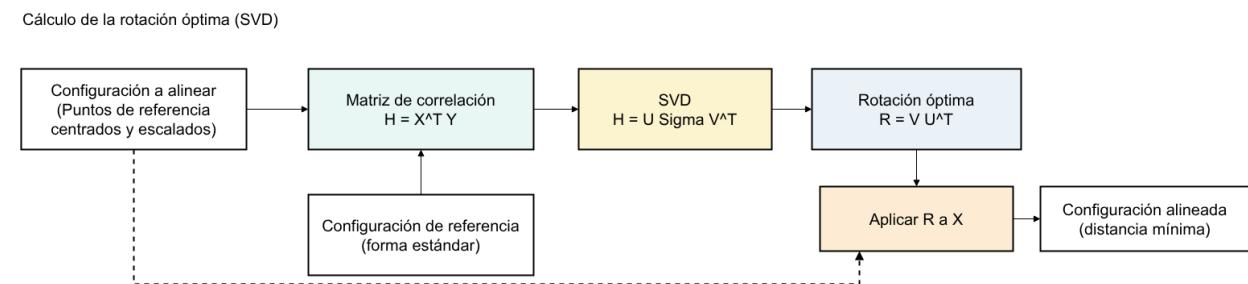


Figura 4.15: Esquema del cálculo de la rotación óptima mediante SVD. A partir de dos configuraciones centradas y escaladas, se construye la matriz de correlación, se aplica la descomposición SVD y se obtiene la rotación que alinea la configuración con la referencia.

El ángulo de rotación óptimo se calcula mediante una técnica de álgebra lineal llamada Descomposición en Valores Singulares (SVD) [44], 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.

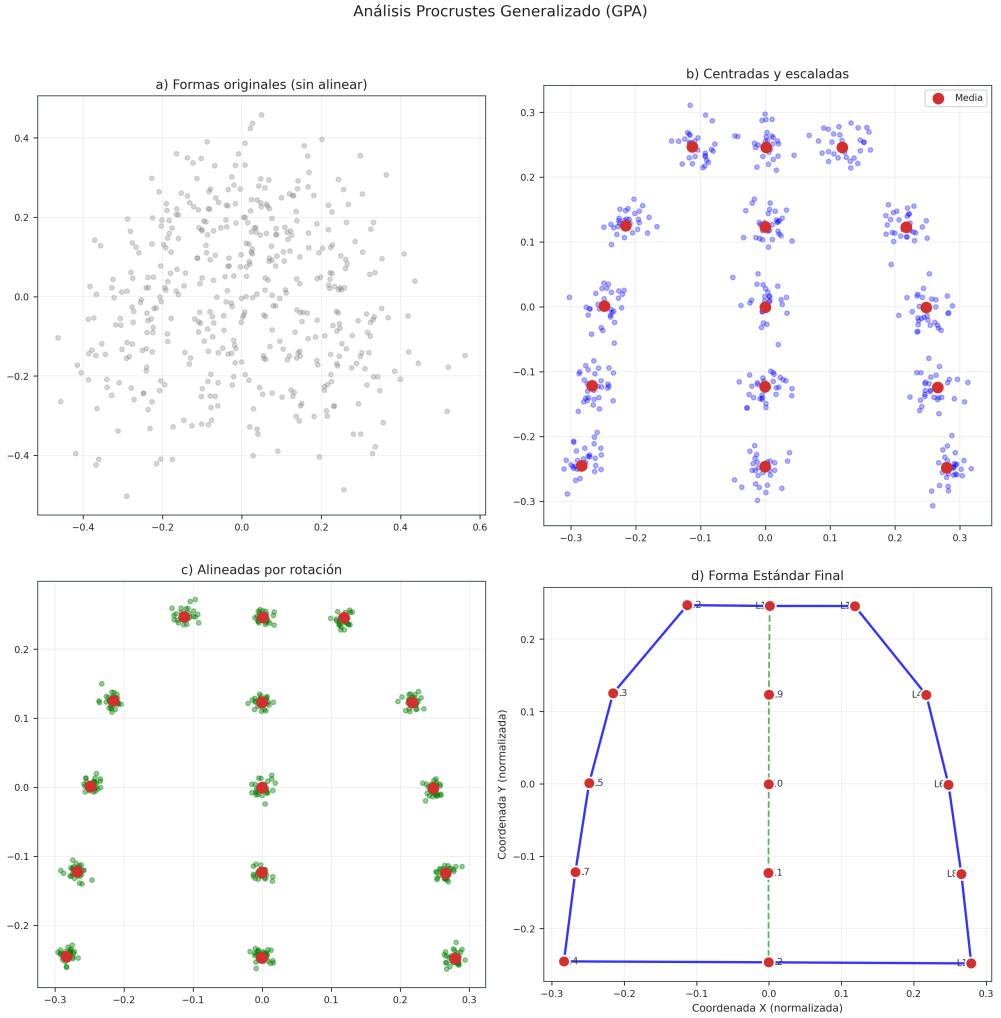


Figura 4.16: Proceso de Análisis Procrustes Generalizado. (a) Las 957 configuraciones de puntos de referencia originales muestran variabilidad en posición, escala y orientación. (b) Despues del centrado y escalado, las formas comparten origen y norma unitaria. (c) La alineación rotacional minimiza las diferencias residuales. (d) La forma estándar representa el consenso estadístico del conjunto.

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 τ , 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 τ **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 224×224 píxeles, dejando un margen del 10% en los bordes.

La Figura 4.16 ilustra el proceso de GPA aplicado al conjunto de landmarks.

4.4.2. 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.

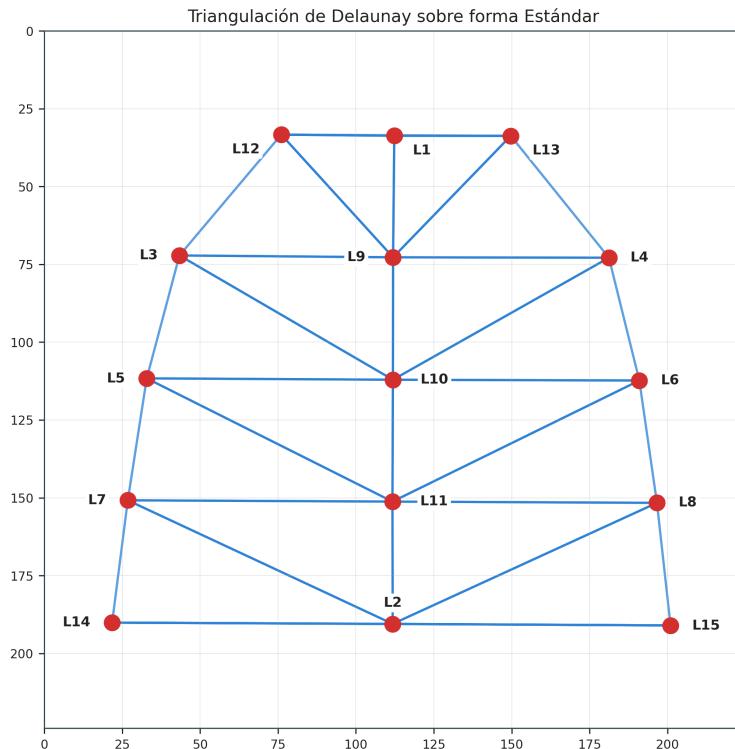


Figura 4.17: Triangulación de Delaunay sobre los 15 puntos de referencia de la forma estándar. Los triángulos definen las regiones donde se aplicarán transformaciones afines independientes durante el proceso de deformación.

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 [45] 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 [29].

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 4.17 muestra la triangulación resultante sobre la forma estándar.

4.4.3. 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. La deformación 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.

La Figura 4.18 muestra la comparación visual entre una radiografía original y su versión normalizada geométricamente.

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

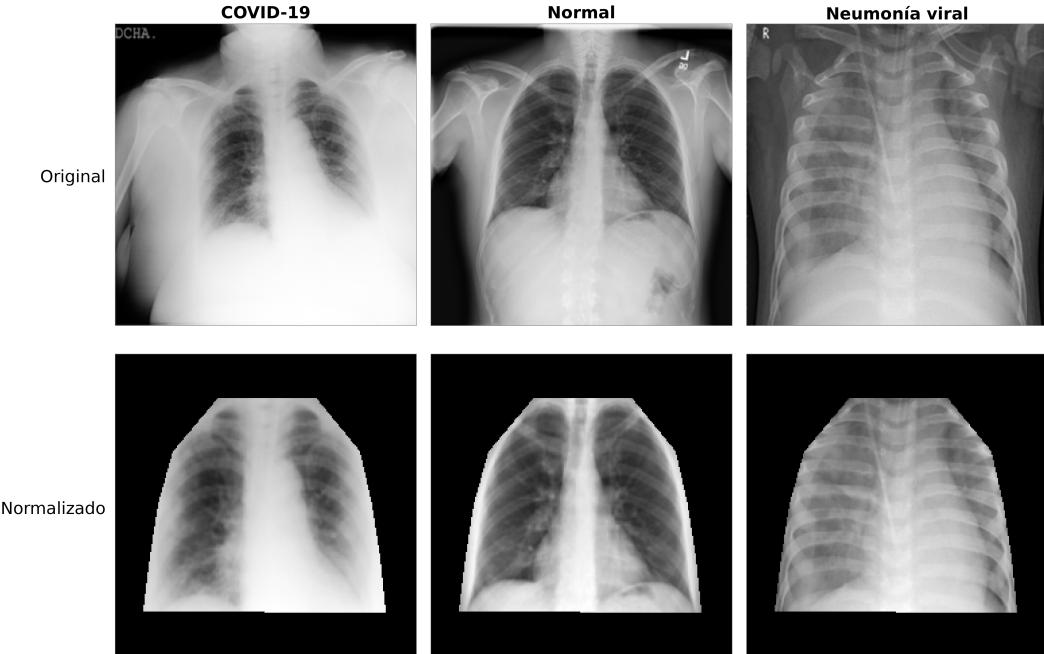


Figura 4.18: Comparación de radiografías originales y normalizadas por clase. Columnas: COVID-19, Normal y Neumonía viral. Filas: Original (arriba) y Normalizado (abajo). La normalización geométrica mediante deformación afín por partes alinea la región pulmonar con la forma estándar y reduce variabilidad de pose y escala.

parallelismo. 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.

Transformación afín de un triángulo

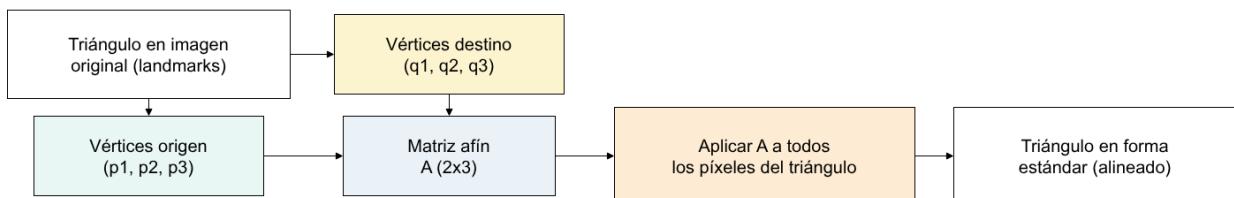


Figura 4.19: Esquema de la transformación afín de un triángulo. A partir de los vértices en la imagen original y sus correspondencias en la forma estándar, se calcula la matriz afín y se aplica a los píxeles del triángulo para obtener la región alineada.

Proceso de Deformación

La 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.

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
-

Durante la deformación, 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.

4.4.4. Proceso Completo de Normalización

Esta sección resume cómo se integran todos los componentes descritos anteriormente para normalizar una radiografía. La Figura 4.20 presenta el diagrama de flujo del proceso completo.

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 4.3) localiza los 15 puntos del contorno pulmonar en la radiografía de entrada.



Figura 4.20: Proceso completo de normalización geométrica. El sistema transforma una radiografía de entrada en una imagen geométricamente normalizada mediante la secuencia de predicción de puntos de referencia, forma estándar, triangulación y deformación afín por partes.

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. **Deformación:** 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 224×224 píxeles donde los pulmones tienen siempre la misma posición, tamaño y orientación.

4.5. 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.

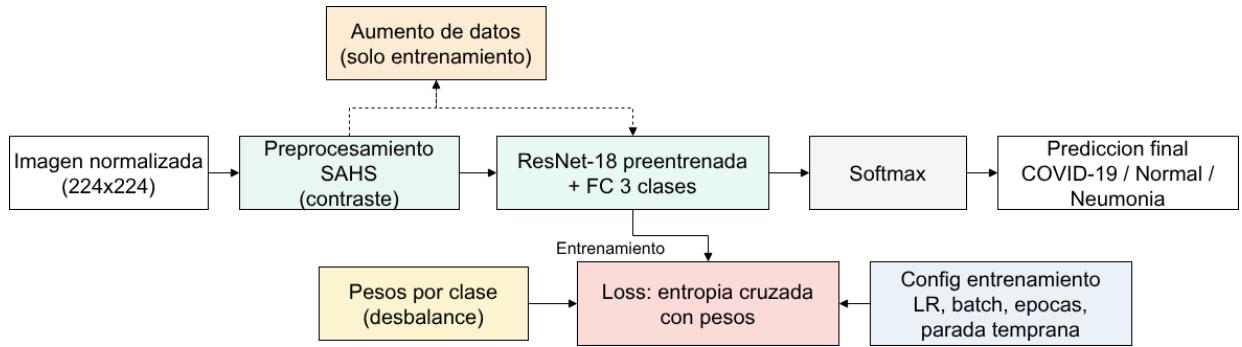


Figura 4.21: Diagrama general del módulo de clasificación. Se muestra el flujo de inferencia desde la imagen normalizada hasta la predicción final, y los componentes usados solo durante entrenamiento (aumento de datos, pesos por clase y entropía cruzada).

4.5.1. 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*) [9] 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.

Efecto del preprocesamiento SAHS sobre imágenes normalizadas

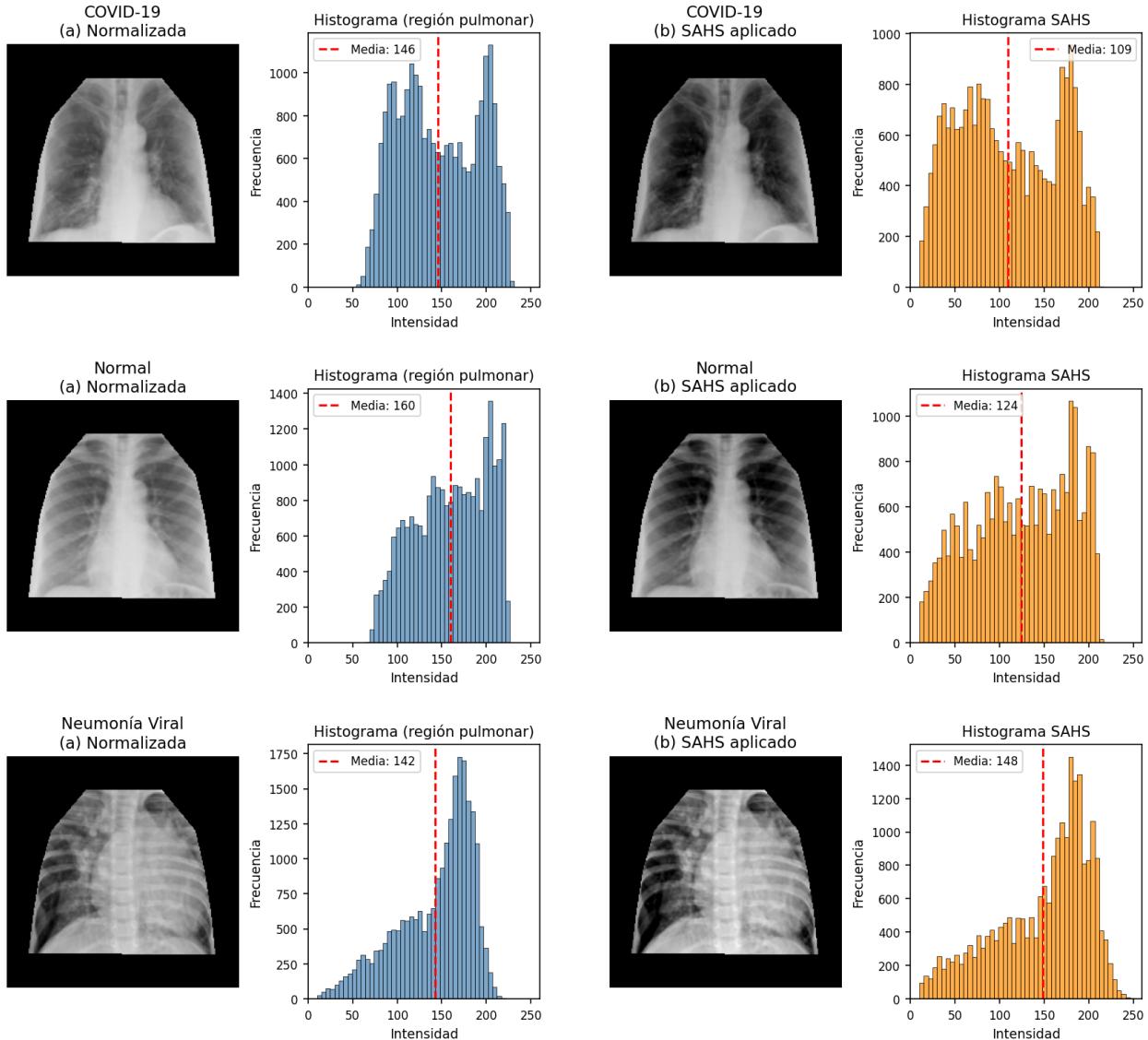


Figura 4.22: Efecto del preprocesamiento SAHS sobre imágenes normalizadas geométricamente. Cada fila muestra una categoría diagnóstica diferente (COVID-19, Normal, Neumonía Viral). Las columnas muestran: (a) imagen después de la normalización geométrica, histograma original, (b) imagen con SAHS aplicado, e histograma resultante. Se observa cómo SAHS redistribuye las intensidades mejorando el contraste sin amplificar artefactos.

- **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 4.22 muestra el efecto de aplicar SAHS a imágenes normalizadas geométricamente de las tres categorías diagnósticas.

4.5.2. Arquitectura del Clasificador

Para la tarea de clasificación se seleccionó la arquitectura ResNet-18 [10], una red neuronal convolucional diseñada específicamente para el análisis de imágenes. El nombre 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 [10].
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 [4].

4.5.3. 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 [21].

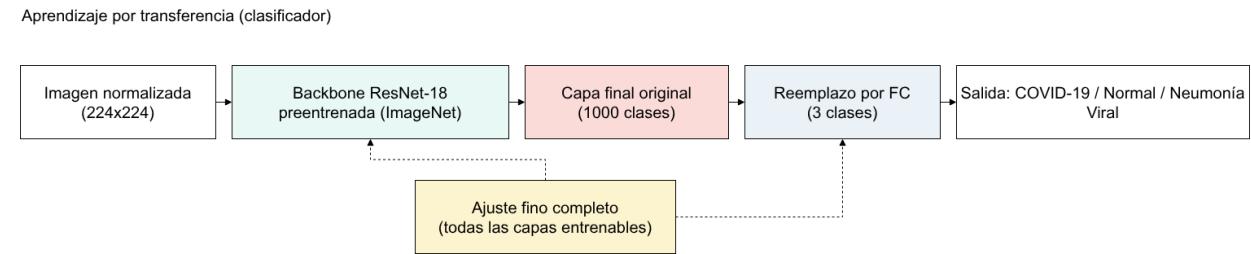


Figura 4.23: Esquema del aprendizaje por transferencia en el clasificador. Se reutiliza el backbone ResNet-18 preentrenado, se reemplaza la capa final por tres salidas y se realiza ajuste fino completo para adaptar el modelo al diagnóstico de COVID-19, Normal y Neumonía Viral.

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.

4.5.4. 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 4.2. La Tabla 4.11 muestra la distribución por categoría diagnóstica.

Cuadro 4.11: 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 %)
Total	11,364	2,271	1,518	15,153

Manejo del Desbalance de Clases

Como se observa en la Tabla 4.11, 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 4.12 muestra los pesos calculados.

Cuadro 4.12: 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 4.13 resume la configuración utilizada.

Cuadro 4.13: Parámetros de configuración del entrenamiento del clasificador.

Parámetro	Valor
Épocas máximas	50
Tamaño de lote	32 imágenes
Tasa de aprendizaje	1×10^{-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

Parada Temprana

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

4.5.5. 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 4.12). 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).

4.5.6. 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 224×224 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 4.24 ilustra ejemplos de las transformaciones aplicadas durante el entrenamiento.

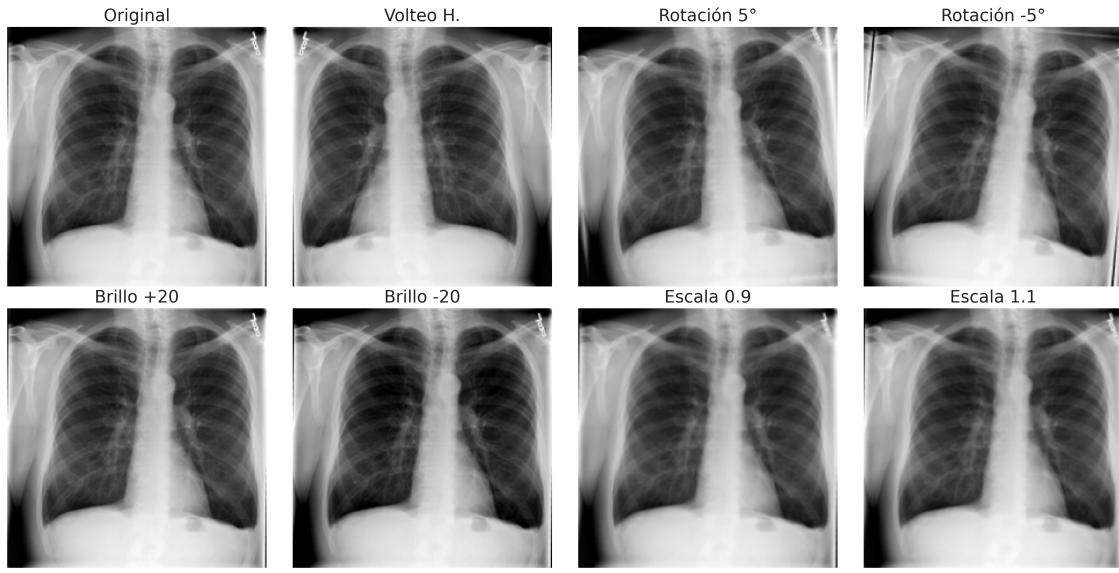


Figura 4.24: Ejemplos de transformaciones de aumento de datos aplicadas durante el entrenamiento del clasificador. Las transformaciones preservan la integridad diagnóstica de las imágenes mientras aumentan la variabilidad del conjunto de entrenamiento.

4.5.7. Resumen de la Configuración

La Tabla 4.14 consolida los aspectos principales del módulo de clasificación descritos en esta sección.

Cuadro 4.14: 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

4.6. 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.

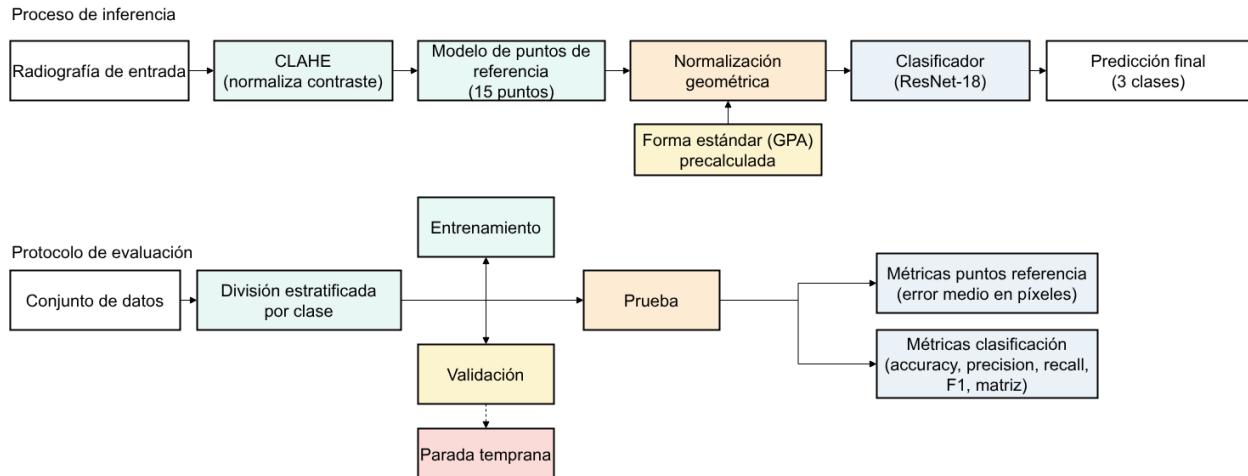


Figura 4.25: Diagrama del flujo de inferencia y del protocolo de evaluación. La parte superior resume el pipeline desde la radiografía de entrada hasta la predicción final; la parte inferior muestra la partición estratificada y el uso de validación y prueba para reportar métricas.

4.6.1. 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.

4.6.2. 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 224×224 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 5 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 5

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.

5.1. 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.

5.1.1. Precisión del Sistema

La Tabla 5.1 presenta los resultados de detección de puntos de referencia. Se compara el mejor modelo individual contra el sistema final que combina cuatro modelos.

Cuadro 5.1: Precisión de detección de puntos de referencia. El error se mide en píxeles sobre imágenes de 224×224 .

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 224×224 píxeles, este error equivale al 1.14 % de la diagonal de la imagen (*Normalized Mean Error*, NME). El error mediano de 3.07 píxeles indica que la mitad de las predicciones tienen un error menor o igual a este valor.

5.1.2. 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 5.1 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.

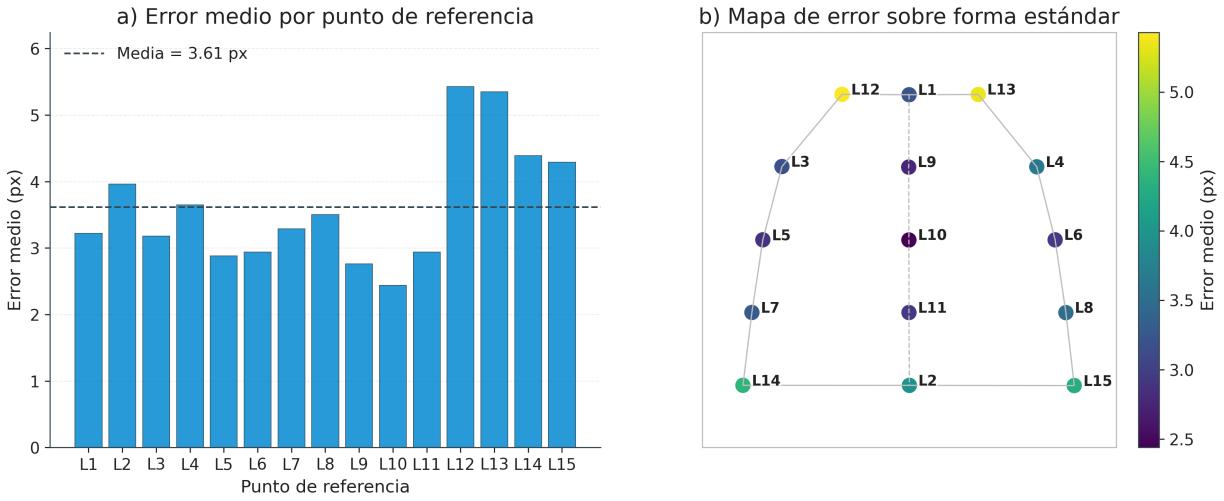


Figura 5.1: Distribución del error de detección por punto de referencia. (a) Error medio en píxeles para cada punto L1–L15. (b) Visualización sobre la forma estándar pulmonar, donde el color indica el error medio (escala continua). Los puntos del eje central presentan la mayor precisión.

La Figura 5.2 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).

5.1.3. Resumen

El sistema de detección de puntos de referencia alcanza una precisión de 3.61 píxeles en promedio, equivalente al 1.14 % NME (error normalizado por la diagonal 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.

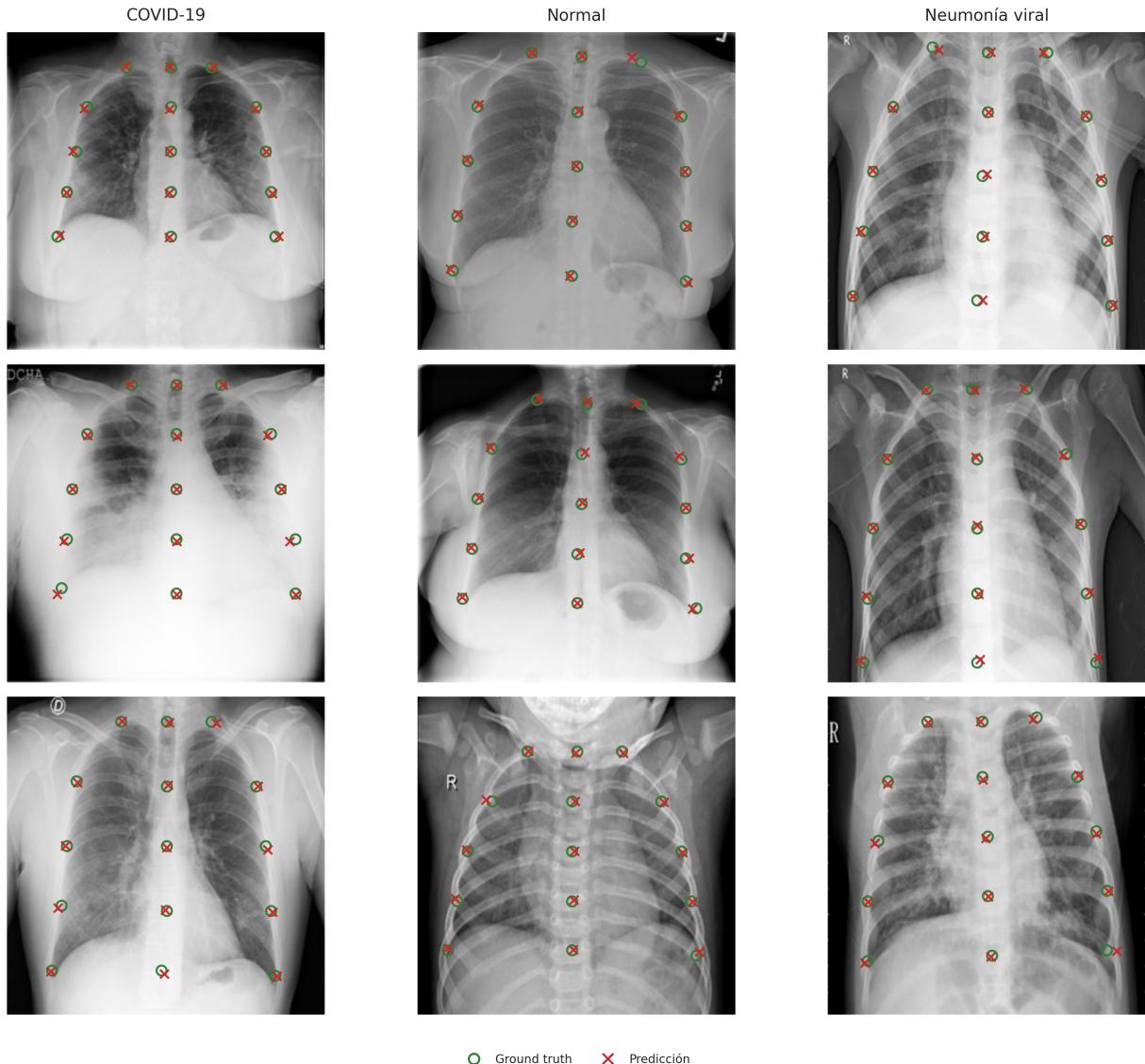


Figura 5.2: Ejemplos de predicción automática de puntos de referencia sobre imágenes reales. Cada fila muestra dos ejemplos de una categoría diagnóstica (Normal, COVID-19, Neumonía Viral). Las cruces indican las predicciones del sistema. El modelo mantiene precisión consistente independientemente del tipo de patología presente.

5.2. 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.

5.2.1. Forma Estándar de Referencia

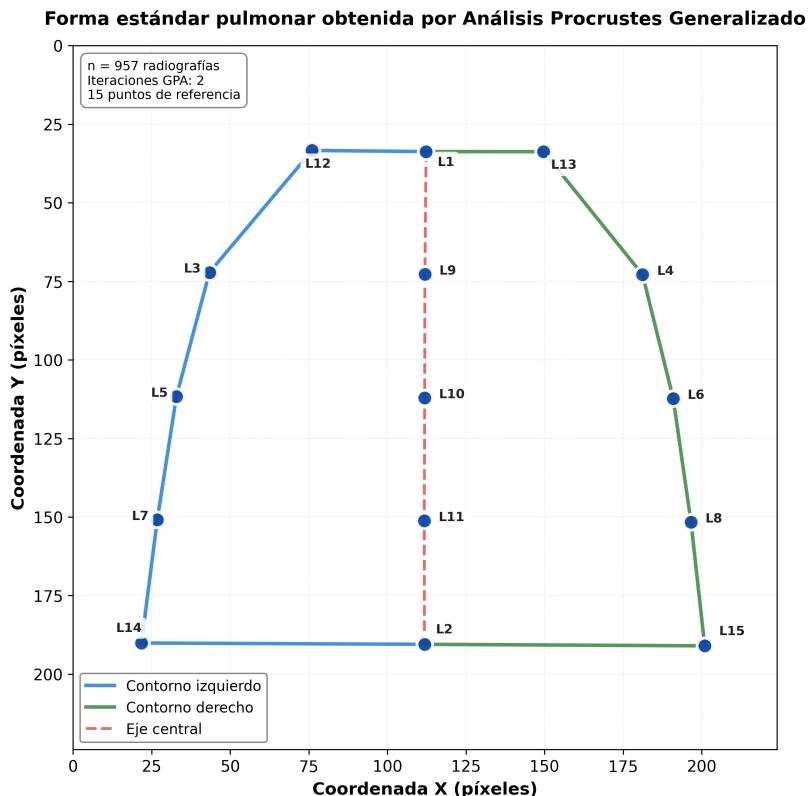


Figura 5.3: Forma estándar pulmonar de referencia obtenida mediante Análisis Procrustes Generalizado. Se muestra la configuración promedio de los 15 puntos de referencia calculada a partir de 957 radiografías anotadas manualmente. El eje central (línea roja discontinua) conecta los puntos L1 → L9 → L10 → 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.

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

5.2.2. División en Triángulos para Transformación

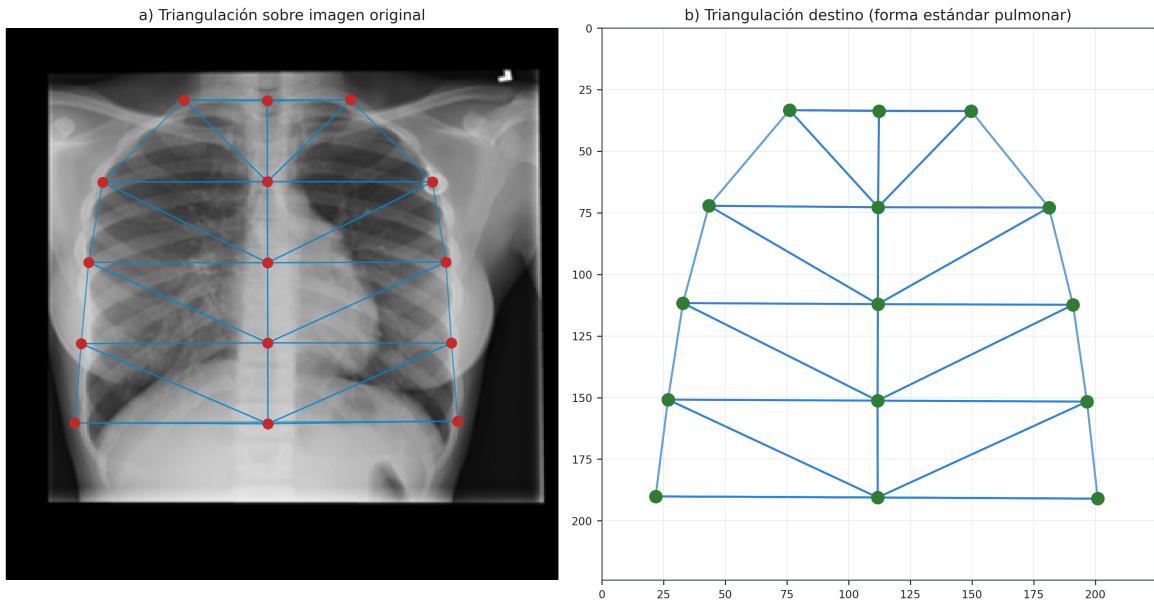


Figura 5.4: Triangulación de Delaunay para la normalización geométrica. (a) 16 triángulos generados sobre una radiografía original utilizando los 15 puntos de referencia detectados automáticamente. (b) Triangulación correspondiente sobre la forma estándar pulmonar. Cada triángulo de la imagen original 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.

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

5.2.3. Ejemplos de Normalización

La Figura 5.5 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.

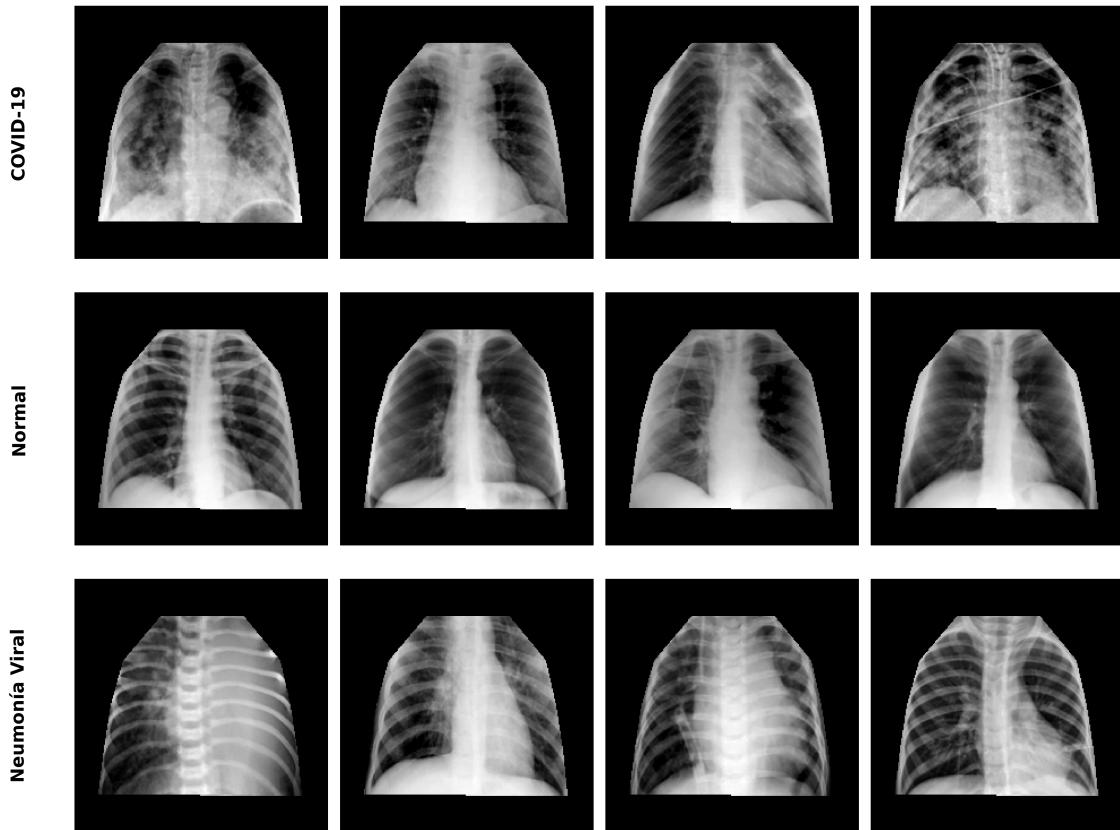


Figura 5.5: Ejemplos de imágenes normalizadas geométricamente y procesadas con SAHS por categoría. Cada fila muestra cuatro ejemplos del conjunto de prueba de una categoría (COVID-19, Normal, Neumonía Viral). Las imágenes resultantes presentan la región pulmonar alineada de forma consistente con contraste 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.

5.2.4. 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.
- 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.

5.3. 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.

5.3.1. Rendimiento General

El clasificador fue evaluado sobre un conjunto de prueba de 1,895 imágenes que no fueron utilizadas durante el entrenamiento. La Tabla 5.2 presenta las métricas principales de rendimiento.

Cuadro 5.2: Rendimiento del clasificador sobre el conjunto de prueba (1,895 imágenes).

Métrica	Valor
Exactitud (Accuracy)	98.10 %
F1-Score Macro	97.17 %
F1-Score Ponderado	98.09 %
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.

5.3.2. 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 5.3 resume el promedio y la desviación estándar de las métricas en validación.

Cuadro 5.3: Resultados de validación cruzada ($k=5$) sobre train+val.

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

5.3.3. Rendimiento por Categoría

La Tabla 5.4 desglosa el rendimiento del clasificador para cada categoría diagnóstica.

Cuadro 5.4: 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.

5.3.4. Análisis de Errores

La Tabla 5.5 muestra la matriz de confusión, que detalla cómo se distribuyen las predicciones correctas e incorrectas entre las categorías.

Cuadro 5.5: 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	436	16	0		452
Normal	4	1,266	4		1,274
Neumonía Viral	0	12	157		169
Total	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 5.6 presenta la matriz de confusión de manera visual, facilitando la identificación de patrones de error mediante un mapa de calor.

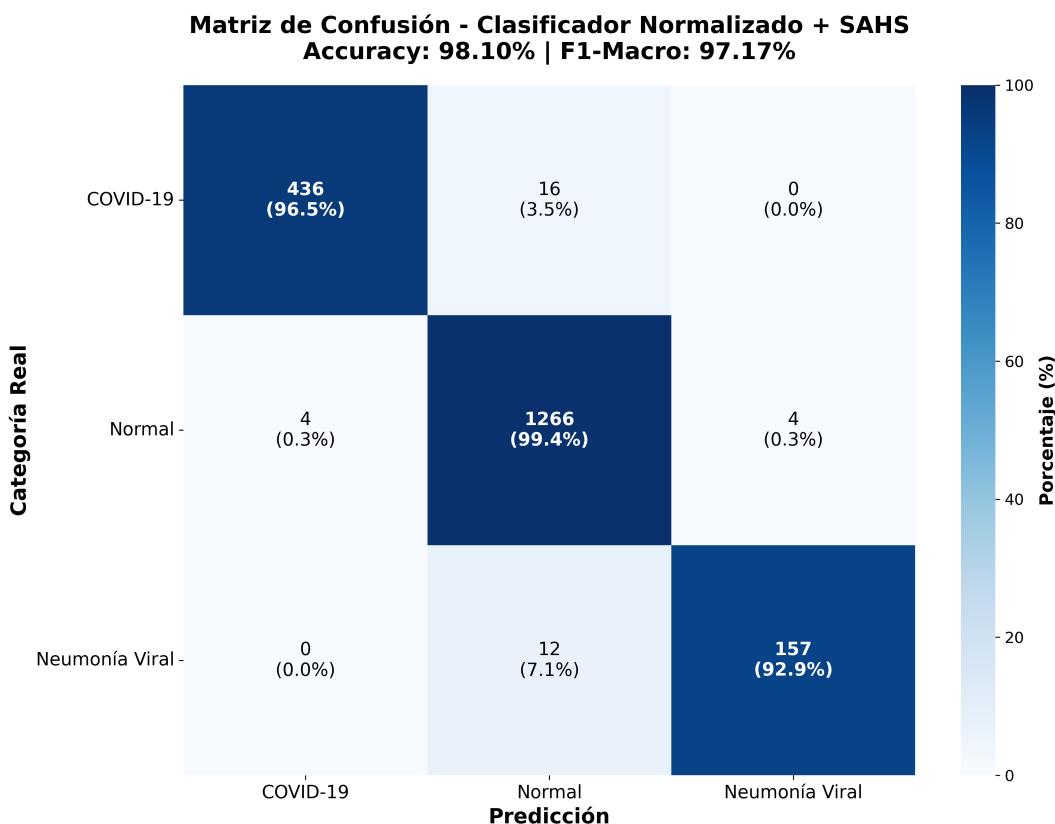


Figura 5.6: Matriz de confusión del sistema de clasificación presentada visualmente. Los valores en la diagonal representan clasificaciones correctas. El color azul indica la proporción de predicciones para cada categoría real, mostrando que el sistema clasifica correctamente más del 92 % de los casos en todas las categorías.

La Figura 5.7 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.

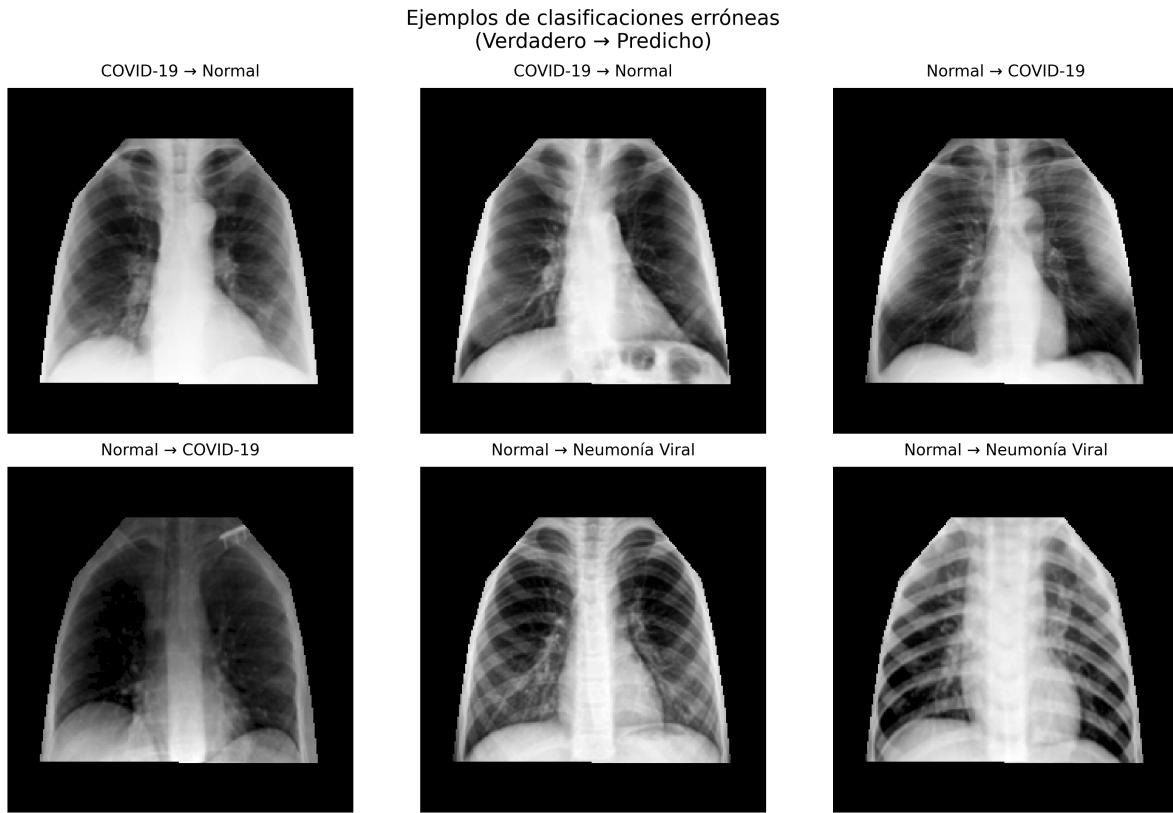


Figura 5.7: Ejemplos de casos mal clasificados del conjunto de prueba usando imágenes normalizadas geométricamente con SAHS. Se muestran 6 casos reales distribuidos entre los cuatro tipos de errores encontrados, con su confianza de predicción. Cada ejemplo indica la clase verdadera y la clase predicha (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 %).

5.3.5. 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 5.8 muestra ejemplos visuales de las tres configuraciones, todas con SAHS aplicado, para las tres categorías diagnósticas.

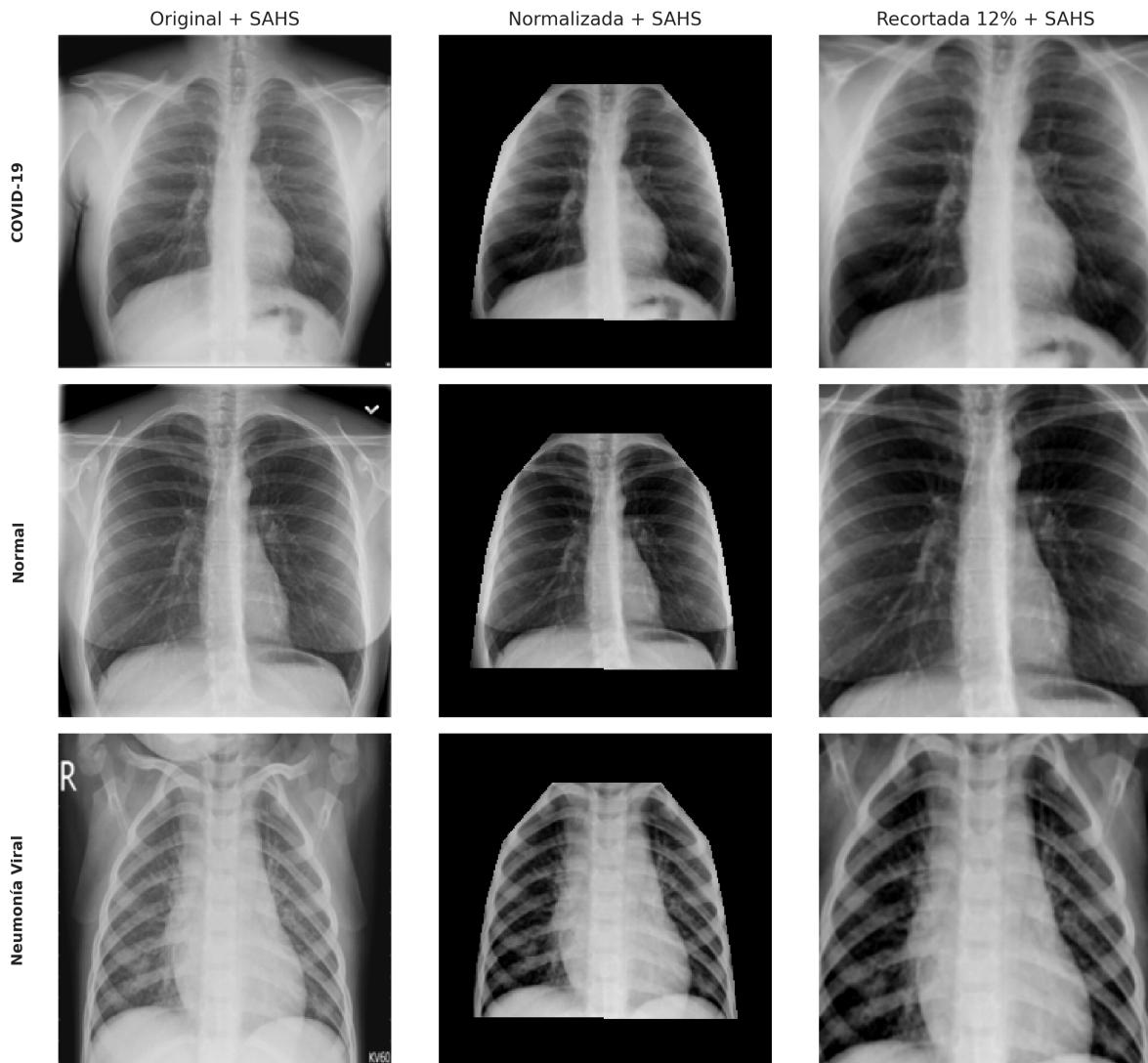


Figura 5.8: Comparación visual del preprocesamiento con SAHS. Filas: categorías diagnósticas (COVID-19, Normal, Neumonía Viral). Columnas: Original + SAHS, Normalizado + SAHS, Recortado 12% + SAHS.

La Tabla 5.6 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 5.6: 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	95.36 %	94.28 %	-3.32 %

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 5.3). 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 5.9 presenta visualmente las matrices de confusión de las tres configuraciones evaluadas.

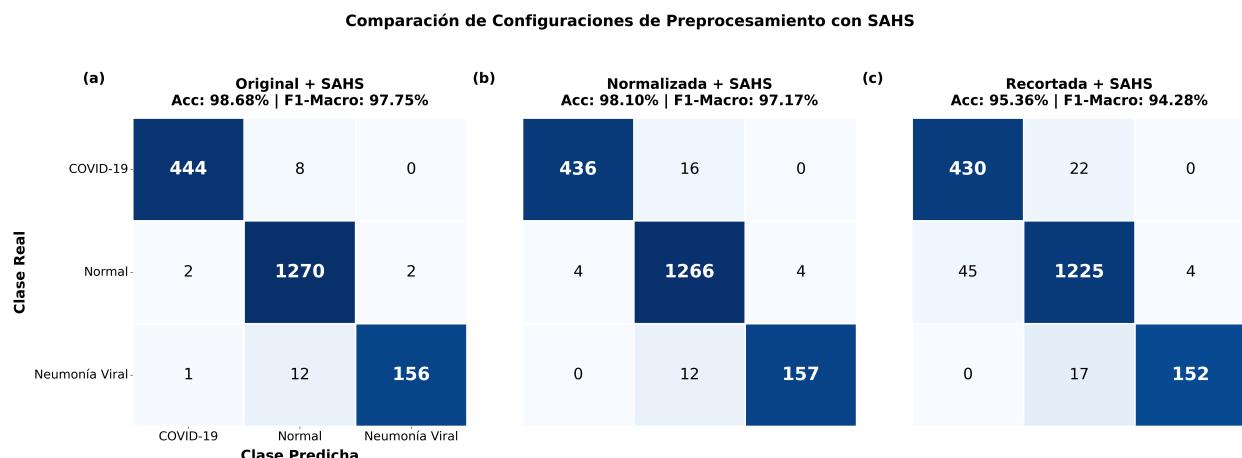


Figura 5.9: Comparación de matrices de confusión para las tres configuraciones de preprocesamiento. (a) Original + SAHS: 98.68 % de exactitud. (b) Normalizado + SAHS (sistema propuesto): 98.10 % de exactitud. (c) Recortado + SAHS: 95.36 % de exactitud. La caída de 3.32 puntos al recortar los bordes evidencia que las imágenes originales utilizan artefactos hospitalarios como características espurias.

5.3.6. 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 6

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.

6.1. 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.

6.1.1. 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.

6.1.2. 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.

6.2. Validación de la Hipótesis

6.2.1. 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.”

6.2.2. 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 (± 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.

6.2.3. 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.

6.2.4. 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.

6.3. Implicaciones del Trabajo

6.3.1. 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.

6.3.2. 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 + deformación afín por partes 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.

6.3.3. 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.

6.4. Limitaciones del Estudio

Las siguientes limitaciones deben considerarse al interpretar los resultados:

6.4.1. 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.

6.4.2. 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.

6.4.3. 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.

6.5. Trabajos Futuros

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

6.5.1. 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.

6.5.2. 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.

6.5.3. 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 (ápex pulmonar, ángulo cardiofrénico, cúpula diafragmática), mejorando la interpretabilidad.
3. **Aprendizaje extremo a extremo:** Explorar un enfoque extremo a extremo 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 [27] 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).

6.5.4. 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.

6.5.5. 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.

6.6. 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

Active Shape Models (ASM)

Método estadístico tradicional de detección de landmarks que aprende patrones de variabilidad de forma a partir de anotaciones manuales, empleando un proceso iterativo de refinamiento análogo a Active Contour Models (Snakes) para ajustar una forma modelo a imágenes. Requieren inicialización cercana a la solución y son sensibles a variaciones significativas de apariencia.

Adaptive Wing Loss

Extensión de Wing Loss para heatmap regression que introduce ponderación adaptativa según el tipo de píxel (foreground vs background), combinando la función Wing con un Weighted Loss Map que asigna mayor peso a píxeles críticos, permitiendo que el entrenamiento se enfoque en regiones relevantes para localización de landmarks. Superó el estado del arte en benchmarks de face alignment (COFW, 300W, WFLW).

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.

AlexNet

Arquitectura pionera de red neuronal convolucional profunda (8 capas, 5 convolucionales + 3 fully connected) que ganó la competencia ImageNet 2012, marcando el inicio de la era moderna del aprendizaje profundo en visión por computadora. Introdujo el uso de ReLU, dropout y procesamiento en GPU para entrenar redes profundas.

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.

Attention Gates

Módulo de atención espacial que aprende máscaras de importancia para ponderar selectivamente diferentes regiones de un mapa de características, permitiendo que la red enfoque recursos computacionales en áreas anatómicamente relevantes. Utilizado en STERN para normalización jerárquica de radiografías.

Attention Maps

Visualizaciones que muestran dónde se enfoca la atención de la red durante predicciones, obtenidas directamente de mecanismos de atención (self-attention, coordinate attention, CBAM), útiles para interpretabilidad de decisiones del modelo y validación de que la red se enfoca en regiones anatómicamente relevantes.

AUC (Area Under the Curve / Área Bajo la Curva)

Métrica que calcula el área bajo la curva ROC (Receiver Operating Characteristic), proporcionando una evaluación de rendimiento de clasificación binaria independiente del threshold de decisión. Valores cercanos a 1.0 indican excelente discriminación entre clases positivas y negativas.

B

BIMCV-COVID19

Conjunto de datos español de COVID-19 en radiografías de tórax con metadata clínica detallada (información demográfica, marcadores de severidad, evolución temporal), desarrollado por el Biosignal Interpretation and Medical Computing Laboratory de Valencia. Facilita investigación clínica mediante información complementaria más allá de imágenes.

B-splines (Basis Splines)

Curvas suaves paramétricas definidas por puntos de control, utilizadas para representar deformaciones no paramétricas complejas en transformaciones de imágenes médicas. Ofrecen extrema flexibilidad para modelar deformaciones locales pero pueden requerir regularización para evitar sobreajuste.

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.

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.

CBAM (*Convolutional Block Attention Module*)

Módulo de atención que combina channel attention y spatial attention de forma secuencial, permitiendo que la red aprenda qué características son importantes (channel attention mediante global pooling) y dónde enfocar (spatial attention mediante agregación de canales). Demostró mejoras consistentes en ImageNet-1K, MS COCO y VOC 2007.

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.

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 = 4×4 píxeles.

ChestX-ray14

Conjunto de datos público de 112,120 radiografías de tórax frontales con anotaciones de 14 patologías torácicas diferentes, proporcionado por NIH Clinical Center. Ampliamente utilizado para preentrenamiento y evaluación de modelos de clasificación de enfermedades pulmonares en imágenes médicas.

CheXNet

Modelo de clasificación de patologías pulmonares basado en DenseNet-121 preentrenado en ImageNet, que excede el desempeño promedio de radiólogos en detección de neumonía en radiografías de tórax según evaluación en ChestX-ray14. Estableció una línea base importante para clasificación de patologías pulmonares mediante transfer learning.

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 Regression

Enfoque de detección de landmarks que predice directamente las coordenadas (x, y) de cada punto de referencia como salida numérica continua de la red, sin representación espacial intermedia. Conceptualmente simple y eficiente computacionalmente, pero enfrenta desafíos de convergencia debido a la naturaleza no lineal del mapeo imagen-coordenadas. Utilizado en este trabajo para detección de 15 landmarks pulmonares.

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.

COVID-Net

Arquitectura CNN diseñada específicamente para detección de COVID-19 mediante principios de diseño humano-en-el-loop, incorporando módulos de expansión-compresión que permiten aprendizaje eficiente de representaciones discriminativas. Alcanzó 93.3 % de exactitud en clasificación de tres clases (COVID-19, neumonía viral, normal) en el conjunto

de datos COVIDx.

COVIDx

Conjunto de datos público de radiografías de tórax específicamente compilado para detección de COVID-19, conteniendo 13,975 imágenes de 13,870 pacientes únicos, recopiladas de múltiples fuentes públicas incluyendo repositorios de investigación y datos hospitalarios. Utilizado para entrenamiento de COVID-Net.

Cosine Annealing

Esquema de ajuste de tasa de aprendizaje que varía el valor siguiendo una función cosenoideal, 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.

DenseNet

Arquitectura CNN que conecta cada capa con todas las capas anteriores mediante conexiones densas (cada capa recibe como entrada la concatenación de salidas de todas las capas previas), mejorando la propagación de gradientes, la reutilización de características y reduciendo el número de parámetros. DenseNet-121 es utilizado en CheXNet para clasificación de patologías pulmonares.

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.

Domain Adaptation (Adaptación de dominio)

Conjunto de técnicas que adaptan un modelo entrenado en dominio fuente a dominio objetivo, con acceso limitado o sin acceso a labels en el dominio objetivo, mediante fine-tuning, adversarial training o feature alignment. Crítico en imágenes médicas para mitigar desplazamiento de dominio entre hospitales con diferentes equipos y protocolos.

Domain Generalization (Generalización de dominio)

Enfoque que entrena modelos para aprender representaciones robustas invariantes a cambios de dominio, sin acceso a datos del dominio objetivo durante entrenamiento, mediante aumento agresivo, meta-learning o aprendizaje de características causales en lugar de correlacionales. Complementario a normalización geométrica para robustez ante variabilidad institucional.

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.

EfficientNet

Familia de arquitecturas CNN que optimiza el escalamiento equilibrado de profundidad, ancho y resolución de entrada mediante Neural Architecture Search (NAS), logrando mejor compromiso entre exactitud y eficiencia computacional que arquitecturas diseñadas manualmente. Introduce el concepto de compound scaling para escalamiento óptimo de redes.

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.

Embeddings (Incrustaciones)

Representaciones vectoriales de dimensión fija que codifican información semántica de entradas (palabras, parches de imagen, características), aprendidas durante entrenamiento para ser significativas para la tarea. En Vision Transformers, cada parche de imagen se proyecta a un embedding antes de procesamiento mediante self-attention.

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 euclíadiana entre las coordenadas predichas y las anotadas manualmente: $\text{Error} = \sqrt{(x_{\text{pred}} - x_{\text{real}})^2 + (y_{\text{pred}} - y_{\text{real}})^2}$.

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

Explainable AI (XAI) (Inteligencia Artificial Explicable)

Campo de investigación enfocado en desarrollar técnicas que permiten interpretar y explicar decisiones de modelos de aprendizaje profundo, crítico en aplicaciones médicas donde la confianza clínica requiere comprensión de razonamiento del modelo. Incluye técnicas como Grad-CAM, saliency maps, attention visualization y métodos de interpretabilidad post-hoc.

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: $F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$.

$\frac{\text{Precisión}\cdot\text{Sensibilidad}}{\text{Precisión}+\text{Sensibilidad}}$, 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.

Federated Learning (Aprendizaje Federado)

Paradigma de aprendizaje distribuido donde múltiples dispositivos o instituciones entran colaborativamente un modelo compartido sin centralizar datos sensibles, preservando privacidad de pacientes. Relevante en contextos médicos donde compartir datos está regulado por leyes de protección de datos (HIPAA, GDPR).

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.

Grad-CAM (*Gradient-weighted Class Activation Mapping*)

Técnica de visualización que genera mapas de importancia espacial calculando gradientes de la clase predicha respecto a activaciones de la última capa convolucional, permitiendo interpretar qué regiones de la imagen influyen en decisiones del modelo. Ampliamente utilizada en diagnóstico asistido por IA para validar que modelos se enfocan en regiones patológicas relevantes.

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

Heatmap Regression

Enfoque de detección de landmarks que predice un mapa 2D de probabilidades para cada punto de referencia, típicamente modelado como una distribución Gaussiana centrada en la ubicación verdadera. La ubicación final se obtiene mediante búsqueda del máximo (argmax) o soft-argmax diferenciable. Proporciona generalización espacial superior a coordinate regression pero introduce sobrecarga computacional.

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.

ImageNet-1K

Versión estándar de ImageNet con 1,000 clases de objetos y aproximadamente 1.2 millones de imágenes de entrenamiento, utilizada como benchmark principal para validación de arquitecturas de visión por computadora. ILSVRC (ImageNet Large Scale Visual Recognition Challenge) impulsa el desarrollo de nuevas arquitecturas mediante competencias anuales.

Inductive Biases (Sesgos inductivos)

Suposiciones incorporadas en la arquitectura de un modelo que facilitan el aprendizaje de ciertos tipos de patrones. CNNs tienen inductive biases de locality (convoluciones procesan regiones locales) y translation equivariance (detección invariante a posición). Vision Transformers carecen de estos biases, requiriendo más datos para convergencia.

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 la deformación afín por partes 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.

J

JFT-300M

Conjunto de datos interno de Google con 300 millones de imágenes etiquetadas de forma semi-supervisada, utilizado para preentrenamiento masivo de Vision Transformers. Demuestra que ViT requiere conjuntos de datos significativamente más grandes que CNNs para alcanzar rendimiento competitivo debido a la ausencia de inductive biases.

K

Kernel (Núcleo convolucional / Filtro)

Matriz de pesos pequeña (típicamente 3×3 , 5×5 o 7×7) 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)

Puntos de control sobre la silueta pulmonar que representan posiciones geométricas, no estructuras anatómicas específicas. En este trabajo, 15 puntos definen el contorno pulmonar bilateral: eje central (L1, L9, L10, L11, L2 a lo largo de la línea media), contorno izquierdo (L12, L3, L5, L7, L14) y contorno derecho (L13, L4, L6, L8, L15). Estos puntos sirven como base para la normalización geométrica mediante deformación afín por partes.

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 la deformación. 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.

MS COCO (*Microsoft Common Objects in Context*)

Conjunto de datos masivo para detección, segmentación y captioning de objetos con múltiples categorías (80 clases) y anotaciones densas (segmentación a nivel de instancia). Ampliamente utilizado para validación de arquitecturas, mecanismos de atención y métodos de detección de objetos en contextos naturales.

Multi-head Attention (Atención multi-cabeza)

Variante de self-attention que ejecuta múltiples operaciones de atención en paralelo con proyecciones lineales diferentes (h cabezas), permitiendo que el modelo capture distintos tipos de dependencias en diferentes subespacios de representación. Las salidas se concatenan y proyectan para obtener resultado final. Componente fundamental de arquitecturas transformer.

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.

NME (*Normalized Mean Error* / Error Medio Normalizado)

Métrica normalizada de error de localización de landmarks que expresa el error promedio en píxeles como porcentaje de una distancia de normalización característica (distancia entre ojos en facial landmarks, diagonal de imagen en chest X-rays). Permite comparación de rendimiento independiente de resolución de imagen.

$$\text{Fórmula: } \text{NME} = \frac{1}{N} \sum_{i=1}^N \frac{\|\mathbf{p}_i - \hat{\mathbf{p}}_i\|_2}{d_{\text{norm}}} \times 100\%$$

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 deformación afín por partes, 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 deformación de radiografías.

Optical Flow (Flujo óptico)

Técnica de visión por computadora que estima el campo vectorial de movimiento de píxeles entre frames consecutivos o entre imágenes relacionadas, utilizada para transformaciones no paramétricas de deformación en registro de imágenes médicas. Proporciona flexibilidad extrema pero puede requerir regularización para evitar deformaciones no realistas.

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.

Patches (Parches)

Divisiones rectangulares no solapadas de la imagen (típicamente 16×16 píxeles en Vision Transformers) que se procesan como tokens individuales. Cada parche se aplana y proyecta a un embedding de dimensión fija antes de procesamiento mediante self-attention, permitiendo que ViT trate imágenes como secuencias similares a NLP.

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.

RegNetX032

Arquitectura CNN de la familia RegNet regularizada mediante búsqueda neural de arquitecturas (NAS), diseñada para optimizar el compromiso entre eficiencia computacional y precisión en tareas de clasificación. Reportada con 98.6 % de exactitud en detección de COVID-19 en literatura reciente.

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: $I_{max} = \mu + 2,5\sigma_+$, $I_{min} = \mu - 2,0\sigma_-$.

Saliency Maps (Mapas de relevancia)

Mapas que visualizan la importancia de píxeles individuales calculando gradientes de la función de pérdida o de la clase predicha respecto a la imagen de entrada, mostrando qué regiones son más influyentes en predicciones. Técnica fundamental de interpretabilidad para diagnóstico médico asistido por IA.

Self-attention (Auto-atención)

Mecanismo en arquitecturas transformer que permite que cada elemento de una secuencia (token, parche de imagen) atienda a todos los demás elementos, capturando dependencias de largo alcance sin restricción de campo receptivo local como en CNNs. Calcula ponderaciones de atención mediante productos escalares entre queries, keys y values derivadas de la entrada.

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:
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.

Spatial Transformer Networks (STN) (Redes de Transformación Espacial)

Módulo diferenciable que permite a redes neuronales aprender transformaciones espaciales globales (affine, perspective, thin-plate spline) de forma de extremo a extremo sin

supervisión adicional, mejorando robustez a variaciones de pose. Limitación: transformaciones son globales, sin capacidad de deformación local adaptativa requerida para estructuras anatómicas deformables.

Squeeze-and-Excitation (SE) (Compresión y Excitación)

Mecanismo de atención a nivel de canal que recalibra feature maps mediante global average pooling (squeeze) seguido de transformación no lineal con dos capas fully-connected (excitation), generando ponderaciones adaptativas por canal. Amplifica canales informativos y suprime irrelevantes con sobrecarga computacional mínima. SE-Net ganó ILSVRC 2017.

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), 2×10^{-5} (backbone fase 2), 2×10^{-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: 4×4 u 8×8 bloques.

Top-5 Error (Error Top-5)

Métrica de evaluación en clasificación multiclas que mide el porcentaje de predicciones

donde la clase correcta NO está entre las 5 predicciones con mayor probabilidad. Utilizada en competencias de ImageNet para evaluar rendimiento en tareas con muchas clases (1,000 categorías). Valores bajos (e.g., $\geq 3\%$) indican excelente discriminación.

Transformación afín

Transformación geométrica que preserva líneas rectas y paralelismo, pudiendo incluir translación, 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.

U

U-Net

Arquitectura de red neuronal con estructura codificador-decodificador simétrica y conexiones saltadas densas (skip connections) entre niveles correspondientes, diseñada originalmente para segmentación médica con capacidad de aprender con pocos datos etiquetados. La estructura en "U" permite combinar información de contexto global (encoder) con localización espacial precisa (decoder).

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.

VGGNet

Arquitectura CNN que demostró la importancia de profundidad en redes neuronales, utilizando bloques de convoluciones 3×3 repetidas para incrementar capacidad de representación mientras se mantiene un diseño simple y homogéneo. VGG-16 y VGG-19 fueron ampliamente adoptadas como extractores de características preentrenados antes de

ResNet.

Vision Transformers (ViT) (Transformadores de Visión)

Arquitectura que aplica mecanismos transformer (originalmente diseñados para NLP) a reconocimiento de imágenes dividiendo la imagen en parches no solapados, proyectándolos a embeddings y procesándolos mediante multi-head self-attention. Captura dependencias de largo alcance globalmente pero requiere conjuntos de datos masivos (JFT-300M) para competir con CNNs debido a ausencia de inductive biases.

VOC 2007 (*Visual Object Classes 2007*)

Conjunto de datos del desafío PASCAL VOC 2007 para detección y segmentación de objetos con 20 categorías, utilizado como benchmark histórico de visión por computadora. Incluye anotaciones de bounding boxes, segmentación semántica y clasificación de acciones.

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
ASM	<i>Active Shape Models</i> Modelos de Forma Activa
AUC	<i>Area Under the Curve</i> Área Bajo la Curva
BIMCV	Biosignal Interpretation and Medical Computing Valencia
CBAM	<i>Convolutional Block Attention Module</i> Módulo de Atención de Bloque Convolucional
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> Completamente 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
MS COCO	<i>Microsoft Common Objects in Context</i>

Acrónimo	Significado
NME	<i>Normalized Mean Error</i> Error Medio Normalizado
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
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
STN	<i>Spatial Transformer Network</i> Red de Transformación Espacial
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>)
ViT	<i>Vision Transformer</i> Transformador de Visión
VOC	<i>Visual Object Classes</i>
XAI	<i>Explainable AI</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] R. Geirhos, J.-H. Jacobsen, C. Michaelis, R. Zemel, W. Brendel, M. Bethge, and F. A. Wichmann, “Shortcut learning in deep neural networks,” *Nature Machine Intelligence*, vol. 2, no. 11, pp. 665–673, 2020.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.

- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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. 13 713–13 722.
- [14] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017, survey fundamental con >300 papers revisados, 11,766+ citas.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, vol. 25, 2012, pp. 1097–1105.
- [16] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [18] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.
- [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. 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.

- [21] 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.
- [22] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, paper seminal demostrando performance comparable a dermatólogos, 11,281+ citas.
- [23] 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.
- [24] J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, “Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study,” *PLoS medicine*, vol. 15, no. 11, p. e1002683, 2018.
- [25] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models—their training and application,” *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
- [26] X. Wang, L. Bo, and L. Fuxin, “Adaptive wing loss for robust face alignment via heatmap regression,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6971–6981, mejora sobre Wing Loss para heatmap regression.
- [27] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial transformer networks,” in *Advances in neural information processing systems*, vol. 28, 2015.
- [28] G. Wolberg, *Digital image warping*. Los Alamitos, CA: IEEE Computer Society Press, 1990.
- [29] M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars, *Computational geometry: algorithms and applications*, 3rd ed. Berlin: Springer-Verlag, 2008.
- [30] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, “Attention mechanisms in computer vision: A survey,” *Computational Visual Media*, vol. 8, no. 3, pp. 331–368, 2022, survey comprensivo de attention mechanisms en computer vision.
- [31] 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.

- [32] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “Cbam: Convolutional block attention module,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
- [33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *International Conference on Learning Representations (ICLR)*, 2021.
- [34] 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.
- [35] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in *Graphics gems IV*. Academic Press Professional, 1994, pp. 474–485.
- [36] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaiier, 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.
- [37] T. G. Dietterich, “Ensemble methods in machine learning,” *Multiple classifier systems*, pp. 1–15, 2000.
- [38] N. Moshkov, B. Mathe, A. Kertesz-Farkas, R. Hollandi, and P. Horvath, “Test-time augmentation for deep learning-based cell segmentation on microscopy images,” *Scientific Reports*, vol. 10, no. 1, p. 5068, 2020, tTA para segmentación de células en microscopía.
- [39] J. C. Gower, “Generalized procrustes analysis,” *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.
- [40] I. L. Dryden and K. V. Mardia, *Statistical shape analysis: with applications in R*, 2nd ed. John Wiley & Sons, 2016.
- [41] Y. Wu and K. He, “Group normalization,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
- [42] I. Loshchilov and F. Hutter, “SGDR: Stochastic gradient descent with warm restarts,” in *International Conference on Learning Representations*, 2017.
- [43] I. L. Dryden and K. V. Mardia, *Statistical shape analysis*. Chichester: John Wiley & Sons, 1998.

- [44] P. H. Schönemann, “A generalized solution of the orthogonal procrustes problem,” *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.
- [45] 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.