



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Reducción de Ruido en Tomografías Computarizadas de
Baja Dosis mediante Redes Neuronales Artificiales

Tesis presentada al

Posgrado en Física Aplicada

como requisito parcial para la obtención del grado de

MAESTRO EN CIENCIAS

por

Uriel Rodríguez Sánchez

Asesorado por

Dr. Cristian Heber Zepeda Fernández

Dr. Juan Moisés Arredondo Velázquez

Puebla Pue.
21 de agosto de 2025



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Reducción de Ruido en Tomografías Computarizadas de
Baja Dosis mediante Redes Neuronales Artificiales

Tesis presentada al

Posgrado en Física Aplicada

como requisito parcial para la obtención del grado de

MAESTRO EN CIENCIAS

por

Uriel Rodríguez Sánchez

Asesorado por

Dr. Cristian Heber Zepeda Fernández
Dr. Juan Moisés Arredondo Velázquez

Puebla Pue.
21 de agosto de 2025

Título: Reducción de Ruido en Tomografías Computarizadas de Baja Dosis mediante Redes Neuronales Artificiales

Estudiante: URIEL RODRÍGUEZ SÁNCHEZ

COMITÉ

Dr. Benito De Celis Alonso
Presidente

Dr. Jorge Velázquez Castro
Secretario

Dr. Lucio Fidel Rebolledo Herrera
Vocal

Dr. Ivan Fuentesilla Carcamo
Suplente

Dr. Cristian Heber Zepeda Fernández
Asesor

Dr. Juan Moisés Arredondo Velázquez
Coasesor

Agradecimientos

Deseo expresar mi más profundo agradecimiento a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por el apoyo económico otorgado durante mis estudios de maestría. Su respaldo fue fundamental para la realización y conclusión de este trabajo académico.

A mis asesores, el Dr. Cristian Heber Zepeda Fernández y el Dr. Juan Moisés Arredondo Velázquez, por su guía, paciencia y por compartir conmigo sus conocimientos. Su acompañamiento fue fundamental en cada etapa de este trabajo.

De manera muy especial, agradezco al Dr. Jorge Velázquez Castro por sus valiosos consejos, observaciones y, sobre todo, por el apoyo moral brindado a lo largo de este proceso. Su disposición para acompañarme académica y personalmente ha sido de gran ayuda.

Agradezco profundamente a mis padres, Armando Rodríguez Cervantes y Gabriela Sánchez López, por su apoyo constante y por enseñarme el valor del trabajo, esfuerzo y la perseverancia. Esta tesis no habría sido posible sin los principios que me inculcaron desde pequeño.

A Jareth Benítez Ochoa, por su comprensión, ánimo y por estar a mi lado en los momentos más exigentes de este proceso. Gracias por tu cariño y por creer en mí.

A José David Pérez Marín, mi mejor amigo, por su amistad sincera y por estar conmigo en los momentos clave de este recorrido. Su apoyo, tanto personal como académico, ha sido invaluable.

Finalmente, extendo mi gratitud a todas las personas que, de una u otra manera, contribuyeron a la realización de este proyecto. Cada gesto de apoyo, palabra de aliento y compañía hizo una diferencia en este logro.

Índice general

1. Introducción	2
1.1. Motivación	2
1.2. Problemática General	3
1.3. Investigaciones Previas	3
1.4. Problema Especifico	4
1.5. Hipótesis	4
1.6. Objetivos	4
1.6.1. General	4
1.6.2. Específicos	4
1.7. Descripción	5
2. Estado del Arte	6
2.1. Método Clásico: Reducción de Ruido en imágenes LDCT	7
2.2. Método Moderno: Redes Neuronales para Reducción de Ruido en imágenes LDCT	9
2.3. Discusión sobre el Estado del Arte	12
3. Marco Teórico	15
3.1. Fundamentos Físicos y Técnicos de la Tomografía Computarizada	15
3.1.1. Rayos X	15
3.1.2. Producción de Rayos X	15
3.1.2.1. Cátodo	16
3.1.2.2. Ánodo	16
3.1.2.3. Radiación de Frenado	16
3.1.2.4. Radiación Característica	17
3.1.3. Interacción de los Rayos X	19
3.1.3.1. Efecto fotoeléctrico	19
3.1.3.2. Efecto Compton (Formula de Klein-Nishina)	21
3.1.3.3. Dispersión Coherente	22
3.1.3.4. Producción de Pares	23
3.1.3.5. Desintegración Fotónica	24
3.1.4. Tomografía Computarizada	26
3.1.4.1. Principios de la Tomografía Computarizada	26
3.1.4.2. Componentes del Tomógrafo Computarizado	29
3.1.5. Tomografía Computarizada de Baja Dosis	31
3.1.5.1. Moteado Cuántico	34
3.2. Teoría de Imágenes Digitales	35
3.2.1. Imagen Digital	35
3.2.2. Ruido en una Imagen	36
3.2.3. Métricas de Calidad de Imagen	37
3.2.3.1. Métricas de Calidad de Imagen con Referencia Completa	37

3.2.3.2.	Relación Señal-Ruido (RSR) y Coeficiente Contraste-Ruido (CCR)	38
3.2.4.	Operador de Convolución 2D	40
3.2.5.	Operador o Filtro Sobel	41
3.3.	Redes Neuronales Artificiales	42
3.3.1.	Redes Neuronales de Convolución (RNC)	42
3.3.2.	Funciones de Activación	43
3.3.3.	Funciones de Costo	43
3.3.4.	Algoritmo de Optimización	44
3.3.5.	Max Pooling	45
3.3.6.	Up Sampling	45
3.3.7.	Autoencoders	46
3.3.8.	Redes Generativas Adversariales (GANs)	46
3.3.8.1.	Redes Generativas Adversariales Semisupervisadas (SSGAN)	47
3.3.9.	Entrenamiento con Datos Desemparejados	47
4.	Metodología	49
4.1.	Base de Datos	49
4.1.1.	Base de datos General	49
4.1.2.	Preprocesamiento de la Base de datos	50
4.1.2.1.	Conjunto de la Base de Datos	50
4.1.2.2.	Emulación de Tomografías Computarizadas de Baja Dosis (LDCT)	50
4.2.	Arquitecturas	51
4.2.1.	Autoencoder	51
4.2.2.	Clasificador	53
4.3.	Reducción de Ruido y Preservación de Bordos	53
4.3.1.	Función de Costo	53
4.4.	Proceso de Entrenamiento del Autoencoder	55
4.4.1.	Entrenamiento - AE	55
4.4.2.	Parámetros de Entrenamiento - AE	56
4.5.	Alineación del Espacio Latente	57
4.5.1.	Entrenamiento - Autoencoder con Modificación en el Codificador (AEMC)	57
4.5.2.	Parámetros de Entrenamiento - AEMC	58
4.6.	Clasificación para cada Reconstrucción	58
4.6.1.	Entrenamiento - Clasificador	58
5.	Análisis de Resultados	61
5.1.	Funciones de Costo	61
5.1.1.	Gráfica de la Función de Costo - Autoencoder	62
5.1.2.	Gráfica de la Función de Costo - AEMC	63
5.1.3.	Gráfica de la Función de Costo - AEMCI	64
5.1.4.	Gráfica de la Función de Costo - Clasificador	64
5.1.5.	Gráfica de la Función de Costo - SSGAN	65
5.2.	Métricas de Calidad en Reconstrucción de Imágenes NDCT	66
5.3.	Alineación del Espacio Latente	67
5.4.	Análisis de Ruido	69
5.5.	Métricas de Calidad en Reconstrucción de Imágenes LDCT	73
5.6.	Relación Señal Ruido (RSR) y Coeficiente Contraste Ruido (CCR)	75
5.7.	Precisión de Clasificación para cada Reconstrucción	80
5.8.	Reconstrucciones - Análisis Cualitativo	81
5.9.	Resultados vs Estado del Arte	84

<i>ÍNDICE GENERAL</i>	VII
6. Conclusiones	86
Limitaciones	87
Trabajo Futuro	87
A. Estrategias AEMCI y SSGAN	88
A) Autoencoder con Modificación en el Codificador Iterativo (AEMCI)	88
Entrenamiento - AEMCI	88
Parámetros de Entrenamiento - AEMCI	89
B) Red Generativa Adversarial Semisupervisada (SSGAN)	89
Entrenamiento - (SSGAN)	90
Parámetros de Entrenamiento - SSGAN	91
Bibliografía	92

Resumen

La Tomografía Computarizada de Baja Dosis (LDCT, por sus siglas en inglés) ha sido ampliamente adoptada en entornos clínicos con el objetivo de minimizar la exposición del paciente a radiación ionizante. Sin embargo, esta reducción de dosis conlleva un aumento significativo en cuanto a cantidad de ruido de las imágenes resultantes, comprometiendo la calidad diagnóstica y dificultando la visualización de estructuras anatómicas finas. Ante esta problemática, el presente trabajo propone una estrategia de entrenamiento novedosa para una Red Neuronal de Convolución (Autoencoder) con la que sea posible obtener un modelo capaz de reducir el ruido presente en imágenes LDCT.

La propuesta se basa en el uso de un autoencoder convolucional, cuya función de costo está compuesta por una suma ponderada de distintas métricas de calidad de imagen, así como funciones que evalúan la preservación de bordes. Este autoencoder fue entrenado inicialmente con Tomografías Computarizadas de Dosis Normal (NDCT), lo que permitió al sistema aprender representaciones latentes propias de las Tomografías Computarizadas de Dosis Normal. Posteriormente, al introducir Tomografías Computarizadas de Baja Dosis (LDCT) en el codificador, se llevó a cabo un proceso de alineación del espacio latente, cuyo objetivo fue acercar las representaciones latentes de las imágenes LDCT a las correspondientes a NDCT. Este enfoque permitió mejorar de forma significativa la reducción de ruido, la preservación de bordes y la conservación de la estructura anatómica en las imágenes LDCT reconstruidas. Cabe resaltar que este enfoque de entrenamiento fue realizado utilizando imágenes desemparejadas, es decir, no fue necesario contar con pares exactos de una misma imagen en su versión de Dosis Normal (NDCT) y su contraparte correspondiente en Baja Dosis (LDCT). Esta característica representa una ventaja significativa, ya que en muchos escenarios clínicos resulta difícil o incluso imposible obtener conjuntos de datos emparejados lo cual no solo contraviene principios éticos fundamentales, sino que además resulta contradictorio frente al objetivo principal de reducir la dosis de radiación en los procedimientos de imagenología médica. El modelo propuesto, por tanto, demuestra ser eficaz incluso bajo estas condiciones.

Asimismo, se incorpora el uso de un clasificador convolucional, cuya función es evaluar el impacto que tiene la mejora en la calidad de las imágenes LDCT en términos de reducción de ruido y preservación de bordes sobre la precisión diagnóstica. A través de este enfoque, se busca demostrar que una mejora en la calidad visual y estructural de las imágenes implica directamente una mayor precisión al determinar si un paciente se encuentra sano o presenta alguna patología, consolidando así la importancia de los procesos de mejora en imágenes médicas para fines clínicos.

Los resultados obtenidos muestran en la reconstrucción una mejora significativa en métricas cuantitativas como PSNR, SSIM, MSE así como una óptima preservación de bordes respecto a las imágenes LDCT, y un desempeño competitivo frente a modelos existentes en el estado del arte. Adicionalmente, se observó que la calidad de las reconstrucciones implica una mejora en la precisión del clasificador, lo que sugiere que las mejoras visuales también aportan valor clínico.

Capítulo 1

Introducción

1.1. Motivación

La tomografía computarizada (TC) es un método para obtener imágenes detalladas del interior del cuerpo. Este procedimiento utiliza rayos X, un tipo de radiación ionizante, para crear dichas imágenes con alta precisión y resolución.

La TC es ampliamente empleada en el ámbito médico para diversas aplicaciones, como:

- Diagnosticar afecciones musculares y óseas, incluyendo tumores y fracturas.
- Identificar la ubicación de tumores, infecciones o coágulos sanguíneos.
- Guiar procedimientos médicos como cirugías, biopsias y radioterapia.
- Detectar y monitorear el progreso de enfermedades como cáncer, enfermedades cardíacas, nódulos pulmonares y tumores hepáticos.
- Evaluar la eficacia de tratamientos específicos, como la terapia contra el cáncer.
- Detectar lesiones internas y hemorragias causadas por traumatismos.

En comparación con una radiografía convencional, la tomografía computarizada expone al paciente a una mayor cantidad de radiación ionizante [42]. Esto se debe a la necesidad de obtener imágenes de mayor calidad, lo que implica el uso de rayos X más energéticos y una mayor dosis de radiación. Por ejemplo, mientras que una radiografía de tórax típica utiliza valores de 70-90 kV y 20-50 mAs, una TC de tórax suele emplear parámetros más elevados, como 100-150 kV y 200-500 mAs, resultando en una dosis significativamente mayor. [66], [12], [3], [13], [14]

Sin embargo, esta exposición a rayos X no está exenta de riesgos. Artículos como el titulado "*Riesgo de neoplasias hematológicas por exposición a la radiación de la TC en niños, adolescentes y adultos jóvenes*" [10] han encontrado una asociación significativa entre la dosis de radiación recibida y el riesgo de desarrollar neoplasias hematológicas.

Esto resalta la importancia de utilizar la TC de manera justificada y optimizada, siguiendo el principio de ALARA (As Low As Reasonably Achievable), para minimizar los riesgos asociados a la radiación ionizante.

Para minimizar la exposición a la radiación en tomografía computarizada (TC), los especialistas han implementado ajustes en tres parámetros principales durante la adquisición de imágenes: la disminución del voltaje del tubo de rayos X (kV), la corriente eléctrica del tubo (mA) y la velocidad de escaneo. Estos cambios permiten reducir significativamente la dosis de radiación recibida por el paciente [12] [66], [14].

Sin embargo, en la práctica clínica, el ajuste más común es la reducción de la corriente del tubo de rayos X, ya que disminuye directamente la cantidad de rayos X producidos. Esta técnica se conoce como Tomografía Computarizada de Baja Dosis (LDCT, por sus siglas en inglés). En este enfoque, el rango de voltaje típico en una TC convencional (100-150 kV) y la corriente del tubo (200-500 mA) se optimizan para mantener una calidad diagnóstica adecuada mientras se reduce la dosis de radiación. [43] [21]

Organizaciones como el American College of Radiology (ACR) han establecido directrices específicas para la implementación de LDCT, señalando que este método puede reducir la corriente del filamento del tubo de rayos X hasta en un 25% en comparación con los estudios convencionales, dependiendo del tipo de equipo y las características del paciente (ACR Appropriateness Criteria, 2022). [48] [47]

La LDCT es especialmente útil en contextos donde se requiere monitoreo frecuente o detección temprana [43] [33]. Por ejemplo:

- Detección temprana de cáncer de pulmón: En personas de alto riesgo, como fumadores, la LDCT es una herramienta esencial que permite realizar estudios repetidos con menor riesgo de efectos acumulativos por la radiación. [33]
- Seguimiento de daño pulmonar: Ha sido empleada en pacientes con enfermedades pulmonares, como aquellos que presentan secuelas tras infecciones por COVID-19, permitiendo un monitoreo continuo sin exponerlos a altas dosis de radiación. [47]

1.2. Problemática General

El uso de Tomografías Computarizadas de Baja Dosis (LDCT) presenta un desafío: la reducción en la cantidad de fotones de rayos X generados. Esto resulta en un menor número de interacciones atómicas con los tejidos del cuerpo humano, lo que, a su vez, disminuye la cantidad de rayos X atenuados que alcanzan los detectores del tomógrafo. Como consecuencia, se reduce la señal útil necesaria para formar la imagen CT. Esta limitación genera una menor calidad de la imagen obtenida, principalmente debido a un aumento en el nivel de ruido. Esto dificulta la visualización de detalles finos, lo que complica la interpretación de las imágenes, pudiendo causar discrepancias diagnósticas, imprecisiones o, en casos extremos, degradar la calidad de la imagen a niveles no diagnósticos. [14], [47] [66], [12]. Esta problemática ha sido ampliamente discutida en varios estudios [52], [74], [15], [6], [43].

1.3. Investigaciones Previas

La reducción de ruido en imágenes de Tomografía Computarizada de Baja Dosis (LDCT) ha sido ampliamente estudiada mediante diversos enfoques. Inicialmente, los métodos tradicionales, basados en filtrado y modelos estadísticos, ofrecieron mejoras en la calidad visual de las imágenes, pero con limitaciones significativas en la preservación de estructuras anatómicas y detalles finos. Con el auge del aprendizaje profundo, las Redes Neuronales Convolucionales (RNCs) han mostrado resultados prometedores, especialmente aquellas basadas en autoencoders y Redes Generativas Adversarias (GANs). Estos modelos han logrado reducir el ruido de manera efectiva mediante arquitecturas especializadas y la optimización de funciones de costo diseñadas para mejorar la fidelidad visual.

Uno de los principales desafíos en la reconstrucción de imágenes médicas es la preservación de bordes, ya que estos contienen información estructural fundamental para la interpretación clínica. Si bien algunos estudios han incorporado métricas perceptuales para mejorar la fidelidad estructural, la mayoría de los enfoques actuales no consideran explícitamente la preservación de bordes como un

criterio central. Esto puede resultar en la pérdida de información anatómica relevante, afectando potencialmente la reconstrucción de la imagen.

Otro aspecto crítico en la evaluación de estos modelos es que, en su mayoría, se han validado utilizando métricas de calidad de imagen, como el PSNR y el SSIM, sin analizar el impacto de la reconstrucción en tareas clínicas más relevantes, como la clasificación de patologías. Esta limitación sugiere la necesidad de un enfoque que no solo optimice la calidad visual de las imágenes LDCT, sino que también garantice la conservación de información clave para el diagnóstico.

En este contexto, resulta imprescindible desarrollar un método que, además de reducir el ruido, preserve de manera efectiva los bordes y realice un análisis sobre el impacto en la clasificación de pacientes sanos y enfermos.

1.4. Problema Especifico

Se identifica que una posible solución para mejorar la calidad de las imágenes LDCT en términos de reducción de ruido son las Redes Neuronales Convolucionales (RNCs). Estos algoritmos han demostrado ser efectivos en la reducción de ruido y, en cierta medida, consistentes en la preservación de características estructurales, como los bordes, que como se menciona en [74] estos son puntos clave al realizar una buena reconstrucción. Generalmente esto lo consiguen mediante la manipulación de la función de costo. Sin embargo, aunque diversos estudios reportan mejoras en la calidad visual de las imágenes, no mencionan si esto contribuye a una mejor clasificación para determinar si una imagen corresponde a una persona sana o enferma.

Por lo que, actualmente, no existe una estrategia de entrenamiento enfocada en la reducción de ruido y la preservación de bordes en imágenes LDCT que, además, analice el impacto de la reconstrucción en la clasificación.

1.5. Hipótesis

Realizando una alineación del espacio latente de las imágenes LDCT a la región del espacio latente de las imágenes NDCT mediante un modelo de entrenamiento basado en un Autoencoder, que use como función de costo métricas de calidad y un método de preservación de bordes, es posible preservar una mayor cantidad de información relevante en las imágenes LDCT. Contribuyendo a una mejor clasificación diagnóstica.

1.6. Objetivos

1.6.1. General

- Reducir el ruido y preservar los bordes en las Tomografías Computarizadas de Baja Dosis (LDCT).

1.6.2. Específicos

1. Encontrar y emular una de base de datos de TC dosis normal y Baja Dosis, (LDCT), respectivamente de pacientes enfermos y sanos.
2. Determinar las configuración de Red Autoencoder y Clasificador para tomar base de estudio.
3. Implementar un método de preservación de bordes.
4. Entrenar el algoritmo por medio de imágenes desparejadas LDCT y NDCT
5. Realizar una alineación en los espacios latentes de las LDCT con las NDCT.

1.7. Descripción

Este trabajo tiene como objetivo principal reducir el ruido y preservar los bordes en las imágenes de tomografías computarizadas de baja dosis (LDCT). Además, se busca analizar si una mejor reconstrucción de las LDCT, en términos de reducción de ruido y preservación de bordes, contribuye a mejorar la precisión en la clasificación diagnóstica.

Para alcanzar este objetivo, se implementará una red neuronal convolucional (RNC) con una arquitectura basada en autoencoders. Una de las principales características de este modelo es la incorporación de funciones de costo diseñadas para minimizar la cantidad de ruido y que pueda preservar estructura de los bordes en la imagen. Esto es crucial, ya que una reducción excesiva del ruido sin una adecuada preservación de los bordes puede comprometer la calidad de la imagen y dificultar la interpretación médica.

Adicionalmente, se llevará a cabo un alineamiento de los espacios latentes de diferentes tipos de tomografías computarizadas, específicamente entre las Tomografías Computarizadas de Dosis Normal (NDCT) y Tomografías Computarizadas de Baja Dosis (LDCT). Este alineamiento tiene como propósito aproximar la representación latente de las imágenes de LDCT a la región del espacio donde se encuentran las imágenes NDCT, que son nuestro objetivo ideal. Al lograr esta correspondencia, se espera que el modelo mejore aún más su capacidad de reconstrucción, proporcionando imágenes de mayor calidad.

El entrenamiento de la red se llevará a cabo utilizando imágenes desparejadas, es decir, no será necesario contar con una tomografía de dosis normal (NDCT) y su contraparte exacta en baja dosis (LDCT). Esto se debe a que obtener ambas versiones implicaría irradiar dos veces al mismo paciente, una con dosis normal y otra con baja dosis, lo que supondría un riesgo considerable para su salud.

Para evaluar la efectividad del modelo, cada imagen reconstruida será sometida a un proceso de clasificación mediante un clasificador entrenado para diferenciar entre TC correspondientes a personas sanas o enfermas. Se analizará si la mejora en la calidad de la reconstrucción conduce a un aumento en la precisión de la clasificación diagnóstica.

El desempeño del modelo se evaluará a través de un estudio estadístico detallado, considerando métricas como la media y desviación estándar del ruido en las imágenes reconstruidas. Asimismo, se emplearán diversas métricas de calidad de imagen, incluyendo métodos para evaluar la preservación de bordes y la fidelidad estructural de las imágenes con respecto a las originales. Finalmente, se analizará las mejoras en la reconstrucción sobre la precisión del clasificador, con el fin de determinar si la reducción del ruido y la preservación de bordes contribuyen efectivamente a un diagnóstico más preciso.

Este estudio no solo permitirá mejorar la calidad de las imágenes de tomografía computarizada de baja dosis, sino que también proporcionará información valiosa sobre la relación entre la calidad de la imagen y la precisión en la clasificación, lo que podría abrir nuevas posibilidades para la optimización de los sistemas de diagnóstico basados en imágenes médicas.

Capítulo 2

Estado del Arte

En los últimos años, la reducción de ruido en imágenes de Tomografía Computarizada de Baja Dosis (LDCT) se ha convertido en un área de investigación de gran relevancia, impulsada por la necesidad de obtener imágenes de alta calidad sin incrementar la exposición a la radiación en los pacientes. Para hacer frente a este desafío, se han propuesto múltiples enfoques, que van desde técnicas tradicionales basadas en filtrado y modelos estadísticos, hasta métodos más recientes fundamentados en redes neuronales profundas.

Estos enfoques pueden agruparse principalmente en dos categorías: los métodos clásicos y los modernos basados en aprendizaje profundo. Cada uno ofrece ventajas particulares y presenta ciertas limitaciones; sin embargo, aunque las técnicas modernas han demostrado un rendimiento superior en diversos estudios, la problemática de la reducción de ruido en LDCT continúa siendo un tema abierto de investigación.

Actualmente, la mayoría de las investigaciones enfocadas en la reducción de ruido en imágenes LDCT han dirigido sus esfuerzos hacia algoritmos basados en redes neuronales como un método de post-procesamiento sobre el dominio de la imagen, ya que estos han mostrado un mayor potencial para abordar esta problemática. Esto ha generado un creciente interés en dicho enfoque, como se puede observar en la Figura 2.1.

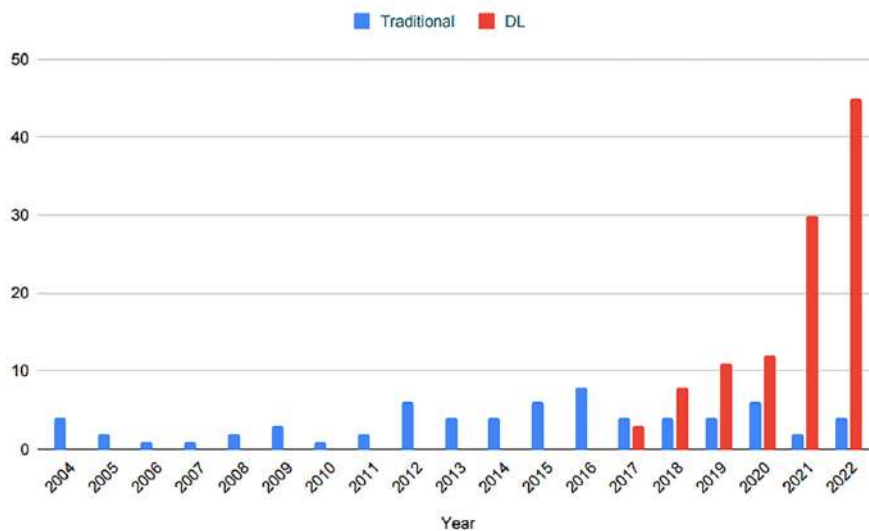


Figura 2.1: Tendencia del número de artículos publicados en los últimos años sobre la reducción de ruido en imágenes LDCT. Gráfico obtenido de [52].

En este capítulo, se presenta un análisis detallado de los enfoques más relevantes en la reducción de ruido en imágenes LDCT, destacando sus ventajas, limitaciones y tendencias actuales. Además, se examinan los principales métodos utilizados en la literatura reciente, evaluando sus métricas de desempeño y su impacto en la calidad de reconstrucción de las imágenes. Este análisis permitirá contextualizar el problema y sentar las bases para el enfoque propuesto en esta investigación. A continuación se muestra un diagrama de las tendencias en esta área de investigación.

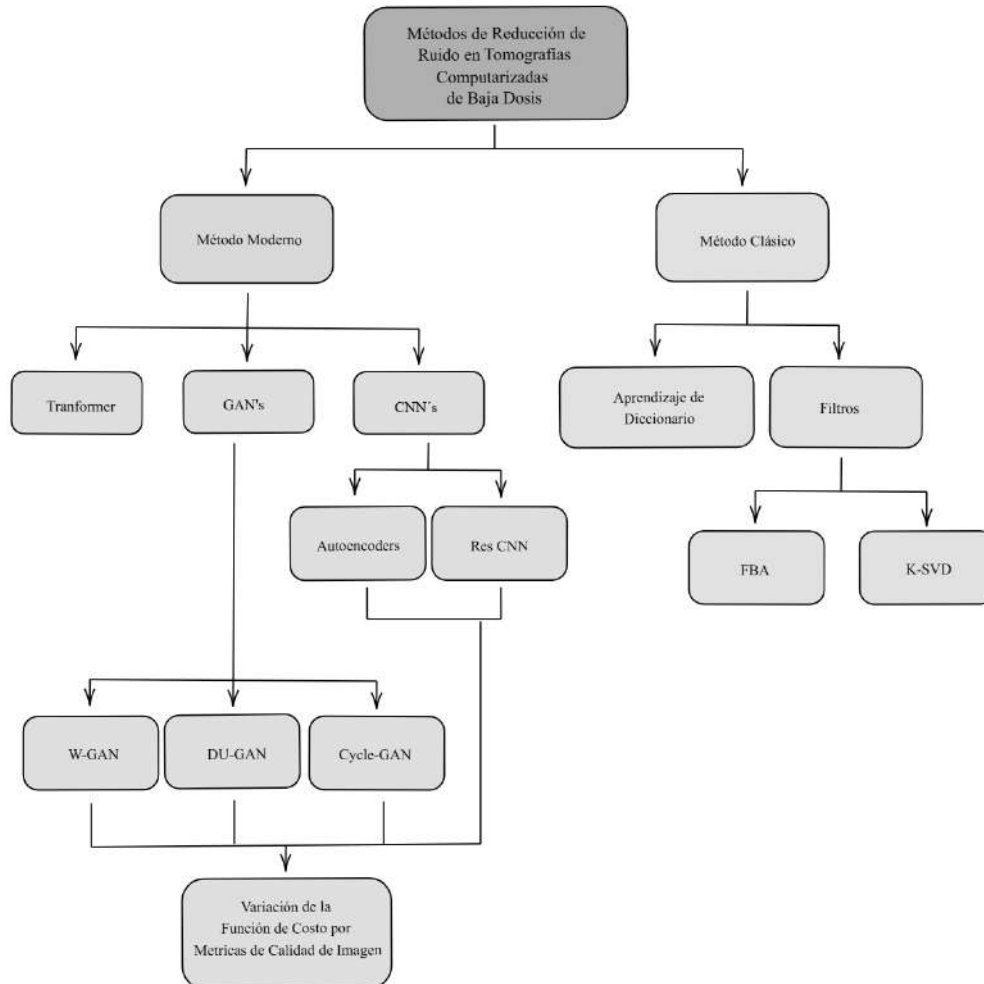


Figura 2.2: Tendencias en el área de investigación para la reducción de ruido en imágenes LDCT

2.1. Método Clásico: Reducción de Ruido en imágenes LDCT

Ramírez Giraldo et al. [50] propusieron el uso de un filtro bilateral anisotrópico (FBA) para la reducción de ruido en imágenes de Tomografía Computarizada (TC). Implementaron el FBA en tres dimensiones, permitiendo ajustar los parámetros del filtro según la resolución de las imágenes en cada eje. Evaluaron su desempeño en imágenes de un fantoma estándar escaneado con dosis de referencia (240 mAs) y dosis reducidas al 50% y 25%, así como en estudios clínicos de tomografía abdominal y torácica con contraste intravenoso.

El modelo fue evaluado utilizando parámetros como la relación entre ruido, resolución espacial

y detectabilidad de bajos contrastes. Los resultados mostraron que el FBA permitió reducir significativamente el ruido sin afectar la resolución espacial ni la capacidad de detectar estructuras de bajo contraste en comparación con la retroproyección filtrada (RPF). Se observó que es posible reducir la dosis de radiación hasta en un 50 % manteniendo la calidad diagnóstica. Sin embargo, se identificó la necesidad de estudios clínicos prospectivos para confirmar que la reducción de ruido permite disminuir significativamente la dosis de radiación en la práctica clínica.

Se concluyó que el FBA es una alternativa viable para la reducción de ruido en TC, con ventajas en su implementación computacional y aplicabilidad a diferentes equipos sin requerir modificaciones en el hardware. Se recomienda continuar investigando su comparación con otros métodos modernos y explorar su uso en aplicaciones específicas que requieran reducción de dosis de radiación.

Chen et al. [17] propusieron un método basado en aprendizaje de diccionarios para la mejora de imágenes de Tomografía Computarizada de dosis baja (LDCT) en pacientes con tumores abdominales. Utilizaron una técnica de aprendizaje de diccionarios dispersos en parches de imágenes, permitiendo la eliminación efectiva de ruido moteado y artefactos de bandas sin afectar la estructura de los tumores.

El modelo fue evaluado en imágenes clínicas de 25 pacientes con tumores abdominales, obtenidas con un escáner Siemens Somatom Sensation 16. Se comparó con el método de filtrado de vecinos no locales a gran escala (LNLN), utilizando métricas como la desviación estándar de regiones de interés y evaluaciones subjetivas por radiólogos. Los resultados mostraron que el método basado en diccionarios logró una mejor supresión de ruido y preservación de detalles estructurales en comparación con LNLN. Sin embargo, se observó que los artefactos de alto contraste persistían en algunas imágenes, y la eficiencia computacional del método aún necesita ser optimizada para su implementación clínica en tiempo real.

Se concluyó que el aprendizaje de diccionarios es una estrategia eficaz para la mejora de imágenes LDCT en pacientes con tumores abdominales, ofreciendo una alternativa viable sin necesidad de modificar los escáneres existentes. Se recomienda continuar investigando la optimización del procesamiento paralelo y la aplicación del modelo en imágenes tridimensionales para mejorar su aplicabilidad clínica.

Ybargollin-Machado et al. [71] compararon 11 métodos de reducción de ruido basados en variación total (TV) en imágenes de Tomografía Computarizada por Emisión de Fotón Único (SPECT). Evaluaron su desempeño en imágenes de un maniquí de Jaszczak simulando lesiones hipocaptantes, analizando métricas de calidad de imagen como la relación señal a ruido (SNR), el contraste imagen (C), la relación contraste a ruido (CNR) y la función de transferencia de la modulación (MTF).

Los resultados mostraron que los métodos FISTA y DCRG-PPXA lograron mantener un alto contraste de imagen, mientras que TVNLR y RTV preservaron mejor la resolución espacial. Por su parte, PDHG y GPCL fueron los más efectivos en la eliminación de ruido, mejorando tanto la SNR como la CNR. Se observó que los métodos que removieron mayor cantidad de ruido también afectaron la resolución espacial y el contraste, lo que sugiere la necesidad de un balance entre estas métricas dependiendo del objetivo clínico.

Se concluyó que los métodos DCRG-PPXA y TVNLR ofrecen el mejor compromiso entre eliminación de ruido, contraste de imagen y preservación de detalles estructurales en imágenes SPECT. Se recomienda evaluar su desempeño en estudios clínicos reales para determinar su aplicabilidad en la práctica médica.

2.2. Método Moderno: Redes Neuronales para Reducción de Ruido en imágenes LDCT

Chen et al. [16] propusieron un modelo de Red Neuronal Convolutiva con Codificador-Decodificador Residual (RED-CNN) para la reducción de ruido en imágenes de Tomografía Computarizada (TC) de baja dosis. Este enfoque combina autoencoders, redes convolucionales y conexiones residuales para mejorar la eliminación de ruido mientras preserva detalles estructurales.

El modelo fue entrenado con imágenes de TC de dosis normal con ruido simulado mediante distribución de Gaussiana, utilizando un conjunto de 7,015 imágenes de 165 pacientes y validado con datos clínicos reales de la "NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge". Se evaluó su desempeño en términos de error cuadrático medio (RMSE), relación señal-ruido pico (PSNR) e índice de similitud estructural (SSIM), comparándolo con técnicas de filtrado en el dominio sinograma, reconstrucción iterativa y métodos de aprendizaje profundo como CNN10 y KAIST-Net.

Los resultados mostraron que RED-CNN superó a los métodos tradicionales en supresión de ruido y preservación estructural, con puntuaciones más altas en PSNR y SSIM. Además, mencionan que una baja cantidad de capas en la red mejora la estabilidad del entrenamiento y reduce la pérdida de detalles.

Se concluyó que RED-CNN es una alternativa eficiente para la reducción de ruido en imágenes de TC de baja dosis, con ventajas en términos de velocidad de procesamiento y calidad de imagen. Se recomienda explorar su aplicación en reconstrucción 3D y en otras modalidades de imagen médica.

Nishio et al. [44] propusieron un modelo de Autoencoder Convolutiva (CAE) para la reducción de ruido en imágenes de Tomografía Computarizada de dosis ultra baja (ULDCT). Entrenaron el CAE utilizando parches de imágenes emparejadas de TC de dosis estándar (SDCT) y ULDCT adquiridas de un fantoma torácico y evaluaron su desempeño en imágenes de prueba.

El modelo fue evaluado utilizando parámetros como el pico de relación señal-ruido (PSNR) y el índice de similitud estructural (SSIM), además de una evaluación visual por parte de cinco radiólogos y tres tecnólogos. Los resultados mostraron que el CAE mejoró significativamente la reducción de ruido y la preservación de estructuras pulmonares en comparación con métodos tradicionales como Large-Scale Nonlocal Mean (LNLN) y Block-Matching and 3D Filtering (BM3D). Sin embargo, la diferencia en la calidad de imagen global entre CAE y BM3D no fue estadísticamente significativa ($p = 0.07272$).

Se concluyó que el uso de CAE permite mejorar la calidad de imagen en ULDCT, no es necesario una cantidad de capas excesivas para el codificador y decodificador, que las capas de convolución transpuesta en el decodificador llevan la tendencia a cambiar información de la imagen lo que no es muy conveniente en esta tarea. También el CAE permite mejorar las ULDCT sin necesidad de acceder a los datos sin procesar del escáner, lo que lo hace aplicable a una amplia gama de equipos. Se recomienda continuar investigando su desempeño en imágenes clínicas y optimizar su arquitectura para mejorar la preservación de detalles estructurales.

Fan et al. [25] propusieron un modelo de Autoencoder (AE) para la reducción de ruido en imágenes de Tomografía Computarizada de dosis baja (LDCT). A diferencia de los autoencoders tradicionales, el Q-AE utiliza neuronas cuadráticas en lugar de neuronas convencionales, lo que mejora la representación y eficiencia del modelo. Entrenaron el Q-AE con imágenes del conjunto de datos de Mayo Clinic y evaluaron su desempeño en múltiples pacientes con imágenes de TC de dosis baja.

El modelo fue evaluado utilizando parámetros como el pico de relación señal-ruido (PSNR), el índice de similitud estructural (SSIM) y el error cuadrático medio (RMSE). Los resultados mostraron que Q-AE obtuvo mayores valores de PSNR y SSIM en comparación con otros modelos

de referencia como RED-CNN, WGAN-VGG, y CPCE-2D, además de reducir significativamente el número de parámetros y operaciones computacionales. Sin embargo, se observó que la inicialización de los parámetros influía en la convergencia del modelo, y que en redes muy complejas la diferencia entre neuronas cuadráticas y convencionales disminuía.

Se concluyó que el uso de neuronas cuadráticas, una cantidad reducida de capas y el uso de alternativas como la función up-sampling en las capas de decodificación mejoran la eficiencia y precisión en la reducción de ruido en LDCT, ofreciendo un mejor equilibrio entre eliminación de ruido y preservación de detalles estructurales. Se recomienda continuar investigando su aplicación en otras tareas de procesamiento de imágenes médicas y mejorar estrategias de entrenamiento para optimizar su rendimiento.

Cokrokusumo et al. [19] propusieron un modelo de Red Neuronal Convolutiva con Codificador-Decodificador Residual (RED-CNN) para la reducción de ruido en imágenes de Tomografía Computarizada (TC) adquiridas con dosis bajas de radiación. Entrenaron la RED-CNN con imágenes de un fantoma cilíndrico de PMMA de 26 cm de diámetro, con distintos niveles de ruido simulado, y evaluaron su desempeño en imágenes de un fantoma elíptico de 21×26 cm con ruido simulado.

El modelo fue evaluado utilizando parámetros como la relación señal-ruido (SNR), el pico de relación señal-ruido comparativo (cPSNR), el índice de similitud estructural (SSIM), la función de transferencia de modulación (MTF 10%) y el espectro de potencia de ruido (NPS). Los resultados mostraron que RED-CNN incrementó SNR, cPSNR y SSIM, mientras que redujo el NPS. Sin embargo, niveles bajos de ruido en las imágenes de prueba resultaron en sobre-suavizado, mientras que niveles altos de ruido en los datos de entrenamiento disminuyeron la nitidez de las imágenes procesadas.

Se concluyó que el desempeño de RED-CNN depende del nivel de ruido en los datos de entrenamiento y prueba, requiriendo un balance adecuado para evitar degradaciones en la calidad de imagen. Se recomienda continuar investigando la selección óptima de parámetros para mejorar su aplicación en escenarios clínicos reales.

Diwakar et al. [23] propusieron un modelo de Red Neuronal Convolutiva con Normalización por Lotes (CNN-BN) para la reducción de ruido en imágenes de Tomografía Computarizada (TC) adquiridas con dosis bajas de radiación en el contexto del diagnóstico de COVID-19. Entrenaron la CNN-BN con imágenes de TC con distintos niveles de ruido gaussiano simulado y evaluaron su desempeño en imágenes reales de pacientes con COVID-19.

El modelo fue evaluado utilizando parámetros como la relación máxima señal-ruido (PSNR) y el índice de calidad de imagen (IQI). Los resultados mostraron que CNN-BN incrementó el PSNR y el IQI en comparación con otros métodos tradicionales de reducción de ruido, como Non-Local Means (NLM) y Transformada Shearlet. Sin embargo, se observó que niveles altos de ruido en las imágenes de prueba reducían la capacidad del modelo para preservar los bordes de las estructuras.

Se concluyó que el desempeño de CNN-BN es superior a métodos convencionales en la reducción de ruido en imágenes de TC de dosis baja, pero su efectividad depende del nivel de ruido presente. Se recomienda continuar investigando la selección óptima de parámetros y evaluar su aplicación en escenarios clínicos reales.

Wang et al. [63] propusieron un modelo de Autoencoder Enmascarado para Visión de Bajo Nivel (LoMAE) para la reducción de ruido en imágenes de Tomografía Computarizada de dosis baja (LDCT). LoMAE se basa en la arquitectura de los autoencoders enmascarados (MAE) y está diseñado para minimizar la dependencia de datos limpios emparejados, lo que lo hace adecuado para escenarios clínicos donde estos datos son limitados.

El modelo fue evaluado utilizando parámetros como el índice de similitud estructural (SSIM), el error cuadrático medio (RMSE) y el espectro de potencia de ruido (NPS). Los resultados mostraron que LoMAE mejoró el desempeño de modelos basados en transformadores como SwinIR y SUNet,

logrando una mayor preservación estructural y reducción de ruido en comparación con métodos previos como RED-CNN y MAPNN. Sin embargo, se observó que la ausencia de un atajo de conexión frontal en la etapa de preentrenamiento afectaba la efectividad del modelo en la fase de ajuste fino.

Se concluyó que LoMAE ofrece un enfoque eficiente y robusto para la reducción de ruido en LDCT, reduciendo la dependencia de datos de referencia y mostrando una alta generalización en distintos niveles de ruido. Se recomienda continuar investigando su integración con otros modelos basados en transformadores y explorar su aplicación en diferentes modalidades de imágenes médicas.

Wang et al. [64] propusieron un modelo de Red Generativa Antagónica Progresiva Wasserstein (PWGAN-WSHL) para la reducción de ruido en imágenes de Tomografía Computarizada de dosis baja (LDCT). Entrenaron el PWGAN-WSHL con un conjunto de imágenes de la base de datos de Mayo Clinic y evaluaron su desempeño en imágenes de prueba de pacientes anónimos.

El modelo fue evaluado utilizando parámetros como el pico de relación señal-ruido (PSNR), el índice de similitud estructural (SSIM) y el error cuadrático medio (RMSE). Los resultados mostraron que PWGAN-WSHL mejoró la calidad de imagen en comparación con otros métodos como BM3D, RED-CNN y WGAN-VGG, logrando mayores valores de SSIM y PSNR, además de reducir artefactos y preservar detalles estructurales. Sin embargo, se observó que la optimización del equilibrio entre la función de pérdida estructural y la pérdida L1 influía en el desempeño final del modelo.

Se concluyó que PWGAN-WSHL ofrece una solución efectiva y simple para la reducción de ruido en LDCT, con ventajas en la preservación de detalles y reducción de artefactos. Se recomienda continuar investigando la integración de información en el dominio de frecuencia y explorar enfoques que permitan entrenar el modelo con imágenes no emparejadas.

Huang et al. [29] propusieron DU-GAN, un modelo basado en redes generativas adversarias (GANs) con discriminadores basados en U-Net para la reducción de ruido en imágenes de Tomografía Computarizada de Baja Dosis (LDCT). Este enfoque busca mejorar la eliminación de ruido preservando estructuras globales y detalles locales en los dominios de imagen y gradiente.

El modelo utiliza un generador basado en RED-CNN y dos discriminadores U-Net: uno en el dominio de imagen y otro en el dominio de gradiente. El discriminador en el dominio de imagen captura diferencias globales y locales entre las imágenes LDCT denoised y NDCT reales, mientras que el discriminador en el dominio de gradiente mejora la preservación de bordes y reduce artefactos de rayos en la imagen. Además, se emplea la técnica de regularización CutMix para mejorar la discriminación a nivel de píxel y generar mapas de confianza que ayuden a los radiólogos en el diagnóstico.

Se evaluó DU-GAN en conjuntos de datos simulados y reales, incluyendo el "NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge" un conjunto de imágenes de un cerdito adquirido con un escáner GE. Se comparó con métodos tradicionales y basados en deep learning como RED-CNN, WGAN-VGG y CPCE-2D, mostrando una mejora en métricas como PSNR, SSIM y RMSE. Además, superó a otros métodos en la reducción de artefactos y preservación de estructuras finas en las imágenes.

Se concluyó que DU-GAN ofrece ventajas sobre métodos previos al proporcionar mejor supresión de ruido y preservar estructuras esenciales en imágenes LDCT. Se recomienda explorar su aplicación en reconstrucción 3D y su integración en tareas clínicas específicas como la detección de lesiones.

Yin et al. [72] propusieron un modelo de Red Generativa Antagónica de Wasserstein (WGAN) con pérdidas multi-perceptuales y de fidelidad para la reducción de ruido en imágenes de Tomografía Computarizada de dosis baja (LDCT). A diferencia de enfoques previos que requieren imágenes emparejadas, este método permite entrenar la red con imágenes no apareadas, facilitando

su aplicación en entornos clínicos donde la obtención de imágenes de referencia es limitada.

El modelo fue evaluado utilizando parámetros como el pico de relación señal-ruido (PSNR) y el índice de similitud estructural (SSIM). Los resultados mostraron que la incorporación de la pérdida perceptual permitió preservar mejor los detalles estructurales y mejorar la calidad visual en comparación con métodos previos como Cycle-GAN y RED-CNN. Sin embargo, se observó que el uso exclusivo de la pérdida perceptual podía generar distorsiones, por lo que se añadió la pérdida de fidelidad basada en L2 para garantizar la coherencia con la imagen original.

Se concluyó que el modelo propuesto mejora significativamente la reducción de ruido en LDCT sin necesidad de imágenes emparejadas, lo que lo hace más viable para aplicaciones clínicas. Se recomienda continuar investigando su integración con modelos híbridos y evaluar su desempeño en conjuntos de datos más diversos.

2.3. Discusión sobre el Estado del Arte

Los métodos clásicos como los mencionados en 2.1 son bastante útiles ya que tienen como ventaja la facilidad de implementación. Estos métodos permiten mejorar la calidad de imagen sin necesidad de modificar el hardware de los escáneres ni contar con grandes volúmenes de datos de entrenamiento. Sin embargo, la capacidad para preservar detalles de estos métodos clásicos es limitada, y tienden a generar imágenes sobre-suavizadas o con artefactos residuales. A medida que la demanda por imágenes médicas de mayor calidad ha aumentado, estos métodos comienzan a quedarse rezagados en comparación con enfoques modernos.

Por otro lado, los métodos basados en redes neuronales han demostrado un avance significativo en la reducción de ruido en imágenes LDCT. Modelos específicos como los autoencoders convolucionales (CAE), redes generativas adversarias (GANs) y redes convolucionales profundas (CNNs) han mostrado una notable mejora en la preservación de detalles estructurales y la reducción de ruido sin perder información crítica. No obstante, estos tienen la desventaja de requerir grandes cantidades de datos para su entrenamiento, y su implementación puede ser costosa en términos computacionales, aunado a que en la mayoría de los casos se suele realizar un entrenamiento con datos emparejados, lo cual es casi imposible tener en un caso médico real.

A pesar de los avances logrados con modelos basados en aprendizaje profundo, el problema aún no está resuelto. Un aspecto clave que se ha observado en múltiples estudios es que la complejidad excesiva en la arquitectura de la red no siempre se traduce en un mejor desempeño. De hecho, uno de los primeros enfoques en este campo, la Red Neuronal Convolucional Residual (RED-CNN) [16], la cual estructuralmente es un Autoencoder, sigue siendo uno de los modelos más efectivos. Esto sugiere que estructuras más simples, como los autoencoders convolucionales, pueden ofrecer un mejor equilibrio entre eliminación de ruido y preservación de detalles.

El hecho de que modelos con arquitecturas de autoencoders y estructuralmente simples continúen obteniendo resultados competitivos como la RED-CNN [16], Residual-CNN [19] y AE [25], indican que la capacidad de reconstrucción no necesariamente mejora con estructuras demasiado profundas o con un número excesivo de parámetros, pero sí que por lo general estas arquitecturas logran resultados eficientes. También en estos casos, la simplicidad ha demostrado ser más efectivo que el aumento indiscriminado de la complejidad del modelo, es por ello que numerosos artículos exploran variaciones de modelos que siguen esta tendencia aun que siempre se proponen nuevas arquitecturas que buscan superar lo reportado, como [29] y [65].

Artículos como [22] nos mencionan que la reducción de ruido de imágenes LDCT debe lograr también los siguientes cuatro puntos:

- Preservar los bordes de la imagen
- Preservar el contraste global
- Evitar la pérdida de detalle de la textura

- Evitar la generación de nuevos artefactos

Algunos de estos aspectos han sido abordados mediante la modificación de las funciones de costo, particularmente en el contexto de las redes generativas adversariales (GANs). Por ejemplo, en [64], se incorporaron a la función de pérdida del generador el Error Medio Absoluto (MAE) y el Índice de Similitud Estructural (SSIM), lo que permitió obtener mejoras cuantitativas significativas en comparación con los modelos que no utilizaban estas métricas.

De igual forma, el estudio presentado en [72], que empleó un enfoque basado en imágenes desemparejadas, integró el Error Cuadrático Medio (MSE) en la función de pérdida del generador, logrando algunos de los mejores resultados en términos de PSNR.

Asimismo, cabe destacar el trabajo de [29], en el que se añadieron el MSE y el MAE_{SOBEL} a la función de pérdida del generador con el objetivo de mejorar la preservación de bordes. Aunque los resultados no superaron a propuestas anteriores, los autores subrayan la importancia de considerar este aspecto en futuras investigaciones.

El Cuadro 2.1 presenta un resumen de los algoritmos utilizados en la tarea de reducción de ruido en imágenes LDCT. Además, se especifica si el entrenamiento se llevó a cabo con datos emparejados o desemparejados, si el método incluye un análisis de bordes, y si evalúa el impacto en la clasificación, indicando el porcentaje de mejora en la clasificación tras la reconstrucción. También se detallan las métricas cuantitativas empleadas para el análisis de resultados y, finalmente, el valor del PSNR (relación señal-ruido pico) obtenido con cada método propuesto.

Un aspecto importante en los resultados del estado del arte es que, en la mayoría de los casos, el análisis se basa en la evaluación de métricas de calidad de imagen, como PSNR, SSIM, RMSE y MSE. Estas métricas proporcionan indicadores cuantitativos claros sobre la calidad de reconstrucción alcanzada por cada método.

Sin embargo, a pesar de los avances y esfuerzos realizados en el desarrollo de métodos para la reducción de ruido en imágenes LDCT, el problema sigue abierto. Si bien las técnicas basadas en redes neuronales profundas han demostrado un desempeño superior en comparación con el método clásico, aún existen desafíos importantes, como la preservación de bordes, el conocer si existe una relación entre reconstrucción y precisión en la clasificación, así como la búsqueda de estrategias más eficientes continúa siendo un área activa de investigación.

Artículo	Método	Función de Costo	Datos	Métricas	Bordes	Clasificación	PSNR
Ramirez et al. [50]	FBA	—	Desemparejados	STD Ruido	No	No	—
Chen et al. [17]	SLD	—	Desemparejados	STD	No	No	—
Ybargollin et al. [71]	VT	—	Desemparejados	RSR, CCR	No	No	—
Zhang et al. [75]	CNN	MSE	Emparejados	PSNR	No	No	26.74 dB
Liang et al. [38]	EDCNN	MSE	Emparejados	PSNR, SSIM, RMSE	No	No	28.12 dB
Chen et al. [16]	RED-CNN (AE)	MSE	Emparejados	PSNR, SSIM, RMSE	No	No	28.39 dB
Nishio et al. [44]	CAE	MSE	Emparejados	SDT	No	No	—
Fan et al. [25]	AE	MSE	Emparejados	PSNR, SSIM, RMSE	No	No	28.59 dB
Cokrokusumo et al. [19]	Residual-CNN (AE)	MSE	Emparejados	PSNR, SSIM	No	No	28.64 dB
Divakar et al. [23]	CNN-BN	MSE	Emparejados	PSNR, IQI	No	No	27.42 dB
Wang et al. [63]	LoMAE	MSE	Emparejados	SSIM, RMSE	No	No	—
Huang et al. [29]	DU-GAN	GAN + MSE + SOBEL	Emparejados	PSNR, SSIM, RMSE	Si	No	27.87 dB
Wang et al. [65]	Uformer	Charbonnier	Emparejados	PSNR	No	No	28.31 dB
Dayang et al. [62]	CTransformer	MSE	Emparejados	PSNR, SSIM, MSE	No	No	27.80 dB
Yang et al. [70]	WGAN-VGG	WGAN + VGG	Emparejados	PSNR, SSIM, STD	No	No	26.70 dB
Wang et al. [64]	PWGAN-WSHL	WGAN + MAE + SSIM	Emparejados	PSNR, SSIM	No	No	28.67 dB
Yin et al. [72]	WGAN	WGAN + MSE _(LDCT, GEN) + MSE _(NDCT, GEN)	Desemparejados	RMSE, PSNR, SSIM	No	No	29.69 dB

Tabla 2.1: Modelos de reducción de ruido del Estado del Arte para imágenes LDCT.

Capítulo 3

Marco Teórico

3.1. Fundamentos Físicos y Técnicos de la Tomografía Computarizada

3.1.1. Rayos X

Los rayos X son una forma de radiación electromagnética con longitudes de onda en el rango de 0.01 a 10 nm, lo que corresponde a energías de 100 eV a 100 keV, caracterizados por su alta capacidad de penetración en la materia. Dentro del espectro electromagnético los rayos X se encuentran entre los rayos ultravioleta y los rayos gamma.

En un enfoque cuántico, los fotones de rayos X son radiación electromagnética con energía dada por la ecuación:

$$E = h\nu \quad (3.1)$$

donde E es la energía del fotón, h es la constante de Planck y ν es la frecuencia de la onda electromagnética.

3.1.2. Producción de Rayos X

La producción de rayos X se lleva a cabo, en términos generales, mediante tres etapas fundamentales:

1. Existencia de una fuente de emisión de electrones.
2. Presencia de un medio para acelerar dichos electrones.
3. Disponibilidad de un medio para frenar los electrones acelerados.

En esencia, los rayos X se generan utilizando dos electrodos: uno cargado negativamente, denominado cátodo, y otro cargado positivamente, conocido como ánodo. Los electrones emitidos por el cátodo, también denominados electrones proyectil, son acelerados mediante una diferencia de potencial. Los electrones proyectil o e^- proyectil adquieren cantidades extremas de energía cinética, la cual puede expresarse mediante la ecuación (3.2):

$$E_C = \frac{1}{2}mv^2 \quad (3.2)$$

donde E_C es la energía cinética del e^- proyectil, m es la masa del e^- proyectil ($9.1 \cdot 10^{-31}$ kg) y v es la velocidad del e^- proyectil en $\frac{m}{s}$, siendo alrededor del 50 % al 80 % la velocidad de la luz.

Estos e^- proyectil interactúan con los átomos del ánodo produciendo Rayos X. Es preciso tener mayores especificaciones sobre del ánodo y cátodo.

3.1.2.1. Cátodo

El Cátodo esta compuesto por una, llamada, copa de enfoque la cual tendrá incrustados filamentos que se calentaran mediante el paso de corriente eléctrica. El fin de este calentamiento es provocar la emisión de electrones a través de un proceso conocido como emisión termoiónica, que no es mas que la liberación de electrones de un electrodo gracias a su temperatura. Al elevar la temperatura de, en este caso, el filamento, los átomos vibran con mayor intensidad debido a la energía térmica, y esa vibración se transfiere a los electrones, los electrones ganan energía cinética y si la temperatura es lo suficientemente alta, algunos electrones ganarán suficiente energía para superar la función de trabajo del material y liberarse de la superficie.

Generalmente los filamentos están echos de Tungsteno Toriado, ya que esta aleación es más eficiente para llevar a cabo la emisión termoiónica, esto es porque al añadirle Torio (entre en 1 % o 2 %) la emisión se produce alrededor de los 1600°C a comparación del Tungsteno solo que emite electrones hasta alrededor de los 2000°C.

3.1.2.2. Ánodo

El ánodo normalmente es un disco grueso rotatorio de Tungsteno-Renio el cual gira alrededor de unas 3400 revoluciones por minuto (rpm) esto ayuda a disipar un poco el calor que producirá el impacto de los e^- proyectil con el ánodo. Los materiales de los que esta echo el ánodo son ideales para el frenado de electrones de alta velocidad ya que al tener un numero atómico Z tan alto como lo es el Tungsteno(W) $Z=74$ y el Renio(Re) $Z=75$, esto es, el numero de protones que tiene su núcleo atómico y al estar en equilibrio tendrá la misma cantidad de electrones distribuido en sus orbitas, lo que hará que haya una mayor probabilidad de interacción del electrón proyectil con los electrones orbitales o con el núcleo del mismo. También es importante no olvidar que los puntos de fusión de ambos materiales es muy elevado ayudando a resistir las altas temperaturas a las que llega el ánodo debido las colisiones de los e^- proyectil, siendo para el Tungsteno 3420°Cy para el Renio 3459°C.

Como bien mencionamos, la interacción de los e^- proyectil con los átomos del ánodo produce radiación de Rayos X, pero dependiendo de como se la interacción de los electrones proyectil con los Átomos del material que este compuesto el ánodo se producirá radiación de frenado o radiación característica.

3.1.2.3. Radiación de Frenado

Se producen debido a la interacción del e^- proyectil con el núcleo atómico. Al pasar cerca del núcleo atómico la carga positiva atraerá al e^- proyectil desviándolo y produciendo una reducción en la velocidad o frenandolo, esto representa una disminución en la energía cinética del e^- proyectil, esta pérdida de energía sera emitida en forma de un fotón de Rayos X, debido a efectos relativistas, como se muestra en la figura 3.1.

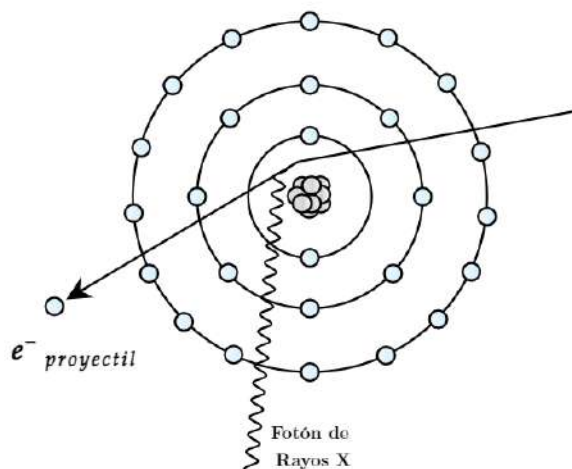


Figura 3.1: Representación de la producción de rayos X por radiación de frenado, ilustrada con el modelo atómico de Bohr como referencia visual.

Los e^- proyectil pueden pasar por el núcleo a diferentes distancias de él. Cuanto más se acerque un electrón al núcleo, mayor será su desaceleración, debido a que aumentara la fuerza de atracción del núcleo. También cuanto mayor sea la desaceleración del e^- proyectil, mayor será la desviación de su dirección original y más energía cinética perderá. Por lo que mientras mayor sea la energía cinética perdida, mayor será la energía del rayo X emitido. La relación que hay entre energía perdida por el e^- proyectil y la energía que tendrá el Fotón de Rayo X representarse en la ecuación (3.3):

$$E_{\text{fotón}} = E_{C_{\text{inicial}}} - E_{C_{\text{perdida}}} \quad (3.3)$$

donde $E_{\text{fotón}}$ es la energía del fotón de Rayos X emitido en kilo-electronvolts (KeV), $E_{C_{\text{inicial}}}$ es la energía cinética inicial del e^- proyectil al entrar al sistema en Joules (J) y $E_{C_{\text{perdida}}}$ es la energía cinética perdida del e^- proyectil al interactuar con el núcleo del átomo en Joules (J). Produciendo una amplia gama de energías de rayos X.

3.1.2.4. Radiación Característica

Se producen debido a la interacción del e^- proyectil con los e^- orbitales de los átomos del ánodo. La colisión entre el e^- proyectil y el e^- orbital puede expulsar al electrón fuera de su orbita siendo dispersados y dejando una vacante en la capa electrónica del átomo. (Fig.3.2)

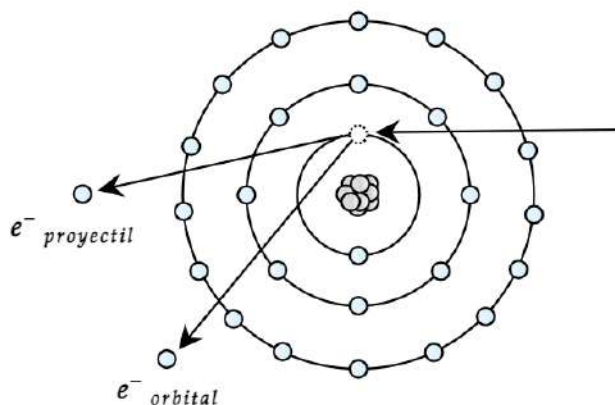


Figura 3.2: Representación de la vacante generada por la interacción entre el e^- proyectil y e^- orbital, ilustrada con el modelo atómico de Bohr como referencia visual.

Electrones de orbitas externas caerán en la vacante del e^- orbital expulsado. Cuando un electrón cae de una orbita externa a una interna (mas cerca del núcleo), hay una perdida de energía potencial en el electrón, ya que la energía potencial del electrón es la energía de enlace de cada capa del átomo, entonces la energía potencial perdida sera la energía del fotón emitido (Fig.3.3), esta relación se puede presentar en la ecuación (3.4):

$$E_{\text{fotón}} = E_{\text{enlace2}} - E_{\text{enlace1}} \quad (3.4)$$

donde $E_{\text{fotón}}$ es la energía del fotón de Rayos X emitido en keV, E_{enlace1} es la energía de enlace de la capa donde está el e^- orbital que tomará la vacante y E_{enlace2} es la energía de enlace de la capa donde está la vacante, ambas en keV.

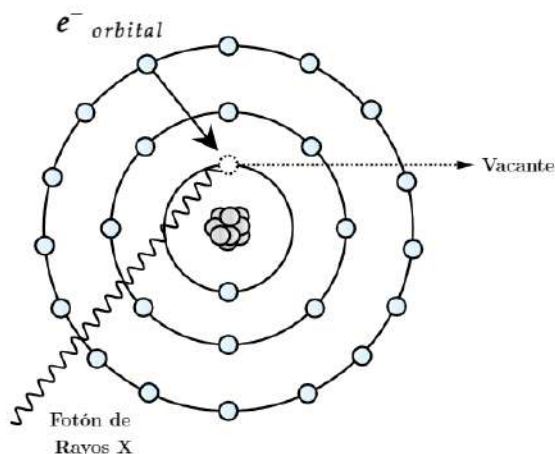


Figura 3.3: Representación de e^- orbital externo cayendo en la vacante y emitiendo un Fotón de Rayos X, ilustrada con el modelo atómico de Bohr como referencia visual.

Los rayos X producidos de esta manera generan energías discretas, en lugar de en un rango de energías como la radiación Bremsstrahlung. Para cualquier capa donde se encuentre la vacante puede ser llenada por un electrón de la capa inmediatamente superior o de dos o tres capas

superiores.

Tanto la radiación característica como la radiación de frenado se van generando aleatoriamente en el impacto de los e^- proyectil con el ánodo, estando en el rango de los KeV y nos es útil ver la forma del espectro completo del haz de rayos X producidos. (Fig.3.4)

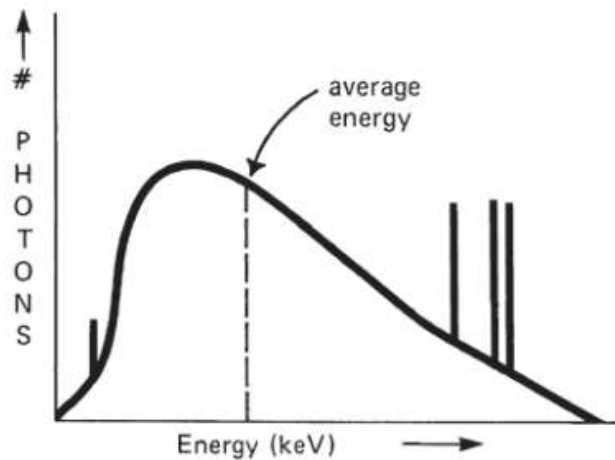


Figura 3.4: Forma del espectro completo del haz de rayos X de diagnóstico, que representa tanto la radiación de frenado como las energías de rayos X características, con filtración presente. Imagen extraída de [13], (Quinn B. Carroll, Practical Radiographic Imaging, 8.^a ed. Springfield, IL: Charles C Thomas, Publisher, Ltd., 2007. (pp. 12)

3.1.3. Interacción de los Rayos X

Los rayos X muestran longitudes de onda aproximadamente 10^{-8} a 10^{-9} m. A mayor energía, menor será su longitud de onda. Los rayos X con baja energía suelen interactuar con átomos completos, cuyo diámetro oscila entre 10^{-9} y 10^{-10} m. Los rayos X de energía intermedia interactúan principalmente con los electrones, mientras que aquellos con energía alta interactúan con los núcleos atómicos.

Existen cinco formas en las que interaccionan los rayos X:

- Efecto fotoeléctrico
- Efecto Compton
- Dispersión coherente
- Producción de pares
- Desintegración fotónica

Para la radiología diagnóstica solo el efecto Compton y el efecto fotoeléctrico son de particular importancia.

3.1.3.1. Efecto fotoeléctrico

El fotón de rayos X interactúa con un electrón de las órbitas internas del átomo. Al interactuar con el electrón, este absorbe completamente la energía del fotón, la cual está dada por la ecuación (3.5):

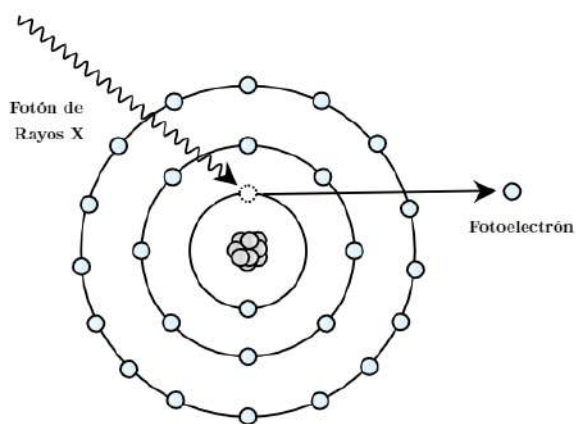


Figura 3.5: Representación de absorción del fotón de rayos X, generando un fotoelectrón, ilustrada con el modelo atómico de Bohr como referencia visual.

$$E = h\nu \tag{3.5}$$

Donde E es la energía del fotón de rayos X, h es la constante de Planck y ν es la frecuencia de la onda.

Tras la interacción, el fotón desaparece, y el electrón es expulsado de su órbita debido a que la energía del fotón supera la función de trabajo del material (es decir, la energía mínima necesaria para liberar un electrón de su órbita o capa). El electrón es entonces eyectado con una energía cinética equivalente a la energía sobrante del fotón, después de superar dicha barrera energética. Ahora es llamado fotoelectrón. (Fig.3.5)

En este punto el átomo queda ionizado dada la vacante generada por el electrón eyectado.

El efecto fotoeléctrico solo ocurre cuando la energía del fotón de rayos X incidente es mayor a la energía de enlace de la capa del electrón con la que interactúa.

Algebraicamente esta interacción puede ser representada por la ecuación (3.6):

$$E_{\text{fotón}} = W + E_C$$

ó de forma equivalente

$$h\nu = W + \frac{1}{2}mv^2 \tag{3.6}$$

Donde $E_{\text{fotón}}$ es la energía del fotón de rayos X, W es la función de trabajo y E_C es la energía cinética del fotoelectrón.

Es importante mencionar que la probabilidad de que se produzca el efecto fotoeléctrico es proporcional al numero atómico del átomo con el que interactúa.(3.7):

$$P_{ef} \propto Z^3$$

ó de forma equivalente

$$P_{ef} \propto \frac{1}{E_f^3} \tag{3.7}$$

Con P_{ef} la probabilidad de que ocurra el efecto fotoeléctrico, Z es el numero atómico del átomo con el que interactúa y E_f^3 es la energía del fotón de rayos X. Esto es, que el efecto fotoeléctrico es mas probable para átomos con numero atómico alto.

3.1.3.2. Efecto Compton (Formula de Klein-Nishina)

El proceso ocurre cuando un rayo X incidente interactúa con un electrón de la capa más externa del átomo. El fotón transfiere parte de su energía al electrón, la suficiente para superar la energía de enlace. Como resultado, el electrón es expulsado de su órbita y pasa a llamarse electrón Compton. Tras la interacción, el fotón se dispersa con una energía menor y cambia de dirección θ con respecto al ángulo de incidencia, variando en ángulos entre 0° y 180° . El átomo, ahora, queda ionizado. (Fig.3.6)

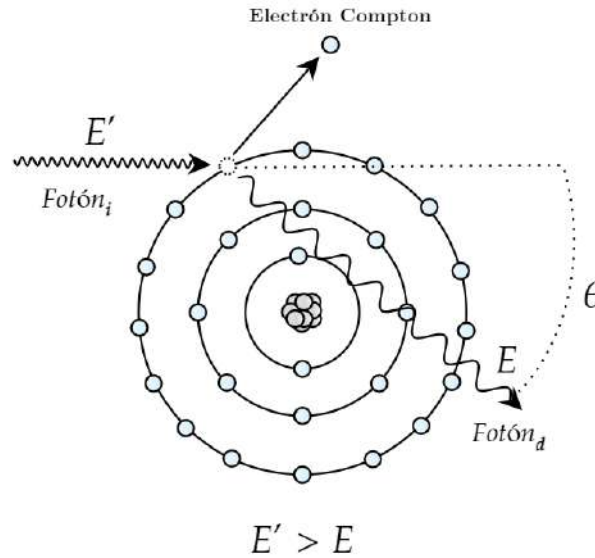


Figura 3.6: Representación de la dispersión de rayo X por medio del Efecto Compton. Donde E' y E son las energías antes de la interacción y después de la interacción respectivamente. Ilustración usando el modelo atómico de Bohr como referencia visual.

La energía del sistema se conserva lo que puede verse en dada la ecuación (3.8).

$$E' = E + W + E_C \tag{3.8}$$

Donde E' es la energía del fotón incidente, E es la energía del fotón disperso, W es la función de trabajo y E_C es la energía cinética del electrón Compton.

El efecto Compton, que describe la dispersión de fotones por electrones, se expresa mediante la ecuación de dispersión de Compton, que relaciona el cambio en la longitud de onda del fotón con el ángulo de dispersión. (Ec.(3.9)).

$$\Delta\lambda = \frac{h}{m_e c} (1 - \cos\theta) \tag{3.9}$$

donde $\Delta\lambda$ es el cambio en la longitud de onda del fotón, h es la constante de Planck, m_e es la masa del electrón, c es la velocidad de la luz y θ es el ángulo de dispersión.

Sin embargo, para una descripción más precisa de esta interacción, es necesario recurrir a la fórmula de Klein-Nishina. Esta fórmula proporciona la sección transversal diferencial, es decir, la probabilidad y la distribución angular, de los fotones dispersados desde un único electrón libre o débilmente ligado. Para un fotón incidente la sección transversal diferencial, (la probabilidad de que una partícula interactúe con un fotón), está dado por la ecuación (3.10).

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2} \left(\frac{\lambda}{\lambda'}\right)^2 \left(\frac{\lambda}{\lambda'} + \frac{\lambda'}{\lambda} - \sin^2 \theta\right) \quad (3.10)$$

donde $\frac{d\sigma}{d\Omega}$ es la sección transversal diferencial, r_e es el radio clásico del electrón, λ' y λ son las longitudes de onda del fotón antes y después de la dispersión, respectivamente, y θ es el ángulo de dispersión.

Por tanto, mientras que la ecuación de Compton es útil para comprender el cambio en la longitud de onda del fotón, la fórmula de Klein-Nishina ofrece una descripción más precisa de la probabilidad de dispersión y la distribución angular de los fotones dispersos, particularmente a energías más altas, como las presentes en la radiación de rayos X.

Aun así, en la literatura, se suele mencionar que la probabilidad de que el Efecto Compton ocurra esta dada por la relación (3.11).

$$P_{EC} \propto \frac{1}{E} \quad (3.11)$$

3.1.3.3. Dispersión Coherente

Los rayos X con energías inferiores a 10 keV interactúan principalmente a través de la dispersión coherente.

Dispersión Thomson

Un fotón de baja energía incidente interactúa con un electrón orbital, siendo momentáneamente absorbido. Como resultado, el electrón se excita debido a la energía recibida. Para recuperar su estado de estabilidad, el electrón emite un fotón de rayos X con la misma energía que el fotón incidente, pero en una dirección diferente. (Fig.3.7)

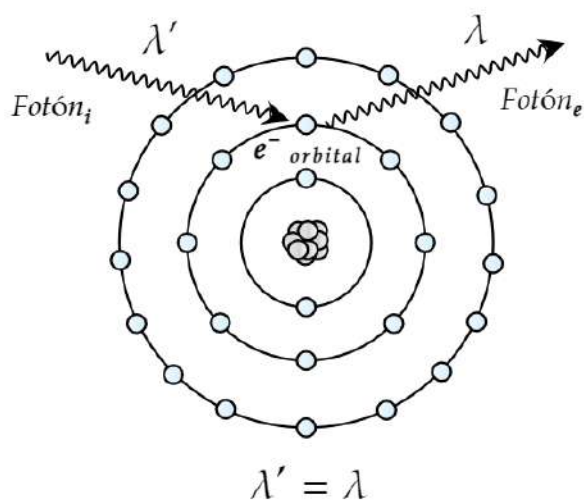


Figura 3.7: Dispersión Thomson. La longitud de onda λ' del fotón de rayos X incidente, F_i , es igual a la longitud de onda λ del fotón de rayos X emitido, F_e . Ilustración usando el modelo atómico de Bohr como referencia visual.

Dispersión Rayleigh

La energía del fotón de rayos X es absorbida temporalmente por toda la nube electrónica del átomo, también denominado átomo diana. Posteriormente, esta energía se reemite como un fotón de rayos X dispersado, manteniendo la misma energía que el fotón incidente, pero en una dirección diferente. (Fig.3.8)

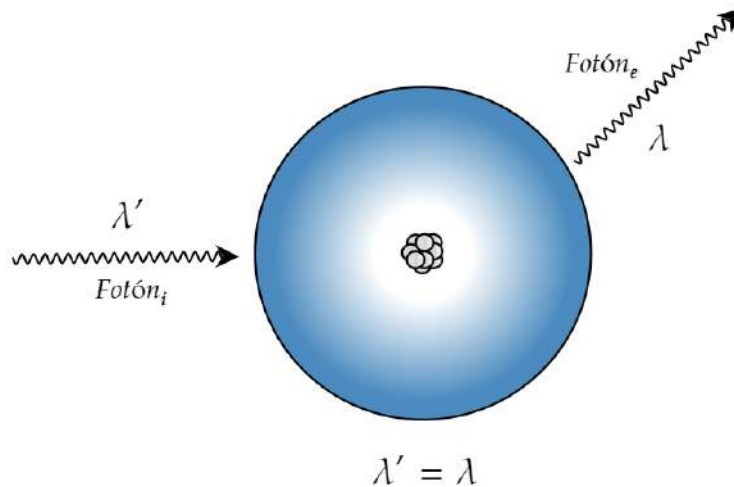


Figura 3.8: Dispersión Rayleigh. La longitud de onda λ' del fotón de rayos X incidente, F_i , es igual a la longitud de onda λ del fotón de rayos X emitido, F_e .

3.1.3.4. Producción de Pares

La producción de pares ocurre cuando un fotón de rayos X tiene suficiente energía, al menos 1.022 MeV, para superar la energía de enlace de los electrones orbitales y acercarse lo suficiente al núcleo del átomo, donde puede interactuar con el campo electromagnético del núcleo.

La energía del fotón se convierte en masa de partículas de acuerdo con la ecuación de Einstein (Ec.(3.12)):

$$E = mc^2 \tag{3.12}$$

donde E es la energía del fón , m es la masa y c es la velocidad de la luz .

Esta interacción da lugar a la creación de un electrón y un positrón (la antipartícula del electrón).(Fig. 3.9) La energía del fotón se distribuye entre ambas partículas en forma de energía cinética, además de cubrir la energía de masa en reposo de cada una. Para que se produzca este fenómeno, el fotón debe tener una energía superior a la suma de las energías de masa en reposo del electrón y del positrón ($2 \cdot 511 \text{ keV} = 1.022 \text{ MeV}$). Cualquier energía adicional se convierte en energía cinética de las partículas creadas.

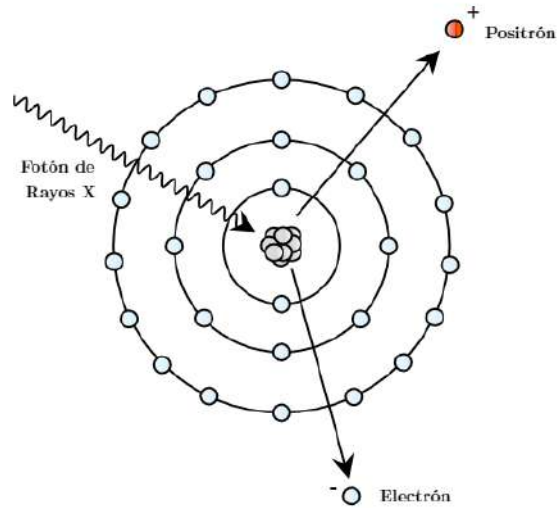


Figura 3.9: Producción de un electrón y un positrón por medio de la interacción de un fotón con el campo electromagnético del núcleo del átomo. Ilustración usando el modelo atómico de Bohr como referencia visual.

3.1.3.5. Desintegración Fotónica

La desintegración fotónica ocurre cuando los fotones de rayos X tienen una energía superior a 10 MeV. En este proceso, el fotón puede ser absorbido por el núcleo del átomo, lo que provoca que el núcleo entre en un estado excitado al absorber la energía del fotón incidente. Como resultado de esta excitación, el núcleo emite de manera casi instantánea un nucleón (ya sea un protón o un neutrón) para volver a un estado más estable. (Fig. 3.10)

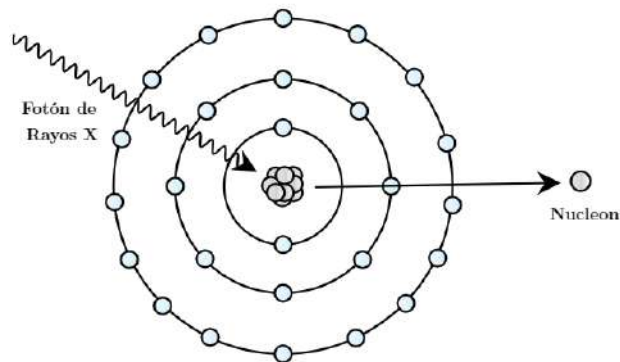


Figura 3.10: Desintegración Fotónica. Ilustración usando el modelo atómico de Bohr como referencia visual.

Es importante mostrar las interacciones predominantes de los rayos X en función de su energía y el número atómico del material absorbente 3.11. Las diferentes regiones del gráfico representan los dominios de las principales interacciones:

- El efecto fotoeléctrico domina en energías bajas y materiales con alto número atómico.
- El efecto Compton predomina en energías intermedias y no es dependiente del número atómico.
- La producción de pares se manifiesta a energías altas y en materiales con mayor número atómico.

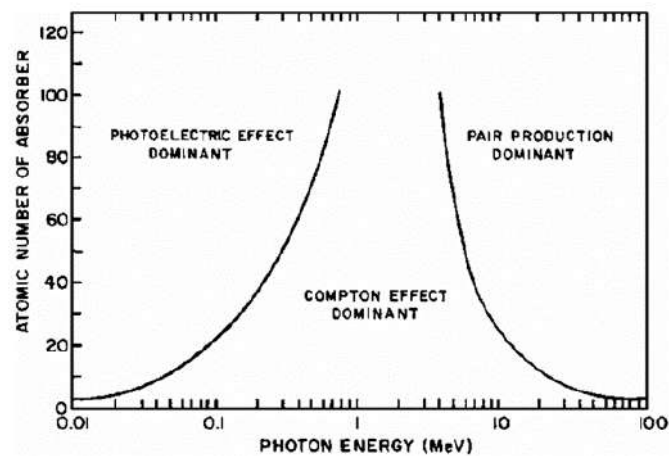


Figura 3.11: Importancia relativa de las tres interacciones principales de los rayos x. Las líneas representan combinaciones de energía/número atómico para las cuales las dos interacciones a cada lado de la línea son igualmente probables. Imagen extraída de [3] (Hendee, W. R., & Ritenour, E. R. (2002). *Medical Imaging Physics* (4th ed.) pp. 63)

Esta representación ilustra cómo la energía de los fotones y la composición del material determinan la interacción predominante en el proceso de atenuación de los rayos X.

La información presentada en esta sección fue recopilada a partir de diversas fuentes bibliográficas, entre ellas: [3, 12–14, 30, 66].

3.1.4. Tomografía Computarizada

La tomografía computarizada (TC) es una técnica de imagen médica que utiliza rayos X y procesamiento informático avanzado para generar imágenes seccionales del cuerpo humano. La TC permite visualizar estructuras internas en tres dimensiones, proporcionando mayor precisión en el diagnóstico y en la evaluación de diversas patologías.

La TC fue desarrollada en los años 70 y revolucionó la medicina diagnóstica al ofrecer imágenes detalladas y de alta resolución, lo que permitió una mejor visualización de tejidos blandos, órganos y huesos.

3.1.4.1. Principios de la Tomografía Computarizada

Absorción Diferencial

De las 5 formas en las que puede interaccionar los rayos X, solo 2 son importantes para la imagenología por TC, las cuales son:

- Efecto Compton (Formula de Klein–Nishina).
- Efecto fotoeléctrico.

Efecto Compton (Formula de Klein–Nishina): La dispersión Compton en el cuerpo ocurre con todos los rayos X, lo que la convierte en un fenómeno relevante dentro de la imagenología médica, aunque su impacto es desfavorable. Los rayos X dispersos no aportan información diagnóstica útil.

Cuando un rayo X disperso por el efecto Compton alcanza el detector, este lo interpreta como si proviniera directamente de la fuente de rayos X. Esto genera ruido en la imagen y una distribución uniforme de la intensidad, lo que reduce las diferencias de tonalidad entre las diversas estructuras observables.

Efecto Fotoeléctrico: Los rayos X que experimentan una interacción fotoeléctrica representan estructuras anatómicas que presentan una alta capacidad de absorción de rayos X, generando las áreas claras en la imagen radiográfica, como las correspondientes al hueso.

Dada la importancia que tiene el número atómico para que ocurra el efecto fotoeléctrico es importante el siguiente cuadro:

Tipo de material	Número atómico efectivo
Materia	
Grasa	6,3
Tejidos blandos	7,4
Pulmón	7,4
Hueso	13,8
Aire	7,6
Hormigón	17
Molibdeno	42
Tungsteno	74
Plomo	82

Tabla 3.1: Número atómico efectivo de diferentes materiales.

Hay rayos X que penetran en el cuerpo y son transmitidos al receptor de imagen sin ningún tipo de interacción. Estos rayos X producen las áreas oscuras.

Una imagen radiológica se genera a partir de la diferencia entre los rayos X absorbidos mediante el efecto fotoeléctrico, aquellos que logran atravesarlo y llegar al receptor de imagen. Esta variación en la interacción de los rayos X se conoce como absorción diferencial.

Atenuación - Ley de Lambert Beer

Los Rayos X generados son emitidos, incidiendo en el paciente y donde estos interactúan de las formas mencionadas anteriormente. Estos Rayos X serán atenuados dada la la forma general (Ec.(3.13)):

$$I = I_0 e^{-\mu x} \quad (3.13)$$

donde I es la intensidad de los rayos X antes de alcanzar el objeto, I_0 es la intensidad de los rayos X después de que penetran a través del objeto, μ es el coeficiente de atenuación lineal de rayos X y x es el espesor del material absorbente en unidades de distancia .

Si a lo largo de su trayectoria los rayos X pasan a través de múltiples objetos con un espesor total x , la intensidad de los rayos X penetrados se puede expresar de la forma (Ec.(3.14)):

$$I = I_0 e^{-(\sum_{i=1}^n \mu_i)x} \quad (3.14)$$

conocida como la ley de Lambert-Beer.

Con una sola medición, no es posible determinar los coeficientes de atenuación por separado ya que hay demasiados valores desconocidos de μ_i en la ecuación. Sin embargo, realizando múltiples mediciones en diferentes orientaciones de la fuente de rayos X y los detectores, generalmente rotando en 360° , los coeficientes se pueden separar de modo que se obtenga una visualización de los coeficientes de atenuación en todo el plano.

Datos de Proyección

Una vez los rayos X son atenuados e inciden en los detectores, los datos adquiridos en diferentes ángulos se envían al sistema de adquisición de datos (DAS). Estos proporcionan datos de proyección para formar imágenes tomográficas. Los datos de proyección son lo que interpreta el detector después de que los rayos X son atenuados y transmitidos a través del objeto. Donde el objeto escaneado que esta en función de $\mu(x, y)$, donde x y y son las coordenadas de la ubicación espacial del objeto, generará una proyección $P(t, \theta)$ que corresponden a la posición t y ángulo θ de los detectores.

Esto abre paso a que la ecuación 3.13 pueda ser reescrita de la siguiente forma:

$$\frac{I}{I_0} = \exp \left[- \int_{-\infty}^{\infty} \mu(t \cos \theta - x \sin \theta, t \sin \theta + x \cos \theta) dx \right] \quad (3.15)$$

Lo que define a la proyección $P(t, \theta)$ como la absorción integrada a través del espesor:

$$P(t, \theta) = -\ln\left(\frac{I(t, \theta)}{I_0}\right) = \int_{-\infty}^{\infty} \mu(t \cos \theta - x \sin \theta, t \sin \theta + x \cos \theta) dx \quad (3.16)$$

Esto es la transformación de Radón, también llamada proyección hacia adelante. En otras palabras, la transformación de Radón convierte la distribución del coeficiente de absorción en datos de proyección.

Sinograma

La forma de poder visualizar estos datos de proyección en conjunto en el espacio (t, θ) es por medio de un sinograma. Al trazar los múltiples $P(t, \theta)$ en el espacio (t, θ) se obtiene un sinograma. Los ejes de un sinograma son el ángulo de los rayos X (θ) y la posición del detector (t).

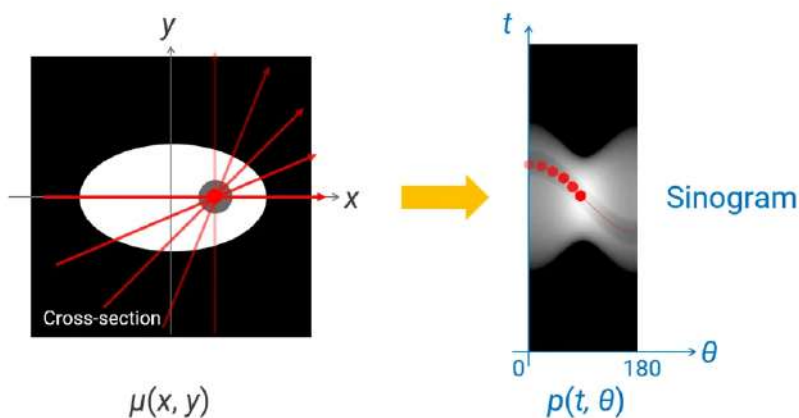


Figura 3.12: Representación del proceso de adquisición de datos en una TC. A la izquierda, se observa la sección transversal del objeto con los rayos X proyectados en diferentes ángulos $\mu(x, y)$. A la derecha, el sinograma $P(t, \theta)$ generado por la recopilación de proyecciones a lo largo de múltiples ángulos. Imagen extraída de [51], (Rigaku. (n.d.). How does CT reconstruction work?

Algoritmo de Reconstrucción

Al tener los datos de proyección en forma de un sinograma se usa lo que se conoce como un algoritmo de reconstrucción para finalmente tener la imagen CT como la solemos conocer. Estos algoritmos procesan los datos obtenidos por los detectores del tomógrafo para generar una representación digital del interior del paciente. A continuación, se describen los principales métodos de reconstrucción:

1. Reconstrucción Basada en Proyección: Algoritmos Analíticos.

Transformada de Radón y Retroproyección Filtrada (FBP): La retroproyección filtrada es el método más utilizado en la reconstrucción de imágenes CT. La FBP emplea la transformada de Radón para convertir las proyecciones en datos espaciales. Luego, utiliza un filtro de convolución para corregir las distorsiones que se producen durante la retroproyección. Es rápido y eficiente, lo que lo hace ideal para aplicaciones clínicas, pero, sensible al ruido, lo que puede afectar la calidad de la imagen fácilmente.

2. Reconstrucción Iterativa.

Método de Máxima Verosimilitud (MLE) y Expectativa-Maximización (EM): Utilizan un enfoque estadístico para iterar entre una estimación inicial de la imagen y los datos medidos, ajustando la solución en cada paso. Es menos sensible al ruido y tiene mejor calidad de imagen, pero, requiere más tiempo de cálculo en comparación con la FBP.

Reconstrucción Iterativa Basada en Regularización: Incorporan términos de regularización en la función objetivo, como la suavidad o la preservación de bordes. Este preserva los detalles anatómicos importantes y reducen los artefactos, pero, requiere una alta capacidad computacional.

3. Métodos Basados en Inteligencia Artificial.

Redes Neuronales Convolucionales (CNN): Las CNN se entrenan para aprender un mapeo directo entre los datos adquiridos y las imágenes reconstruidas, eliminando la necesidad de pasos intermedios. La capacidad para reducir ruido y artefactos, generando imágenes de alta calidad, pero, requieren grandes cantidades de datos de entrenamiento y pueden ser sensibles a datos no vistos.

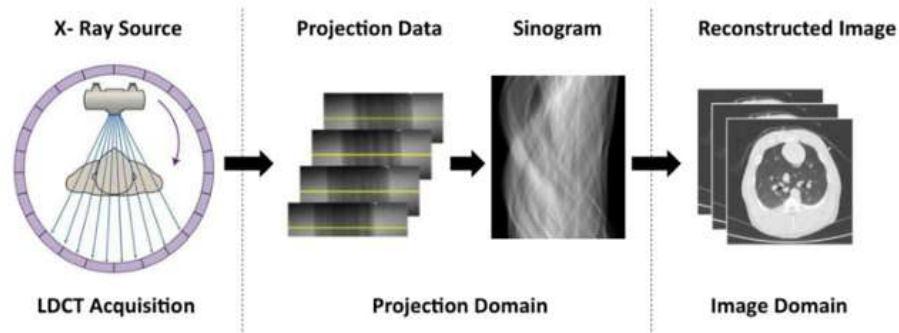


Figura 3.13: El proceso de adquisición y obtención de imágenes TC. Imagen extraída de [36], (Kulathilake, K. A. S. H., Abdullah, N. A., Sabri, A. Q. M., & Lai, K. W. (2023). A review on Deep Learning approaches for low-dose Computed Tomography restoration. Complex & intelligent systems.

3.1.4.2. Componentes del Tomógrafo Computarizado

Un equipo de Tomografía Computarizada consiste principalmente en:

- Gantry
 - Tubo de Rayos X
 - Colimadores
 - Detectores

Gantry

El gantry en un tomógrafo computarizado es una estructura rotatoria en forma de anillo que alberga los componentes esenciales del sistema, como el tubo de rayos X, los detectores y los colimadores. Durante la generación de rayos X y la obtención de imágenes, el gantry gira a una velocidad angular que varía entre 2 y 3.7 revoluciones por segundo (rps). Este parámetro también se puede expresar como el tiempo de rotación del gantry, que corresponde a la cantidad de segundos por rotación, y se sitúa entre 0.27 y 0.5 segundos por vuelta, de acuerdo con la relación ((3.17)):

$$\text{Segundos por rotación} = \frac{1}{rps} \quad (3.17)$$

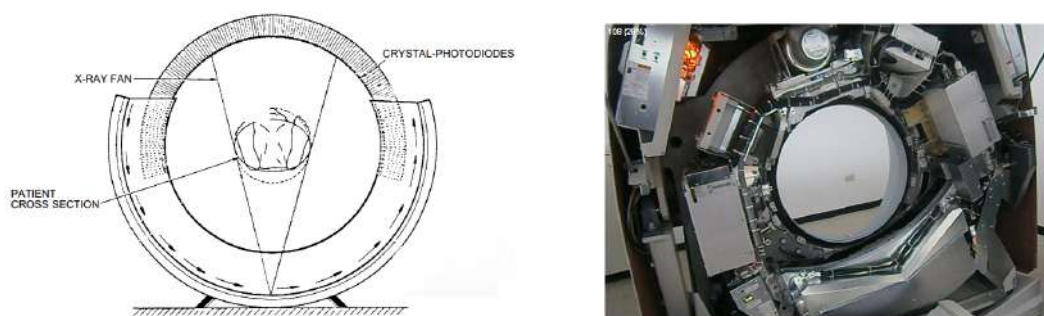


Figura 3.14: Gantry de un equipo TC. A la izquierda, se muestra un esquema general del gantry; a la derecha, un gantry real con sus componentes internos. Imágenes tomadas de [3] (Hendee, W. R., & Ritenour, E. R. 2002. *Medical Imaging Physics*, 4th ed. Wiley-Liss pp. 257) y [57] (Timo Bio Tomógrafos. 2013, diciembre. *Mantenimiento tomógrafo computarizado*, 24366-TOMOGRAFOS - MANUAL).

Tubo de Rayos X

El tubo de rayos X es una ampolla de vidrio sellada al vacío. En el, se encuentra un filamento de Tungsteno Toriado, que emite electrones por procesos mencionados anteriormente. Cambien contiene un target conocido como ánodo rotatorio y una fuente de energía de alto voltaje, llamada voltaje del tubo usada para acelerar los electrones. Los rayos X producidos por el tubo de rayos X tienen un rango de energía de 20 a 150 keV estos son conocidos como Rayos X de diagnóstico.

El tubo de rayos X está rodeado por aceite dieléctrico, que actúa como medio de enfriamiento. El aceite absorbe el calor generado por el ánodo y el tubo, transfiriéndolo a una carcasa que, a su vez, lo disipa hacia el ambiente. Los componentes del tubo de Rayos X se describen en las subsecciones 3.1.2.1 y 3.1.2.2, respectivamente.

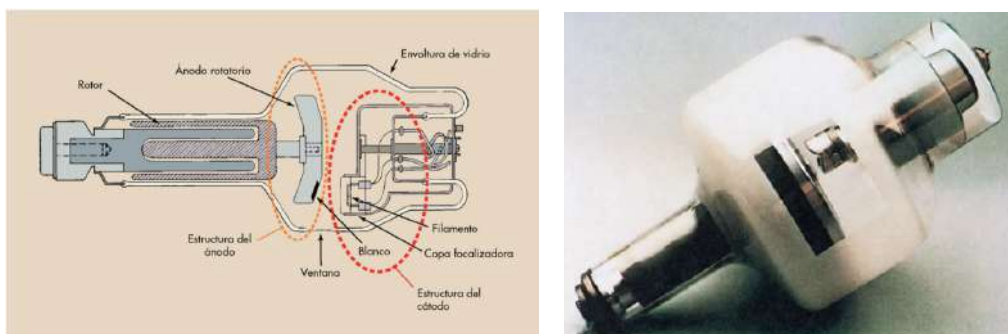


Figura 3.15: Tubo de rayos X de un equipo TC. A la izquierda, se encuentra un diagrama de las partes principales del tubo de rayos X. A la derecha, un tubo de rayos X real. Imagen extraída de [12] (Bushong, S. C., 2019, *Manual de radiología para técnicos* (9ª ed.), pp.120 y 125, Elsevier).

Colimadores

Existen dos tipos de colimadores en un equipo de tomografía computarizada:

- Colimadores pre-paciente: Ayudan a definir el grosor mínimo del haz de rayos X que se emite hacia el paciente, controlando así la sección del cuerpo que será escaneada.
- Colimadores post-paciente: Reducen la radiación dispersa que llega a los detectores.

Los colimadores normalmente están fabricados de Plomo (Pb) o Tungsteno (W). Suelen ser usados estos materiales debido a su alta densidad atómica, la probabilidad de que un fotón de rayos X interactúe de alguna manera con el átomo de uno de estos materiales es más alta. También son útiles estos materiales debido a que su punto de fusión es alto, lo que lo hace resistente a las altas temperaturas que puede provocarse debido a las interacciones de los rayos X.

Plomo (Pb)
 Número atómico: 82
 Punto de fusión: 327°C

Tungsteno (W)
 Número atómico: 74
 Punto de fusión: 3420°C

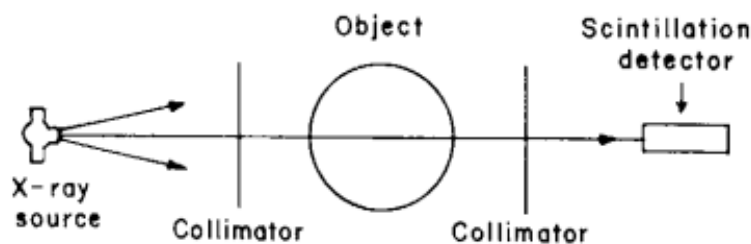


Figura 3.16: Diagrama de posición de los colimadores primario y secundario en un gantry. Imagen extraída de [3] (Hendee, W. R., & Ritenour, E. R. (2002). Medical Imaging Physics (4th ed.). Wiley-Liss pp. 253

Detectores

Los detectores en un equipo TC pueden ser de estado sólido o centelleadores, pero son mayormente usados los llamados centelleadores. Estos detectores tienen la capacidad de transformar los rayos X incidentes en luz visible. Generalmente, están fabricados con ioduro de cesio, un material que, al interactuar con la radiación, se ioniza y emite luz en el proceso.

Los detectores están conectados a foto-diodos de silicio, los cuales convierten la luz visible en una señal eléctrica para su posterior procesamiento.

El número de filas de detectores está directamente relacionado con el número de cortes que el tomógrafo es capaz de adquirir. Por ejemplo, un tomógrafo de 16 cortes contará con 16 filas de detectores, con aproximadamente 800 a 1000 detectores en cada fila. Actualmente, existen tomógrafos con capacidades de 16, 64, 128, 256 y hasta 320 cortes, permitiendo un rango diverso de aplicaciones clínicas y diagnósticas.

La información presentada en esta sección fue recopilada a partir de diversas fuentes bibliográficas, entre ellas: [3, 9, 12, 14, 26, 28, 36, 51, 57, 61]

3.1.5. Tomografía Computarizada de Baja Dosis

La Tomografía Computarizada de Baja Dosis (LDCT) utiliza una cantidad significativamente menor de rayos X para generar imágenes médicas. Esta modalidad surgió como respuesta a los riesgos asociados con la exposición a la radiación ionizante en la tomografía computarizada convencional, entre los que se incluye un aumento en la probabilidad de desarrollar ciertas enfermedades como leucemias, mielomas o linfomas.

El objetivo principal de la LDCT es reducir la dosis de radiación administrada al paciente, manteniendo una calidad de imagen adecuada para el diagnóstico clínico. Este enfoque se alinea

con el principio ALARA (As Low As Reasonably Achievable), que promueve el uso de la menor cantidad de radiación posible sin comprometer la utilidad diagnóstica.

Actualmente, la LDCT se emplea de manera frecuente en diversas aplicaciones, tales como la detección temprana de cáncer de pulmón, la evaluación de daños pulmonares en poblaciones vulnerables (como niños, pacientes con enfermedades crónicas o antecedentes de tabaquismo) y en estudios que requieren exploraciones repetitivas.

La LDCT se basa en la reducción de la cantidad de rayos X utilizados durante el escaneo, lo que disminuye la dosis absorbida por el paciente. Esto se logra mediante ajustes en los parámetros del equipo, como:

- Miliamperaje (mA): Controla la cantidad de electrones liberados por emisión termoiónica, lo que influye directamente en la cantidad de rayos X que se generarán mediante los procesos descritos en 3.1.2. Al variar el mA, obtendremos como resultado una variación proporcional en la amplitud del espectro de rayos X.

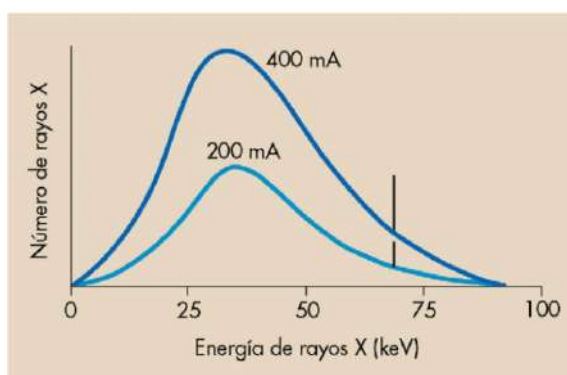


Figura 3.17: Cambio en el espectro de rayos X debido a la variación del mA. Imagen extraída de [12] (Bushong, S. C., 2019, *Manual de radiología para técnicos* (9ª ed.), p. 146, Elsevier).

- Kilovoltaje (kV): El cambio en el kV afecta tanto la amplitud como la posición del espectro de emisión de rayos X. Es decir, se producen rayos X más energéticos y en mayor cantidad.

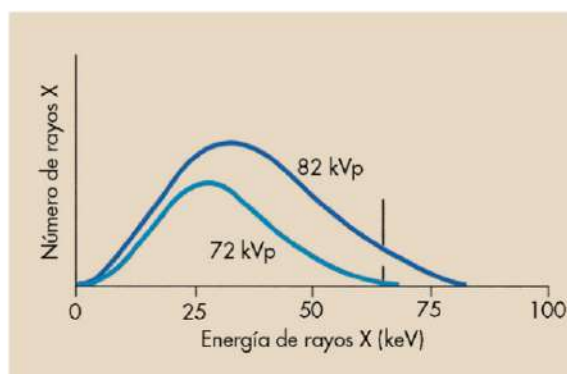


Figura 3.18: Cambio en el espectro de rayos X debido a la variación del kV. Imagen extraída de [12] (Bushong, S. C., 2019, *Manual de radiología para técnicos* (9ª ed.), pp.147, Elsevier).

En la práctica, esto se logra gracias a un circuito regulador el cual se presenta en la siguiente figura.

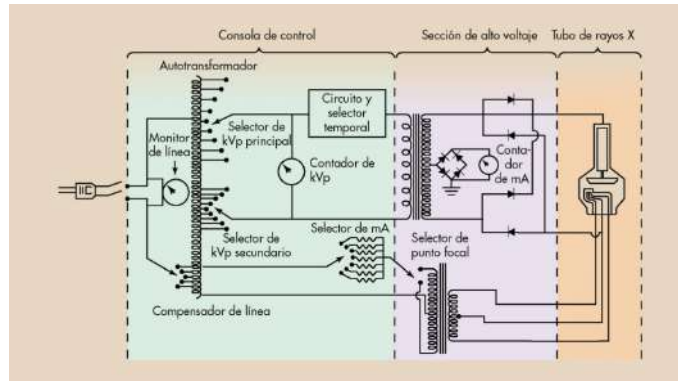


Figura 3.19: Circuito esquemático de un sistema de imagen por rayos X. Imagen extraída de [12] (Bushong, S. C., 2019, *Manual de radiología para técnicos* (9ª ed.), pp.117, Elsevier).

- Tiempo de escaneo (s): Reducir el tiempo de adquisición de la imagen implica una disminución directa en la cantidad de radiación de rayos X incidente en el paciente.

El Colegio Americano de Radiología (ACR) proporciona recomendaciones y directrices específicas para garantizar la reducción de la dosis en diferentes tipos de estudios, pero generalmente esto se basa en la disminución del mA para realizar estudios LDCT.

La disminución del mA depende del tipo de estudio que se vaya a realizar y es determinada por el especialista. Sin embargo, se ha reportado que es posible reducir el mA hasta $\frac{1}{4}$ del valor original, como lo señala la American Association of Physicists in Medicine (n.d.). Low Dose CT Grand Challenge. [7]

En tomografía computarizada, la dosis absorbida se expresa comúnmente mediante el índice dosimétrico $CTDI_{vol}$ (Computed Tomography Dose Index volumen), el cual se reporta en mili-gray (mGy). Este parámetro permite comparar de manera estandarizada la cantidad de radiación utilizada en distintos protocolos de adquisición. En un estudio de tórax convencional en adultos, los valores típicos de $CTDI_{vol}$ se encuentran en un rango de aproximadamente 8 a 12 mGy, dependiendo del tamaño del paciente y de los parámetros técnicos empleados. En contraste, en un protocolo de tomografía computarizada de baja dosis (LDCT) para tórax el $CTDI_{vol}$ recomendado es ≤ 3 mGy, con valores reportados habitualmente entre 1 y 3 mGy. [46], [7]

El uso de las LDCT se enfrenta a un problema en específico el cual es el aumento en la presencia de **ruido** en la imagen resultante, lo que puede llevar a discrepancias o inexactitudes al momento del diagnóstico esto es debido principalmente a lo que se conoce como moteado cuántico.

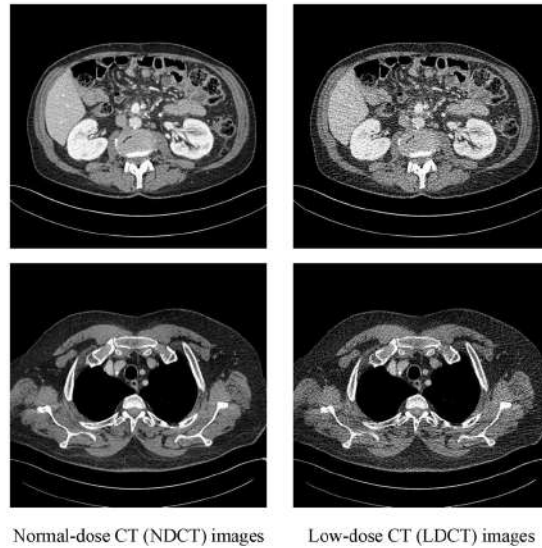


Figura 3.20: Imágenes de TC de la base de dayos de Mayo Clinic para el AAPM Low Dose CT Grand Challenge. La columna izquierda muestra las imágenes de TC de dosis normal y la columna derecha, las de dosis baja. Imagen extraída de [64].

3.1.5.1. Moteado Cuántico

El efecto del moteado cuántico ocurre cuando se utilizan niveles bajos de miliamperaje (mA), lo que produce una clase de granulado en la imagen, constituyendo la principal fuente de ruido en las imágenes de tomografía computarizada (TC). Este fenómeno se debe a la naturaleza aleatoria de la interacción de los fotones con el receptor de imagen y puede explicarse mediante una analogía presentada en el libro *Radiography in the Digital Age* de Quinn B. Carroll (2018) [14] Cap. 15 pp. 242:

"Durante una lluvia ligera, se pueden observar las gotas de agua individuales al caer sobre una acera. Si se cuentan las gotas de lluvia en cada cuadrado de cemento, se puede notar que no están distribuidas de manera uniforme; algunas áreas reciben más gotas que otras. Este fenómeno es aleatorio y depende de la probabilidad estadística de que una región específica sea impactada por más gotas. En cambio, durante una lluvia intensa, aunque persiste la distribución desigual de las gotas, esta se vuelve imperceptible porque la acera queda completamente saturada de agua."

De forma similar, el haz de rayos X se comporta como una "lluvia" de fotones con una distribución aleatoria. Cuando se utilizan valores muy bajos de mAs, el número de fotones que inciden sobre el receptor es reducido, esto es fácil de observar dadas las siguientes relaciones:

$$\begin{aligned} \text{LDCT} : N_{e_{baja}^-} \text{ y NDCT} : N_{e_{normal}^-} \\ \text{tal que } N_{e_{baja}^-} < N_{e_{normal}^-} \\ \therefore N_{Fotones_{e_{baja}^-}} < N_{Fotones_{e_{normal}^-}} \end{aligned}$$

donde $N_{e_{baja}^-}$ es el numero de electrones emitidos en una tomografía computarizada de baja dosis, $N_{e_{normal}^-}$ es el numero de electrones emitidos en una tomografía computarizada de dosis normal, $N_{Fotones_{e_{baja}^-}}$ es el numero de fotones incidente en los detectores por una tomografía de baja dosis y $N_{Fotones_{e_{normal}^-}}$ es el numero de fotones incidente en los detectores por una tomografía de dosis normal. Cabe recalcar que, tal como se ve en la figura 3.17, al variar el mA se reduce el numero de rayos X producidos, mas no la energía promedio que tendrán. En otras palabras, la

energía promedio de los fotones es similar tanto en tomografías de dosis normal como en aquellas de baja dosis siempre y cuando la variación se realice en el mA.

El moteado cuántico en la imagen es consecuencia de fluctuaciones estadísticas en la cantidad de fotones absorbidos por el detector. Suponiendo que cada fotón que interactúa con el detector es completamente absorbido y que la eficiencia del receptor (ϵ) es la misma para fotones primarios y secundarios, la única variable que cambia es el número de fotones detectados en una región determinada del área A del detector.

El problema radica en la variabilidad estadística de la medición de la energía de los fotones incidentes en el detector. Esto provoca que, en una región que debería ser homogénea en la tomografía computarizada (TC), los valores de los píxeles presenten oscilaciones, generando inconsistencias en la imagen resultante.

Lo que permite visualizar una distribución desigual de la exposición, generando un patrón de lo que suele llamarse moteado. Por el contrario, al incrementar el mA, el receptor se satura con fotones, disminuyendo el efecto visible del moteado cuántico.

En general un moteado visible en la imagen indica que una cantidad insuficiente de rayos X ha llegado al receptor de imagen.

La información presentada en esta sección fue recopilada a partir de diversas fuentes bibliográficas, entre ellas: [7, 12, 14, 34, 37, 43, 66]

3.2. Teoría de Imágenes Digitales

3.2.1. Imagen Digital

Una imagen digital es una representación visual de datos, generalmente en forma de una matriz bidimensional donde cada elemento (píxel) tiene un valor que representa su intensidad o color.

Matemáticamente, una imagen digital en escala de grises puede representarse como una función $f(x, y)$:

$$f : \mathbb{N}^2 \rightarrow \mathbb{R} \tag{3.18}$$

donde los ejes x e y se usan para denotar las filas y columnas de la matriz.

Existen diversas formas de representar una imagen digital $f(x, y)$, para procesamiento informático, en forma de ecuación, escribimos la representación de una matriz numérica de tamaños $M \times N$, dado un dominio discreto de píxeles como:

$$f(x, y) = F(i, j) = F_{M \times N} = \begin{bmatrix} f(1, 1) & f(1, 2) & \dots & f(1, N) \\ f(2, 1) & f(2, 2) & \dots & f(2, N) \\ \vdots & \vdots & \ddots & \vdots \\ f(M, 1) & f(M, 2) & \dots & f(M, N) \end{bmatrix} \tag{3.19}$$

Un píxel (abreviatura de "picture element") es la unidad más pequeña de una imagen digital representado como un único valor numérico en imágenes en escala de grises.

Generalmente usamos $f(i, j)$ cuando nos referimos a un píxel con coordenadas (i, j) .

Las imágenes digitales tienen formatos específicos. El formato es la profundidad de bits indica cuántos bits se usan para representar cada píxel en la imagen.

Existen diferentes formatos, algunos son los siguientes:

Enteros sin signo (Unsigned Integer)

- uint8 (8 bits): valores entre 0 y 255 (usado en imágenes estándar en escala de grises y color RGB).
- uint16 (16 bits): valores entre 0 y 65,535 (usado en imágenes médicas como DICOM, y en imágenes HDR)

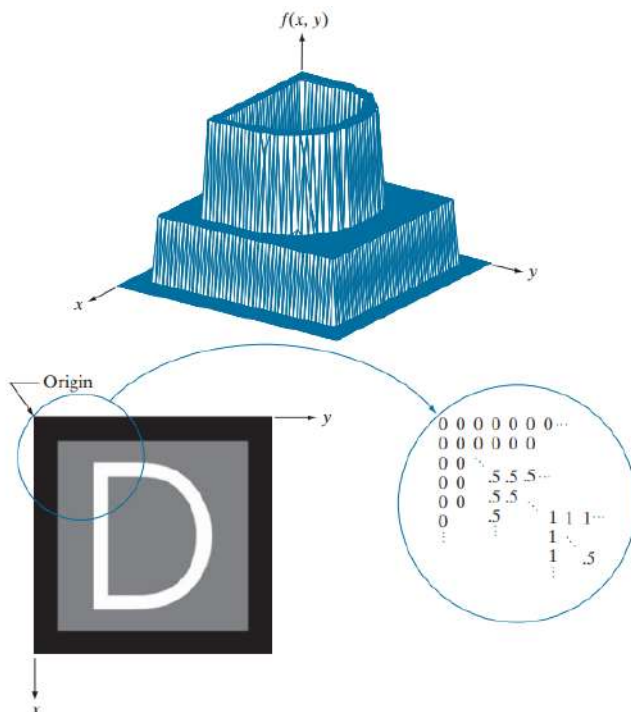


Figura 3.21: a) Imagen representada como una superficie. b) Imagen mostrada como una matriz de intensidad visual. c) Imagen mostrada como una matriz numérica 2D. Imagen extraída de [26](Gonzalez, R. C., & Woods, R. E. (2018). Digital image processing (4th ed.). pp. 67, Pearson.

La información presentada en esta sección fue recopilada de [26]

3.2.2. Ruido en una Imagen

Es posible definir el ruido en una imagen como variaciones aleatorias en los valores de píxeles aleatorios. Este valor en el píxel será anómalo al de sus 8 vecinos.(Fig. 3.22)

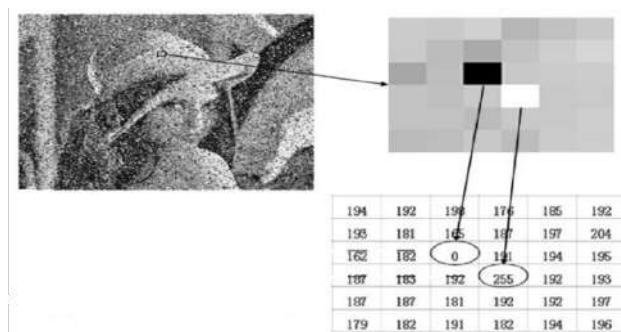


Figura 3.22: Ruido en una imagen. Imagen extraída de [5] (Ahmed, Mahmoud , Eslam and Al-turki, Fahd. (2017). Image noise Detection and Removal based on Enhanced GridLOF Algorithm. International Journal of Advanced Computer Science and Applications. 8.)

Existen diversas maneras de agregar ruido a una imagen, pero, en general, todas implican la adición de valores aleatorios a los píxeles.

En particular, para añadir ruido gaussiano, el modelo de ruido en el dominio de la imagen se puede expresar de la siguiente manera:

Se genera un conjunto de valores aleatorios con una distribución normal (o gaussiana), con media μ y desviación estándar σ . En cada píxel de la imagen, el ruido tiene la forma (3.20):

$$N(i, j) \sim N(\mu, \sigma) \quad (3.20)$$

donde $N(i, j)$ es el ruido que se suma a cada valor (i, j) de la matriz que representa a la imagen, y se obtiene de una distribución normal con media μ y desviación estándar σ .

Este ruido se añade por medio de una suma matricial (3.21)

$$Y(i, j) = X(i, j) + N(i, j) \quad (3.21)$$

donde $Y(i, j)$ es la imagen distorsionada por el ruido, $X(i, j)$ es la imagen original y $N(i, j)$ es el ruido aditivo.

La información presentada en esta sección fue recopilada a partir de las siguientes fuentes bibliográficas: [5, 26]

3.2.3. Métricas de Calidad de Imagen

3.2.3.1. Métricas de Calidad de Imagen con Referencia Completa

La calidad de una imagen puede verse afectada por distorsiones que ocurren tanto durante su captura como en el procesamiento. Ejemplos de estas distorsiones incluyen la presencia de ruido, el desenfoque o la aparición de artefactos.

Existen métricas cuantitativas para evaluar la calidad de las imágenes.

Cuando se dispone de una imagen sin distorsiones, esta puede utilizarse como referencia para evaluar la calidad de otras imágenes. Las métricas de calidad con referencia completa permiten realizar comparaciones directas entre la imagen objetivo y la imagen de referencia.

Usualmente estas métricas suelen ser usadas para comparar diferentes algoritmos de procesamiento de imágenes.

Los algoritmos de calidad de imagen suelen ser con referencia completa. Nos referimos a referencia completa cuando la imagen de entrada como la imagen distorsionada son en esencia las mismas.

Algunos de estos algoritmos de métricas de calidad de imagen con referencia completa son:

- Error cuadrático medio (MSE) mide la diferencia cuadrática promedio entre los valores de píxeles reales y distorsionados. Aunque esta métrica es muy útil, no suele corresponder con la percepción humana de calidad. (Ec. (3.22))

$$MSE = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N (X(i, j) - Y(i, j))^2 \quad (3.22)$$

Donde M y N son el alto y ancho de las matrices asociadas a las imágenes, $X(i, j)$ es la matriz asociada a la imagen original o de entrada y $Y(i, j)$ es la matriz asociada a la imagen distorsionada.

- Error medio Absoluto (MAE) mide la diferencia promedio entre los valores de píxeles reales y distorsionados. (Ec. (3.23))

$$MAE = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N |X(i, j) - Y(i, j)| \quad (3.23)$$

Donde M y N son el alto y ancho de las matrices asociadas a las imágenes, $X(i, j)$ es la matriz asociada a la imagen original o de entrada y $Y(i, j)$ es la matriz asociada a la imagen distorsionada.

- Relación señal-ruido máxima (PSNR) indica la relación entre la intensidad máxima de píxeles y la potencia de la distorsión. Cuantifica la potencia de la señal y el ruido en decibelios, en una escala de 0 a 100, mientras mas alto sea el valor mejor sera la calidad de imagen.(Ec. (3.24))

$$PSNR = 20 \log_{10} \left(\frac{MAX_{pix}}{\sqrt{MSE}} \right) \quad (3.24)$$

Donde MAX_{pix} es el valor máximo de pixel de la matriz y MSE es el error cuadrático medio definido anteriormente.

- Índice de similitud estructural (SSIM) es una métrica utilizada para evaluar la similitud entre dos imágenes. Este índice se basa en un enfoque perceptual, que interpreta la degradación de una imagen como una alteración en la información estructural que percibe el ojo humano. Tiene una escala de 0 a 1, mientras mas alto sea el valor mejor sera la calidad de imagen.(Ec. (3.25))

$$SSIM = \frac{(2\mu_{X(i,j)}\mu_{Y(i,j)} + c_1)(2\sigma_{X(i,j)Y(i,j)} + c_2)}{(\mu_{X(i,j)}^2 + \mu_{Y(i,j)}^2 + c_1)(\sigma_{X(i,j)}^2 + \sigma_{Y(i,j)}^2 + c_2)} \quad (3.25)$$

Donde $X(i, j)$ es la matriz asociada a la imagen original, $Y(i, j)$ es la matriz asociada a la imagen distorsionada, $\mu_{X(i,j)}$ es la media de los valores de los píxeles de la matriz $X(i, j)$, $\mu_{Y(i,j)}$ la media de los valores de los píxeles de la matriz $Y(i, j)$, $\sigma_{X(i,j)}^2$ es la varianza de $X(i, j)$, $\sigma_{Y(i,j)}^2$ la varianza de $Y(i, j)$, $\sigma_{X(i,j)Y(i,j)}$ la covarianza de $X(i, j)$ y $Y(i, j)$, $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ son dos variables para estabilizar la expresión, L el rango dinámico de los valores de los píxeles (normalmente esto es 255), $k_1 = 0,01$ y $k_2 = 0,03$ por defecto.

3.2.3.2. Relación Señal-Ruido (RSR) y Coeficiente Contraste-Ruido (CCR)

En todas las señales siempre está presente cierta contribución del ruido. Este ruido puede considerarse, en muchos casos, como una señal aleatoria que se superpone a la señal real, como se muestra en la figura 3.23. De manera convencional, se utiliza la desviación estándar del ruido como medida cuantitativa de su nivel.

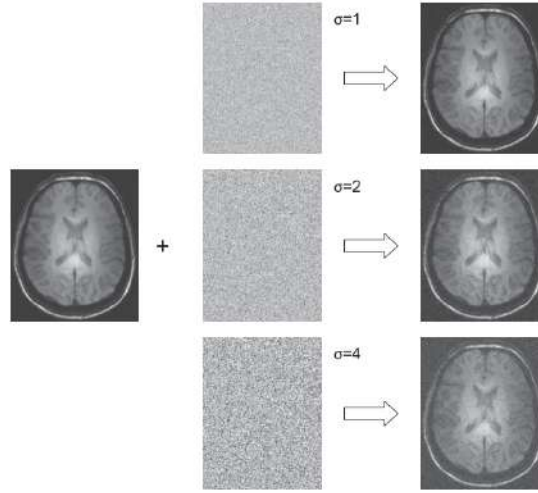


Figura 3.23: Efectos del ruido en la calidad de imagen en una imagen médica, tomada de [54].

La **Relación Señal-Ruido (RSR)** es una métrica que compara el promedio de una señal deseada o útil con el ruido de fondo en una imagen. Se define matemáticamente como:

$$RSR = \frac{\mu_s}{\sigma_n} \quad (3.26)$$

Donde:

- μ_s es la media de la señal en una región de interés (ROI).
- σ_n es la desviación estándar del ruido de fondo.

De este modo, la media μ_s se calcula a partir de una ROI correspondiente a un tejido homogéneo, mientras que la desviación estándar del ruido σ_n se obtiene de una región oscura de la imagen, considerada como fondo.

Los valores de RSR pueden variar en órdenes de magnitud muy grandes. Por lo que se suele usar la siguiente expresión, la cual ayuda a representar los valores en un rango más compacto, manejable y facilita la comparación.

$$RSR_{dB} = 20 \log_{10} \left(\frac{\mu_s}{\sigma_n} \right) \quad (3.27)$$

Se dice que mientras mayor sea el valor numérico del RSR, se tiene una mejor relación señal-ruido. Pero incluso si una imagen en particular tiene una Relación Señal-Ruido (RSR) muy alta, no es muy útil a menos que exista un Coeficiente de Contraste-Ruido (CCR) lo suficientemente alto como para distinguir entre diferentes tejidos. Por ello, es útil acompañar la RSR con el **Coefficiente de Contraste-Ruido (CCR)**.

El **Coefficiente de Contraste-Ruido (CCR)** mide cuán diferentes son dos tejidos y se define como:

$$CCR_{AB} = |RSR_A - RSR_B| \quad (3.28)$$

Donde:

- RSR_A es la relación señal ruido de un tejido A.
- RSR_B es la relación señal ruido de un tejido B.

También mientras mayor sea el valor numérico, mejor es el coeficiente.

Dada la definición que se establece en textos como [4] y [54], es importante señalar algunos aspectos clave para el cálculo de la Relación Señal-Ruido (RSR) y el Coeficiente Contraste-Ruido (CCR) en imágenes médicas ya reconstruidas o generadas, cuyo objetivo sea únicamente la evaluación de la calidad de imagen.

- La **señal** se define como cualquier región de la imagen que involucre al paciente. Para su análisis, se selecciona una región representativa de un tejido homogéneo, como músculo, grasa o hueso, sobre la cual se realiza el cálculo de la media. Esta región se conoce como región de interés para la señal, denotada como $ROI_{señal}$.
- El **ruido**, por otro lado, se evalúa en una zona de la imagen que no contiene información del paciente, típicamente en el fondo. Esta área debe estar libre de artefactos provocados por el cuerpo del paciente y corresponder a regiones oscuras. A esta se le denomina ROI_{ruido} .
- Ambas regiones de interés (ROIs) deben tener el mismo tamaño para asegurar una comparación adecuada entre los valores extraídos de cada una.

La información presentada en esta sección fue recopilada a partir de las siguientes fuentes bibliográficas: [4, 31, 41, 54]

3.2.4. Operador de Convolución 2D

La convolución 2D (\odot) consiste en desplazar un pequeño kernel o filtro, generalmente de tamaño reducido como 3×3 , sobre cada posición de una imagen o matriz de entrada. En cada posición, se realiza un producto punto entre el kernel y el área de la imagen bajo él, y el resultado es la suma de todos estos productos. La salida es una nueva matriz o "mapa de características" que contiene información relevante extraída por el kernel. Generalmente se maneja en el dominio espacial, es decir, directamente en las coordenadas de la imagen (x, y) en cada posición de la matriz de entrada. (Fig. 3.24)

1. **Kernel o ventana:** El kernel es una matriz pequeña (por ejemplo, 3×3 o 5×5) que contiene los valores de los pesos que determinan cómo se extraerán las características de la imagen de entrada.
2. **Desplazamiento del Kernel:** El kernel se "desliza" sobre la matriz de entrada, posicionándose en cada submatriz de tamaño equivalente. En cada posición:
 - Se toma el producto elemento a elemento entre los valores del kernel y los valores correspondientes de la submatriz de la entrada.
 - Estos productos se suman para obtener un solo valor.
 - El valor calculado se coloca en la matriz de salida en la posición correspondiente.

Para una matriz de entrada I y un kernel K , la salida en una posición (i, j) se calcula como, (3.29):

$$O(i, j) = \sum_{m=1}^M \sum_{n=1}^N K(m, n) \cdot I(i + m, j + n) \quad (3.29)$$

donde M y N son las dimensiones del kernel, y O es la matriz de salida.

Esta operación debe incluir pasos adicionales como:

- *Stride* (paso): Determina cuántos píxeles se desplaza el kernel cada vez. Un stride de 1 significa que el kernel se desplaza un píxel a la vez; un stride mayor genera una salida de menor tamaño.
- *Padding* (relleno): Se utiliza para controlar el tamaño de la salida. Con padding same o ceros, se agregan ceros alrededor de los bordes de la imagen de entrada para mantener el tamaño original de la imagen en la salida.

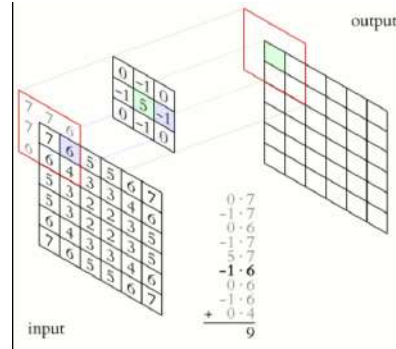


Figura 3.24: Ilustración del operador convolución obtenida de [67]

La información presentada en esta sección fue recopilada a partir de las siguientes fuentes bibliográficas: [18, 67]

3.2.5. Operador o Filtro Sobel

El operador es utilizado en el procesamiento de imágenes, particularmente para algoritmos de detección de bordes donde crea una imagen que enfatiza estos.

La formulación para la obtención de una imagen a la cual se le aplico el operador Sobel es la siguiente:

Sea $X(i, j)$ una matriz que representa a la imagen original, usando 2 ventanas de 3x3 que convolucionan a $X(i, j)$ con padding same. Se obtendrán 2 imágenes $G_{X \rightarrow}(i, j)$ y $G_{X \uparrow}(i, j)$ correspondientes a los cambios horizontales y verticales del mismo tamaño que la imagen original X .

$$G_{X \rightarrow}(i, j) = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \odot X(i, j)$$

$$G_{X \uparrow}(i, j) = \begin{bmatrix} -1 & -2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \odot X(i, j)$$

Elevando al cuadrado cada uno de los componentes de las 2 imágenes, $(g_{x \rightarrow}(i, j))^2$ y $(g_{x \uparrow}(i, j))^2$. Luego sumando los valores uno a uno y calculando la raíz cuadrada a cada uno de los resultados de la suma.

$$G_X(i, j) = \begin{bmatrix} \sqrt{g_{x \rightarrow}(1, 1)^2 + g_{x \uparrow}(1, 1)^2} & \sqrt{g_{x \rightarrow}(1, 2)^2 + g_{x \uparrow}(1, 2)^2} & \dots & \sqrt{g_{x \rightarrow}(1, N)^2 + g_{x \uparrow}(1, N)^2} \\ \sqrt{g_{x \rightarrow}(2, 1)^2 + g_{x \uparrow}(2, 1)^2} & \sqrt{g_{x \rightarrow}(2, 2)^2 + g_{x \uparrow}(2, 2)^2} & \dots & \sqrt{g_{x \rightarrow}(2, N)^2 + g_{x \uparrow}(2, N)^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{g_{x \rightarrow}(M, 1)^2 + g_{x \uparrow}(M, 1)^2} & \sqrt{g_{x \rightarrow}(M, 2)^2 + g_{x \uparrow}(M, 2)^2} & \dots & \sqrt{g_{x \rightarrow}(M, N)^2 + g_{x \uparrow}(M, N)^2} \end{bmatrix} \quad (3.30)$$

Obtenemos el operador sobel aplicado a una imagen (Fig.A.1).

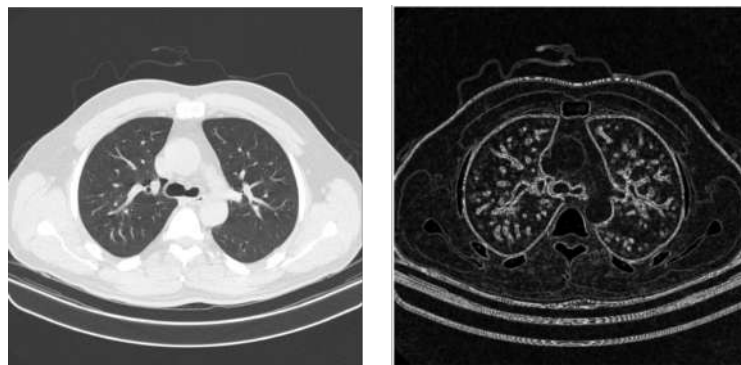


Figura 3.25: TAC de Tórax vs TAC de Tórax aplicando el operador Sobel

La información presentada en esta sección fue recopilada a partir de las siguientes fuentes bibliográficas: [26, 35, 69]

3.3. Redes Neuronales Artificiales

Las redes neuronales artificiales (RNAs) son modelos computacionales inspirados en la estructura y funcionamiento del cerebro humano. Su diseño se basa en la interconexión de unidades de procesamiento simples, denominadas neuronas artificiales, que trabajan en conjunto para aprender patrones complejos a partir de los datos. Debido a su capacidad de generalización y aprendizaje, las RNAs han demostrado ser altamente eficaces en tareas como reconocimiento de patrones, procesamiento de imágenes, predicción de series temporales y clasificación de datos.

Es un modelo matemático que imita la activación de las neuronas biológicas. Una neurona artificial recibe múltiples entradas ponderadas, realiza una combinación lineal de ellas y aplica una función de activación para generar una salida.

Matemáticamente, una neurona artificial puede expresarse como:

$$y = \sigma \left(\sum_{i=1}^n w_i x_i + b \right) \quad (3.31)$$

Donde y es la salida de la neurona, x_i representa las entradas de la neurona, w_i son los pesos asignados a cada entrada, b es el sesgo (*bias*) que permite ajustar la activación de la neurona. σ es la función de activación, que introduce no linealidad en la red.

3.3.1. Redes Neuronales de Convolución (RNC)

Las Redes Neuronales Convolucionales son un tipo de red neuronal artificial diseñada para procesar datos con una estructura de cuadrícula, como las imágenes, se suele mencionar que esta estructura de red aprende directamente de los datos. Esta clase de redes usan el operador convolución (3.2.4) en lugar de solo conexiones densas para extraer características locales y compartir pesos w , lo que reduce la complejidad computacional. Los aspectos importantes de las redes neuronales incluyen las funciones de activación, las funciones de costo, las diferentes capas y la optimización.

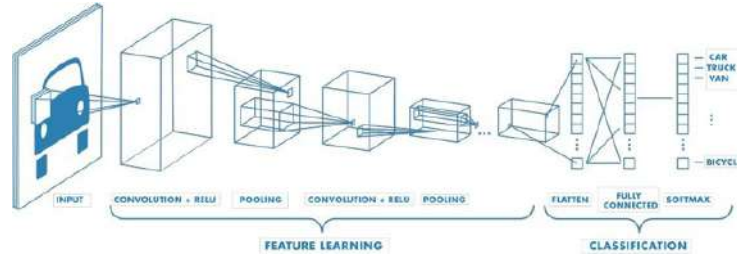


Figura 3.26: Ejemplo de red con múltiples capas convolucionales. Se aplican filtros a las imágenes de entrenamiento y la salida resultante se emplea como entrada para la siguiente capa. Imagen extraída de [40], (MathWorks. (s.f.). Convolutional neural network.)

3.3.2. Funciones de Activación

Las funciones de activación son usadas para aprender las características abstractas a través de transformaciones no lineales. Esto es debido a que se debe agregar una curvatura no lineal en el panorama de optimización para mejorar la convergencia de entrenamiento de la red.

Algunas funciones de Activación son:

- **Sigmoide** utilizada en problemas de clasificación binaria:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.32)$$

La función sigmoide comprime los valores en el rango (0,1), lo que facilita la interpretación como una probabilidad.

- **ReLU** (Rectified Linear Unit):

$$ReLU(x) = \max(0, x) = \begin{cases} x, & \text{si } x \geq 0 \\ 0, & \text{otro caso} \end{cases} \quad (3.33)$$

3.3.3. Funciones de Costo

Las funciones de costo miden cuán bien el modelo está realizando su tarea y guían el proceso de optimización. Básicamente mide el error entre las predicciones y los valores reales. El objetivo del entrenamiento es minimizar esta diferencia.

Existen diferentes funciones de costo también dependiendo de la tarea que se busca realizar, algunas de ellas son:

- **Error Cuadrático Medio** (MSE por sus siglas en inglés) es comúnmente usado para tarea de regresión.

$$C(w) = \frac{1}{m} \sum_i^m (P_i(w) - \hat{P}_i)^2 \quad (3.34)$$

donde m es el número total de las muestras, $P_i(w)$ es la salida predicha por la red y \hat{P}_i es el valor real o verdadero.

- **Entropía Cruzada Binaria** (BCE) es comúnmente usada para clasificación. Mide la disimilitud entre la probabilidad predicha y la etiqueta de clase verdadera. La entropía cruzada se origina en la teoría de la información, donde mide la disimilitud entre dos distribuciones de probabilidad.

$$C(w) = -\frac{1}{m} \sum_{i=1}^m [y_i \ln(P_i(w)) + (1 - y_i) \ln(1 - P_i(w))] \quad (3.35)$$

Esta fórmula se puede entender intuitivamente considerándola en dos partes:

$$\begin{cases} -\ln(P_i(w)), & \text{si } y_i = 1 \\ -\ln(1 - P_i(w)), & \text{si } y_i = 0 \end{cases} \quad (3.36)$$

donde y_i es la etiqueta de clase verdadera y $P_i(w)$ es la probabilidad predicha de que la clase sea 1. El costo se minimiza cuando la probabilidad predicha $P_i(w)$ coincide con la etiqueta real y_i .

3.3.4. Algoritmo de Optimización

La función principal de los algoritmos de optimización es minimizar una función de costo ajustando los parámetros del modelo de manera eficiente. Estos algoritmos buscan encontrar el conjunto de parámetros óptimos que minimicen el costo.

Algunos de los métodos de optimización son:

- **Descenso de Gradiente Estocástico (SGD)** donde se actualizan los pesos del modelo utilizando una muestra aleatoria del conjunto de datos en cada iteración, en lugar de usar todo el conjunto de datos como en el descenso de gradiente batch.

El SGD actualiza los parámetros del modelo según la siguiente ecuación:

$$w_{t+1} = w_t - \eta \frac{\partial C(w)}{\partial w_t} \quad (3.37)$$

Donde:

w_t representa los parámetros en la iteración actual.

η es la tasa de aprendizaje.

$\frac{\partial C(w)}{\partial w_t}$ es el gradiente de la función de costo respecto a los parámetros.

- **Optimizador Adam** ajusta la tasa de aprendizaje de manera adaptativa para cada parámetro del modelo. Adam utiliza dos momentos:

Promedio exponencial de los gradientes (momento de primer orden):

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \frac{\partial C(w)}{\partial w_t} \quad (3.38)$$

Promedio exponencial de los cuadrados de los gradientes (momento de segundo orden):

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) \left(\frac{\partial C(w)}{\partial w_t} \right)^2 \quad (3.39)$$

Los momentos se corrigen con:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (3.40)$$

Finalmente, los parámetros se actualizan con:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t \quad (3.41)$$

Donde:

β_1 y β_2 son factores de decaimiento ($\beta_1 = 0.9$ y $\beta_2 = 0.999$ por defecto).

ϵ es un pequeño valor para evitar divisiones por cero.

3.3.5. Max Pooling

Es una operación de agrupación que enfatiza el valor máximo de un mapa de características generando un mapa de características mas pequeño.

Supongamos una matriz de entrada X :

$$X = \begin{bmatrix} 2 & 3 & 2 & 3 \\ 4 & 9 & 5 & 1 \\ 2 & 6 & 2 & 3 \\ 1 & 2 & 5 & 9 \end{bmatrix}$$

Aplicamos MaxPooling con los siguientes parámetros:

- Tamaño de ventana $k = 2 \times 2$,
- Stride $s = 2$.

$$\text{Región 1: } \begin{bmatrix} 2 & 3 \\ 4 & 9 \end{bmatrix}, \quad \text{Región 2: } \begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix},$$

$$\text{Región 3: } \begin{bmatrix} 2 & 6 \\ 1 & 2 \end{bmatrix}, \quad \text{Región 4: } \begin{bmatrix} 2 & 3 \\ 5 & 9 \end{bmatrix}.$$

$$\text{máx} \begin{bmatrix} 2 & 3 \\ 4 & 9 \end{bmatrix} = 9, \quad \text{máx} \begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix} = 5,$$

$$\text{máx} \begin{bmatrix} 2 & 6 \\ 1 & 2 \end{bmatrix} = 6, \quad \text{máx} \begin{bmatrix} 2 & 3 \\ 5 & 9 \end{bmatrix} = 9.$$

Entonces

$$Y = \begin{bmatrix} 9 & 5 \\ 6 & 9 \end{bmatrix}.$$

3.3.6. Up Sampling

Es una técnica utilizada en redes neuronales convolucionales para aumentar la resolución espacial de un mapa de características. Este proceso es clave en tareas como segmentación semántica, generación de imágenes y reconstrucción, donde se necesita obtener una salida con dimensiones mayores que las de la entrada.

Up Sampling copia o repite cada elemento del mapa de entrada.

Dado un mapa de entrada X de tamaño 2×2 con un factor de escala $r = 2$:

$$X = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$

el mapa de salida Y generado será:

$$Y = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ 3 & 3 & 4 & 4 \\ 3 & 3 & 4 & 4 \end{bmatrix}.$$

La información presentada en esta sección fue recopilada a partir de las siguientes fuentes bibliográficas: [1, 2, 18, 24, 32, 40, 56, 60].

3.3.7. Autoencoders

Un Autoencoder es un tipo de Red Neuronal utilizada principalmente para la compresión y descompresión de datos por medio del aprendizaje de una representación compacta y eficiente de los datos.

Un Autoencoder consiste en tres partes principales (Fig. 3.27)

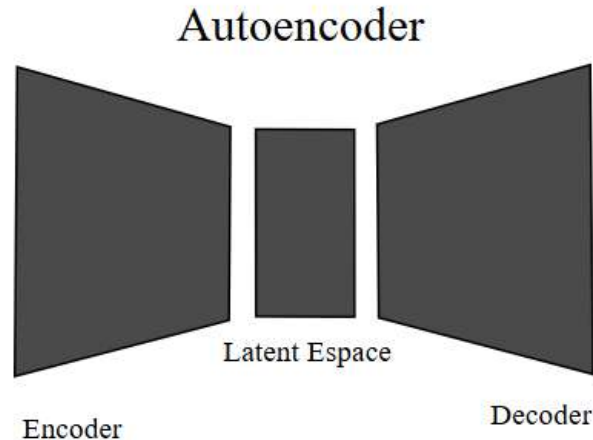


Figura 3.27: Esquema de un Autoencoder.

- Encoder (Codificador): Comprime los datos de entrada en una representación más pequeña, conocida como el espacio latente.
- Espacio latente: Es la representación mas compacta de los datos de entrada.
- Decoder (Decodificador): Reconstruye los datos originales a partir de la representación comprimida, tratando de minimizar la diferencia entre la entrada original y la salida reconstruida.

Un autoencoder es generalmente simétrico en su arquitectura, es decir, el encoder y el decoder tienen un número similar de capas.

La idea central es que el modelo aprenda una codificación que preserve la mayor cantidad de información relevante posible, generalmente reduciendo la dimensionalidad espacial de los datos de entrada. Esto permite que, a partir de los datos codificados, el modelo pueda reconstruir de forma eficiente la información original.

El entrenamiento de un autoencoder se lleva a cabo minimizando una función de costo que mide la diferencia entre la entrada original y la reconstrucción como se vio en la sección 3.3.3.

Generalmente este tipo de redes son usadas con el objetivo de reducir la dimensionalidad espacial o para eliminación de ruido.

La información presentada en esta sección fue recopilada a partir de las siguientes fuentes bibliográficas: [18, 39, 49, 68].

3.3.8. Redes Generativas Adversariales (GANs)

Las Redes Generativas Adversariales (GANs), por sus siglas en inglés) representan un enfoque de aprendizaje no supervisado, permitiendo generar datos sintéticos realistas a partir de distribuciones aleatorias.

La arquitectura GAN está formada por dos redes neuronales, una denominada discriminadora (D) y otra generadora G. La red G se encarga de generar nuevas instancias del mismo dominio

que el del conjunto de datos reales. La red D se encarga de discriminar si los datos de entrada son reales o son generados. Tanto el discriminador como el generador entrenan de manera conjunta de manera que G maximice sus posibilidades de no ser detectada por D y D pueda detectar los datos generados por G .

Para una GAN se genera un vector R que es muestreo de una distribución de probabilidad aleatoria P_R , $R \sim P_R$, esta es la entrada de G . El propósito de la optimización es conseguir que $G(R)$ termine siendo una estimación de la distribución de probabilidad P_{data} .

El discriminador tienen como función de costo

$$C_D(w) = \frac{1}{m} \sum_{i=1}^m [\ln(D(x_i)) + \ln(1 - D(G(R_i)))] \quad (3.42)$$

Y el generador tendrá como función de costo:

$$C_G(w) = \frac{1}{m} \sum_{i=1}^m [\ln(1 - D(G(R_i)))] \quad (3.43)$$

Durante el proceso de optimización la red D recibirá como entrada datos reales y otros del generador G . Se optimizará su funcionamiento para que su discriminación sea efectiva. Al mismo tiempo, cuando D reciba una entrada procedente de G , G se optimizará para mejorar sus predicciones y hacer cada vez más difícil el papel de D . Esto únicamente se puede conseguir mejorando la calidad de los datos generados.

Lo que nos da la función de costo de una GAN, que suele tener la siguiente notación:

$$\min_G \max_D C(D, G) = \mathbb{E}_{x \sim P_{data}} [\ln(D(x))] + \mathbb{E}_{R \sim P_R} [\ln(1 - D(G(R)))] \quad (3.44)$$

3.3.8.1. Redes Generativas Adversariales Semisupervisadas (SSGAN)

Esta clase de redes contienen una variación en el discriminador que le permite aprovechar las ventajas de contar con datos etiquetados.

Consiste en añadir la predicción de la clase de pertenencia de cada dato real. En casos que se tiene mas de una clase se agrega un costo supervisado generalmente es el llamado entropía cruzada categórica y en aquellos que se requiera una clasificación binaria típica, se usa BCE.

3.3.9. Entrenamiento con Datos Desemparejados

Uno de los objetivos específicos más relevantes es implementar un método de entrenamiento basado en imágenes desemparejadas.

Generalmente, al entrenar una Red Neuronal Convolutiva (RNC) para la reconstrucción de imágenes, se dispone de pares de imágenes: una imagen de entrada (Input) y una imagen objetivo (Ground Truth).

Imagen de entrada (Input): Es la imagen que se alimenta a la red. Dependiendo de la tarea específica, esta imagen puede ser:

- Una imagen ruidosa, para modelos de eliminación de ruido.
- Una imagen degradada (por ejemplo, de baja resolución), para tareas de superresolución.
- Una imagen incompleta, para tareas de relleno (inpainting).
- Una imagen comprimida o con artefactos, para restauración de calidad.

Imagen objetivo (Ground Truth): Es la imagen con la que se compara la imagen reconstruida, y debe cumplir con las siguientes características:

- Ser una imagen limpia y sin ruido, para tareas de reducción de ruido.
- Representar una versión completa o sin defectos de la imagen de entrada.

La función de costo de la red neuronal mide la diferencia entre la imagen reconstruida y la imagen objetivo. Una menor diferencia indica una mejor reconstrucción de la imagen.

En el enfoque tradicional, conocido como entrenamiento con *imágenes emparejadas*, se requiere que la imagen de entrada (distorsionada) y la imagen objetivo (original) sean, en esencia, la misma imagen.

La información presentada en esta sección fue recopilada a partir de las siguientes fuentes bibliográficas: [11, 20, 27, 45, 53, 55, 58, 59].

Capítulo 4

Metodología

Como se mencionó en la Sección 1.4, actualmente no existe un enfoque que realice un proceso de entrenamiento especializado en la reducción de ruido y la preservación de bordes en imágenes LDCT que, además, analice el impacto de cada reconstrucción en la clasificación.

Por lo que, se plantea que, mediante un proceso de entrenamiento que realice una alineación del espacio latente de las imágenes LDCT a la región del espacio latente de las imágenes NDCT mediante un Autoencoder, que use como función de costo métricas de calidad y un método de preservación de bordes, es posible preservar una mayor cantidad de información relevante en las imágenes LDCT. Contribuyendo a una mejor clasificación diagnóstica, facilitando la identificación precisa de pacientes sanos y enfermos.

Con el objetivo general de que la reconstrucción de las Tomografías Computarizadas de Baja Dosis (LDCT) presenten una reducción de ruido y preservación de bordes.

En esta tesis se han planteado cinco objetivos específicos:

1. Encontrar y emular una de base de datos de TC dosis normal y Baja Dosis, (LDCT), respectivamente de pacientes enfermos y sanos.
2. Determinar las configuración de Red Autoencoder y Clasificador para tomar base de estudio.
3. Implementar un método de preservación de bordes.
4. Entrenar el algoritmo por medio de imágenes desparejadas LDCT y NDCT
5. Realizar una alineación en los espacios latentes de las LDCT con las NDCT.

Los cuales se abordarán progresivamente a lo largo de la metodología.

4.1. Base de Datos

Para realizar el primer objetivo, se llevaron a cabo las actividades descritas en esta sección.

4.1.1. Base de datos General

Se encontró la base de datos COVID-CT-MD la cual es de uso publico y contiene tomografías computarizadas de tórax de 301 pacientes, distribuidos de la siguiente manera: 242 pacientes con infección por COVID-19 y Neumonía Adquirida en la Comunidad (NAC) y 76 pacientes sin patologías. En total, la colección cuenta con 23,280 imágenes. Los estudios fueron recopilados entre febrero y abril de 2020. La adquisición y clasificación tuvo lugar en el Centro de Imágenes Babak, Teherán, Irán por 3 radiólogos expertos. En la documentación de la base de datos mencionan que

se excluyeron las imágenes CT de mala calidad y también aquellas que contenían alguna clase de artefactos visibles.

Todas las imágenes de encuentran en formato PNG (Portable Network Graphics) y fueron capturadas con un escáner SIEMENS SOMATOM Scope en vista axial, utilizando una técnica de adquisición helicoidal, la reconstrucción se realizó mediante el algoritmo de retroproyección filtrada (FBP), teniendo dimensiones de 512 x 512 píxeles. Las configuraciones de exploración usadas para adquirir la base de datos fueron de 110 a 130 kV con una corriente del tubo de rayos X de 200 a 343 mA. Estos valores se encuentran dentro de los parámetros estándar para una tomografía axial computarizada (TAC) de tórax con dosis normal.

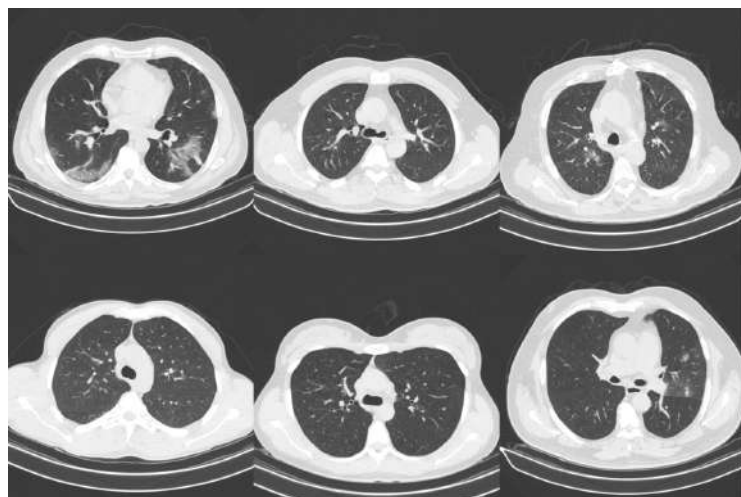


Figura 4.1: Tomografías Computarizadas Axiales de Tórax, extraídas de la base de datos COVID-CT-MD.

4.1.2. Preprocesamiento de la Base de datos

4.1.2.1. Conjunto de la Base de Datos

Debido al gran tamaño de la base de datos, se tomó un conjunto de solo 10,000 imágenes de 245 pacientes:

5,000 imágenes corresponden a 169 pacientes enfermos (aproximadamente 32-29 imágenes por paciente) 5,000 a 76 pacientes sanos (aproximadamente 68-65 imágenes por paciente).

Se normalizaron las imágenes en una escala de $[0,1]$ y se escalaron a 256 x 256 píxeles.

4.1.2.2. Emulación de Tomografías Computarizadas de Baja Dosis (LDCT)

Para emular las Tomografías Computarizadas de Baja Dosis (LDCT) se añadió ruido gaussiano, de la misma forma mencionada en (3.20), ya que este es considerado una buena aproximación [16] [23] [29]. El Ruido Gaussiano añadido tiene los siguientes parámetros:

- Desviación estándar $\sigma = 0.06$
- Media $\mu = 0$

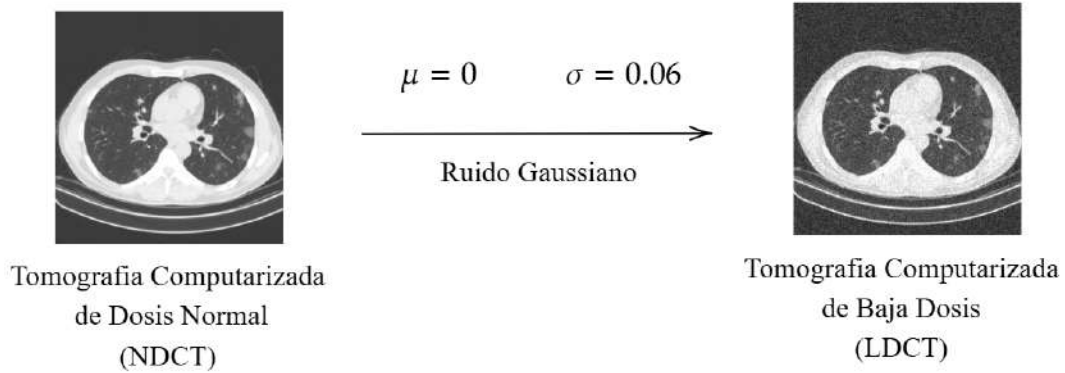


Figura 4.2: Emulación de Tomografías Computarizadas de Baja Dosis (LDCT)

Ya que, en investigaciones que buscan realizar tareas similares de reducción de ruido en imágenes LDCT, se ha reportado que las imágenes LDCT reales adquiridas con $\frac{1}{4}$ de la dosis original presentan valores de PSNR entre 24.4 y 24.6 dB, [52], [74]. Con los parámetros de ruido gaussiano añadido en este estudio, se obtuvieron valores de PSNR dentro de este rango, alcanzando específicamente una media de 24.676803 ± 0.004439 dB.

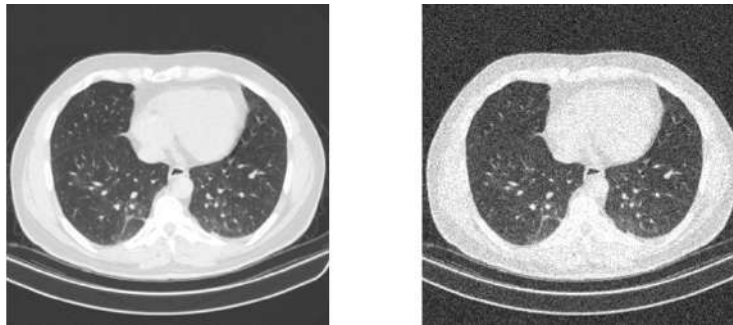


Figura 4.3: Tomografía Computarizada Axiales de Tórax de Dosis Normal (NDCT) vs Tomografía Computarizada Axial de Tórax con ruido gaussiano añadido con $\mu = 0$ y $\sigma = 0.06 \sim$ (LDCT).

4.2. Arquitecturas

Ahora, para poder completar nuestro segundo objetivo específico, se llevaron a cabo las actividades descritas a continuación.

4.2.1. Autoencoder

Se decidió explorar una red neuronal convolucional (RNC) con estructura de Autoencoder, ya que, como se menciona en la sección 3.3.7, este tipo de redes es ampliamente reconocido por su efectividad en realizar tareas relacionadas a la reducción de ruido. Diversos estudios han demostrado el potencial de los Autoencoders para llevar a cabo estas tareas con éxito [15, 16, 25, 44, 52, 63, 74]. Además, un Autoencoder resulta especialmente valioso cuando se requiera realizar una tarea de extracción de puntos latentes del espacio latente.

El Autoencoder implementado tiene la estructura del diagrama (Fig.4.4).

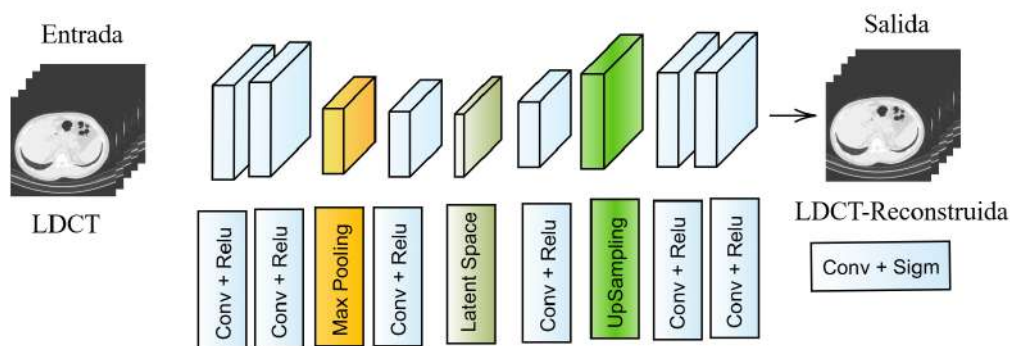


Figura 4.4: Estructura de Red Autoencoder

Donde:

- El codificador está compuesto por tres capas de convolución y una de maxpooling. La primera capa de convolución tiene 50 filtros de 3x3, con activación ReLU y padding same. La segunda capa de convolución cuenta con 30 filtros de 3x3, activación ReLU y padding same, seguida de una capa de maxpooling de 2x2 con padding same. La tercera capa de convolución contienen 20 de 3x3 con activación ReLU y padding same.
- El espacio latente surge de una capa de convolución que cuenta con un solo filtro de 3x3, activación ReLU y padding same.
- El decodificador consta de cuatro capas de convolución y una de upsampling. La primera capa del decoder tiene 20 filtros de 3x3, activación ReLU y padding same. La segunda capa de convolución contiene 30 filtros de 3x3, activación ReLU y padding same, seguida de una capa de upsampling de 2x2. La tercera capa de convolución cuentan con 50 de 3x3, activación ReLU y padding same. Finalmente la cuarta capa de convolución cuenta con 1 filtro de 3x3, función de activación sigmoide y padding same.

La estructura Autoencoder fue inspirado por varios puntos clave mencionados en diferentes Artículos y Foros.

En [15, 16, 74], mencionan que a través de experimentos exhaustivos, se ha demostrado que las redes neuronales convolucionales (RNC) con una estructura secuencial presentan un mayor rendimiento y potencial para resolver tareas de reducción de ruido en imágenes, en comparación con redes que emplean diferentes tipos de conexiones.

En [16, 25, 44] explican que utilizar más de 10 capas de convolución puede llevar a una pérdida o desvanecimiento de detalles en las imágenes. Por ello, se recomienda mantener la cantidad de capas por debajo de este umbral. Además, aunque el número de filtros en cada capa puede variar, se ha observado que los Autoencoders logran mejores resultados en la reconstrucción de imágenes cuando se mantiene una cantidad simétrica de filtros entre las capas del codificador y el decodificador.

Ahora, para reducir y ampliar la dimensionalidad espacial, se tomó como referencia lo siguiente:

En [1, 2, 73], se ha demostrado que las capas de submuestreo con MaxPooling son especialmente útiles para la identificación de características relevantes en las imágenes, ya que permiten conservar las características más significativas mientras se reduce la dimensionalidad.

Por otro lado, en [8, 49, 60], se ha observado que la función UpSampling es más adecuada para tareas de reconstrucción de imágenes. Esto se debe a que no ajusta parámetros, sino que simplemente aumenta la dimensionalidad manteniendo los mismos valores de los parámetros de entrada, lo cual permite conservar la integridad de las características originales sin introducir distorsiones adicionales en la imagen reconstruida.

4.2.2. Clasificador

Aprovechando el hecho de que la base de datos ya incluía información que indicaban el estado de salud de cada persona (Sana o enferma), se decidió usar el siguiente clasificador.

El clasificador implementado tiene la estructura del diagrama (Fig.4.9).

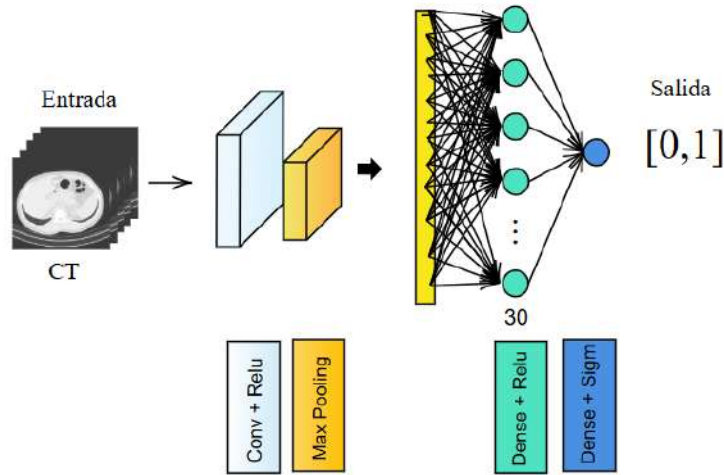


Figura 4.5: Estructura de Clasificador.

Donde el clasificador se compone por una capa de convolución que consta de 16 filtros de 3x3, tiene activación relu y padding same, una de Maxpooling de 2x2 con padding same, una red densa de 30 neuronas con activación relu conectada a una sola neurona con activación sigmoide que corresponde a la clasificación 0 = Sana o 1 = Enferma respectivamente.

4.3. Reducción de Ruido y Preservación de Bordes

Con el objetivo de mejorar la reconstrucción de las imágenes LDCT, especialmente en términos de reducción de ruido y preservación de bordes, se exploraron diversas estrategias en este proceso.

Esto nos conduce al desarrollo del tercer objetivo específico:

3. Implementar un método de preservación de bordes.

4.3.1. Función de Costo

El entrenamiento de una red Autoencoder se basa en la minimización de la diferencia entre los datos de entrada y los datos reconstruidos, mediante una función de costo o pérdida $C(w)$. Generalmente, se utiliza el Error Cuadrático Medio (MSE), como se muestra en la ecuación (4.1). Según lo señalado en [29,72], la adición de diferentes métricas que evalúan la calidad de la imagen, como una suma ponderada en la función de costo de una RNC, ha demostrado mejorar significativamente la reconstrucción de imágenes LDCT.

$$C(w) = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N (X(i, j) - Y_w(i, j))^2 \quad (4.1)$$

Donde M y N son el alto y ancho de las matrices asociadas a las imágenes, $X(i, j)$ es la matriz asociada a la imagen original o de entrada y $Y_w(i, j)$ es la matriz asociada a la imagen reconstruida.

Por ello, se incorporó una suma ponderada de dos funciones adicionales relacionadas con las métricas de calidad de imagen. En concreto, se utilizaron la Relación Señal-Ruido Máxima (PSNR) y el Índice de Similitud Estructural (SSIM).

$$\begin{aligned}
 C(w) = & \alpha \cdot \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N (X(i, j) - Y_w(i, j))^2 \\
 & + \beta \cdot \left(1 - (0.01) \left(20 \log_{10} \left(\frac{MAX_{pix}}{\sqrt{MSE}} \right) \right) \right) \\
 & + \gamma \cdot \left(1 - \frac{(2\mu_{X(i,j)}\mu_{Y_w(i,j)} + c_1)(2\sigma_{X(i,j)Y_w(i,j)} + c_2)}{(\mu_{X(i,j)}^2 + \mu_{Y_w(i,j)}^2 + c_1)(\sigma_{X(i,j)}^2 + \sigma_{Y_w(i,j)}^2 + c_2)} \right)
 \end{aligned} \tag{4.2}$$

Donde:

- α , β y γ son valores asignados a cada función en la combinación ponderada.
- MAX_{pix} es el valor máximo del píxel en la matriz correspondiente a la imagen original.
- MSE es el Error Cuadrático Medio, definido en la ecuación (4.1).
- $X(i, j)$ y $Y_w(i, j)$ son las matrices asociadas a la imagen original (o de entrada) y la imagen reconstruida, respectivamente.
- $\mu_{X(i,j)}$ y $\mu_{Y_w(i,j)}$ representan las medias de los valores de los píxeles de las matrices $X(i, j)$ y $Y_w(i, j)$.
- $\sigma_{X(i,j)}^2$ y $\sigma_{Y_w(i,j)}^2$ son las varianzas de $X(i, j)$ y $Y_w(i, j)$, respectivamente.
- $\sigma_{X(i,j)Y_w(i,j)}$ es la covarianza entre $X(i, j)$ y $Y_w(i, j)$.
- $c_1 = (k_1L)^2$ y $c_2 = (k_2L)^2$ son constantes de estabilización, donde L es el rango dinámico de los píxeles (normalmente 255), y los valores por defecto son $k_1 = 0.01$ y $k_2 = 0.03$.

Es importante destacar que, tanto para las métricas PSNR (Relacion Señal-Ruido Máxima) como para SSIM (Índice de Similitud Estructural), un valor numérico alto indica una mejor reconstrucción de la imagen. Esto implica que estas dos funciones deberían maximizarse, a diferencia del MSE, que se minimiza.

Además, debido a que el PSNR tiene un rango de valores entre [0, 100] y el SSIM entre [0, 1], las transformaciones aplicadas a estas métricas tienen como objetivo:

- Convertirlas en funciones a minimizar, alineándolas con el comportamiento del MSE.
- Ajustar las escalas comparables entre sí.

De este modo, se garantiza que la reconstrucción de la imagen LDCT no solo reduzca el ruido, sino que también preserve la estructura y los bordes esenciales.

Un aspecto crucial fue la implementación del **método de preservación de bordes**, llevado a cabo mediante el operador Sobel (ver Sección 3.2.5).

Dado que el operador Sobel genera una imagen que resalta los bordes de la imagen original, es posible aplicarlo tanto a la imagen de entrada como a la imagen reconstruida. Posteriormente, se puede utilizar una métrica de distancia o error, tal como se propone en [29], donde se calculó el Error Medio Absoluto (MAE) tras aplicar el operador Sobel a las imágenes, obteniendo así la función definida en la ecuación (4.3)

$$MAE_{SOBEL} = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N |G_X(i, j) - G_{Y_w}(i, j)| \tag{4.3}$$

Donde:

- $X(i, j)$ es la matriz de la imagen original.
- $G_X(i, j)$ es la matriz resultante tras aplicar el operador Sobel a la imagen original.
- $Y_w(i, j)$ es la matriz de la imagen reconstruida.
- $G_{Y_w}(i, j)$ es la matriz de la imagen reconstruida después de aplicar el operador Sobel.

La ecuación (4.3) permite evaluar la preservación de bordes en las imágenes procesadas. Al aplicar el filtro Sobel tanto a la imagen original como a la reconstruida y luego calcular el MAE entre ambas, se obtiene una medida de cuán bien se han conservado los bordes. Un valor bajo de este error indica una buena preservación de bordes. Además, esta métrica puede utilizarse no solo como un criterio de evaluación, sino también como una función de costo durante el entrenamiento del modelo.

Por lo que, finalmente se añadió la función (4.3) a la función de costo (4.2), obteniendo:

$$\begin{aligned}
 C(w) = & \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N \left[\alpha \cdot (X(i, j) - Y_w(i, j))^2 + \beta \cdot |G_X(i, j) - G_{Y_w}(i, j)| \right] \\
 & + \gamma \cdot \left\{ 1 - (0.01) \cdot 20 \log_{10} \left(\frac{MAX_{pix}}{\sqrt{MSE}} \right) \right\} \\
 & + \delta \cdot \left[1 - \frac{(2\mu_{X(i,j)}\mu_{Y_w(i,j)} + c_1) (2\sigma_{X(i,j)}\sigma_{Y_w(i,j)} + c_2)}{(\mu_{X(i,j)}^2 + \mu_{Y_w(i,j)}^2 + c_1) (\sigma_{X(i,j)}^2 + \sigma_{Y_w(i,j)}^2 + c_2)} \right]
 \end{aligned} \tag{4.4}$$

4.4. Proceso de Entrenamiento del Autoencoder

4.4.1. Entrenamiento - AE

Con el objetivo de entrenar la red Autoencoder utilizando imágenes desparejadas, se optó por un enfoque basado exclusivamente en Tomografías Computarizadas de Dosis Normal (NDCT). Durante el proceso de entrenamiento, tanto las imágenes de entrada como las imágenes objetivo correspondieron a NDCT, es decir, la red aprendió a reconstruir imágenes NDCT a partir de sí mismas.

El entrenamiento se llevó a cabo principalmente utilizando la función de costo (4.4), dado que esta fue la propuesta diseñada para cumplir con uno de los objetivos específicos planteados. Sin embargo, también se realizaron distintos experimentos variando esta función de costo, los cuales se describirán más adelante.

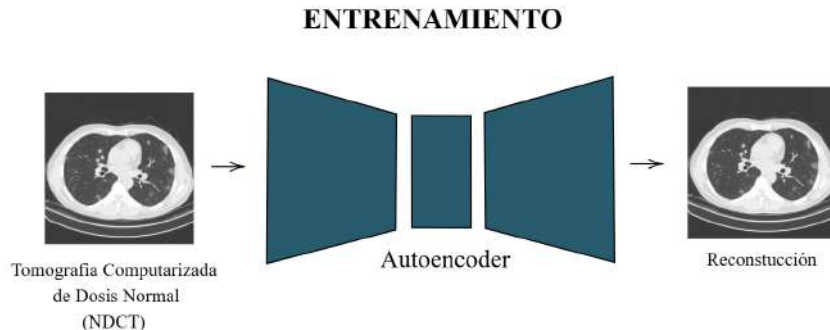


Figura 4.6: Diagrama del entrenamiento del Autoencoder 4.2.1 mediante imágenes NDCT.

Una vez finalizado el entrenamiento, la red se evaluó con un conjunto de prueba compuesto por Tomografías Computarizadas de Baja Dosis (LDCT). Esto permitió cuantificar la capacidad del modelo para reducir el ruido y preservar los bordes al reconstruir las imágenes LDCT.

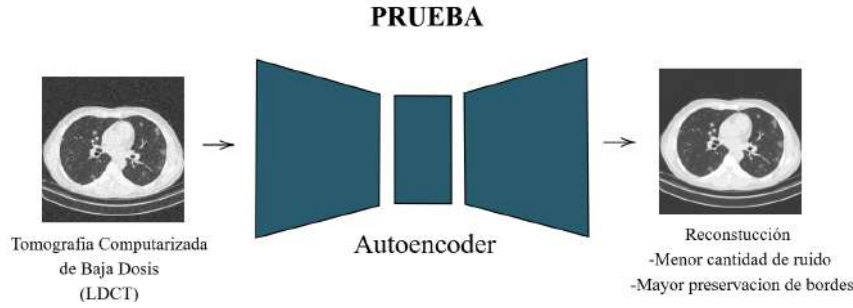


Figura 4.7: Diagrama del Autoencoder 4.2.1 entrenado, sometido al conjunto de prueba de imágenes LDCT.

4.4.2. Parámetros de Entrenamiento - AE

Para el entrenamiento del Autoencoder se dividió la base de datos 4.1.2.1 en tres subconjuntos:

1. Entrenamiento: 7000 imágenes de Tomografías Computarizadas de Dosis Normal (NDCT).
2. Validación: 2000 imágenes de Tomografías Computarizadas de Dosis Normal (NDCT).
3. Prueba: 1000 imágenes de Tomografías Computarizadas de Baja Dosis (LDCT).

Se llevaron a cabo diferentes experimentos de entrenamiento de la red Autoencoder, variando la función de costo utilizada en cada caso. Se probaron las siguientes configuraciones:

1. AE_{MSE} : Función de costo MSE (Error Cuadrático Medio).
2. AE_{PSNR} : Función de costo compuesta por una suma ponderada del MSE y PSNR (Relación Señal-Ruido Máxima).
3. AE_{SSIM} : Función de costo compuesta por una suma ponderada del MSE, PSNR y SSIM (Índice de Similitud Estructural).
4. AE_{SOBEL} : Se utilizó la función de costo propuesta en la (4.4)

En nuestro trabajo, todos los experimentos se llevaron a cabo en el entorno de TensorFlow en una computadora portátil con una CPU Intel Core i7-13620H, 16 GB de RAM y una tarjeta gráfica NVIDIA RTX 4060. El entrenamiento se realizó durante 200 épocas utilizando un tamaño de Batch de 4.

En particular, aun que si bien, se utilizo en todos los casos el algoritmo de optimización Adam 3.41, los parámetros no fueron los mismos.

- AE_{MSE} : Adam($\eta = 0.001, \beta_1=0.9, \beta_2=0.999, \epsilon=1e-07$)
- AE_{PSNR} : Adam($\eta = 0.0005, \beta_1=0.001, \beta_2=0.999, \epsilon=1e-07$)
- AE_{SSIM} : Adam($\eta = 0.0005, \beta_1=0.001, \beta_2=0.999, \epsilon=1e-07$)
- AE_{SOBEL} : Adam($\eta = 0.0005, \beta_1=0.001, \beta_2=0.999, \epsilon=1e-07$)

4.5. Alineación del Espacio Latente

Para completar el último objetivo específico, se llevó a cabo el siguiente procedimiento:

Partiendo de la premisa de que el espacio latente es una representación de menor dimensionalidad que captura las características más relevantes de los datos de entrada, eliminando detalles menos significativos, se aprovechó esta propiedad de los autoencoders. Esta capacidad de los autoencoders es la razón por la cual resultan altamente eficaces en la eliminación de ruido, ya que retienen únicamente la información más importante de los datos.

Con base en lo anterior, se realizó la alineación o acercamiento de la región del espacio latente de las imágenes LDCT al espacio latente de las imágenes NDCT. Dicho espacio, correspondiente a las imágenes NDCT, contiene una menor cantidad de ruido, por lo que esta alineación en un principio mejoraría la reconstrucción de las imágenes LDCT.

El propósito de este procedimiento fue lograr que la extracción de características para las LDCT, si bien no fuera exactamente la misma, fuera similar al de las NDCT y se pudiera realizar de mejor manera la reconstrucción, por ello el uso de las palabras alineación o acercamiento a la región del espacio latente.

Esta alineación o acercamiento puede lograrse mediante la extracción de puntos latentes aleatorios que permiten una reconstrucción eficiente de tomografías computarizadas de dosis normal (NDCT). Posteriormente, a través de un proceso de entrenamiento, se busca que los puntos latentes correspondientes a las tomografías computarizadas de baja dosis (LDCT) sean similares a los de las NDCT.

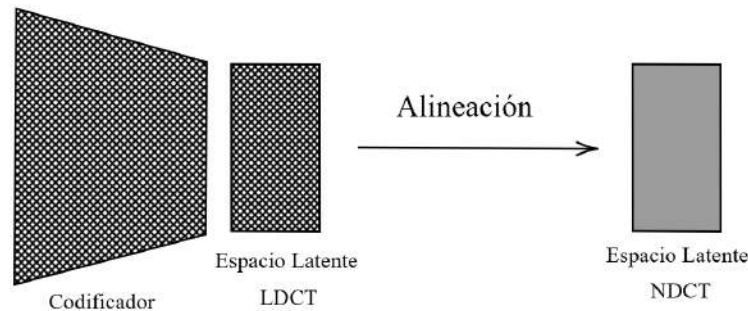


Figura 4.8: Diagrama de la Alineación o Acercamiento a la región del espacio latente de las imágenes codificadas con una menor cantidad de ruido.

4.5.1. Entrenamiento - Autoencoder con Modificación en el Codificador (AEMC)

Una vez completado el entrenamiento descrito en la sección 4.4, se extrajeron 100 puntos latentes aleatorios.

Posteriormente, utilizando el codificador del Autoencoder (ver sección 4.2.1), se generaron los puntos latentes correspondientes a las imágenes LDCT, y se llevó a cabo el entrenamiento para su alineación con los puntos latentes de las imágenes NDCT.

Este proceso consiste en alimentar el codificador con tomografías computarizadas de baja dosis (LDCT) y utilizar como objetivo los puntos latentes previamente extraídos de las tomografías computarizadas de dosis normal (NDCT).

El objetivo del entrenamiento es ajustar los pesos del codificador para que los puntos latentes generados a partir de imágenes LDCT sean cada vez más similares a los de las NDCT, logrando así lo que nombramos como una alineación o acercamiento. Esto logra que ahora en el Autoencoder con Modificación en el Codificador (AEMC) pueda tener como entrada Tomografías Computarizadas

de Baja Dosis (LDCT) y que al realizar la codificación los valores del espacio latente obtenidos sean similares a los de una imagen NDCT, esto es en esencia, que la extracción de valores omitirá características irrelevantes de las imágenes LDCT logrando conservar la información necesaria para realizar una óptima decodificación.

4.5.2. Parámetros de Entrenamiento - AEMC

Para el entrenamiento del codificador con el objetivo de realizar la alineación o acercamiento del espacio latente, se utilizaron los siguientes parámetros:

Entrada:

- Entrenamiento: 50 imágenes de Tomografías Computarizadas de Baja Dosis (LDCT).
- Validación: 50 imágenes de Tomografías Computarizadas de Baja Dosis (LDCT).

Objetivo:

- Entrenamiento: 50 puntos latentes correspondientes a 50 imágenes de Tomografías Computarizadas de Baja Dosis (LDCT).
- Validación: 50 puntos latentes correspondientes a 50 imágenes de Tomografías Computarizadas de Baja Dosis (LDCT).

El entrenamiento se llevó a cabo durante 15 épocas, con un batch size de 32, utilizando el optimizador Adam con los siguientes hiperparámetros:

$$\eta = 0.001, \quad \beta_1 = 0.9, \quad \beta_2 = 0.999, \quad \epsilon = 1e-07$$

La función de costo empleada fue (4.4).

4.6. Clasificación para cada Reconstrucción

Como se mencionó en capítulos anteriores, un aspecto fundamental de esta investigación es determinar si la mejora en la reconstrucción de imágenes LDCT en términos de reducción de ruido y preservación de bordes se traduce en un aumento en la precisión del diagnóstico, es decir, en la capacidad de clasificar correctamente si una imagen corresponde a un paciente sano o enfermo. Por ello, se decidió hacer uso de un clasificador con el fin de evaluar si existe una relación directa entre la calidad de la reconstrucción y la precisión en la clasificación.

4.6.1. Entrenamiento - Clasificador

Se entrenó el clasificador (ver sección 4.2.2) utilizando el conjunto de datos descrito en la sección 4.1.2.1, en el cual se asignaron etiquetas **1** a los pacientes enfermos y **0** a los pacientes sanos, contando así con las imágenes y sus respectivas etiquetas.

El conjunto de datos se dividió de la siguiente manera:

1. Entrenamiento: 7,000 imágenes NDCT con sus respectivas etiquetas.
2. Validación: 2,000 imágenes NDCT con sus respectivas etiquetas.
3. Prueba: 1,000 imágenes LDCT con sus respectivas etiquetas.

Durante el entrenamiento, el clasificador recibió como entrada imágenes de Tomografías Computarizadas de Dosis Normal (NDCT) y tuvo como objetivo las etiquetas correspondientes a cada imagen. Los subconjuntos de entrenamiento y validación se utilizaron durante el entrenamiento del clasificador.

El modelo se entrenó durante 5 épocas con un *batch size* de 32, utilizando el algoritmo de optimización Adam con los siguientes parámetros: $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e - 07$. La función de costo empleada fue Binary Cross Entropy, definida como:

$$C(w) = -\frac{1}{m} \sum_{i=1}^m [y_i \ln(P_i(w)) + (1 - y_i) \ln(1 - P_i(w))] \quad (4.5)$$

donde y_i representa la etiqueta real de la muestra, $P_i(w)$ es la probabilidad predicha de que la muestra pertenezca a su respectiva clase, y m es el número total de muestras utilizadas en el entrenamiento.

Antes de realizar diferentes análisis con el clasificador, se verificó su capacidad para distinguir correctamente entre personas sanas y enfermas utilizando un conjunto de tomografías de dosis normal (NDCT). En esta prueba inicial, el modelo alcanzó una precisión del **100 %** en la clasificación de imágenes NDCT.

Posteriormente, las 1,000 imágenes LDCT del conjunto de prueba fueron procesadas mediante cada una de las diferentes reconstrucciones propuestas en esta investigación. Esto permitió analizar si una mejor reconstrucción, en términos de reducción de ruido y preservación de bordes en las imágenes LDCT, implica una mejora en la precisión del diagnóstico del clasificador.

Finalmente, se generó el siguiente conjunto de prueba:

Prueba:

- LDCT: 1,000 imágenes y 1,000 etiquetas.
- AE_{MSE} : 1,000 imágenes y 1,000 etiquetas.
- AE_{PSNR} : 1,000 imágenes y 1,000 etiquetas.
- AE_{SSIM} : 1,000 imágenes y 1,000 etiquetas.
- AE_{SOBEL} : 1,000 imágenes y 1,000 etiquetas.
- $AEMCI$: 1,000 imágenes y 1,000 etiquetas.
- $AEMC$: 1,000 imágenes y 1,000 etiquetas.
- $SSGAN$: 1,000 imágenes y 1,000 etiquetas.

Cabe mencionar que se dividió cada uno de los conjuntos de prueba en 5 partes, cada uno con 200 imágenes y sus respectivas etiquetas para realizar 5 mediciones en la precisión por cada modelo y calcular un error estadístico en la precisión.(Fig.4.9).

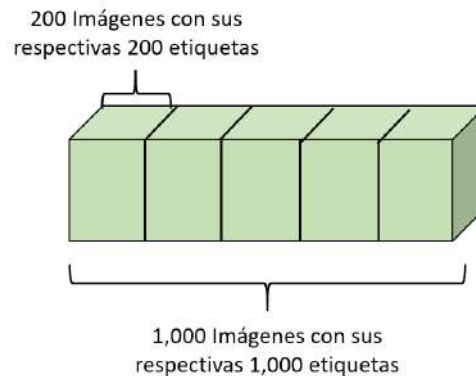


Figura 4.9: División de cada conjunto de prueba.

Nota:

Es importante mencionar que en cada uno de los experimentos aplicados en esta investigación, como se menciona en cada sección de entrenamiento, no se realizó el entrenamiento mediante imágenes emparejadas. Esto es, no se requiere la Tomografía Computarizada de Dosis Normal (NDCT) y su contraparte la misma Tomografía Computarizada, del mismo paciente pero de Baja Dosis (LDCT). Dado que obtener una base de datos real de imágenes emparejadas sería extremadamente raro y éticamente problemático. Esto se debe a que implicaría irradiar al paciente dos veces:

- *Una primera vez con una dosis normal para obtener la imagen NDCT.*
- *Una segunda vez con una dosis baja para obtener la imagen LDCT.*

El uso de una base de datos así para el entrenamiento no solo sería contradictorio con la motivación de reducir la exposición a la radiación, sino que también violaría principios éticos al aumentar innecesariamente la dosis de radiación para los pacientes. Por ello, cada uno de nuestros enfoques se basa en el entrenamiento con imágenes desemparejadas. En su lugar, se usa:

- *Imágenes NDCT y LDCT de diferentes pacientes.*
- *Asegurando siempre que las imágenes sean de la misma región anatómica y correspondan al mismo tipo de corte tomográfico.*

Capítulo 5

Análisis de Resultados

En este capítulo se presentan diversos análisis de los resultados obtenidos en esta investigación, abarcando distintos aspectos del entrenamiento y desempeño de las redes desarrolladas. Se inicia con el estudio de las funciones de costo a lo largo del entrenamiento en los diferentes experimentos, a través de sus respectivas gráficas, lo que permite evaluar la convergencia y estabilidad del modelo.

Posteriormente, se analiza la calidad de las reconstrucciones obtenidas para imágenes NDCT, lo que proporciona una referencia visual y cuantitativa sobre el desempeño de los métodos propuestos. También se examinan los puntos latentes en el proceso de alineación del espacio latente, con el objetivo de entender mejor cómo se estructuran y qué sucede en esta alineación.

Además, se lleva a cabo un análisis estadístico del ruido presente en las reconstrucciones generadas, permitiendo comparar su efectividad en la reducción del ruido. Para complementar este estudio, se realiza una evaluación mediante métricas de calidad de imagen, con el fin de determinar en qué medida los objetivos planteados en esta tesis fueron alcanzados y si la hipótesis inicial es validada o refutada.

Finalmente, se comparan los resultados obtenidos con aquellos reportados en el estado del arte, utilizando la métrica de calidad de imagen PSNR como criterio de referencia. Esto permitirá cuantificar si el mejor método de reconstrucción desarrollado en esta investigación es competitivo frente a los enfoques previamente propuestos en la literatura.

Antes de llevar a cabo un análisis detallado de los diferentes métodos de reconstrucción de imágenes LDCT, es fundamental destacar lo siguiente:

- Cada análisis de resultados se realizó sobre el conjunto de prueba, que es de 1,000 imágenes de Tomografías Computarizadas de Baja Dosis (LDCT), las cuales, es importante recordar que, no fueron utilizadas durante el entrenamiento de los diferentes experimentos.
- Las imágenes del conjunto de prueba fueron reconstruidas utilizando los diferentes métodos propuestos, con el objetivo de determinar cuál de ellos logró la mejor reconstrucción. Dado que se realizaron siete experimentos (AE_{MSE} , AE_{PSNR} , AE_{SSIM} , AE_{SOBEL} , AEMCI, AEMC y SSGAN), el análisis de resultados se centrará en la evaluación comparativa de estos modelos.

5.1. Funciones de Costo

El análisis de las funciones de costo, aunque no puede predecir con certeza la calidad de la reconstrucción para el conjunto de prueba con imágenes LDCT, es un buen indicador para evaluar la eficiencia del entrenamiento de la red y verificar si se está llevando a cabo de manera adecuada. Por ello, es útil realizar este análisis antes de profundizar en la evaluación de las reconstrucciones obtenidas.

5.1.1. Gráfica de la Función de Costo - Autoencoder

Uno de los primeros análisis de resultados se realiza mediante la visualización de la gráfica de la función de costo durante el entrenamiento. Este análisis inicial permite evaluar el desempeño de la red, recordando que el entrenamiento de un Autoencoder (ver Sección 3.3.7) se basa en minimizar una función de costo que generalmente mide el error entre la imagen reconstruida y la original. Por ello, es común que estas redes utilicen como función de costo el Error Cuadrático Medio (MSE), descrito en la Ecuación (4.1). Para la función de costo propuesta en la Ecuación (4.4), también se busca minimizar su valor, a pesar de que se trata de una suma ponderada de diferentes funciones (ver implementación en la Sección 4.3.1). Esto implica que el objetivo principal es reducir al mínimo posible la función de costo, logrando así una reconstrucción más efectiva. La interpretación de la función de costo frente al número de épocas de entrenamiento, que en este caso fue de 200, permite evaluar si el modelo se está ajustando adecuadamente y si los componentes individuales de la función de costo ponderada también se están minimizando correctamente. A continuación, se muestra tanto la evolución de la función de costo total (4.4) durante el entrenamiento, como el monitoreo del costo de cada uno de los términos ponderados, tanto para los datos de entrenamiento como para los de validación.

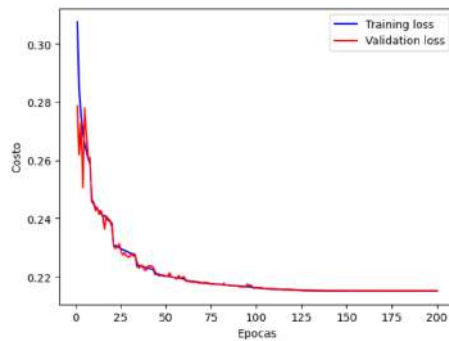


Figura 5.1: Función de Costo (4.4) vs Épocas

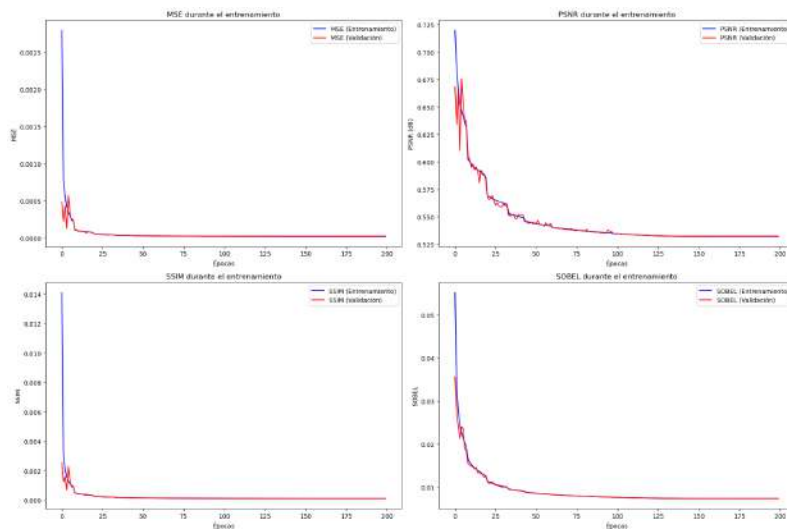


Figura 5.2: Monitoreo de cada función por separado de la función de costo (4.4) vs Épocas

En general 5.1, la función de costo disminuye rápidamente en las primeras épocas y luego

se estabiliza alrededor de la época 150. El Autoencoder se entrenó durante 200 épocas, ya que la función de costo tanto para el conjunto de entrenamiento como para el de validación se ha estabilizado.

Individualmente, gracias a la figura 5.2, se logra visualizar que para cada función se alcanza una convergencia o estabilidad. Siendo más específicos, las funciones de costo MSE y SSIM implementadas logran estabilizarse alrededor de la época 50. Sin embargo, el PSNR implementado y el MAE_{SOBEL} , que recordemos es nuestra implementación como método de preservación de bordes durante el entrenamiento, logran estabilizarse en un mayor número de épocas, aproximadamente alrededor de la época 150.

El hecho de que las funciones de costo para los datos de entrenamiento y validación sea muy cercana sugiere que el modelo no está sobreajustándose (*overfitting*). La ausencia de una brecha significativa entre ambas curvas indica una buena generalización del modelo.

Cabe mencionar que, aunque se lograron buenos resultados durante el entrenamiento, esto no implica que sean los mejores. Si se buscara mejorar el desempeño, se podrían explorar técnicas como el ajuste de hiperparámetros.

5.1.2. Gráfica de la Función de Costo - AEMC

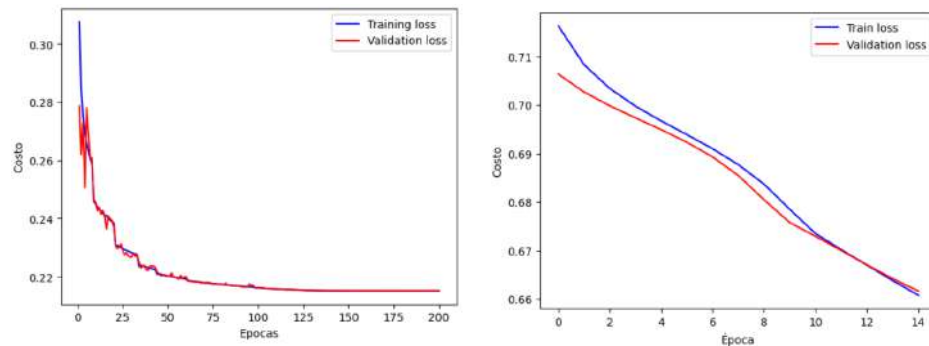


Figura 5.3: Función de costo vs Épocas para el experimento AEMC: (Izquierda) durante el entrenamiento del autoencoder, y (Derecha) durante la alineación del espacio latente.

Al realizar la alineación del espacio latente (ver sección 4.5), el entrenamiento del autoencoder es el mismo, por lo que la gráfica de la función de costo durante este entrenamiento será la misma que la vista en la figura 5.1.

Sin embargo, al llevar a cabo la alineación de los puntos latentes, se entrenó el codificador para realizar dicha alineación. En este contexto, monitorear la función de costo durante este entrenamiento es de suma importancia, ya que dicho proceso es altamente volátil. La gráfica de la función de costo es un indicador clave que señala el punto óptimo en el que se debe detener el entrenamiento. Este punto se alcanza cuando las funciones de costo para los datos de entrenamiento y validación son muy cercanas. Lo que nos lleva a la figura 5.3

Si el entrenamiento se prolonga más allá de este punto, la función de costo para los datos de entrenamiento comenzará a disminuir más rápidamente que la función de costo para los datos de validación, lo que resultará en una reconstrucción ineficiente para el conjunto de prueba.

5.1.3. Gráfica de la Función de Costo - AEMCI

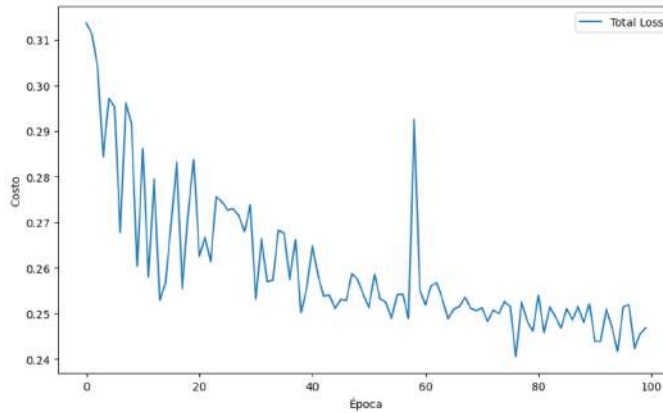


Figura 5.4: Función de costo vs Épocas

La Figura 5.4 muestra la evolución de la función de costo a lo largo de 100 épocas. Se puede observar una tendencia general decreciente, lo cual indica que el modelo fue capaz de ajustar sus parámetros progresivamente, reduciendo el error en la reconstrucción de las imágenes. En las primeras épocas, la gráfica presenta una mayor variabilidad, probablemente asociada a la fase inicial del entrenamiento. No obstante el costo se estabiliza y las oscilaciones disminuyen, lo que sugiere una convergencia progresiva del modelo. A partir de la época 30, la disminución se vuelve más suave, con pequeñas fluctuaciones, manteniéndose en un rango cercano a 0.24 en las últimas épocas. Aunque se presenta un pico aislado alrededor de la época 60, no representa una pérdida general de aprendizaje.

5.1.4. Gráfica de la Función de Costo - Clasificador

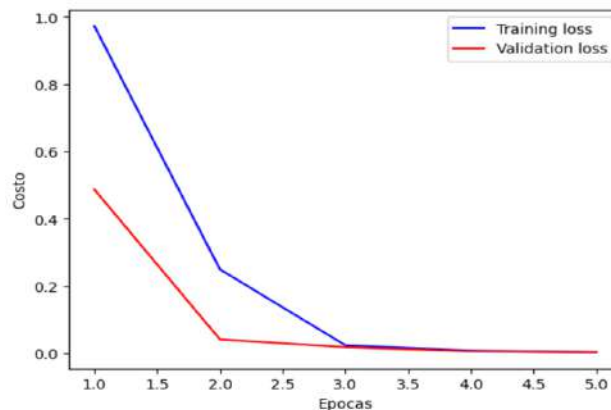


Figura 5.5: Función de Costo vs Épocas del clasificador

La figura 5.5 muestra un aprendizaje rápido en las primeras épocas, logrando una convergencia temprana. Desde la tercera época, la pérdida de entrenamiento y validación casi llegan a cero, lo que sugiere que el modelo ha aprendido a clasificar los datos con gran precisión en muy pocas épocas.

Debido a esta rápida convergencia, el entrenamiento del clasificador no requirió más de 5 épocas. Además, no se observa sobreajuste, ya que la función de costo para los datos de entrenamiento y validación son extremadamente cercanas.

Ambas pérdidas disminuyen de forma similar y se estabilizan en valores muy bajos, lo que indica que el modelo generaliza bien y es capaz de realizar una clasificación efectiva sin comprometer su capacidad predictiva.

5.1.5. Gráfica de la Función de Costo - SSGAN

La figura 5.6 nos muestra el comportamiento de 3 gráficos respectivamente del generador, discriminador (clasificador rea/fake) y discriminador (clasificador sano/enfermo).

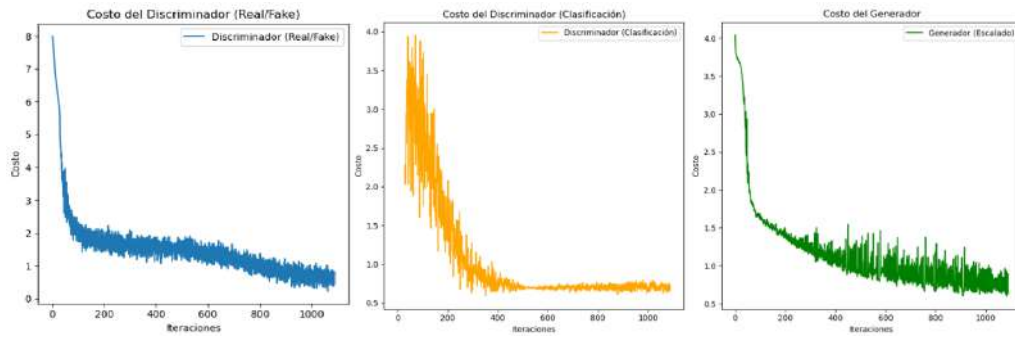


Figura 5.6: Función de Costo vs Épocas del modelo SSGAN

En general el discriminador (clasificador rea/fake) tienen como función de costo:

$$C_{D_{unsupervised}}(w) = -\frac{1}{m} \sum_{k=1}^m [\ln(D(x_k)) + \ln(1 - D(G(R_k)))] \quad (5.1)$$

Donde $D(x_k)$ es la salida del discriminador cuando se introduce una imagen real, por lo que x_k representa una muestra correspondiente a una imagen $X_k(i, j)$. Dado que la última función de activación del modelo es sigmoide, los valores resultantes están confinados en el intervalo $[0, 1]$. Al calcular $\ln(D(x_k))$, se obtiene un valor en el rango $(-\infty, 0]$, razón por la cual toda la expresión se multiplica por un signo negativo, asegurando que los valores sean positivos.

De manera análoga, la expresión $(1 - D(G(R_k)))$ también se encuentra en el intervalo $[0, 1]$, y tras aplicar el logaritmo y las operaciones correspondientes, los valores resultantes estarán en $[0, \infty)$. Esto implica que cuando el generador logra engañar al discriminador generando una muestra similar a una imagen real, el valor resultante de la función de costo será menor en comparación con los casos donde el discriminador distingue correctamente entre imágenes reales y generadas, permitiendo así la minimización de la función de costo.

Generalmente, en los textos [18, 20, 58] se menciona que esta función debe maximizarse. Sin embargo, vemos que para esta implementación no es así.

Ahora, la función de costo supervisado es:

$$C_{D_{supervised}}(w) = -\frac{1}{m} \sum_{k=1}^m [y_k \ln(D(x_k)) + (1 - y_k) \ln(1 - D(x_k))] \quad (5.2)$$

Esta es la conocida entropía cruzada binaria (Binary Cross Entropy, BCE), la cual también busca minimizarse de tal manera que el error sea pequeño cuando el valor obtenido corresponde a la etiqueta correcta. Como se puede observar en la figura 5.6, el entrenamiento logra este objetivo.

Finalmente, la función de costo para el generador es:

$$C_G(w) = -\frac{1}{m} \sum_{k=1}^m \ln(1 - D(G(R_k))) \tag{5.3}$$

Esta función indica que cuando el error es pequeño, el valor numérico es bajo, mientras que cuando el error es grande, el valor numérico es alto. Esto implica que el objetivo es minimizar esta expresión, lo que se ve reflejado en su respectiva gráfica en la figura 5.6, mostrando que el modelo cumple con su propósito.

5.2. Métricas de Calidad en Reconstrucción de Imágenes NDCT

Un aspecto fundamental radica en los primeros cuatro experimentos: AE_{MSE} , AE_{PSNR} , AE_{SSIM} y AE_{SOBEL} . Estos experimentos proporcionan la base para la alineación del espacio latente.

El modelo AE_{SOBEL} destaca por ser el que mejor realiza la tarea de reconstrucción de imágenes LDCT. Sin embargo, su desempeño no debe limitarse únicamente a esta clase de imágenes, sino que también debe ser el modelo más eficiente en la reconstrucción de imágenes NDCT.

Esta observación es crucial, ya que indicaría que el modelo AE_{SOBEL} es el mejor capturando las características más relevantes en su espacio latente para la tarea de reconstrucción. Por lo tanto, es esencial validar si efectivamente este modelo mantiene su superioridad al reconstruir imágenes NDCT, lo cual se detalla en la siguiente tabla.

Reconstrucciones	PSNR ± Error	SSIM ± Error	MSE ± Error	MAE _{Sobel} ± Error
AE_{MSE}	44.3247 ± 1.2988	0.9877 ± 1.46e-5	3.82e-5 ± 1.02e-10	0.0344 ± 1.01e-6
AE_{PSNR}	46.3877 ± 1.3803	0.9921 ± 5.74e-6	2.38e-5 ± 4.23e-11	0.0344 ± 1.02e-6
AE_{SSIM}	45.4921 ± 1.3446	0.9905 ± 8.42e-6	2.93e-5 ± 6.2e-11	0.0344 ± 1.00e-6
AE_{SOBEL}	49.5790 ± 1.3039	0.9964 ± 1.12e-6	1.14e-5 ± 9.12e-12	0.0342 ± 1.00e-6

Tabla 5.1: Resultados de métricas de calidad de imagen para la reconstrucción de Tomografías Computarizadas de Dosis Normal (NDCT).

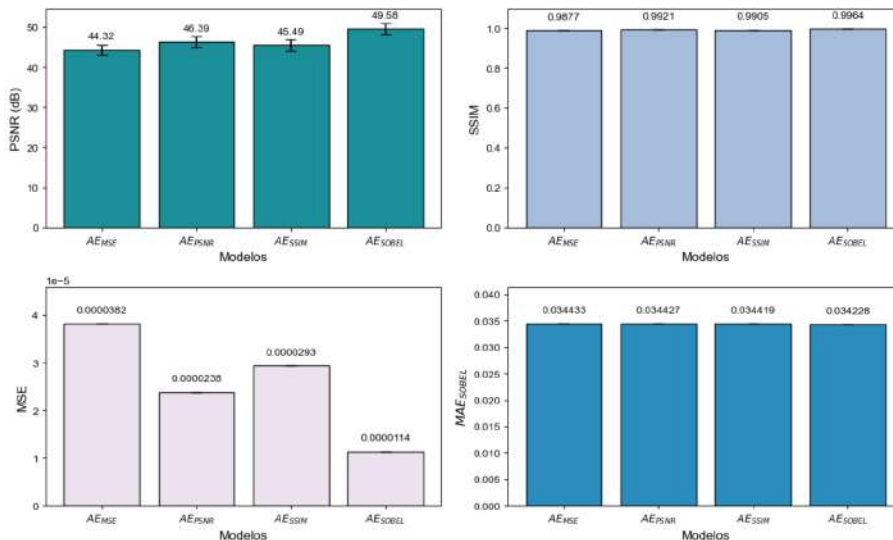


Figura 5.7: Gráficos correspondientes al Cuadro 5.1

Cuantitativamente AE_{SOBEL} es ,en todo aspecto, superior para realizar esta tarea de reconstrucción de imágenes NDCT, indicándonos que es el mejor capturando las características más relevantes en su espacio latente.

5.3. Alineación del Espacio Latente

Lo que se buscó lograr en la alineación del espacio latente fue que los valores latentes correspondientes a las imágenes LDCT se acercaran o fueran similares a los valores latentes de las imágenes NDCT. Es decir, después de realizar la alineación, los valores latentes de ambas modalidades deberían seguir una misma tendencia.

Durante el entrenamiento en el que se realizó la alineación del espacio latente (ver Sección 4.5), se extrajeron 100 puntos latentes correspondientes a imágenes NDCT. Posteriormente, al codificar imágenes LDCT, se obtuvieron sus respectivos puntos latentes. Mediante el proceso de alineación descrito en dicha sección, se ajustaron los valores latentes de las imágenes LDCT para acercarlos a la distribución de los valores latentes de las imágenes NDCT.

Para analizar si esta alineación fue efectiva y si en efecto los puntos latentes correspondientes a las imágenes LDCT lograron seguir la misma tendencia que los correspondientes a las imágenes NDCT, se visualizaron los histogramas de los puntos latentes antes de hacer la alineación (Esp-Lat LDCT), después de realizar su alineación (Esp-Lat Alineado) y compáralos con histogramas de puntos latentes correspondientes a imagenes NDCT (Esp-Lat NDCT).

Aquí es importante mencionar que, un punto latente en nuestro Autoencoder (ver Sección 4.2.1) está representado por un mapa de características de menor dimensionalidad. Dado el diseño implementado, estos puntos latentes tienen una dimensión de $128 \times 128 \times 1$, lo que se expresa matemáticamente como:

$$\text{Un Punto Latente} = (L_{128 \times 128})$$

De este modo, al codificar una imagen NDCT, se genera un punto latente de dichas dimensiones y de la misma manera con las imágenes LDCT. Entonces cuando nos referimos a los valores de un puto latente nos referimos a los valores $l(i, j)$ de la matriz $L_{128 \times 128}$, por lo que al realizar la alineación los valores de cada punto latente debería de seguir la misma tendencia que los valores correspondientes a los puntos de las imágenes NDCT.

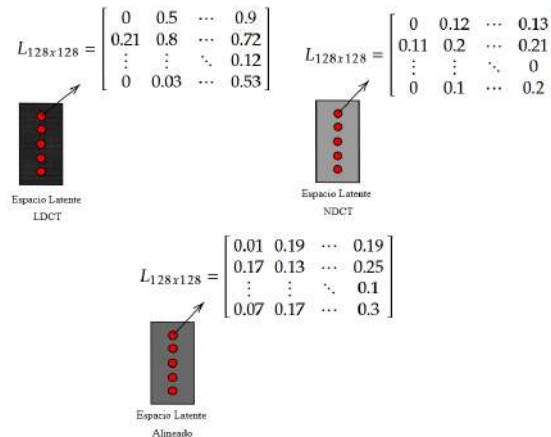


Figura 5.8: Diagrama de espacio latente y puntos latentes correspondientes a el Espacio Latente LDCT, Espacio Latente NDCT y Espacio Latente Alineado

Análisis de Resultados

5.3 Alineación del Espacio Latente

Por lo que se visualizo el histograma de 200 valores correspondientes a 10 puntos latentes aleatorios. Obteniendo la figura 5.12.

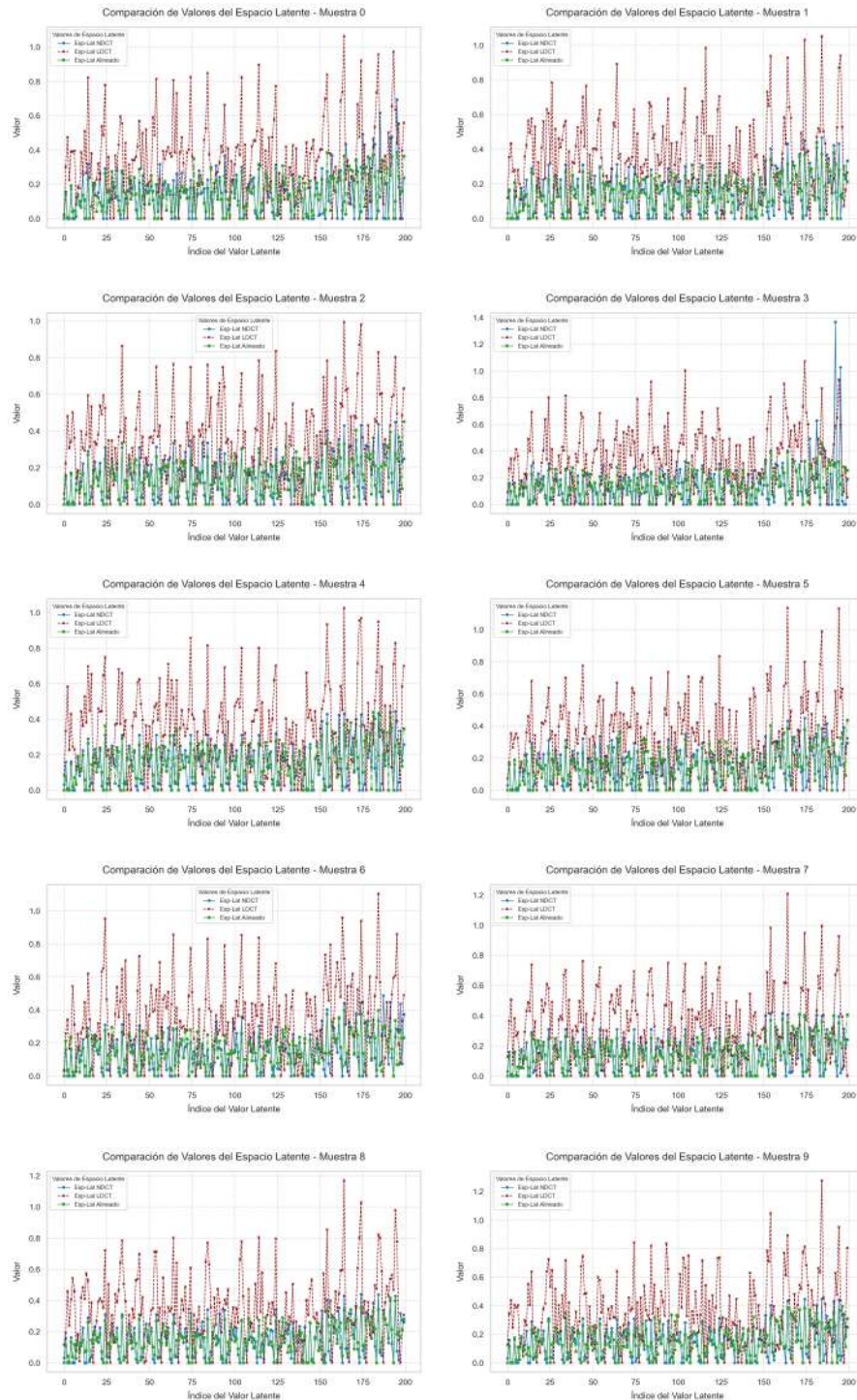


Figura 5.9: Histogramas de los valores del espacio latente de imágenes NDCT, LDCT y después de realizar el acercamiento o alineación

En la figura 5.12 se muestra la tendencia de 200 valores correspondientes a 10 puntos latentes aleatorios de imágenes LDCT (línea roja), la tendencia de los valores en la representación latente de imágenes NDCT (línea azul) y, finalmente, la tendencia de los valores en la representación latente de las imágenes después de la alineación (línea verde).

Se observa que los valores latentes de LDCT (línea roja) presentan mayor dispersión y amplitud en comparación con NDCT (línea azul). Esto sugiere que el ruido en las imágenes LDCT introduce una mayor variabilidad en el espacio latente. En contraste, los valores de NDCT parecen mantenerse en un rango más estable y compacto, lo que indica que su representación latente es más homogénea.

Después del alineamiento, los valores latentes ajustados (línea verde) muestran una distribución más similar a la de NDCT. Aquí es importante mencionar que no se busca se los valores sean los mismos ya que si ese fuera el caso estaríamos transformando la imagen en otra totalmente diferente ya que si bien la línea azul corresponde a los valores de un punto latente correspondiente a una imagen NDCT no corresponde a la versión limpia de la imagen LDCT analizada y esto es porque el entrenamiento no se realizó mediante imágenes emparejadas.

Se realizó este análisis para 10 puntos latentes aleatorios con el objetivo de evaluar si el comportamiento observado era generalizable a distintas muestras. A partir de los resultados obtenidos, podemos concluir que la alineación del espacio latente ha sido efectiva, logrando reducir la discrepancia entre las representaciones latentes de LDCT y NDCT, lo que ha permitido que los valores latentes de imágenes LDCT se acerquen a los de las imágenes NDCT.

Este resultado sugiere que el modelo fue capaz de aprender una representación más fiel de las imágenes NDCT, lo que podría beneficiar la posterior reconstrucción de imágenes LDCT al preservar mejor la estructura y características relevantes de las imágenes originales.

5.4. Análisis de Ruido

Un análisis importante se puede realizar mediante el estudio del ruido. Para ello, es fundamental partir de lo ya conocido:

De acuerdo con la sección 4.1.2.2, sabemos que se añadió ruido gaussiano con $\mu = 0$ y $\sigma = 0.06$ a las imágenes NDCT. Esto implica que, al calcular la diferencia entre la matriz $NDCT(i, j)$, que representa una Tomografía Computarizada de Dosis Normal (NDCT), y la matriz $LDCT(i, j)$, que corresponde a su versión distorsionada con ruido gaussiano (la Tomografía Computarizada de Baja Dosis LDCT), se debería obtener una nueva matriz $N(i, j)$, que represente exclusivamente el ruido añadido. En otras palabras, esta matriz debería mostrar una imagen únicamente de ruido.

Si se calcula la media μ y la desviación estándar σ de la matriz $N(i, j)$, los valores resultantes deberían coincidir con los parámetros del ruido añadido inicialmente ($\mu = 0$ y $\sigma = 0.06$).

Matemáticamente, esto se expresa de la siguiente manera:

$$NDCT(i, j) - LDCT(i, j) = N(i, j) \sim N(\mu = 0, \sigma = 0.06) \quad (5.4)$$

Una representación visual de este cálculo se muestra en la Figura 5.10, donde se observa la matriz $N(i, j)$ resultante de la resta píxel a píxel entre las imágenes $NDCT(i, j)$ y $LDCT(i, j)$.

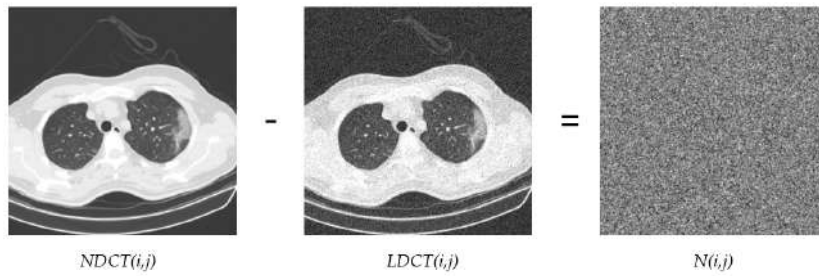


Figura 5.10: Visualización de la matriz de ruido $N(i, j)$ obtenida al restar la imagen $LDCT(i, j)$ de la imagen $NDCT(i, j)$.

Idealmente, el histograma de $N(i, j)$ debería seguir una distribución gaussiana con los parámetros establecidos ($\mu = 0, \sigma = 0.06$). La Figura 5.11 muestra este histograma, confirmando que el ruido añadido cumple con la distribución esperada, por medio de la superposición de una curva gaussiana con media $\mu = 0$ y desviación estándar $\sigma = 0.06$.

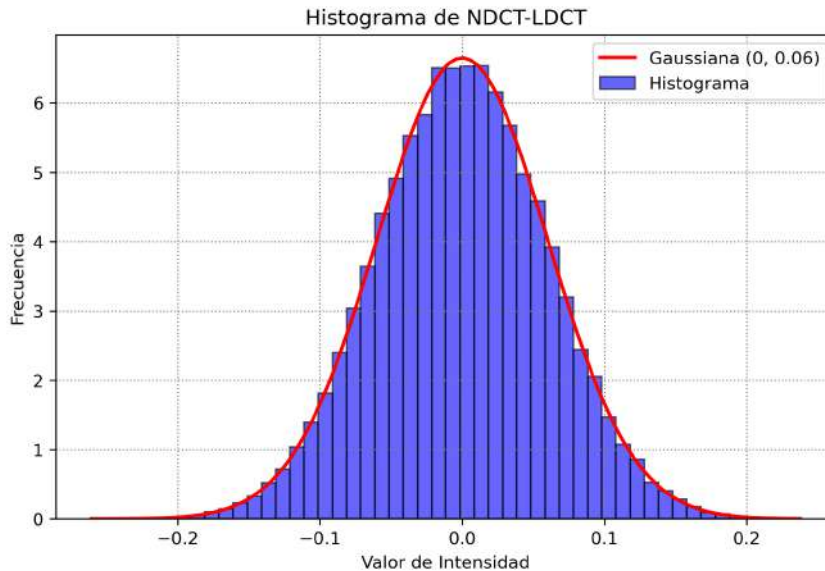


Figura 5.11: Histograma de la matriz de ruido $N(i, j)$ superpuesta a una curva gaussiana con $\mu = 0$ y $\sigma = 0.06$.

Dado que este es el comportamiento esperado del ruido añadido, los modelos de reconstrucción deberían tender a aproximar esta distribución. Para evaluar este aspecto, se realizó un análisis de la media μ y la desviación estándar σ de la diferencia entre cada imagen $LDCT(i, j)$ y su respectiva reconstrucción en cada experimento. Los resultados obtenidos se presentan en la Tabla 5.2.

A continuación, se muestran algunos histogramas correspondientes a la diferencia entre una imagen $LDCT(i, j)$ y sus distintas reconstrucciones obtenidas en cada experimento, junto con la superposición del ajuste de su curva gaussiana.

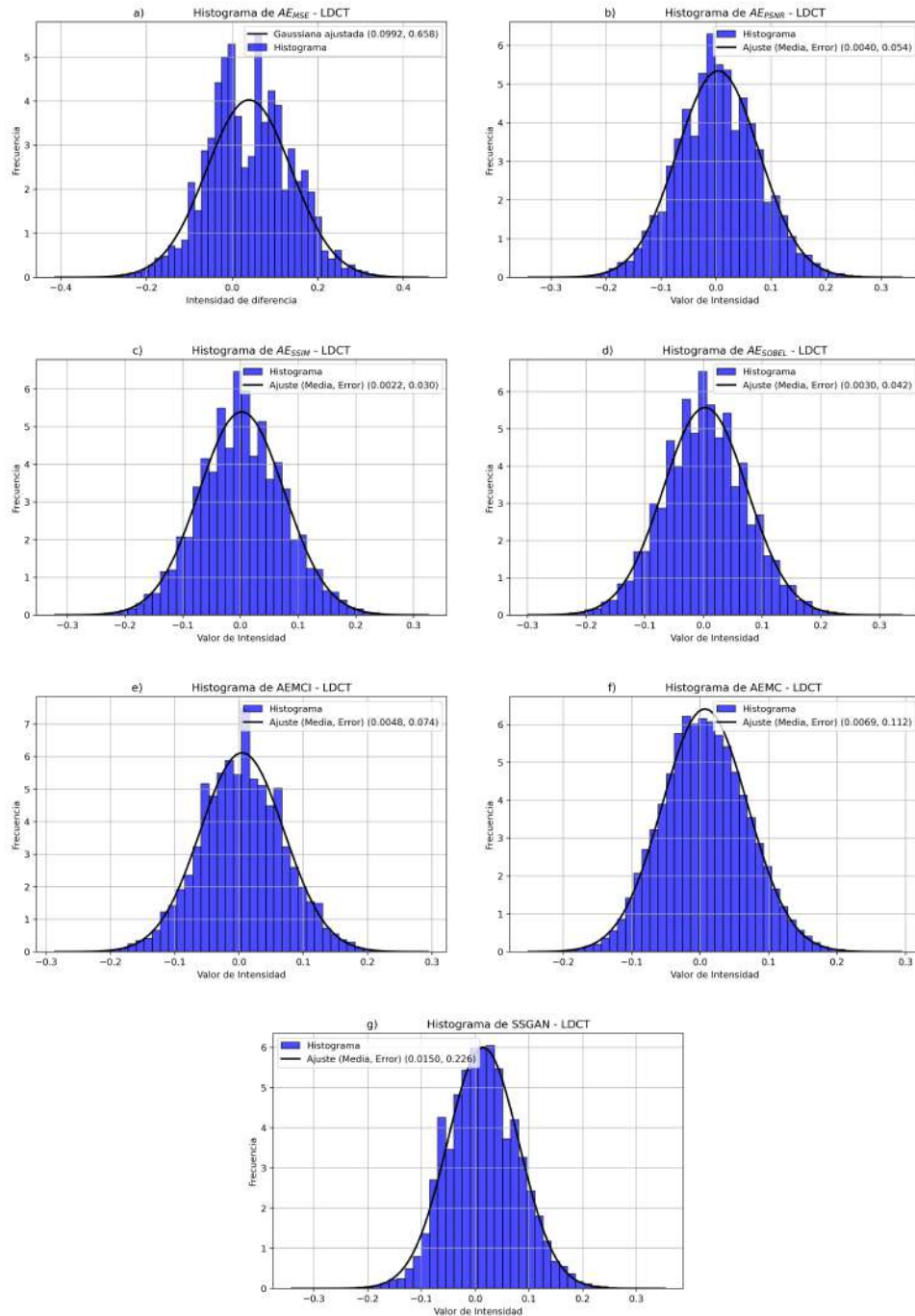


Figura 5.12: Histogramas de la diferencia entre una imagen $LDCT(i, j)$ y sus respectivas reconstrucciones obtenidas en distintos experimentos. Cada histograma se superpone con se respectivo ajuste de la curva gaussiana. a) Diferencia entre una imagen reconstruida por AE_{MSE} y su correspondiente $LDCT(i, j)$. b) Diferencia por AE_{PSNR} . c) Diferencia por AE_{SSIM} . d) Diferencia por AE_{SOBEL} , e) Diferencia por AE_{MCI} , f) Diferencia por AE_{MC} y g) Diferencia por $SSGAN$.

Realizando el análisis de las desviaciones estándar (σ) y medias (μ), se obtienen los valores

promedio correspondientes a cada método de reconstrucción 5.2.

Reconstrucciones	Promedio de $\sigma \pm$ Error	Promedio de $\mu \pm$ Error
AE_{MSE}	0.0992 ± 0.006139	0.039 ± 0.000989
AE_{PSNR}	0.0747 ± 0.002454	0.0040 ± 0.001726
AE_{SSIM}	0.0740 ± 0.002327	0.0022 ± 0.0005846
AE_{SOBEL}	0.0716 ± 0.001925	0.0035 ± 0.00073839
AEMCI	0.0652 ± 0.000873	0.0048 ± 0.0003464
AEMC	0.0622 ± 0.000369	0.0062 ± 0.0002749
SSGAN	0.0665 ± 0.001091	0.0150 ± 0.0007134
Ideal	0.0600 ± 0.000000	0.000 ± 0.000000

Tabla 5.2: Comparación de la media de la desviación estándar σ y la media de μ de la diferencia de las diferentes reconstrucciones vs la LDCT, incluyendo el error.

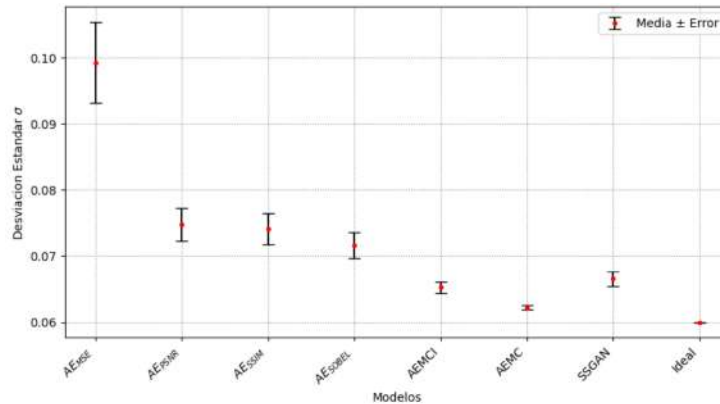


Figura 5.13: Gráfico correspondiente las desviaciones estándar σ de los diferentes modelos de reconstrucción.

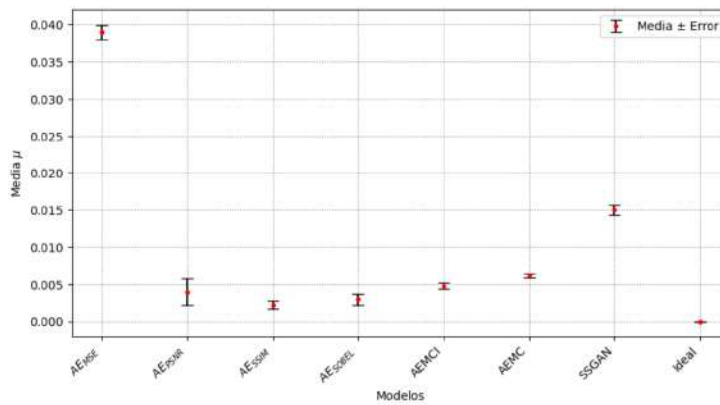


Figura 5.14: Gráfico correspondiente a las medias μ de los diferentes modelos de reconstrucción.

Gracias al cuadro 5.2 y a la figura 5.13, podemos identificar que el modelo que más se aproxima a nuestro caso ideal ($\sigma = 0.06$) es AEMC, con un valor promedio de desviación estándar $\sigma_{prom} = 0.0622$. Esto es consistente con el histograma f) en la figura 5.12, el cual muestra una de las mejores

coincidencias con la curva gaussiana. Los siguientes modelos con mejor desempeño en este criterio son AEMCI y SSGAN. En contraste, el modelo que más se aleja del caso ideal es AE_{MSE} , lo que indica que las imágenes resultantes de la diferencia presenta una mayor variabilidad. Es decir, las $N(i, j)$ resultantes contienen una cantidad de ruido superior a la esperada, lo que sugiere que este método de reconstrucción no es eficiente en la reducción de ruido. Esto se refleja en el hecho de que, al calcular la diferencia entre las imágenes reconstruidas con este método y sus respectivas LDCT, las imágenes resultantes muestran una mayor dispersión en los valores de los píxeles, lo que contribuye al aumento de la desviación estándar.

La figura 5.14 muestra los valores promedio de μ , evidenciando que el modelo AEMC no es completamente óptimo, ya que su estimación de la media no es la más precisa. Esto es consistente con su histograma, en el que se observa un ligero desplazamiento hacia la derecha. Un valor de μ mayor a 0 indica un desplazamiento del histograma hacia valores positivos, lo que sugiere que los valores de los píxeles en la imagen reconstruida son superiores a los esperados. En este aspecto, el modelo AE_{SSIM} es el que más se apega a la media ideal ($\mu = 0$), obteniendo un valor promedio de $\mu_{prom} = 0.0022$. Por otro lado, al igual que en el análisis de σ , el modelo AE_{MSE} es el que más se aleja del comportamiento deseado, lo que refuerza la inferencia de que su reconstrucción no es adecuada.

Sin embargo, este análisis no es suficiente para determinar cuál de los modelos es el más adecuado para la tarea de reducción de ruido y preservación de bordes. Por ello, se llevó a cabo un análisis adicional basado en diversas métricas de calidad de imagen, así como un método de evaluación para la preservación de bordes.

5.5. Métricas de Calidad en Reconstrucción de Imágenes LDCT

Para evaluar las diferentes reconstrucciones obtenidas en los distintos experimentos, se realizó un análisis cuantitativo utilizando métricas de calidad de imagen. Específicamente, se calcularon las siguientes métricas:

- Relación Señal-Ruido Pico (PSNR)
- Índice de Similitud Estructural (SSIM)
- Error Cuadrático Medio (MSE)
- Error Medio Absoluto sobre Gradientes de Sobel (MAE_{Sobel})

Cabe mencionar que se empleó la métrica MAE_{Sobel} para evaluar la preservación de bordes en la reconstrucción. Se calculó el error medio absoluto después de aplicar el filtro de Sobel, dado que esta técnica fue implementada para mejorar la preservación de bordes durante el entrenamiento.

Los resultados obtenidos se presentan a continuación:

Reconstrucciones	PSNR(dB) \pm Error	SSIM \pm Error	MSE \pm Error	MAE_{Sobel} \pm Error
LDCT	24.676803 \pm 0.004439	0.498428 \pm 0.00109	0.003407 \pm 2.69e-09	0.034429 \pm 1.01e-06
AE_{MSE}	25.681483 \pm 0.016313	0.536543 \pm 0.000955	0.002704 \pm 6.36e-09	0.034126 \pm 5.16e-07
AE_{PSNR}	26.41676 \pm 0.005477	0.565194 \pm 0.000951	0.002282 \pm 1.5e-09	0.032533 \pm 7.42e-07
AE_{SSIM}	26.744281 \pm 0.008677	0.577901 \pm 0.000866	0.002117 \pm 2.06e-09	0.031786 \pm 5.1e-07
AE_{SOBEL}	27.548469 \pm 0.02029	0.615346 \pm 0.000693	0.001759 \pm 3.31e-09	0.029684 \pm 4.94e-07
AEMCI	27.809272 \pm 0.02443	0.635549 \pm 0.000736	0.001657 \pm 3.53e-09	0.027666 \pm 1.26e-06
AEMC	29.731123 \pm 0.049247	0.764993 \pm 0.00018	0.001065 \pm 2.92e-09	0.018605 \pm 4.96e-07
SSGAN	28.268310 \pm 0.541511	0.848018 \pm 0.000338	0.001512 \pm 7.18e-08	0.026198 \pm 2.56e-06

Tabla 5.3: Resultados de PSNR(dB), SSIM, MSE y MAE_{Sobel} con sus respectivos errores para diferentes reconstrucciones.

Análisis de Resultados
5.5 Métricas de Calidad en Reconstrucción de Imágenes LDCT

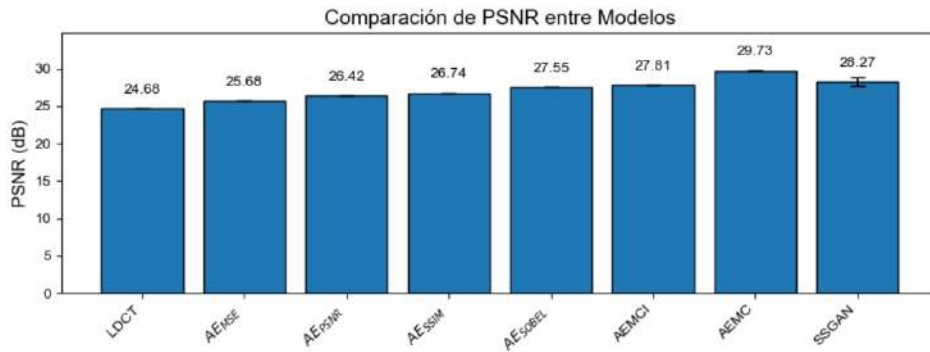


Figura 5.15: Gráfico de valores de PSNR de diferentes reconstrucciones.

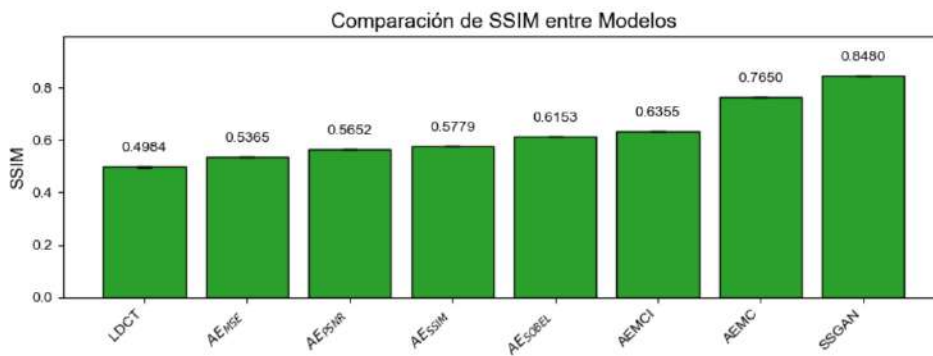


Figura 5.16: Gráfico de valores de SSIM de diferentes reconstrucciones.

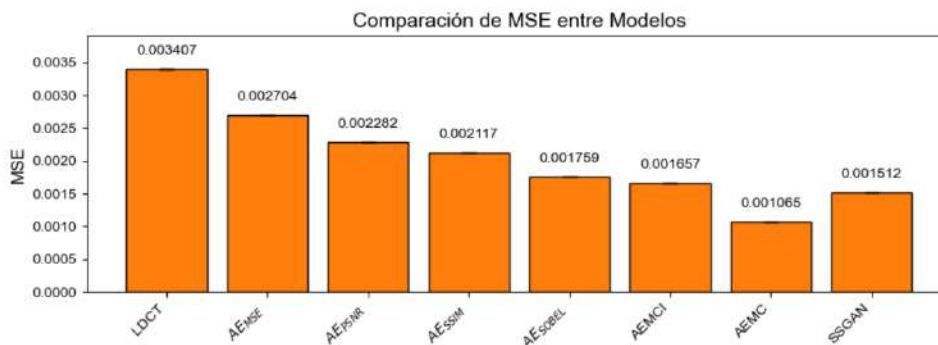


Figura 5.17: Gráfico de valores de MSE de diferentes reconstrucciones.

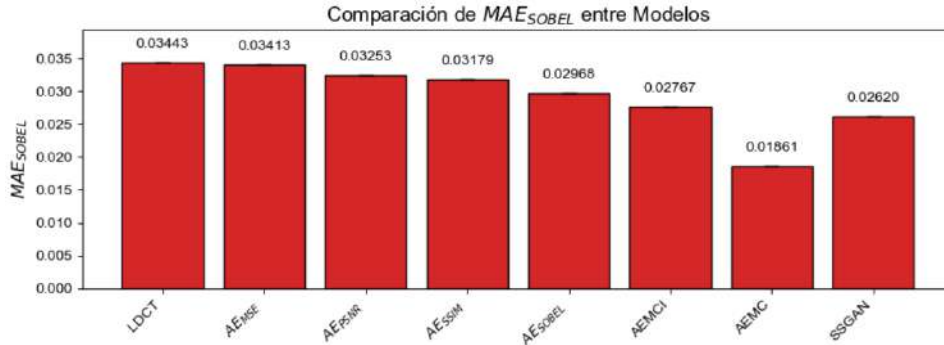


Figura 5.18: Gráfico de valores de MAE aplicado el filtro SOBEL de diferentes reconstrucciones.

Al analizar los resultados obtenidos en la métrica PSNR, es importante recordar que un valor mayor indica una mejor reconstrucción en términos de reducción de ruido. En este sentido, las mejores reconstrucciones fueron obtenidas mediante el Autoencoder con Modificación en el Codificador (AEMC), alcanzando un valor promedio de 29.73 dB, seguido por las imágenes reconstruidas mediante la Red Generativa Adversarial Semisupervisada (SSGAN).

De manera similar, en la métrica SSIM, un valor más alto indica una mejor preservación de la estructura de la imagen, se observa que las reconstrucciones con mejores resultados fueron aquellas generadas por la SSGAN, alcanzando un valor promedio de 0.8480, seguidas por las reconstrucciones obtenidas mediante el AEMC.

Ahora bien, en los resultados obtenidos para el MSE y el MAE_{Sobel} , un valor más bajo indica una menor cantidad de error, lo que equivale a una mejor reconstrucción.

En el caso del MSE, las reconstrucciones con mejores resultados fueron aquellas generadas mediante el Autoencoder con Modificación en el Codificador (AEMC), seguidas, al igual que en los casos anteriores, por las reconstrucciones obtenidas a través de la SSGAN.

Por último, un análisis fundamental es la cuantificación de la preservación de bordes, evaluada mediante el MAE_{Sobel} . Los resultados indican que el AEMC se posiciona como el mejor modelo de reconstrucción para imágenes LDCT, logrando los mejores resultados en 3 de las 4 métricas evaluadas.

También es importante mencionar que, a partir de estos resultados, se confirma que el Autoencoder con función de costo MSE (AE_{MSE}) es el modelo que obtiene las peores reconstrucciones. Esto se evidencia en el hecho de que en cada una de las métricas evaluadas, fue el que presentó los resultados más bajos.

5.6. Relación Señal Ruido (RSR) y Coeficiente Contraste Ruido (CCR)

Se realizó la evaluación de la Relación Señal-Ruido (RSR) y el Coeficiente Contraste-Ruido (CCR) de 200 tomografías computarizadas de corte axial de tórax de cada diferente reconstrucción realizada. Para ello, se llevaron a cabo mediciones en dos regiones de interés (ROI) denominadas **A** y **B**, correspondientes a dos tejidos distintos: músculo del pecho y hueso, respectivamente. Estas regiones se utilizaron para calcular la denominada $ROI_{señal}$, junto con las mediciones realizadas en el fondo, identificadas como ROI_{ruido} .

Las ROI's utilizadas son cuadros de 11×11 píxeles en todas las imágenes. La figura 5.19 muestra la visualización de las ROI's de 3 imágenes diferentes.

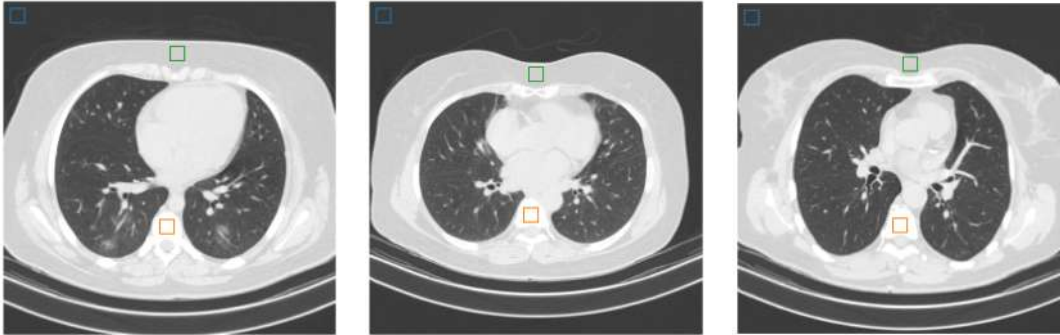


Figura 5.19: Visualización de las regiones de interés (ROI). Región A (color verde), correspondiente al músculo del pecho; region B (color naranja), correspondiente al hueso; y la ROI_{ruido} (color azul), correspondiente al fondo, en tres tomografías computarizadas de corte axial de tórax diferentes.

La Relación Señal Ruido (RSR) de las Regiones A y B se obtuvieron según las expresiones (3.26) y (3.27). Los resultados fueron los siguientes:

Reconstrucción	RSR-A	RSR-A (dB)
LDCT	14.79 ± 0.89	23.39 ± 0.52
<i>AE_{MSE}</i>	17.86 ± 1.18	25.02 ± 0.58
<i>AE_{PSNR}</i>	18.22 ± 1.31	25.19 ± 0.62
<i>AE_{SSIM}</i>	18.87 ± 1.32	25.49 ± 0.62
<i>AE_{SOBEL}</i>	20.65 ± 1.04	26.29 ± 0.44
AEMCI	54.97 ± 9.99	34.66 ± 1.56
AEMC	119.34 ± 4.76	41.53 ± 0.34
SSGAN	158.20 ± 17.59	43.93 ± 1.01
NDCT	125.62 ± 6.64	41.97 ± 0.46

Tabla 5.4: Media ± error de la RSR-A y RSR-A (dB) para cada reconstrucción.

Reconstrucción	RSR-B	RSR-B (dB)
LDCT	17.60 ± 1.00	24.90 ± 0.49
<i>AE_{MSE}</i>	21.34 ± 1.36	26.56 ± 0.55
<i>AE_{PSNR}</i>	21.56 ± 1.31	26.66 ± 0.53
<i>AE_{SSIM}</i>	22.34 ± 1.58	26.96 ± 0.63
<i>AE_{SOBEL}</i>	24.64 ± 1.04	27.83 ± 0.37
AEMCI	62.51 ± 11.84	35.77 ± 1.61
AEMC	143.38 ± 4.90	43.12 ± 0.30
SSGAN	173.03 ± 19.39	44.71 ± 1.01
NDCT	152.65 ± 7.59	43.66 ± 0.43

Tabla 5.5: Media ± error de la RSR-B y RSR-B (dB) para cada reconstrucción.

Comparación de Métricas RSR

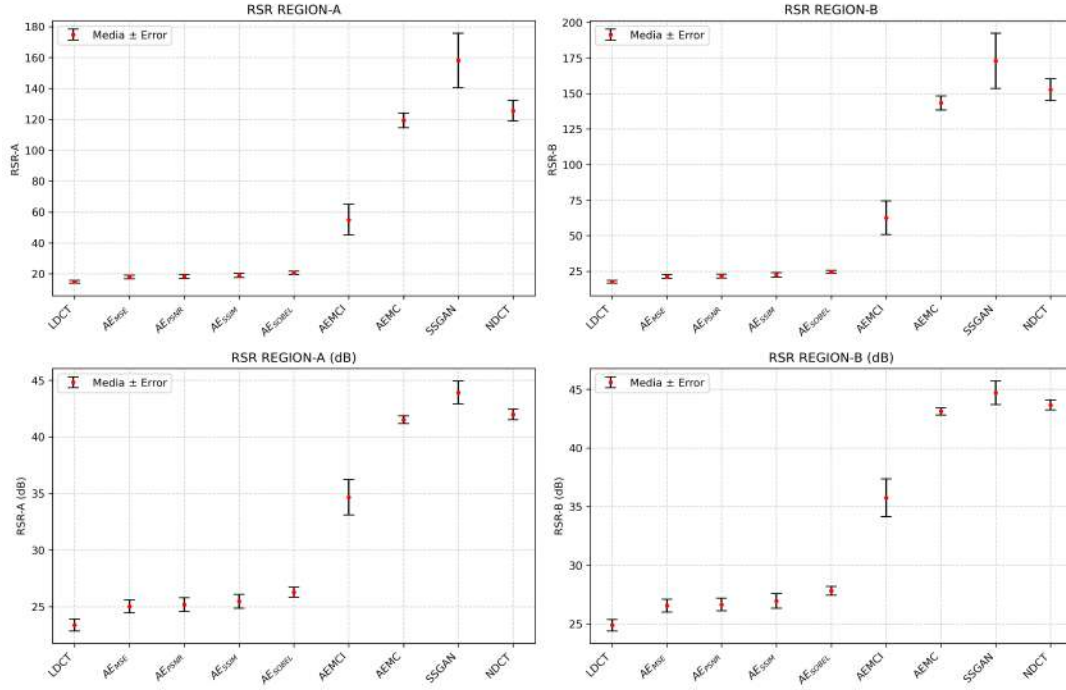


Figura 5.20: Gráfico correspondiente a la Relación Señal Ruido (RSR) de las regiones A y B

A su vez, también se realizó el cálculo del Coeficiente Contraste-Ruido (CCR) entre las regiones **A** y **B** de cada una de las imágenes, de acuerdo con la expresión (3.28). Los resultados obtenidos se muestran a continuación:

Reconstrucción	CCR _{AB}
LDCT	2.80 ± 0.15
AEMSE	3.47 ± 0.24
AEPN	3.26 ± 0.18
AESSIM	3.47 ± 0.35
AESOBEL	4.04 ± 0.16
AEMCI	7.54 ± 2.34
AEMC	24.04 ± 0.56
SSGAN	14.83 ± 2.22
NDCT	27.04 ± 1.24

Tabla 5.6: Media ± error del Coeficiente de Contraste-Ruido (CCR) entre las regiones A y B para cada reconstrucción.

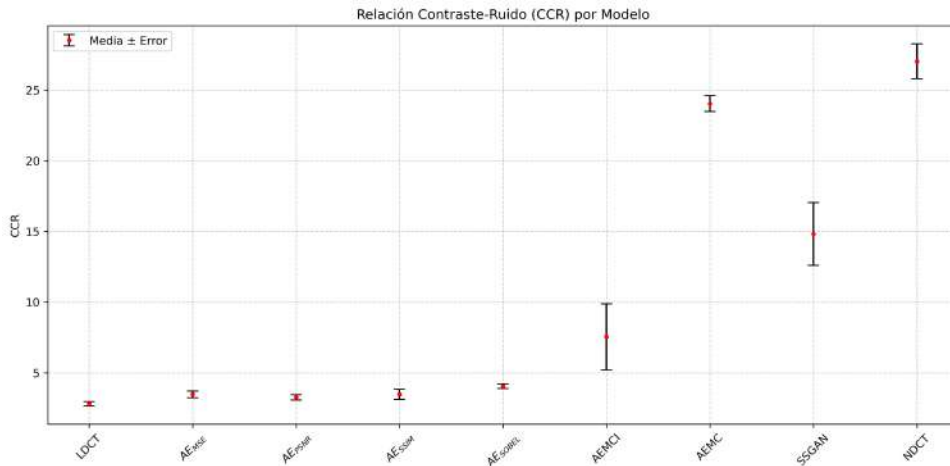


Figura 5.21: Gráfico correspondiente al Coeficiente Contraste-Ruido (CCR) entre las regiones A y B

Analizando los resultados de la Relación Señal-Ruido (RSR), figura 5.20, tanto de la región **A** como de la región **B**, se puede observar en primera instancia que las RSR correspondientes a las Tomografías Computarizadas de Baja Dosis (LDCT) presentan, en general, los valores más bajos obtenidos en este cálculo. Conforme se logran mejores reconstrucciones, este valor tiende a incrementarse, siguiendo una tendencia coherente con los resultados reportados anteriormente.

Un punto interesante es que las RSR obtenidas mediante las reconstrucciones generadas por la SSGAN son, en ambos casos (región **A** y región **B**), superiores a las de cualquier otra reconstrucción, incluso mayores que las RSR correspondientes a las Tomografías Computarizadas de Dosis Normal (NDCT), las cuales, en teoría, deberían representar la referencia ideal a alcanzar.

Este comportamiento puede explicarse considerando la forma en que se calcula la RSR, según la ecuación (3.26). Dicha relación establece que, mientras mayor sea la media de los valores de los píxeles en la región de señal ($ROI_{señal}$), mayor será el valor de la RSR, siempre y cuando la región de fondo (ROI_{ruido}) sea la misma. Esta misma razón explica por qué la Relación Señal-Ruido (RSR) de la región **A** (músculo) es menor que la de la región **B** (hueso), dado que los valores de los píxeles en el hueso son, en promedio, mayores que los del músculo.

Esto sugiere que las reconstrucciones generadas por la SSGAN tienden a presentar una mayor intensidad en los valores de sus píxeles, lo cual se manifiesta en imágenes visiblemente más claras.

Esta es una de las razones por las cuales la Relación Señal-Ruido (RSR) debe ser acompañada por el Coeficiente de Contraste-Ruido (CCR), ya que, como se mencionó anteriormente, el CCR debe ser lo suficientemente alto como para permitir distinguir entre diferentes tejidos. Esto nos lleva a la figura 5.21.

Los resultados del CCR complementan los obtenidos con la RSR, ya que el CCR cuantifica qué tan distinguibles son dos tejidos distintos, permitiendo evaluar si es posible diferenciarlos adecuadamente en la imagen reconstruida. En este sentido, se observa nuevamente que el Autoencoder con Modificación en el Codificador (AEMC) obtuvo los mejores resultados en términos de consistencia entre la RSR y el CCR.

Es importante destacar que una RSR alta no garantiza necesariamente un CCR elevado. Esto se hace evidente en el caso de las reconstrucciones generadas por la SSGAN, las cuales, a pesar de tener una RSR superior a la de todas las demás reconstrucciones, muestran un valor de CCR significativamente menor, lo que indica que podría ser difícil distinguir entre ambos tejidos.

Consideraciones

Algo sumamente relevante en el análisis de la Relación Señal-Ruido (RSR) es la **validez estadística** de las medidas utilizadas para representar la señal, particularmente cuando se recurre a la **media** $\mu_{señal}$ como valor representativo. En general, se toma como “señal” convenientemente a aquella región homogénea que involucra tejido del paciente, y si dicha región es completamente homogénea, las mediciones pueden aproximarse a una distribución normal. En estos casos, la media μ es un estimador estadísticamente válido, reflejando así el valor más probable del conjunto de datos.

Sin embargo, la situación cambia cuando se introduce ruido aditivo con distribución normal que no necesariamente se superpone con la distribución del tejido original. Esta combinación provoca que los datos resultantes no sigan una distribución normal (5.22). Esto invalida estadísticamente el uso de la media μ como una representación fiable del valor central, dado que ya no estaríamos hablando del valor más probable de los datos.

En estos escenarios, numéricamente se estaría simplemente calculando un promedio, su interpretación cambia.

Así mismo para el fondo se suele tomar una región oscura, esta región en la mayoría de los casos al añadirle ruido gaussiano sigue conservando una distribución normal, en estos casos si tiene sentido usar la σ pero no en todos los casos es así, cuando deja de tener una distribución normal los datos numéricamente lo que se calcula es

Así como la media $\mu_{señal}$ representa una medida de tendencia central, la desviación estándar σ es una medida de dispersión que cuantifica qué tan alejados están los valores de los datos respecto a su media. En el cálculo de la Relación Señal-Ruido (RSR), el valor de σ se asocia generalmente con la región de fondo, es decir, una región oscura de la imagen donde se espera que no haya señal del paciente.

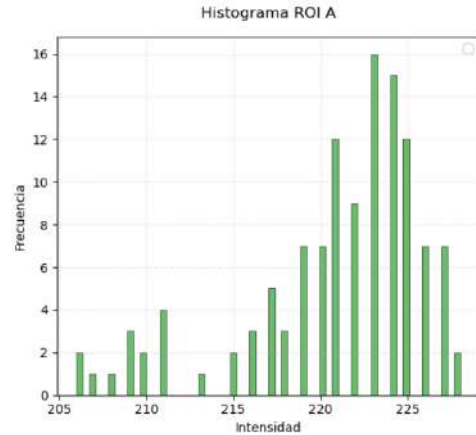


Figura 5.22: Histograma de ROI_{señal} región A

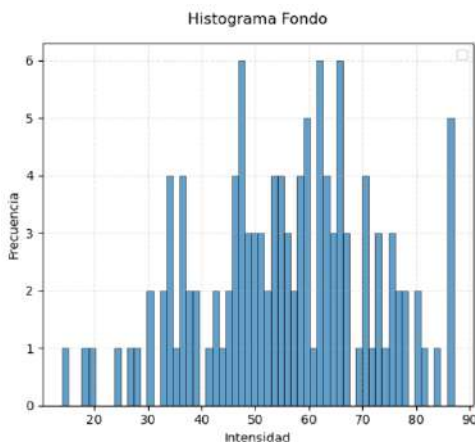


Figura 5.23: Histograma de ROI_{ruido} del fondo

Cuando esta región de fondo conserva una distribución normal, lo cual suele suceder al añadirle ruido gaussiano, el uso de la desviación estándar σ_{ruido} como estimador de la variabilidad del ruido es estadísticamente correcto. En este contexto, σ representa no solo la dispersión, sino también una probabilidad bien definida de qué tan probable es que un valor se aleje de la media, en función de la campana de Gauss.

Sin embargo, no en todos los casos el fondo mantiene una distribución normal. Hay casos en donde la región no tiene distribución normal 5.23. En tales casos, aunque se calcule numéricamente una desviación estándar, su interpretación estadística se ve comprometida. Ya no puede considerarse un estimador de confianza para caracterizar la dispersión, y simplemente se convierte en una medida descriptiva de variabilidad sin la validez probabilística que tiene en el caso normal.

Por ello, es importante tener en cuenta que el uso de σ en la RSR debe ir acompañado de una evaluación de la distribución de los datos en la región de fondo. Si esta no es aproximadamente normal, el valor de σ deja de ser un reflejo fiel de la variabilidad del ruido.

5.7. Precisión de Clasificación para cada Reconstrucción

Uno de los factores de mayor relevancia en esta investigación es analizar si una mejor reconstrucción de imágenes LDCT, en términos de una mayor reducción de ruido, preservación de estructura y conservación de bordes, conlleva una mejora en la precisión de clasificación diagnóstica.

Para evaluar este aspecto, se sometieron distintos subconjuntos de imágenes reconstruidas mediante los diferentes experimentos realizados en esta investigación, (AE_{MSE} , AE_{PSNR} , AE_{SSIM} , AE_{SOBEL} , AEMCI, AEMC y SSGAN), al clasificador descrito en las subsecciones 4.2.2 y 4.6.1.

Los resultados obtenidos se resumen en el cuadro 5.7 y figura 5.24:

Reconstrucciones	Precisión \pm Error
LDCT	0.654 ± 0.029
AE_{MSE}	0.768 ± 0.027
AE_{PSNR}	0.859 ± 0.030
AE_{SSIM}	0.888 ± 0.015
AE_{SOBEL}	0.957 ± 0.007
AEMCI	0.968 ± 0.0093
AEMC	1.000 ± 0.0000
SSGAN	0.926 ± 0.029

Tabla 5.7: Precisión para cada reconstrucción.

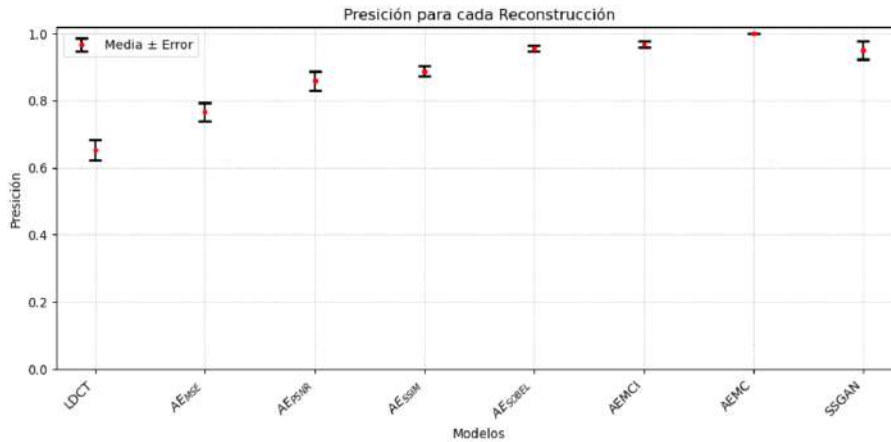


Figura 5.24: Gráfico de la precisión dada la diferente reconstrucción.

Estos resultados son imprescindibles, ya que indican que, en efecto, una mejor reconstrucción de imágenes LDCT conlleva una mejora en la precisión de la clasificación diagnóstica. De acuerdo con los resultados obtenidos, la precisión promedio al clasificar imágenes LDCT en pacientes sanos o enfermos es de un 65.4%, lo cual no es un diagnóstico confiable. Esto significa que, en promedio, solo 65 de cada 100 imágenes son clasificadas correctamente, lo que resulta insuficiente para fines clínicos.

De manera similar el Autoencoder con función de costo MSE (AE_{MSE}), mostró una mejora de apenas 11.4%, alcanzando un nivel aún insuficiente para el diagnóstico, demostrando que es una

forma muy ineficiente de reconstrucción de imágenes LDCT.

Por otro lado, es crucial destacar que las imágenes reconstruidas mediante el Autoencoder con Modificación en el Codificador (AEMC), que obtuvo los mejores resultados cuantitativos, alcanzaron una precisión del 100 %. Esto representa una mejora del 34.6% en comparación con las imágenes LDCT, logrando clasificar correctamente todas las imágenes según corresponda a un paciente sano o enfermo.

Gracias a estos resultados, es evidente que una mejor reconstrucción de imágenes LDCT, en términos de reducción de ruido, preservación de la estructura y conservación de bordes, conduce a una mejora en la precisión del diagnóstico. En la Figura 5.24, se observa que las imágenes reconstruidas con una menor cantidad de ruido y mejor preservación estructural y de bordes tienden a mejorar la precisión en la clasificación, presentado un mayor porcentaje de aciertos.

5.8. Reconstrucciones - Análisis Cualitativo

Un aspecto de suma importancia, aunque no tiene la virtud de ser cuantitativo, es la visualización de las diferentes reconstrucciones.

Esta visualización permite obtener una evaluación preliminar de la calidad de la reconstrucción mediante una inspección rápida de aspectos relevantes en las Tomografías Computarizadas de Corte Axial (TAC) de tórax. Entre estos aspectos se incluyen:

- Visualización de regiones en la zona pulmonar, como la presencia y claridad de los vasos pulmonares.

- Detección de atenuación pulmonar, es decir, la presencia de áreas con alteraciones en la densidad del tejido pulmonar.

- Diferenciación clara entre tejidos, como hueso, tejido blando, grasa y aire.

- Preservación de bordes, lo que resulta fundamental para distinguir estructuras anatómicas y detectar posibles patologías.

Por lo que se realizó una inspección visual de imágenes aleatorias del conjunto de prueba, reconstruidas mediante los distintos experimentos realizados en esta investigación.

Los resultados obtenidos se presentan a continuación:

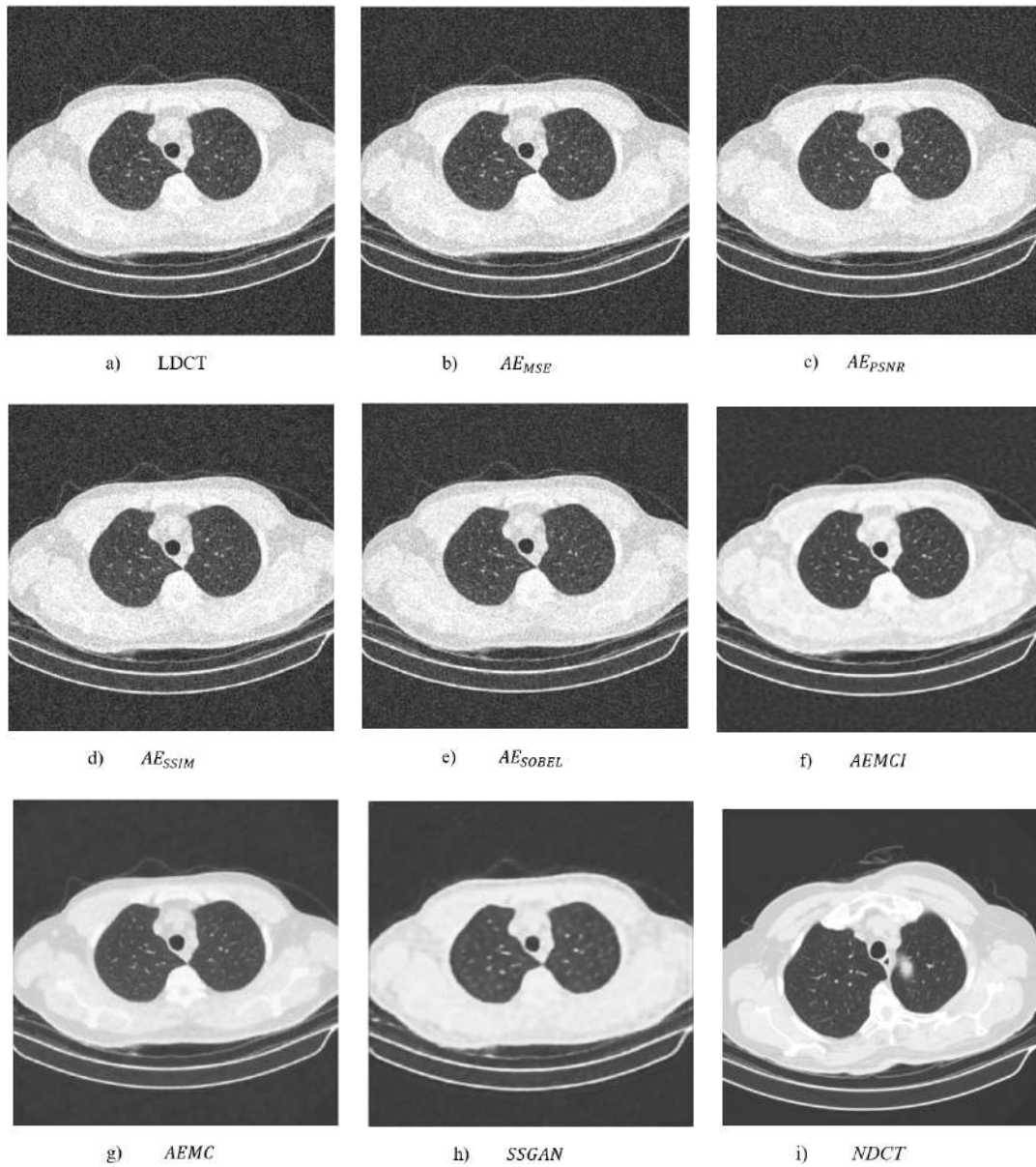


Figura 5.25: Comparación de diferentes reconstrucciones de una TAC de Tórax. a) LDCT, b) AE_{MSE} , c) AE_{PSNR} , d) AE_{SSIM} , e) AE_{SOBEL} , f) $AEMCI$, g) $AEMC$, h) $SSGAN$ e i) $NDCT$.

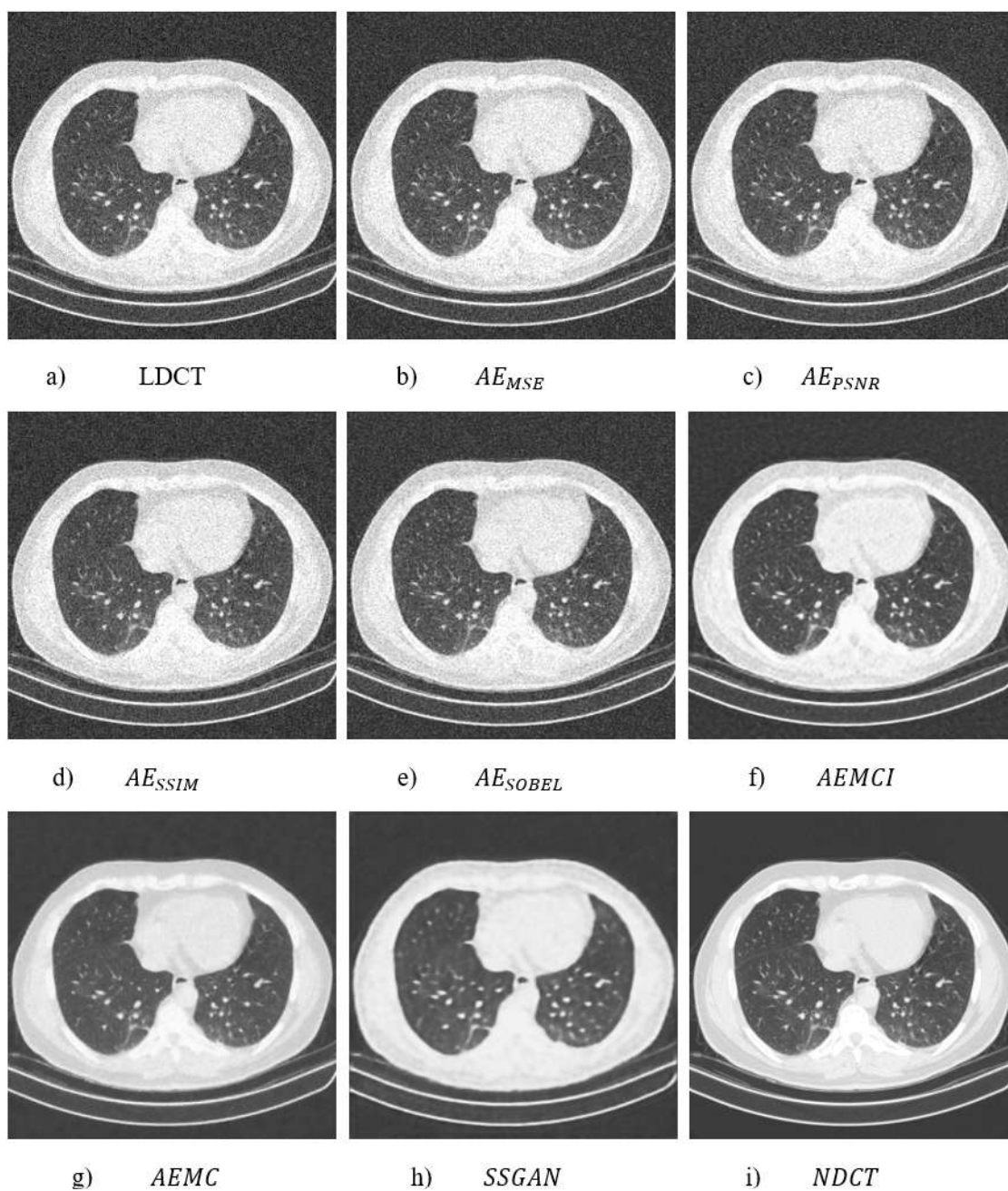


Figura 5.26: Comparación de diferentes reconstrucciones de una TAC de Tórax. a) LDCT, b) AE_{MSE} , c) AE_{PSNR} , d) AE_{SSIM} , e) AE_{SOBEL} , f) $AEMCI$, g) $AEMC$, h) $SSGAN$ e i) $NDCT$.

Las reconstrucciones obtenidas mediante los distintos métodos muestran que algunos enfoques logran una reducción efectiva del ruido sin comprometer en exceso la estructura de la imagen, así como la diferenciación y separación entre tejidos. Sin embargo, no todos los métodos empleados presentan un desempeño óptimo, lo que resalta la importancia de visualizar las reconstrucciones obtenidas de imágenes LDCT con fines diagnósticos.

Entre estos, la imagen a), reconstruida mediante el Autoencoder con Modificación en el Codificador ($AEMC$), destaca notablemente al proporcionar una representación clara de las estructuras

internas. Se observa una aceptable separación entre tejidos, permitiendo una mejor identificación de regiones pulmonares, como los vasos pulmonares y áreas de atenuación pulmonar. Además, esta reconstrucción logra separar regiones como hueso, tejido blando, grasa y aire, lo que la convierte en una de las mejores reconstrucciones para imágenes LDCT.

Por otro lado, es importante reconocer que las imágenes f) y h) presentan reconstrucciones visualmente aceptables. Aunque logran una notable reducción del ruido, muestran un cierto suavizado en las regiones de mayor textura, lo que sugiere que el modelo podría estar eliminando información relevante en su intento por generar una imagen más uniforme. Este efecto puede ser particularmente problemático en el contexto clínico, donde la pérdida de detalles puede dificultar la identificación de anomalías sutiles. No obstante, la imagen h), reconstruida por la SSGAN, ha demostrado ser una alternativa prometedora. Aunque no es la mejor reconstrucción obtenida, su rendimiento sugiere que, con ciertas modificaciones podría alcanzar un desempeño prometedor.

En contraste, los casos de las imágenes b), c), d) y e), al menos desde una perspectiva visual, no presentan mejoras significativas respecto a la imagen LDCT. La reducción de ruido en estos métodos parece ser insuficiente y deficiente. Esto indica que estas estrategias no son viables para cumplir con los objetivos de esta investigación, ya que una reconstrucción deficiente podría afectar la capacidad de diagnóstico en aplicaciones médicas reales.

Estos resultados resaltan la importancia de equilibrar la reducción de ruido con la preservación estructural y de bordes. Si bien eliminar el ruido es crucial para mejorar la calidad visual de la imagen, una reconstrucción excesivamente suavizada podría comprometer la interpretación clínica, al ocultar detalles esenciales. Por ello, el modelo AEMC representa una opción viable.

5.9. Resultados vs Estado del Arte

Se llevó a cabo una comparación de los valores de PSNR obtenidos por diferentes métodos de reconstrucción reportados en el estado del arte, contrastándolos con el desempeño del Autoencoder con Modificación en el Codificador (AEMC). Los resultados se presentan en la Tabla 5.8 y la Figura 5.27.

Modelo	PSNR(dB)
LDCT	24.6884
CNN	26.7486
EDCNN	28.1206
RED-CNN	28.3997
AE	28.5900
Res-CNN	28.6400
DUGAN	27.8717
Uformer	28.3197
CTformer	27.8008
DRWGAN-PF	29.6700
AEMC	29.7311

Tabla 5.8: Comparación de valores PSNR entre distintos modelos de reconstrucción de imágenes de TAC.

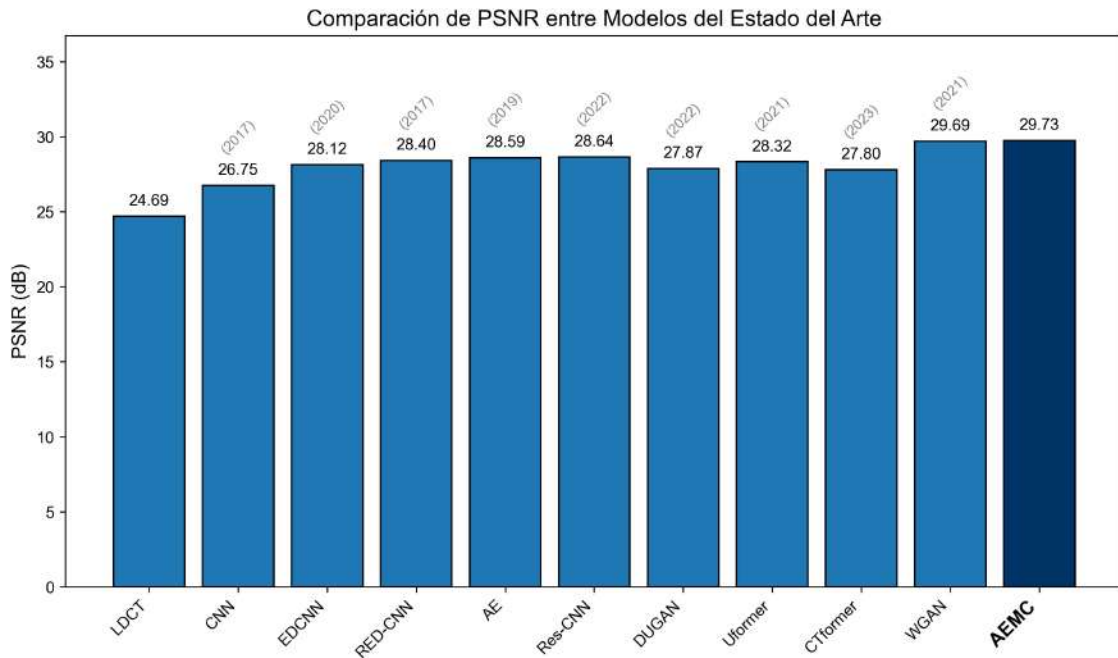


Figura 5.27: Comparación de valores de PSNR de Reconstrucciones obtenidas del estado del arte vs AEMC

A partir de los resultados obtenidos, es posible mencionar que el modelo AEMC logra un desempeño altamente competitivo, alcanzando un valor de PSNR de 29.7311 dB, lo que representa el mejor resultado dentro de la comparación.

Aunque la diferencia con el segundo mejor modelo, DRWGAN-PF, es marginal (0.0611 dB), este resultado demuestra que el enfoque propuesto es capaz de competir con técnicas de reconstrucción actuales. Además, el AEMC supera significativamente a métodos tradicionales y redes neuronales convolucionales estándar, como CNN, consolidando su eficacia en la reconstrucción de imágenes LDCT.

Es importante destacar que, aunque el PSNR no es la única métrica relevante para evaluar la calidad de la reconstrucción, es por excelencia una de las métricas con mayor relevancia en investigaciones donde se busca realizar esta clase de tareas.

Por lo tanto, estos resultados validan la efectividad del AEMC como una alternativa viable para la reconstrucción de imágenes LDCT, con un desempeño comparable e incluso superior a algunos modelos publicados.

Capítulo 6

Conclusiones

La presente investigación demostró que el uso de una arquitectura basada en autoencoders, combinada con funciones de costo que contienen métricas de calidad de imagen y métodos de preservación de bordes, así como una estrategia de alineación del espacio latente entre Tomografías Computarizadas de Baja Dosis (LDCT) y Tomografías Computarizadas de Dosis Normal (NDCT), denominada Autoencoder con Modificación en el Codificador (AEMC), constituye una metodología efectiva para la reducción de ruido, la preservación de bordes y la conservación de la estructura en imágenes LDCT. Asimismo, se evidenció que contar con imágenes de mayor calidad, en términos de menor cantidad de ruido y mayor fidelidad estructural, se traduce en una mejora sustancial en la precisión de la clasificación diagnóstica.

El Autoencoder con Modificación en el Codificador (AEMC), permitió mejoras significativas tanto en aspectos cuantitativos como cualitativos, abarcando la reducción del ruido, la preservación de bordes, la conservación de la estructura y la calidad visual de las imágenes reconstruidas; aspectos esenciales para una adecuada interpretación clínica.

Los resultados obtenidos, evaluados mediante métricas cuantitativas como el Peak Signal-to-Noise Ratio (PSNR), el Structural Similarity Index Measure (SSIM), el Mean Square Error (MSE), y métricas orientadas a la evaluación de bordes como el MAE_{SOBEL} , evidencian una mejora sustancial en comparación con las imágenes LDCT. Estos valores alcanzaron niveles comparables a los reportados en el estado del arte, con la ventaja adicional de no requerir imágenes emparejadas. Esta capacidad se atribuye al proceso de alineación del espacio latente, el cual permitió al codificador aprender representaciones latentes características de las imágenes NDCT, incrementando así la fidelidad de las reconstrucciones generadas a partir de entradas degradadas (LDCT). Es importante recalcar que estos resultados fueron obtenidos mediante datos no vistos durante el entrenamiento lo que indica que el AEMC mantiene un buen grado de generalización.

Concluyendo, los resultados obtenidos validan el enfoque propuesto como una alternativa eficiente y viable para el posprocesamiento de imágenes LDCT, contribuyendo a la mejora de la calidad visual y estructural en contextos clínicos donde se busca reducir la exposición del paciente a radiación ionizante, sin comprometer la información anatómica esencial para el diagnóstico.

Un punto importante a destacar son los resultados obtenidos mediante el uso de la SSGAN, la cual se propuso como un método conjunto capaz de mejorar la calidad de imágenes LDCT generadas y, de manera simultánea, clasificar si corresponden a pacientes sanos o enfermos. Los resultados obtenidos fueron altamente prometedores, lo cual refuerza la razón por la cual en el estado del arte se exploran con frecuencia variantes de las redes GAN. Sin duda, este enfoque representa un campo de gran interés, con amplio potencial de desarrollo, lo cual nos lleva al siguiente apartado.

Limitaciones

A pesar de los resultados obtenidos, este trabajo presenta ciertas limitaciones que deben ser consideradas. En primer lugar, no se incorporan técnicas específicas de mejora de contraste o resolución espacial en las imágenes reconstruidas, lo cual podría potenciar la calidad visual en general y la utilidad clínica de los resultados.

Asimismo, es importante señalar que los experimentos se realizaron con imágenes simuladas a partir de la adición de ruido gaussiano a estudios de dosis normal. En consecuencia, no puede garantizarse que el modelo mantenga el mismo desempeño frente a imágenes LDCT reales, ya que el ruido y los artefactos presentes en éstas, aunque sean de naturaleza similar, no son exactamente idénticos a los simulados. Esta diferencia podría generar variaciones significativas en la calidad de la reconstrucción cuando se apliquen a datos clínicos reales.

De igual forma, aunque las reconstrucciones logradas muestran una reducción significativa del ruido, no se está eliminando el ruido por completo. Esto es que, aun que si bien, las imágenes reconstruidas se acercan visualmente a las TC de dosis normal, aún persisten diferencias visibles que limitan la equivalencia plena entre ambas modalidades y esto es fácilmente notable en los resultados cuantitativos al señalar que no es una reconstrucción perfecta.

Estos aspectos representan áreas de mejora y motivan el desarrollo de investigaciones futuras que incluyan pruebas con bases de datos clínicas reales, la incorporación de técnicas de realce de contraste y la evaluación de métricas perceptuales más avanzadas.

Trabajo Futuro

A partir de los hallazgos de este estudio, se plantean diversas líneas de trabajo que permitirían ampliar el alcance y la aplicabilidad del modelo propuesto.

Una dirección inmediata consiste en la integración de funciones de costo que incorporen métricas de calidad de imagen, así como métodos orientados a la preservación de bordes dentro de la arquitectura SSGAN (véase la sección A). La implementación de esta estrategia no es tan trivial como lo pudiera parecer, particularmente debido a la ausencia de datos emparejados durante el entrenamiento, lo cual exige mecanismos más sofisticados de alineación semántica y supervisión implícita. No obstante, este enfoque representa una línea de investigación con un alto potencial, tanto en términos de innovación metodológica como de aplicabilidad clínica.

Por otro lado, si bien las estructuras utilizadas en esta investigación demostraron ser funcionales, los hiperparámetros empleados no fueron objeto de una optimización exhaustiva. Resulta pertinente, por tanto, investigar de manera más sistemática la influencia de estos parámetros sobre el rendimiento del modelo, mediante técnicas como búsqueda en malla (grid search), optimización bayesiana o algoritmos evolutivos, con el objetivo de maximizar la estabilidad y la eficiencia del entrenamiento.

También, se propone explorar la posibilidad de realizar transformaciones sobre las imágenes antes de realizar el entrenamiento, tales como la representación en el dominio de las frecuencias mediante la transformada de Fourier. Esta estrategia permitiría analizar si una representación alternativa de los datos contribuye a mejorar la capacidad del modelo para distinguir y reconstruir patrones estructurales, especialmente en Tomografías Computarizadas de Baja Dosis (LDCT), lo que podría derivar en mejoras sustanciales respecto a los resultados obtenidos en el presente estudio.

Finalmente, sería de interés validar el modelo en escenarios clínicos reales mediante estudios retrospectivos o prospectivos, comparando su desempeño con métodos de reducción de dosis actualmente implementados en estaciones de trabajo comerciales, para así establecer su viabilidad práctica y su impacto potencial en la rutina clínica de radiología diagnóstica.

Apéndice A

Estrategias AEMCI y SSGAN

En este apartado se presentan estrategias de entrenamiento alternativas desarrolladas durante la investigación para reducir el ruido en Tomografías Computarizadas de Baja Dosis (LDCT). Una de estas estrategias se basa en la técnica de alineación del espacio latente implementada en el Autoencoder con Modificación en el Codificador (AEMC), pero con variaciones en el método de entrenamiento. Aunque los resultados obtenidos no superaron a los del enfoque principal, se consideran relevantes y, por ello, se incluyen en este apéndice.

Adicionalmente, se desarrolló una propuesta de red capaz de generar imágenes con menor cantidad de ruido y, de manera simultánea, clasificar si las imágenes corresponden a pacientes sanos o enfermos. Aunque esta propuesta aborda algunos de los aspectos centrales de esta investigación, no se alinea completamente con los objetivos específicos ni con la hipótesis planteada. Aun así, los resultados obtenidos por estas estrategias fueron reportadas en el capítulo 5.

A) Autoencoder con Modificación en el Codificador Iterativo (AEMCI)

Para llevar a cabo la alineación o aproximación del espacio latente, se emplearon dos enfoques. El primero fue el presentado anteriormente en la sección 4.5. El segundo corresponde a una variante de este, denominada Autoencoder con Modificación en el Codificador Iterativo. Este nombre se debe a que, a diferencia del primer enfoque donde se realizan dos entrenamientos por separado, en esta variante la alineación se lleva a cabo de manera iterativa, actualizando los pesos w del codificador en cada minibatch durante el entrenamiento.

Entrenamiento - AEMCI

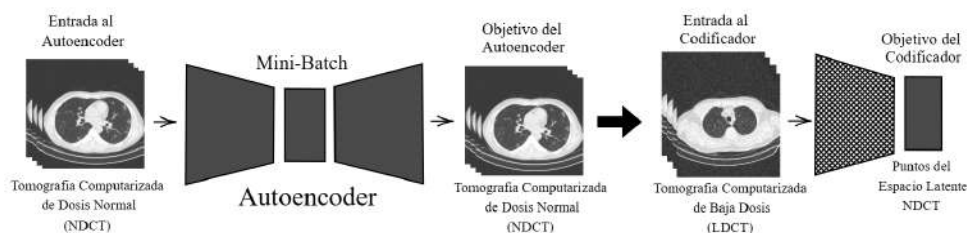


Figura A.1: Esquema general de entrenamiento de Autoencoder con Modificación en el Codificador Iterativo (AEMCI)

La propuesta consistió en realizar una alineación iterativa del espacio latente entre imágenes NDCT y LDCT. Específicamente, para cada minibatch de imágenes NDCT se generaron puntos latentes mediante el autoencoder. Posteriormente, se utilizó un minibatch de imágenes LDCT para producir sus correspondientes representaciones latentes, y se llevó a cabo un proceso de alineación entre ambos conjuntos latentes. Este procedimiento se repitió iterativamente durante el entrenamiento, actualizando los pesos del codificador dos veces por minibatch y los del decodificador una sola vez. Gracias a esta estrategia, el autoencoder sería capaz de aprender a codificar las imágenes LDCT de forma similar a como lo haría con imágenes NDCT, lo que permitiría obtener reconstrucciones de mayor calidad, con una mejor reducción de ruido y una adecuada preservación de bordes.

Durante el entrenamiento, el autoencoder recibe como entrada un minibatch de imágenes NDCT y tiene como objetivo reconstruirlas a sí mismas. A partir de este proceso, se extraen los puntos latentes generados por dicho minibatch. Posteriormente, estos puntos latentes se utilizan como objetivos para entrenar el codificador, el cual ahora recibe como entrada un minibatch del mismo tamaño, pero compuesto por imágenes LDCT.

Es decir, por cada minibatch que pasa por el autoencoder, se obtienen nuevas representaciones latentes de imágenes NDCT, que luego sirven como referencia para alinear iterativamente las representaciones latentes generadas a partir de las imágenes LDCT. Este proceso se repite constantemente, renovando los objetivos latentes del codificador en cada iteración.

Para garantizar una alineación efectiva, en cada ciclo de entrenamiento se actualizan dos veces los pesos del codificador y una sola vez los pesos del decodificador. Esta estrategia permite que el codificador aprenda a mapear imágenes LDCT hacia un espacio latente más cercano al de las NDCT, favoreciendo así una reconstrucción de mayor calidad.

Parámetros de Entrenamiento - AEMCI

El conjunto de datos mencionado en 4.1 se dividió en 3 subconjuntos.

1. Entrenamiento Autoencoder: 4500 imágenes de Tomografías Computarizadas de Dosis Normal (NDCT).
2. Entrenamiento Codificador: 4500 imágenes de Tomografías Computarizadas de Baja Dosis (LDCT).
3. Prueba: 1000 imágenes de Tomografías Computarizadas de Baja Dosis (LDCT).

El AEMCI fue entrenado durante 100 épocas con un batch size de 8, usando para el Autoencoder la función de costo (4.4) y para el codificador en la alineación de los espacios latentes (4.1).

Se usó el algoritmo de optimización Adam ($\eta = 0.0005$, $\beta_1 = 0.001$, $\beta_2 = 0.999$, $\epsilon = 1e-07$).

B) Red Generativa Adversarial Semisupervisada (SSGAN)

El uso de una red convolucional Autoencoder, así como de un clasificador convolucional (ver sección 4.2), abre la posibilidad de emplear estas dos redes, que anteriormente fueron utilizadas de forma individual, en un único método de entrenamiento. Este método permitiría generar una imagen y, al mismo tiempo, clasificar si la imagen generada corresponde a un paciente sano o enfermo. Este enfoque es conocido como una Red Generativa Adversarial Semisupervisada (SSGAN) (ver sección 3.3.8.1).

Recordemos que el Autoencoder utilizado (ver sección 4.2.1) tiene la tarea de reconstruir imágenes, es decir, recibe como entrada una imagen y produce como salida una imagen reconstruida. Por otro lado, el clasificador (ver sección 4.2.2) recibe como entrada una imagen y genera como salida un escalar: 0 si la imagen corresponde a un paciente sano y 1 si corresponde a un paciente enfermo.

En este contexto el Autoencoder (4.2.1) fungirá como el generador, mientras que el clasificador (4.2.2) actuará como el discriminador.

Es importante mencionar que, al tratarse de un entrenamiento semisupervisado, el clasificador deberá realizar dos tareas:

1. Tarea no supervisada: Determinar si la imagen es real o generada (fake).
2. Tarea supervisada: Clasificar si la imagen corresponde a un paciente sano o enfermo, para lo cual se anexará el etiquetado correspondiente.

Este enfoque permite aprovechar la capacidad del Autoencoder para generar imágenes con una menor cantidad de ruido y, al mismo tiempo, utilizar la información de clasificación para optimizar el proceso de reconstrucción y mejorar la precisión del diagnóstico basado en imágenes LDCT.

Entrenamiento - (SSGAN)

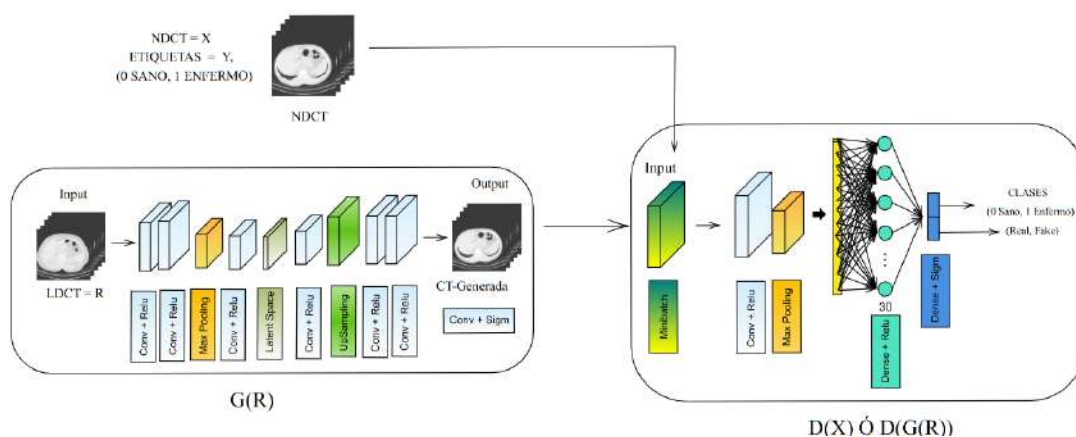


Figura A.2: Diagrama de red SSGAN

El flujo de entrenamiento de una Red Generativa Adversarial Semisupervisada (SSGAN) sigue un proceso iterativo en cada época. Durante cada iteración, se procesan minibatches de imágenes reales y generadas de la siguiente manera:

- Entrada de imágenes reales al discriminador: Un minibatch de imágenes NDCT (denotadas como x) junto con sus respectivas etiquetas y , ingresan al discriminador D , lo que produce $D(x)$, esto es para aprender a distinguir los datos reales.
- Generación de imágenes a partir de LDCT: Un minibatch de imágenes LDCT (denotadas como R) del mismo tamaño que el minibatch anterior, se introduce al generador G . Como resultado, se obtiene un minibatch de imágenes generadas o reconstruidas: $G(R)$.
- Clasificación de imágenes generadas por el discriminador: Las imágenes generadas $G(R)$ se introducen al discriminador D , obteniendo $D(G(R))$. El discriminador evalúa si las imágenes generadas son reales o falsas (fake) y adicionalmente, el discriminador clasifica si la imagen corresponde a un paciente sano o enfermo.

Este proceso se repite para todos los minibatches del conjunto de entrenamiento, y cuando se han procesado todos, se completa una época. De esta manera, se lleva a cabo el entrenamiento de una SSGAN, permitiendo que el generador mejore en la reconstrucción de imágenes LDCT y que el discriminador refine su capacidad de clasificación así como su detección de imágenes sintéticas.

Parámetros de Entrenamiento - SSGAN

Para el entrenamiento de la SSGAN, se utilizaron los siguientes parámetros:

Un aspecto importante durante el entrenamiento de la SSGAN es la función de costo, ya que consta de tres partes:

- La función de costo del discriminador no supervisado $C_{D_{unsupervised}}(w)$.
- La función de costo del discriminador supervisado $C_{D_{supervised}}(w)$.
- La función de costo del generador $C_G(w)$.

El discriminador no supervisado (clasificador rea/fake) tienen como función de costo:

$$C_{D_{unsupervised}}(w) = -\frac{1}{m} \sum_{k=1}^m [\ln(D(x_k)) + \ln(1 - D(G(R_k)))] \quad (\text{A.1})$$

Donde m es el numero de muestras totales, es decir, de imágenes, $D(x_k)$ es la salida del discriminador cuando se introduce una imagen real, por lo que x_k representa una muestra correspondiente a una imagen real y $D(G(R_k))$ es la salida del discriminador cuando entran datos del generador $G(R_k)$ por lo que en esencia R_k corresponde a la muestra k-esima de una imagen LDCT.

La función de costo supervisado es:

$$C_{D_{supervised}}(w) = -\frac{1}{m} \sum_{k=1}^m [y_k \ln(D(x_k)) + (1 - y_k) \ln(1 - D(x_k))] \quad (\text{A.2})$$

Donde y_k es la etiqueta k-esima correspondiente a la imagen k-esima. Aquí solo se evalúa con los datos reales ya que se busca que aprenda a clasificar imágenes correspondientes a pacientes sanos o enfermos.

La función de costo para el generador es:

$$C_G(w) = -\frac{1}{m} \sum_{k=1}^m \ln(1 - D(G(R_k))) \quad (\text{A.3})$$

Obteniendo finalmente el el costo total para la red SSGAN (A.4)

$$C_{SSGAN}(w) = -\frac{1}{m} \sum_{k=1}^m \left[[\ln D(x_k) + \ln (1 - D(G(R_k)))] + [y_k \ln D(x_k) + (1 - y_k) \ln(1 - D(x_k))] \right] \quad (\text{A.4})$$

La base de datos utilizada (4.1.2.1) se dividió en dos subconjuntos principales:

1. Entrenamiento: 4500 imágenes de Tomografías Computarizadas de Dosis Normal (NDCT) con sus respectivas 4500 etiquetas y 4500 imágenes de Tomografías Computarizadas de Baja Dosis (LDCT).
2. Prueba: 1000 imágenes de Tomografías Computarizadas de Baja Dosis (LDCT).

En el contexto del entrenamiento de la SSGAN, cada una de las 4500 imágenes NDCT con su respectiva etiqueta corresponde a x_k y y_k , mientras que cada una de las 4500 imágenes LDCT corresponde a R_k .

El entrenamiento de la SSGAN se realizó durante 20 épocas, con un tamaño de batch (batch size) de 32, utilizando el optimizador Adam con los siguientes hiperparámetros:

$$\eta = 0.000001, \quad \beta_1 = 0.005, \quad \beta_2 = 0.999, \quad \epsilon = 1e-07$$

Bibliografía

- [1] Convolutional neural network. <https://es.innovatiana.com/post/convolutional-neural-network>. Accedido el 5 de marzo de 2025.
- [2] Max pooling. <https://paperswithcode.com/method/max-pooling>. Accedido el 5 de marzo de 2025.
- [3] *Production of X Rays*, chapter 5, pages 69–90. John Wiley Sons, Ltd, 2002.
- [4] AAPM Task Group 150. Acceptance testing and quality control of digital radiographic imaging systems. Technical report, American Association of Physicists in Medicine, 2015.
- [5] Mahmoud Ahmed, Eslam Mahmoud, and Fahd Alturki. Image noise detection and removal based on enhanced gridlof algorithm. *International Journal of Advanced Computer Science and Applications*, 8, 01 2017.
- [6] Ali Alyami, Munassar Lasloom, Marzouq Lesloun, Ali Mutarid, Saleh Alyami, Hamad Batnain, and Salem Batnain. Simple physical principles and medical uses of computed tomography (ct) scan a new assessment. *International Journal of Advanced Research*, 11:1045–1054, 07 2023.
- [7] American Association of Physicists in Medicine (AAPM). Low dose ct grand challenge, 2025. Accessed: 2025-04-04.
- [8] Anonymous. Downsampling and upsampling of images: Demystifying the theory. <https://medium.com/analytics-vidhya/downsampling-and-upsampling-of-images-demystifying-the-theory-4ca7e21db24a>, 2020. Accedido el 5 de junio de 2020.
- [9] J. Beatty. *The radon transform and the mathematics of medical imaging*. Colby College, 2012. Accessed: 2021-02-15.
- [10] M. Bosch de Basea Gomez, I. Thierry-Chef, R. Harbron, et al. Risk of hematological malignancies from ct radiation exposure in children, adolescents and young adults. *Nature Medicine*, 29(12):3111–3119, 2023.
- [11] Jason Brownlee. Semi-supervised generative adversarial network, 2025. Accessed: 2025-04-04.
- [12] S.C. Bushong. *Manual de radiología para técnicos*. Elsevier.
- [13] Q.B. Carroll. *Practical Radiographic Imaging*. Charles C Thomas, 2007.
- [14] Q.B. Carroll. *Radiography in the Digital Age: Physics - Exposure - Radiation Biology (Third Edition)*. Charles C. Thomas, Publisher, Limited, 2018.
- [15] Hongchi Chen, Qiuxia Li, Lazhen Zhou, and Fangzuo Li. Deep learning-based algorithms for low-dose ct imaging: A review. *European Journal of Radiology*, 172:111355, 02 2024.

- [16] Hu Chen, Yi Zhang, Mannudeep K. Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*, 36(12):2524–2535, 2017.
- [17] Yang Chen, Xindao Yin, Luyao Shi, Huazhong Shu, Limin Luo, Jean-Louis Coatrieux, and Christine Toumoulin. Improving abdomen tumor low-dose ct images using a fast dictionary learning based processing. *Physics in medicine and biology*, 58:5803–5820, 08 2013.
- [18] François Chollet. *Deep Learning with Python*. Manning, second edition, 2021.
- [19] H. Cokrokusumo, Ika Hariyati, Lukmanda Lubis, Prawito Prajitno, and Djarwani Soejoko. Noise suppression of computed tomography (ct) images using residual encoder-decoder convolutional neural network (red-cnn). *Atom Indonesia*, 48:171, 11 2022.
- [20] Jordi de la Torre. Redes generativas adversarias (gan) fundamentos teóricos y aplicaciones, 02 2023.
- [21] Gobierno de México. Protección radiológica, s.f. Accedido: 19 de enero de 2025.
- [22] Manoj Diwakar and Manoj Kumar. A review on ct image noise and its denoising. *Biomed. Signal Process. Control.*, 42:73–88, 2018.
- [23] Manoj Diwakar, Prabhishkek Singh, Girija Rani Karetla, Preeti Narooka, Arvind Yadav, Rajesh Kumar Maurya, Reena Gupta, José Luis Arias-González, Mukund Pratap Singh, Dasharathraj K. Shetty, Rahul Paul, and Nithesh Naik. Low-dose covid-19 ct image denoising using batch normalization and convolution neural network. *Electronics*, 11(20), 2022.
- [24] Shiv Ram Dubey, Satish Kumar Singh, and Bidyut Baran Chaudhuri. Activation functions in deep learning: A comprehensive survey and benchmark, 2022.
- [25] Fenglei Fan, Hongming Shan, Mannudeep Kalra, Ramandeep Singh, Guhan Qian, Matthew Getzin, Yueyang Teng, Juergen Hahn, and Ge Wang. Quadratic autoencoder (q-ae) for low-dose ct denoising. *IEEE Transactions on Medical Imaging*, PP:1–1, 12 2019.
- [26] R.C. Gonzalez and R.E. Woods. *Digital Image Processing, Global Edition*. Pearson Education, 2018.
- [27] Ayaan Haque. Ec-gan: Low-sample classification using semi-supervised algorithms and gans, 2021.
- [28] J. Hornegger, A. Maier, and M. Kowarschik. Ct image reconstruction basics, 2018. Accessed: 2021-02-15.
- [29] Zhizhong Huang, Junping Zhang, Yi Zhang, and Hongming Shan. Du-gan: Generative adversarial networks with dual-domain u-net-based discriminators for low-dose ct denoising. *IEEE Transactions on Instrumentation and Measurement*, 71:1–12, 2022.
- [30] John H. Hubbell. Photon cross sections, attenuation coefficients, and energy absorption coefficients from 10 kev to 100 gev. pages 20–40, 1969.
- [31] P. Jagalingam and Arkal Vittal Hegde. A review of quality metrics for fused image. *Aquatic Procedia*, 4:133–142, 2015. INTERNATIONAL CONFERENCE ON WATER RESOURCES, COASTAL AND OCEAN ENGINEERING (ICWRCOE'15).
- [32] Ruinan Jin, Xiao Li, Yaoliang Yu, and Baoxiang Wang. A comprehensive framework for analyzing the convergence of adam: Bridging the gap with sgd, 2025.

- [33] Daniel E. Jonas, Daniel S. Reuland, Shivani M. Reddy, Max Nagle, Stephen D. Clark, Rachel Palmieri Weber, Chineme Enyioha, Teri L. Malo, Alison T. Brenner, Charli Armstrong, Manny Coker-Schwimmer, Jennifer Cook Middleton, Christiane Voisin, and Russell P. Harris. Screening for lung cancer with low-dose computed tomography: Updated evidence report and systematic review for the us preventive services task force. *JAMA*, 325(10):971–987, 03 2021.
- [34] Haijo Jung. Basic physical principles and clinical applications of computed tomography. *Progress in Medical Physics*, 32:1–17, 03 2021.
- [35] N. Kanopoulos, N. Vasanthavada, and R.L. Baker. Design of an image edge detection filter using the sobel operator. *IEEE Journal of Solid-State Circuits*, 23(2):358–367, 1988.
- [36] Saneera Kulathilake, Nor Abdullah, Aznul Sabri, and Khin Wee Lai. A review on deep learning approaches for low-dose computed tomography restoration. *Complex Intelligent Systems*, 9, 05 2021.
- [37] H. L. Lancaster, M. A. Heuvelmans, and M. Oudkerk. Low-dose computed tomography lung cancer screening: Clinical evidence and implementation research. *Journal of Internal Medicine*, 292(1):68–80, 2022.
- [38] Tengfei Liang, Yi Jin, Yidong Li, Tao Wang, Songhe Feng, and Congyan Lang. Edcnn: Edge enhancement-based densely connected network with compound loss for low-dose ct denoising. *2020 15th IEEE International Conference on Signal Processing (ICSP)*, 1:193–198, 2020.
- [39] MathWorks. Autoencoder - matlab & simulink, 2025. Accessed: 2025-04-04.
- [40] MathWorks. Convolutional neural network, 2025. Accedido el 27 de febrero de 2025.
- [41] MathWorks. Image quality metrics, n.d. Accessed: 2025-01-18.
- [42] Mayo Clinic. Tomografía computarizada (ct scan), 2025. Accedido el 27 de febrero de 2025.
- [43] Cynthia H. McCollough, Michael R. Bruesewitz, and James M. Kofler. Ct dose reduction and dose management tools: overview of available options. *Radiographics : a review publication of the Radiological Society of North America, Inc*, 26 2:503–12, 2006.
- [44] Mizuho Nishio, Chihiro Nagashima, Saori Hirabayashi, Akinori Ohnishi, Kaori Sasaki, Tomoyuki Sagawa, Masayuki Hamada, and Tatsuo Yamashita. Convolutional auto-encoder for image denoising of ultra-low-dose ct. *Heliyon*, 3(8):e00393, 2017.
- [45] Augustus Odena. Semi-supervised learning with generative adversarial networks, 2016.
- [46] American Association of Physicists in Medicine. Acr–aapm practice parameter for adult routine chest ct protocols. Technical report, American College of Radiology / AAPM, 2019. Accedido en agosto 2025.
- [47] American Association of Physicists in Medicine (AAPM). Low dose ct grand challenge, s.f. Accedido: 19 de enero de 2025.
- [48] American College of Radiology. Frequently asked questions: Lung cancer screening with low-dose ct (ldct), n.d.
- [49] Jordi Pons, Santiago Pascual, Giulio Cengarle, and Joan Serrà. Upsampling artifacts in neural audio synthesis. 2021.
- [50] J. C. Ramírez Giraldo, J. J. Fletcher, and C. H. McCollough. Reducción del ruido en imágenes de tomografía computarizada usando un filtro bilateral anisotrópico - noise reduction in computed tomography images using an anisotropic bilateral filter. *Revista Ingeniería Biomédica*, 4(7):62–68, 2011.

- [51] Rigaku Corporation. How does ct reconstruction work? Accessed: 2025-01-18.
- [52] R. T. Sadia, J. Chen, and J. Zhang. Ct image denoising methods for image quality improvement and radiation dose reduction. *Journal of Applied Clinical Medical Physics*, 25(2):e14270, 2024.
- [53] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans, 2016.
- [54] Nadine Barrie Smith and Andrew Webb. *Introduction to Medical Imaging: Physics, Engineering and Clinical Applications*. Cambridge University Press, Cambridge, 2010.
- [55] Jost Tobias Springenberg. Unsupervised and semi-supervised learning with categorical generative adversarial networks, 2016.
- [56] Juan Terven, Diana M. Cordova-Esparza, Alfonso Ramirez-Pedraza, Edgar A. Chavez-Urbiola, and Julio A. Romero-Gonzalez. Loss functions and metrics in deep learning, 2024.
- [57] Timo. Mantenimiento de un tomógrafo computarizado, 2013. Accessed: 2025-01-18.
- [58] Jamal Toutouh, Subhash Nalluru, Erik Hemberg, and Una-May O’Reilly. Semi-supervised generative adversarial networks with spatial coevolution for enhanced image generation and classification. *Applied Soft Computing*, 148:110890, 2023.
- [59] Jos van de Wolfshaar. Semi-supervised learning with gans, 2025. Accessed: 2025-04-04.
- [60] C. Versloot. How to use upsampling2d with keras: A simple explanation. <https://github.com/christianversloot/machine-learning-articles/blob/main/upsampling2d-how-to-use-upsampling-with-keras.md>, n.d. Accedido el 5 de junio de 2020.
- [61] Robert Villegas Flores. Tomografía computada, 2018. Accessed: 2025-01-18.
- [62] Dayang Wang, Fenglei Fan, Zhan Wu, Rui Liu, Fei Wang, and Hengyong Yu. Ctformer: convolution-free token2token dilated vision transformer for low-dose ct denoising. *Physics in Medicine amp; Biology*, 68(6):065012, March 2023.
- [63] Dayang Wang, Yongshun Xu, Shuo Han, Zhan Wu, Li Zhou, Bahareh Morovati, and Hengyong Yu. Lomae: Low-level vision masked autoencoders for low-dose ct denoising. 2023.
- [64] Guan Wang and Xueli Hu. Low-dose ct denoising using a progressive wasserstein generative adversarial network. *Computers in Biology and Medicine*, 135:104625, 2021.
- [65] Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. *CoRR*, abs/2106.03106, 2021.
- [66] Steve Webb. *Webb’s Physics of Medical Imaging*. CRC Press, second edition, 2003.
- [67] Wikimedia Commons. 2d convolution animation, 2023. Accedido el 27 de febrero de 2025.
- [68] Wikipedia. Autocodificador, s.f. Accedido: 19 de enero de 2025.
- [69] Wikipedia. Sobel operator, s.f. Accedido: 19 de enero de 2025.
- [70] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Manudeep K. Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, 37(6):1348–1357, 2018.

- [71] Sergio E. Ybargollin-Machado, Marlen Pérez-Díaz, Rubén Orozco-Morales, and Reinaldo Roque-Díaz. Reducción de ruido en imágenes de spect utilizando métodos de variación total. *Brazilian Journal of Experimental Design, Data Analysis and Inferential Statistics*, 2(1):60–71, 2022.
- [72] Zhixian Yin, Kewen Xia, Ziping He, Jiangnan Zhang, Sijie Wang, and Baokai Zu. Unpaired image denoising via wasserstein gan in low-dose ct image with multi-perceptual loss and fidelity loss. *Symmetry*, 13(1), 2021.
- [73] Afia Zafar, Muhammad Aamir, Nazri Mohd Nawi, Ali Arshad, Saman Riaz, Abdulrahman Alruban, Ashit Kumar Dutta, and Sultan Almotairi. A comparison of pooling methods for convolutional neural networks. *Applied Sciences*, 12(17), 2022.
- [74] Ju Zhang, Weiwei Gong, Lieli Ye, Fanghong Wang, Zhibo Shangguan, and Yun Cheng. A review of deep learning methods for denoising of medical low-dose ct images. *Computers in biology and medicine*, 171:108112, 2024.
- [75] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.