



SÍNTESIS DE NEUROIMÁGENES DIAGNÓSTICAS MEDIANTE EL USO DE UNA
ESTRATEGIA DE APRENDIZAJE PROFUNDO GENERATIVO

OSCAR ANDRÉS RAMÍREZ SERNA
YEFERSON VALENCIA VANEGAS

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2024

SÍNTESIS DE NEUROIMÁGENES DIAGNÓSTICAS MEDIANTE EL USO DE UNA
ESTRATEGIA DE APRENDIZAJE PROFUNDO GENERATIVO

OSCAR ANDRÉS RAMÍREZ SERNA
YEFERSON VALENCIA VANEGAS

Tesis presentada en cumplimiento de los requisitos para optar por el título de:
Ingeniero de Sistemas

Director:

Fabio Martínez Carrillo, Ph.D.

Codirector:

Gustavo Adolfo Garzón Villamizar, MSc.

Codirector:

Santiago Gómez Hernández, MSc.

Codirector:

Franklin Samuel Sierra Jerez, MSc.

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2024

DEDICATORIA

A mis padres, Oscar y Elizabeth, por ser mi mayor guía e inspiración. Por enseñarme que no hay sueños demasiado grandes y por brindarme su apoyo incondicional en cada etapa del camino.

A mis abuelos, Víctor y Yolanda, cuya sabiduría y amor han sido el pilar fundamental en mi vida. Por enseñarme el valor del esfuerzo y recordarme siempre la importancia de nunca rendirse.

Oscar

A mis padres, Hilda y José, quienes siempre me han apoyado y ayudado; gracias por su sabiduría incondicional, que me ha guiado hasta donde estoy y me permite soñar con el lugar a donde quiero llegar. Les agradezco profundamente por su cariño incondicional.

Yeferson

AGRADECIMIENTOS

A mi director, el profesor Fabio, y a mis codirectores, Santiago, Franklin y Gustavo, mi más sincero agradecimiento por su valiosa guía y dedicación en este apasionante camino de la ciencia. Sus enseñanzas han dejado una huella importante en mi desarrollo personal y profesional.

A mis compañeros del grupo de investigación BIVL²ab por su apoyo y colaboración. Cada idea, conversación y proyecto que hemos compartido han hecho de este camino una experiencia memorable.

A mi familia, por ser mi fuente constante de amor y confianza. Gracias por ser mi inspiración, por creer en mí incluso en los momentos en que yo mismo dudaba, y por ser el faro de guía a lo largo de mi vida.

A mis amigos, en especial a Laura, Santiago, Juan David y Daniel, por estar a mi lado y ser un refugio en el camino. Por brindarme su apoyo en todo momento y compartir cada pequeño logro en este trayecto.

Oscar

Quiero expresar mi más profundo agradecimiento a todas las personas que me han acompañado y apoyado a lo largo de mi proceso formativo. Agradezco especialmente a mi Director Fabio Martínez, por su guía y dedicación incansable, y a mis codirectores Franklin , Santiago y Gustavo, cuya sabiduría y consejos fueron fundamentales para alcanzar este logro. Cada uno de ellos ha aportado de manera significativa, enriqueciendo este proceso con su conocimiento y compromiso.

Mi gratitud se extiende también al grupo de investigación BIVL²ab, que me abrió sus puertas y me acogió como parte de su comunidad. En este espacio, no solo adquirí un vasto conocimiento, sino también el calor de la familiaridad y el apoyo de un grupo que se convirtió en una segunda familia. Los momentos que compartí con ellos han dejado una huella imborrable en mi vida, y su compañía ha sido una fuente constante de inspiración y motivación.

Yeferson

CONTENIDO

	pág.
INTRODUCCIÓN	13
1. FUNDAMENTOS Y TRABAJOS PREVIOS	17
1.1. Neuroimágenes diagnósticas en el ACV isquémico	17
1.1.1 Tomografía computarizada sin contraste (NCCT).	18
1.1.2 Resonancia magnética ponderada por difusión.	19
1.1.3 FLAIR.	19
1.2. Neuroimágenes diagnósticas en la EP	20
1.2.1 Resonancia magnética (RM).	22
1.2.2 Tomografía computarizada de emisión monofotónica (SPECT).	23
1.3. Estrategias para la traducción entre secuencias diagnósticas	24
1.3.1 Arquitecturas codificador-decodificador.	24
1.3.2 Redes Generativas Adversarias.	26
1.4. Trabajos previos	28
2. PROBLEMA DE INVESTIGACIÓN	34
3. OBJETIVOS	36
3.1. Objetivo general	36
3.2. Objetivos específicos	36
4. MÉTODO PROPUESTO	37
4.1. Difusión progresiva Gaussiana	38
4.2. Traducción de neuroimágenes mediante Difusión Inversa Condicional	39
4.2.1 Enmascarado aleatorio de la imagen condición	41

4.2.2	Arquitectura U-Net	41
4.3.	Síntesis de secuencias diagnósticas para ACV isquémico	42
4.4.	Síntesis de secuencias diagnósticas de la EP	43
5.	DISEÑO EXPERIMENTAL	45
5.1.	Conjuntos de datos	45
5.2.	Configuración de la arquitectura desarrollada	48
5.3.	Validación	49
6.	EVALUACIÓN Y RESULTADOS	53
6.1.	Síntesis de secuencias DWI y mapas de ADC para el ACV isquémico	53
6.2.	Síntesis de secuencias SPECT para la EP	59
7.	CONCLUSIONES Y TRABAJO FUTURO	66
	BIBLIOGRAFÍA	69

LISTA DE FIGURAS

	pág.
Figura 1. Estudios imagenológicos en el ACV isquémico	18
Figura 2. Estudios de neuroimagen en pacientes control y la EP	22
Figura 3. Arquitectura codificador-decodificador	25
Figura 4. Arquitectura U-Net	25
Figura 5. Arquitectura GAN	26
Figura 6. Arquitectura Pix2Pix	27
Figura 7. Arquitectura CycleGAN	28
Figura 8. Esquema de la estrategia implementada.	37
Figura 9. Esquema de un modelo de difusión	38
Figura 10. Esquema de un modelo de difusión condicional	40
Figura 11. Configuración experimental en ACV.	43
Figura 12. Configuración experimental en la EP.	44
Figura 13. Resultados cualitativos de la síntesis de DWI a partir de NCCT.	54
Figura 14. Resultados cualitativos de la síntesis de ADC a partir de NCCT.	55
Figura 15. Comparación de los valores obtenidos para la métrica del SSIM en las imágenes sintéticas de DWI y ADC según los criterios de visibilidad de la lesión en NCCT, el tamaño de la lesión y la fase del ACV.	57
Figura 16. Resultado cualitativos de estudios de neuroimagen en pacientes control y la EP	61

LISTA DE TABLAS

	pág.
Tabla 1. Información demográfica para el conjunto de datos para el ACV isquémico. .	46
Tabla 2. Características demográficas y clínicas de los pacientes con Enfermedad de Parkinson (EP) y controles, en los conjuntos de entrenamiento y prueba.	47
Tabla 3. Evaluación de las secuencias DWI y mapas ADC sintéticos generados a partir de distintas arquitecturas generativas.	56
Tabla 4. Estudio de ablación sobre el conjunto de validación en la tarea de clasificación Parkinson/Control.	58
Tabla 5. Métricas de similitud sobre la generación de muestras SPECT sintéticas generadas a partir de distintas arquitecturas generativas.	62
Tabla 6. Métricas de evaluación estructural y relación señal/ruido de secuencias SPECT generadas a partir de RM-T1.	63
Tabla 7. Resultados de clasificación en estudios de SPECT generados a partir de RM-T1 en sujetos control y pacientes con EP	64

RESUMEN

TÍTULO: Síntesis de neuroimágenes diagnósticas mediante el uso de una estrategia de aprendizaje profundo generativo *

AUTORES: Oscar Andrés Ramírez Serna, Yeferson Valencia Vanegas. **

PALABRAS CLAVE: Enfermedades cerebrovasculares (ACV), enfermedades neurodegenerativas, síntesis de neuroimágenes, aprendizaje generativo, modelos de difusión.

DESCRIPCIÓN: Las neuroimágenes son esenciales para la caracterización de alteraciones estructurales y funcionales de enfermedades neurológicas. Para la asistencia y caracterización efectiva de hallazgos radiológicos, en algunas enfermedades, es imprescindible el uso complementario de imágenes, permitiendo asociar patrones estructurales con funcionales o brindando una mayor definición y seguimiento de las afecciones detectadas. Sin embargo, en la rutina clínica, muchas veces es desafiante contar con todas las secuencias, debido a contraindicaciones de pacientes o la disponibilidad de los dispositivos de captura. Particularmente, en este trabajo abordaremos este problema para el accidente cerebrovascular (ACV) y la enfermedad del Parkinson (EP), teniendo en cuenta el carácter complementario requerido para la caracterización efectiva de estas enfermedades. En cuanto al ACV, primero se hace un tamizaje para determinar el origen de la interrupción del flujo sanguíneo con tomografías (NCCT) y luego se cuantifica el volumen de la lesión desde secuencias de resonancia ponderada por difusión (DWI, ADC). En cuanto a la EP las resonancias T1 permiten hacer un análisis estructural del cerebro, mientras que el SPECT complementa este análisis brindando información sobre la deficiencia de neurotransmisores dopaminérgicos. En este trabajo se implementó una estrategia de aprendizaje de difusión condicional para la síntesis de neuroimágenes complementarias tanto en el dominio del ACV como en el dominio de la EP. En cuanto al ACV, la estrategia fue entrenada y validada para realizar síntesis de resonancias DWI y ADC desde imágenes NCCT, utilizando un conjunto propio de 95 pacientes y logrando puntajes de reconstrucción SSIM de 0.55 para DWI y 0.45 para ADC. En cuanto a la EP, la estrategia fue ajustada para realizar síntesis de secuencias SPECT desde resonancias T1, utilizando un conjunto público de 115 pacientes y logrando un SSIM de 0.98.

* Trabajo de investigación

** Facultad de Ingenierías Físicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez, PhD. Codirectores: Gustavo Adolfo Garzón Villamizar, MSc., Santiago Gómez Hernández, MSc., Franklin Samuel Sierra Jerez, MSc.

ABSTRACT

TITLE: Diagnostic neuroimaging synthesis using a generative deep learning strategy *

AUTHORS: Oscar Andrés Ramírez Serna, Yeferson Valencia Vanegas. **

KEYWORDS: Cerebrovascular diseases, neurodegenerative diseases, neuroimaging synthesis, generative learning, diffusion models.

DESCRIPTION: Neuroimaging is essential for the characterization of structural and functional alterations in neurological diseases. For effective assistance and characterization of radiological findings in some diseases, the complementary use of images is crucial, as it allows the association of structural and functional patterns or provides greater definition and monitoring of detected conditions. However, in clinical practice, it is often challenging to obtain all the necessary sequences due to patient contraindications or the availability of capture devices. Specifically, this work addresses this issue for stroke and Parkinson's disease (PD), considering the complementary nature required for the effective characterization of these diseases. For stroke, screening is first performed to determine the origin of the blood flow interruption using CT scans (NCCT), and then the lesion volume is quantified using diffusion-weighted MRI sequences (DWI, ADC). In the case of PD, T1-weighted MRIs allow for structural brain analysis, while SPECT complements this by providing information on dopaminergic neurotransmitter deficiencies. In this work, a conditional diffusion learning strategy was implemented for the synthesis of complementary neuroimages in both the stroke and PD domains. For stroke, the strategy was trained and validated to synthesize DWI and ADC MRIs from NCCT images, using a proprietary dataset of 95 patients and achieving SSIM reconstruction scores of 0.55 for DWI and 0.45 for ADC. For PD, the strategy was adjusted to synthesize SPECT sequences from T1 MRIs, using a public dataset of 115 patients and achieving an SSIM of 0.98.

* Research work

** Faculty of Physics-Mechanics Engineering. School of Systems Engineering and Informatics. Advisor: Fabio Martínez, PhD. Co-advisors: Gustavo Adolfo Garzón Villamizar, MSc., Santiago Gómez Hernández, MSc., Franklin Samuel Sierra Jerez, MSc.

INTRODUCCIÓN

Los trastornos neurológicos representan la principal causa de discapacidad a nivel mundial, afectando a más del 43% de la población en el año 2021 ¹. Dentro de estas afecciones neurológicas se destacan el accidente cerebrovascular (ACV) y la enfermedad de (EP). Por un lado, el ACV es la afección cerebrovascular más prevalente, con más de 12.2 millones de nuevos casos anuales y representa la segunda causa de muerte más común y la principal causa de discapacidad en adultos a nivel mundial ²³. Por otro lado, la EP se destaca como la enfermedad neurodegenerativa de mayor crecimiento, con un aumento proyectado del 20% en el total de casos para 2040 ⁴.

Los estudios imagenológicos constituyen una de las principales herramientas para caracterizar, soportar el diagnóstico y definir tratamientos de estas enfermedades. En cuanto al ACV, los estudios más utilizados son los de tomografía computarizada (TC) y de resonancia magnética (RM). Específicamente, en el diagnóstico del ACV, la TC sin contraste se considera la técnica de primera línea debido a su rapidez, amplia disponibilidad y capacidad para

-
- ¹ STEINMETZ et al. “Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021”. In: *The Lancet Neurology* (2024).
 - ² FEIGIN et al. “World Stroke Organization (WSO): global stroke fact sheet 2022”. In: *International Journal of Stroke* 17.1 (2022), pp. 18–29.
 - ³ Vijay K. SHARMA. “Cerebrovascular Disease”. In: *International Encyclopedia of Public Health (Second Edition)*. Ed. by Stella R. QUAH. Second Edition. Oxford: Academic Press, 2017, pp. 455–470. DOI: <https://doi.org/10.1016/B978-0-12-803678-5.00058-8>.
 - ⁴ DORSEY et al. “The emerging evidence of the Parkinson pandemic”. In: *Journal of Parkinson's disease* 8.s1 (2018), S3–S8.

descartar hemorragias cerebrales y otras patologías con síntomas similares ⁵. Sin embargo, esta modalidad muestra una sensibilidad limitada para detectar signos tempranos del ACV isquémico ⁶. Por su parte, la RM ofrece detalles precisos sobre las estructuras y regiones cerebrales, demostrando una mayor sensibilidad frente a cambios tempranos en el tejido cerebral relacionados con lesiones del ACV, aunque su uso es menos frecuente debido a su mayor coste, tiempo de adquisición prolongado y restricciones para pacientes con implantes metálicos o dispositivos médicos.

En cuanto a la EP, en el contexto diagnóstico, las secuencias más utilizadas son la RM y la TC de emisión de positrones (SPECT). Específicamente, las RM estructurales se destacan como una alternativa viable para cuantificar cambios volumétricos en las regiones cortical y subcortical, así como las alteraciones en la sustancia negra que están correlacionadas con las alteraciones dopaminérgicas ⁷. Por otro lado, la SPECT se ha establecido como una técnica prometedora que ofrece una mayor sensibilidad en la detección de la EP. Esta técnica permite medir el anabolismo y el transporte de dopamina, facilitando la distinción entre la EP y otras enfermedades neurodegenerativas ⁸. Sin embargo, el uso de SPECT implica ciertos desafíos, como su naturaleza invasiva debido al uso de un agente de contraste especial (*radiotracer*), y la consecuente exposición de los pacientes a la radiación. Además, la interpretación de los hallazgos radiológicos de la EP a partir de una única modalidad puede resultar subjetiva, lo

⁵ WINTERMARK et al. “Imaging recommendations for acute stroke and transient ischemic attack patients: a joint statement by the American Society of Neuroradiology, the American College of Radiology, and the Society of NeuroInterventional Surgery”. In: *American Journal of Neuroradiology* 34.11 (2013), E117–E127.

⁶ Michelle P LIN and David S LIEBESKIND. “Imaging of ischemic stroke”. In: *Continuum: Lifelong Learning in Neurology* 22.5 (2016), p. 1399.

⁷ S SIVARANJINI and CM SUJATHA. “Deep learning based diagnosis of Parkinson’s disease using convolutional neural network”. In: *Multimedia tools and applications* 79.21 (2020), pp. 15467–15479.

⁸ Hajer KHACHNAOUI; Rostom MABROUK, and Nawres KHLIFA. “Machine learning and deep learning for clinical data and PET/SPECT imaging in Parkinson’s disease: a review”. In: *IET Image Processing* 14.16 (2020), pp. 4013–4026.

que subraya la necesidad de un análisis complementario para una comprensión más profunda de la enfermedad.

En este contexto, tanto para el ACV como para la EP, el uso de imágenes de múltiples modalidades es crucial, puesto que cada una proporciona información complementaria esencial para un diagnóstico efectivo. Sin embargo, la disponibilidad de estas secuencias se ve limitada en la práctica clínica debido a factores como el costo, la accesibilidad y el tiempo necesario para adquirir cada una de las secuencias ⁶. Computacionalmente, se han desarrollado diversas estrategias de aprendizaje profundo generativo para sintetizar neuroimágenes faltantes, con el fin de modelar el mapeo entre una modalidad de origen y una modalidad objetivo. Estas técnicas incluyen las arquitecturas de codificador-decodificador, las redes adversarias generativas (GAN) y, más recientemente, los modelos basados en difusión. A pesar de sus avances, estas estrategias presentan desafíos significativos en el aprendizaje del mapeo entre modalidades, ya que cada modalidad captura características anatómicas y funcionales distintas. Además, la relación entre modalidades implica una complejidad no lineal, lo que puede resultar en la pérdida de patrones de intensidad y de información crítica relacionada con la enfermedad en estudio ⁹.

En este trabajo se implementó una estrategia de aprendizaje profundo generativo para la síntesis de secuencias diagnósticas complementarias, demostrando su eficacia tanto en el contexto del ACV isquémico como en el de la EP. La estrategia consiste en un modelo de difusión condicional que utiliza las imágenes del dominio de origen como condición para generar imágenes sintéticas del dominio objetivo. Además, se emplea un mecanismo de enmascarado aleatorio durante el proceso generativo para mejorar la robustez del modelo ante las características distintivas de cada modalidad de neuroimagen. La estrategia desarrollada se implementó en los siguientes escenarios:

- Para el ACV, en cuanto a la generación de imágenes sintéticas de RM ponderada por

⁹ YANG et al. “MRI cross-modality image-to-image translation”. In: *Scientific reports* 10.1 (2020), p. 3753.

difusión (DWI y ADC) a partir de secuencias NCCT. Para ello se utilizó un conjunto de datos de 95 pacientes diagnosticados con la enfermedad y que contaban con las secuencias capturadas en una ventana temporal no mayor a 24 horas.

- Para la EP, se codificó la estrategia para generar imágenes SPECT a partir de RM de tipo T1. Para ello, se consideró un conjunto de datos de 195 pacientes, estratificados en diferentes estadios motores de la enfermedad.

1. FUNDAMENTOS Y TRABAJOS PREVIOS

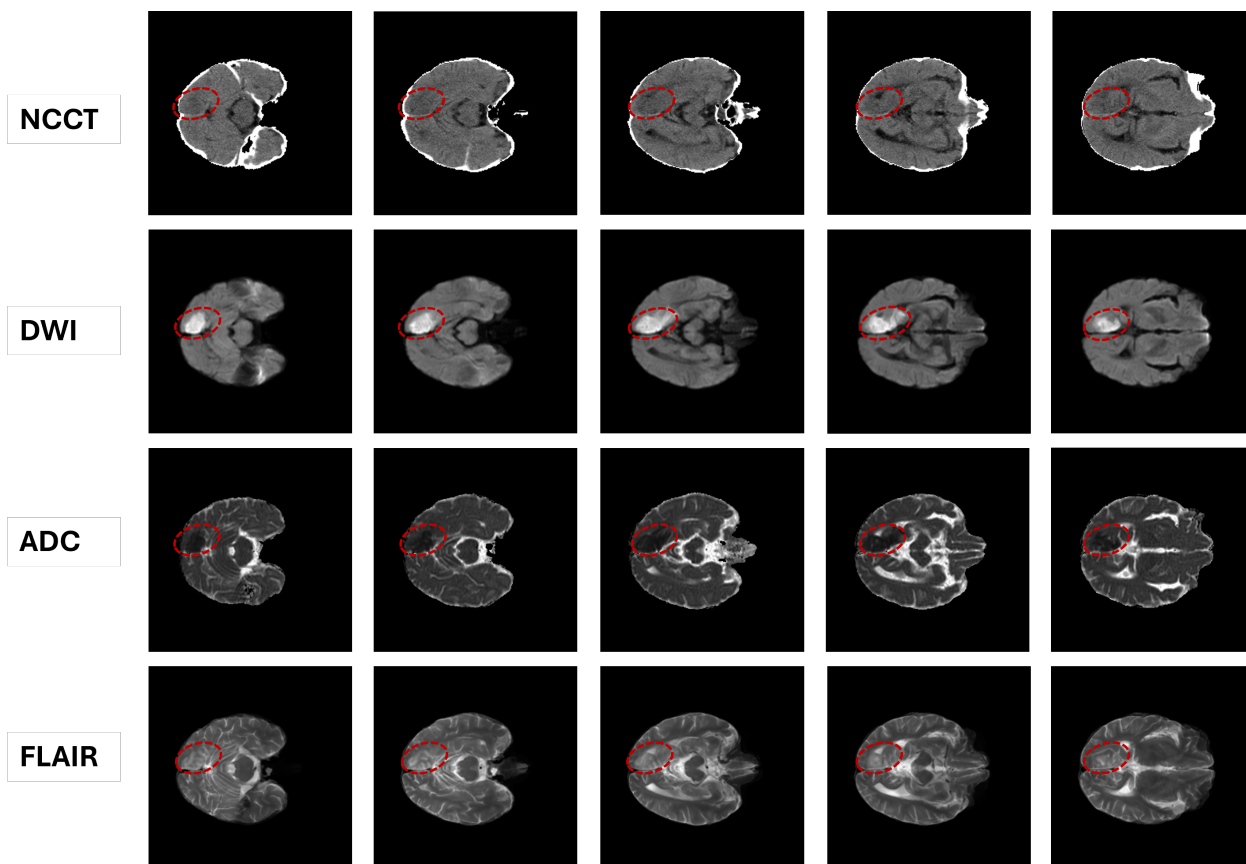
1.1. Neuroimágenes diagnósticas en el ACV isquémico

El accidente cerebrovascular isquémico constituye aproximadamente el 80% del total de casos de ACV a nivel mundial. Este tipo de ACV se produce por la disminución o interrupción del flujo sanguíneo hacia el cerebro debido a la obstrucción o estrechamiento de los vasos sanguíneos, lo cual compromete el suministro de oxígeno y nutrientes a las células cerebrales. En las áreas afectadas por una restricción severa del flujo sanguíneo, la necrosis celular sucede en cuestión de minutos, generando daños irreversibles en el parénquima cerebral ¹⁰. Por consiguiente, el diagnóstico rápido y preciso es crítico en la caracterización de la lesión isquémica y en la mitigación de los daños cerebrales a largo plazo.

Los estudios imagenológicos son el pilar fundamental para realizar una caracterización morfológica de las lesiones de ACV isquémico, ya que permiten localizar la lesión, determinar la extensión del daño tisular y diferenciar entre las áreas de núcleo isquémico (tejido afectado irreversiblemente) y penumbra (tejido potencialmente salvable) ⁵. Entre las técnicas de neuroimagenología principalmente empleadas en el diagnóstico del ACV isquémico se encuentran la tomografía computarizada (TC) y la resonancia magnética (RM). En cuanto a la TC, esta secuencia imagenológica ofrece una mayor disponibilidad, rapidez y sensibilidad en la detección de hemorragias cerebrales. En el caso de la RM, esta ofrece información detallada de las estructuras cerebrales y de los cambios en el tejido asociados a la lesión. La selección de una técnica de neuroimagenología depende principalmente de factores asociados con los recursos disponibles en los centros médicos y del tiempo asociado al protocolo de atención de los pacientes ⁵. Estas modalidades de neuroimagen, comúnmente utilizadas para el diagnóstico del ACV isquémico, se describen con mayor detalle a continuación.

¹⁰ GONZÁLEZ et al. *Acute ischemic stroke*. Springer, 2011.

Figura 1. Estudios de NCCT (A), DWI (B), ADC (C) y FLAIR (D) de un mismo paciente con un ACV isquémico. En A se observa una leve hipoatenuación en la zona de la lesión, mientras que en B y en C se observa un alto contraste entre la lesión y el tejido cerebral sano. Por otro lado, en D no es visible la lesión debido a que el tiempo desde la presentación de los síntomas es menor a 6 horas.



1.1.1. Tomografía computarizada sin contraste (NCCT). La TC sin contraste constituye el primer estudio realizado para la evaluación de un paciente con síntomas de ACV debido a su amplia disponibilidad y a su corto tiempo de adquisición. Su función principal es la detección de hemorragias cerebrales, con lo cual es posible diferenciar un ACV isquémico de uno hemorrágico. De igual forma, provee información de los signos tempranos de isquemia. Sin embargo, estos signos son sutiles, por lo que presentan una sensibilidad limitada en la ventana hiperaguda (primeras 6 horas desde el inicio de los síntomas del ACV)

⁶. Un ejemplo de esta secuencia se ilustra en la Figura 1A.

1.1.2. Resonancia magnética ponderada por difusión. Las secuencias de RM ponderadas por difusión (DWI, por sus siglas en inglés) aportan información única sobre el estado del tejido cerebral, ofreciendo un mayor contraste entre el tejido sano e hipoperfundido, el cual es dependiente del movimiento molecular del agua en el cerebro ⁶. Las DWI son más sensibles en la detección de cambios isquémicos en comparación al estudio de NCCT, reportando una sensibilidad del 99% y una especificidad del 92% ⁵. Por otra parte, los mapas del coeficiente aparente de difusión (ADC), obtenidos a partir de las secuencias de DWI, representan la magnitud de la difusión en el tejido cerebral y ayudan a identificar las lesiones isquémicas agudas, solventando problemas típicos del DWI como el *T2-shine through*, que es causado por señales altas en imágenes de T2 que se manifiestan en los estudios de DWI, pero no están relacionadas con la restricción de la difusión. En la Figura 1B y 1C se muestran estas secuencias.

1.1.3. FLAIR. De forma complementaria al DWI y ADC, la imagenología por recuperación de la inversión atenuada del fluido (FLAIR) permite realizar un diagnóstico más preciso del ACV isquémico al aumentar la visibilidad de la lesión y resaltar características distintivas en las imágenes, tal como se muestra en la Figura 1D. Esta técnica, basada en la supresión de la señal del líquido cefalorraquídeo, genera imágenes ponderadas en T2 que atenúan la intensidad de la señal proveniente del agua agrupada en masa ¹¹. Junto con las imágenes de DWI, estas secuencias desempeñan un rol esencial en la identificación de lesiones isquémicas en la ventana trombolítica del ACV.

¹¹ MAKKAT et al. "Signs of acute stroke seen on fluid-attenuated inversion recovery MR imaging". In: *American Journal of Roentgenology* 179.1 (2002), pp. 237–243.

1.2. Neuroimágenes diagnósticas en la EP

La enfermedad de Parkinson (EP) es la segunda enfermedad neurodegenerativa más común a nivel mundial. Esta enfermedad no tiene cura y su desarrollo se encuentra íntimamente ligado al daño progresivo de las neuronas dopaminérgicas (encargadas de la coordinación del movimiento) presentes en la sustancia negra (SN). Esta pérdida progresiva resulta evidente en los síntomas motores típicamente asociados a la enfermedad, como temblores, inestabilidad postural, ralentización de la marcha (bradicinesia) y, en general, la afectación motora de quien la padece ¹².

El diagnóstico clínico de la EP se basa principalmente en la valoración netamente visual de los síntomas motores. Durante dicha valoración, el experto puntúa el nivel funcional del paciente junto con el historial clínico para determinar el grado de severidad de la enfermedad. Estas puntuaciones vienen dadas por escalas internacionales de referencia que asignan valores numéricos según sea el grado de funcionalidad comprometido durante la ejecución de actividades que involucran la coordinación de los movimientos. Una de estas escalas es la Hoehn and Yahr (H&Y), que dada su sencillez y practicidad, es una de las escalas de mayor uso dentro de la rutina clínica. Esta escala estima la severidad de la enfermedad en cinco estadios. Para el estadio uno, los pacientes no experimentan una disfunción significativa en la ejecución de actividades diarias, pero los síntomas se manifiestan de manera unilateral. Para el estadio dos, los síntomas se presentan bilateralmente sin llegar a limitar la independencia funcional del paciente. Un estadio nivel tres implica para el paciente un nivel marcado de inhabilidad en la ejecución de las actividades, puesto que los problemas de estabilidad y rigidez empiezan a ser evidentes. Durante el estadio cuatro, los síntomas se recrudecen, llegando al punto de requerirse una asistencia constante en las actividades de la vida diaria. Por último, en el nivel más severo de la enfermedad, el paciente no solo presenta el mayor

¹² AYAZ et al. “Automated methods for diagnosis of Parkinson’s disease and predicting severity level”. In: *Neural Computing and Applications* 35.20 (2023), pp. 14499–14534.

grado disfuncional a nivel motor (uso de silla de ruedas o postración en la cama), sino que también el declive cognitivo es evidente, requiriendo una asistencia de tiempo completo. A pesar de estos avances, esta valoración netamente visual de los síntomas es altamente subjetiva incluso entre expertos radiólogos ¹³. Por otra parte, dada la naturaleza multilateral de la enfermedad, no existe un patrón motriz definido, dificultando su diferenciación entre estadios e incluso frente a otras patologías motoras ¹⁴. Dramáticamente, se ha documentado que la manifestación motora de la enfermedad se da cuando cerca del 80% de las neuronas dopaminérgicas han sufrido un daño ¹². Dado lo anterior, las técnicas de neuroimagen surgen como alternativa al diagnóstico oportuno de la enfermedad, evidenciando patrones asociados al déficit dopaminérgico.

Dentro de estas técnicas de neuroimagen se encuentran la resonancia magnética (RM) y la tomografía computarizada por emisión de fotón único (SPECT). Dentro de la rutina clínica, la RM ha sido empleada como triaje, permitiendo descartar otras enfermedades cerebrales con sintomatología similar a la EP (parkinsonismos) ¹⁵. Por otra parte, la modalidad de SPECT ha sido catalogada como la modalidad de neuroimagen referencia para lograr una diferenciación adecuada entre sujetos control y EP, y dada su alta sensibilidad, ha permitido la categorización incluso entre los estadios de la misma ¹⁶. A continuación se aborda cada una de estas modalidades.

¹³ Eduardo TOLOSA et al. “Challenges in the diagnosis of Parkinson’s disease”. In: *The Lancet Neurology* 20.5 (2021), pp. 385–397.

¹⁴ POEWE et al. “Parkinson disease”. In: *Nature reviews Disease primers* 3.1 (2017), pp. 1–21.

¹⁵ HEIM et al. “Structural imaging in atypical parkinsonism”. In: *International Review of Neurobiology* 142 (2018), pp. 67–148.

¹⁶ BIDESI et al. “The role of neuroimaging in Parkinson’s disease”. In: *Journal of neurochemistry* 159.4 (2021), pp. 660–689.

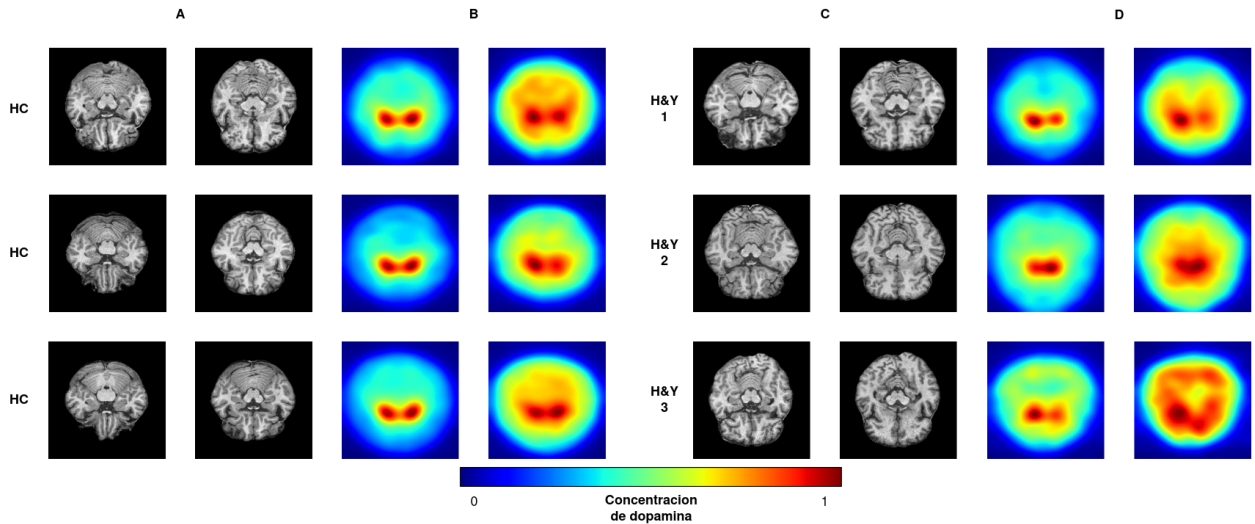


Figura 2. Estudios de resonancia magnética (RM-T1) y tomografía por emisión de fotón único (SPECT) en sujetos control (HC) y pacientes con enfermedad de Parkinson en diferentes estadios de Hoehn y Yahr (H&Y 1-3). Las columnas (A) y (C) muestran las imágenes de RM-T1 usadas como triaje en la rutina clínica para descartar afecciones similares a la EP. Las columnas (B) y (D) corresponden a las imágenes de SPECT, donde se visualiza la concentración de dopamina en el cerebro, evidenciando un patrón asimétrico en la concentración de dopamina en los pacientes con EP.

1.2.1. Resonancia magnética (RM). Esta modalidad de neuroimagen constituye la primera línea de diagnóstico imagenológico de la EP, permitiendo descartar otras patologías que presentan desórdenes de movimiento similares¹⁵. Particularmente, la diferenciación de pacientes EP y poblaciones control se logra mediante la identificación de cambios estructurales y de volumen en el cerebro. Sin embargo, esta valoración resulta evidente en los estadios avanzados de la enfermedad, siendo tediosa y subjetiva en las primeras fases donde no existen

diferencias apreciables frente a poblaciones control ¹⁷¹⁸¹⁹.

1.2.2. Tomografía computarizada de emisión monofotónica (SPECT). La tomografía por emisión monofotónica (SPECT) es una técnica de imagen nuclear que se utiliza para cuantificar la concentración dopaminérgica en el cerebro mediante el uso de un radiofármaco específico a través del estriado cerebral ²⁰. En el contexto de la EP, el SPECT se utiliza para examinar la integridad del sistema dopaminérgico presináptico evidenciado como dos regiones simétricas brillantes con forma de “coma”. El radiotrazador se une selectivamente a los transportadores de dopamina (DAT), relacionándose estrechamente con la presencia de la EP inclusive en sus estadios iniciales ²¹. La figura 2 evidencia este patrón simétrico de personas sanas frente a sujetos afectados por la EP. Aunque el SPECT es una técnica de neuroimagen considerada como referencia para el diagnóstico de la EP, su uso es altamente restringido debido principalmente a los costes de uso asociados, así como el empleo de sustancias radiactivas que pueden ser perjudiciales para ciertos grupos poblacionales. Por último, dada su naturaleza funcional, lograr una relación anatómica directa con la actividad dopaminérgica evidenciada resulta ser una tarea desafiante y subjetiva, inclusive entre grupos

¹⁷ SAEED et al. “Imaging biomarkers in Parkinson’s disease and Parkinsonian syndromes: current and emerging concepts”. In: *Translational neurodegeneration* 6 (2017), pp. 1–25.

¹⁸ SCHULZ et al. “Magnetic resonance imaging–based volumetry differentiates idiopathic Parkinson’s syndrome from multiple system atrophy and progressive supranuclear palsy”. In: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 45.1 (1999), pp. 65–74.

¹⁹ CHOUGAR et al. “The role of magnetic resonance imaging for the diagnosis of atypical parkinsonism”. In: *Frontiers in Neurology* 11 (2020), p. 665.

²⁰ Mehrzad COHENPOUR and Haim GOLAN. “Nuclear neuroimaging of dopamine transporter in Parkinsonism—role in routine clinical practice”. In: *Harefuah* 146.9 (2007), pp. 698–702.

²¹ WANG et al. “SPECT molecular imaging in Parkinson’s disease”. In: *BioMed Research International* 2012 (2012).

de expertos ²²

1.3. Estrategias para la traducción entre secuencias diagnósticas

Las estrategias de aprendizaje profundo generativo se han consolidado como herramientas prometedoras en el ámbito médico debido a su capacidad para estimar secuencias diagnósticas faltantes. Estas técnicas son especialmente útiles en contextos donde las relaciones entre dominios de diferentes secuencias diagnósticas son complejas y exhiben dinámicas no lineales ²³. En las siguientes subsecciones se detallan las arquitecturas de aprendizaje profundo para la traducción entre secuencias diagnósticas.

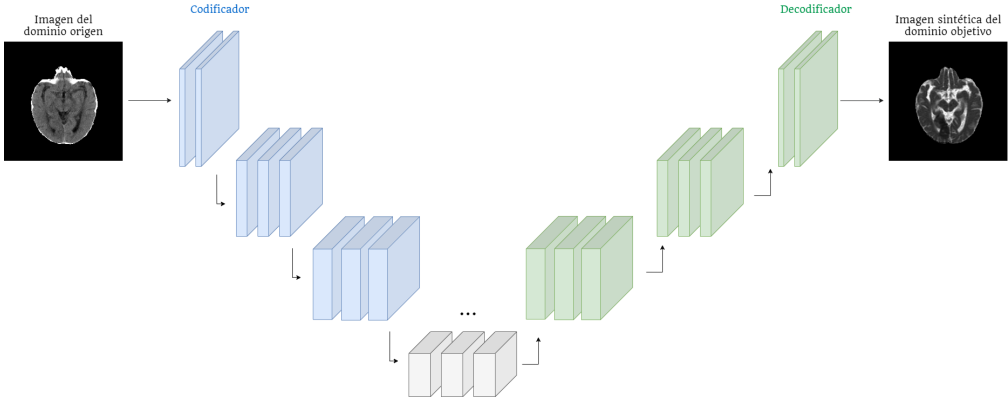
1.3.1. Arquitecturas codificador-decodificador. Estas arquitecturas utilizan un codificador que resume los patrones visuales más relevantes de la entrada en una representación embebida de baja dimensionalidad ²⁴. Luego, el decodificador, específicamente en la tarea de síntesis, restaura la espacialidad que se reduce en el proceso de codificación y realiza una estimación detallada de las intensidades para cada píxel en el dominio de destino. Una representación visual de una arquitectura codificador-decodificador se muestra en la Figura 3.

²² MORTEZAZADEH et al. “Imaging modalities in differential diagnosis of Parkinson’s disease: opportunities and challenges”. In: *Egyptian journal of radiology and nuclear medicine* 52 (2021), pp. 1–12.

²³ NIE et al. “Medical image synthesis with deep convolutional adversarial networks”. In: *IEEE Transactions on Biomedical Engineering* 65.12 (2018), pp. 2720–2730.

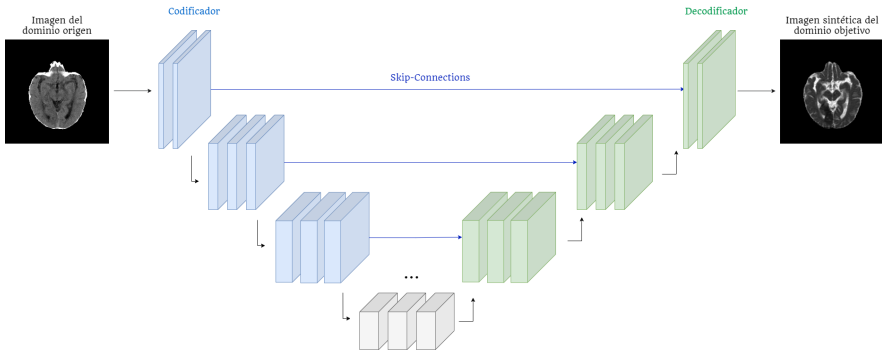
²⁴ YOO et al. “Image-to-image translation using a cross-domain auto-encoder and decoder”. In: *Applied Sciences* 9.22 (2019), p. 4780.

Figura 3. Proceso de codificación y decodificación en un codificador-decodificador. La imagen de origen, en este caso una TC, se procesa a través de una serie de capas de codificación que reducen la dimensionalidad y extraen características de la imagen. El decodificador reconstruye la salida a partir de las características codificadas, incrementando la resolución para generar una imagen sintética de RM.



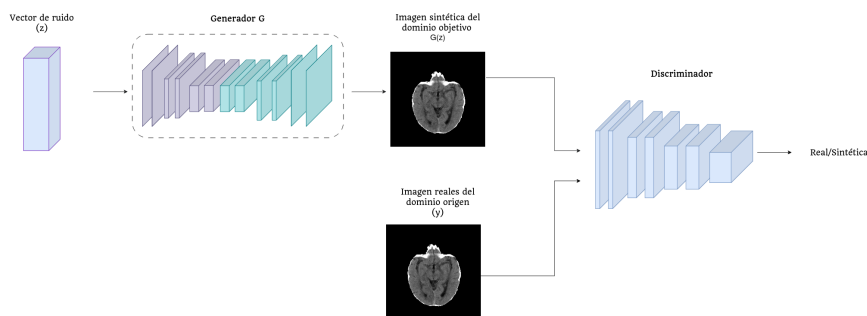
En la actualidad, la red codificador-decodificador más utilizada es la U-Net. Esta red reincorpora información del codificador en el decodificador en distintos niveles de profundidad, por medio de conexiones denominadas *skip-connections* (ver Figura 4). De esta manera, logra recuperar información perdida durante el proceso de codificación y aliviar problemas típicos de las arquitecturas codificador-decodificador como el desvanecimiento del gradiente.

Figura 4. Arquitectura U-Net. La imagen de entrada, en este caso una TC, se procesa a través de una serie de capas de codificación que reducen la dimensionalidad y extraen características de la imagen. El decodificador reconstruye la salida a partir de las características codificadas y de las proporcionadas por las *skip connections*, incrementando la resolución para generar una imagen sintética de RM.



1.3.2. Redes Generativas Adversarias. Estas estrategias de aprendizaje profundo generativo están compuestas por dos redes neuronales, un generador G y un discriminador D , que se entrenan de paralelamente y compiten entre sí para mejorar su desempeño. Por un lado, el generador recibe como entrada un vector de ruido z proveniente de una distribución $p(z)$. Entonces el generador produce muestras sintéticas. Por otro lado, el objetivo del discriminador es recibir las imágenes sintéticas $G(z)$ del generador y las imágenes reales y estimar el realismo de cada conjunto de imágenes. Esta dinámica es fundamental en el proceso de entrenamiento de las GANs, como se muestra en la Figura 5. Simultáneamente, el generador intenta producir imágenes cada vez más similares a las reales, dificultando la tarea del discriminador de distinguir entre datos falsos y reales.

Figura 5. GAN básica: El generador G toma una distribución de ruido z y produce una imagen sintética denotada por $G(z)$ en el dominio Y donde se presentan imágenes de TC. El discriminador D compara estas imágenes sintéticas con imágenes reales para determinar su autenticidad.

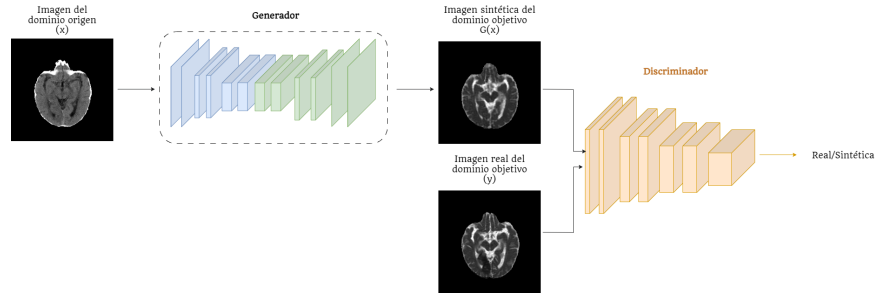


Dentro de la familia de las GANs, existen dos estrategias destacadas para la traducción entre imágenes desde conjuntos de datos con diferentes configuraciones. Las dos estrategias se detallan a continuación.

Pix2Pix. La arquitectura Pix2Pix es una GAN condicional, compuesta por un generador G y un discriminador D , que requiere un conjunto de datos emparejados para realizar la tarea de traducción entre dominios. El generador recibe como entrada una imagen x del dominio de origen y tiene como objetivo sintetizar una imagen $G(x)$ que se asemeje a la imagen correspondiente del dominio objetivo y . Por su parte, el discriminador es entrenado

por medio de una estrategia *min-max* para diferenciar entre la imagen sintética $G(x)$ y la imagen real y . La Figura 6 muestra una representación gráfica de la arquitectura Pix2Pix.

Figura 6. Arquitectura Pix2Pix. En este modelo, una imagen de TC (x) ingresa al generador para sintetizar una imagen de RM (y). Posteriormente, el discriminador recibe la imagen generada ($G(x)$) y la imagen real, con el objetivo de distinguir entre la imagen sintética y la real.



Para el ajuste de esta arquitectura, se emplea una función de pérdida que incluye un componente de reconstrucción (ver ecuación 1, el cual busca maximizar la similitud entre las imágenes sintéticas y las reales. Adicionalmente, un componente adversario (ver ecuación 2) tiene como objetivo que las imágenes generadas sean indistinguibles de las imágenes reales.

$$\mathcal{L}_{\text{rec}}(G) = \mathbb{E}_{x,y}[\|y - G(x)\|_1] \quad (1)$$

$$\mathcal{L}_{\text{adv}}(D, G) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_x[\log(1 - D(x, G(x)))] \quad (2)$$

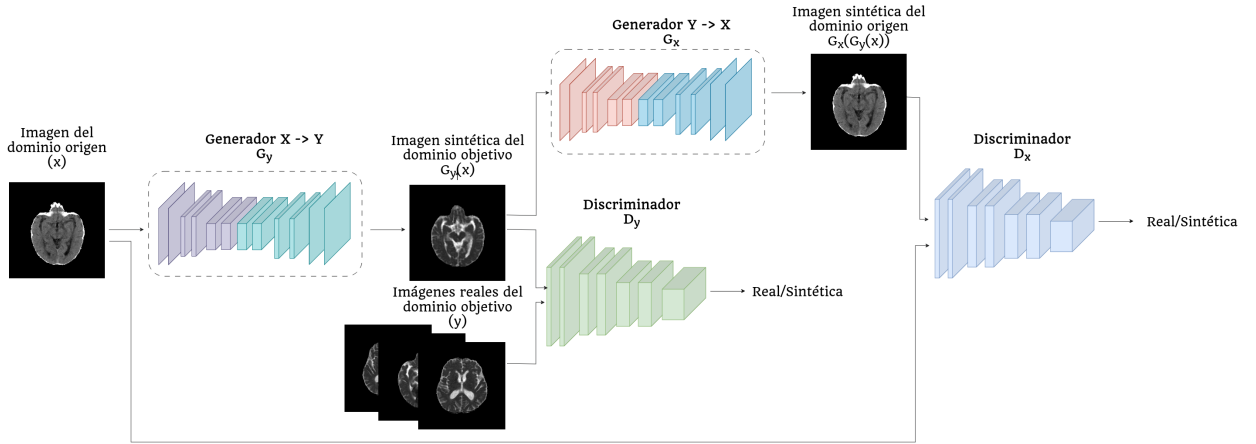
Donde $\mathbb{E}_{x,y}$ representa el valor esperado sobre las distribuciones de las imágenes reales x y y , y \mathbb{E}_x representa el valor esperado sobre las imágenes de entrada x .

CycleGAN Esta arquitectura fue diseñada para la traducción entre imágenes de dominios no emparejados, por medio del uso de dos generadores y dos discriminadores ²⁵. Por un lado, el generador G_y se encarga de realizar la traducción desde el dominio X al dominio Y ,

²⁵ DAYARATHNA et al. “Deep learning based synthesis of MRI, CT and PET: Review and analysis”. In: *Medical Image Analysis* (2023), p. 103046.

mientras que su respectivo discriminador D_y se encarga de diferenciar entre las imágenes generadas $G_y(x)$ y las imágenes reales y . De igual forma, el generador G_x y el discriminador D_x se encargan de realizar el mismo proceso para la traducción inversa. De esta manera, un generador recibe una imagen de un dominio y sintetiza una nueva imagen en otro dominio. Esta imagen resultante se emplea entonces como entrada para otro generador, proceso que se ilustra en la Figura 7.

Figura 7. Arquitectura CycleGAN, en la cual una imagen de TC (x) se transforma en una imagen sintética del dominio objetivo (Y) a través de G_y , la cual es evaluada por D_y . El modelo también incluye un proceso inverso donde las imágenes de RM (y) son transformadas de vuelta al dominio origen (X) mediante G_x y luego evaluadas por D_x .



Esta arquitectura es ajustada por medio de una función de pérdida de consistencia cíclica que fuerza que las imágenes generadas sean lo más similares a las del dominio de origen. En la ecuación 3 se describe la función de pérdida para ajustar esta arquitectura:

$$\mathcal{L}_{cyc}(G_x, G_y) = E_{x \sim P_{data}(x)}[\|G_x(G_y(x)) - x\|_1] + E_{y \sim P_{data}(y)}[\|G_y(G_x(y)) - y\|_1] \quad (3)$$

1.4. Trabajos previos

Las estrategias generativas han suplido falencias relacionadas con la escasa disponibilidad de datos y las anotaciones relacionadas que resultan ser limitadas, sobre todo, dentro del

contexto clínico. Sus aplicaciones están relacionadas con la estimación de modalidades complementarias y la mejora en la calidad de la captura a través de técnicas de súper resolución de las secuencias y la corrección de artefactos, permitiendo maximizar su utilidad clínica y facilitando la adquisición de diversos estudios imagenológicos. Por ejemplo, Wang *et al.* propone una red codificador-decodificador 3D y una GAN de consistencia de características para evaluar la generación de mapas sintéticos del volumen sanguíneo cerebral a partir de múltiples modalidades de RM en pacientes con tumores cerebrales, obteniendo resultados comparables para ambas arquitecturas ²⁶. Por otra parte, Li *et al.* propone la complementariedad de las modalidades empleando arquitecturas U-Net, Pix2Pix y CycleGAN para la síntesis de imágenes de T2 de RM a partir de imágenes de TC del cerebro. Este trabajo compara la capacidad de cada arquitectura al emplear distintas configuraciones para reproducir las neuroimágenes complementarias, destacando la superioridad de la U-Net en la conservación de detalles anatómicos entre los dominios ²⁷. En este mismo sentido, Abu-Srhan *et al.* propone el uso de una GAN no supervisada guiada por atención (uagGAN) para la traducción pareada bidireccional entre estudios de RM y de TC cerebrales empleando un esquema de entrenamiento en cascada ²⁸. La primera etapa proporciona una inicialización efectiva a partir de una configuración de datos emparejados, mientras que la segunda etapa logra una generalización al ajustarse sobre un conjunto de datos no emparejados. La inclusión del mecanismo de atención permite la síntesis de imágenes con un mayor sentido clínico centrado en los hallazgos radiológicos asociados. Por otra parte, Zeng *et al.* propone un modelo CVT-GAN que consiste en una GAN que utiliza un *transformer* de visión convolucional (CVT)

²⁶ WANG et al. “Quantitative Cerebral Blood Volume Image Synthesis from Standard MRI Using Image-to-Image Translation for Brain Tumors”. In: *Radiology* 308.2 (2023), e222471.

²⁷ LI et al. “Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy”. In: *Quantitative imaging in medicine and surgery* 10.6 (2020), p. 1223.

²⁸ ABU-SRHAN et al. “Paired-unpaired Unsupervised Attention Guided GAN with transfer learning for bidirectional brain MR-CT synthesis”. In: *Computers in Biology and Medicine* 136 (2021), p. 104763.

para capturar características espaciales locales y contextos globales en la reconstrucción de imágenes SPECT del cerebro de pacientes con deterioro cognitivo leve superando así las limitaciones de los métodos que dependen exclusivamente de operaciones convolucionales o de atención ²⁹. A pesar de estos resultados prometedores, la preservación de detalles finos propios de la anatomía cerebral se ven comprometidos durante la traducción de cada modalidad. En este sentido, Armanious *et al.* propone una traducción entre secuencias de PET y TC y la corrección de distorsiones asociadas al movimiento, mediante una arquitectura MedGAN que emplea un generador CasNet para mejorar la nitidez de las imágenes sintéticas mediante un refinamiento progresivo a través de bloques codificador-decodificador (*U-blocks*). La adaptación al dominio de interés se logra mediante un discriminador PatchGAN que clasifica parches individuales, permitiendo la captura de la textura y los detalles finos locales propios de las modalidades ³⁰.

Recientemente, Özbey *et al.*, propuso un modelo de difusión adversaria para la síntesis entre imágenes de RM cerebrales, mejorando los aspectos generativos en cuanto a la calidad de la traducción y la convergencia prematura del discriminador en las GANs ³¹. Para lograr esto, se propuso el uso de un módulo difusivo y uno no difusivo, cada uno compuesto por una CycleGAN. En el módulo no difusivo, el modelo se encarga de sintetizar imágenes emparejadas en el dominio objetivo a partir de las imágenes del dominio origen. El módulo difusivo realiza el proceso de difusión inversa, tomando como condición la imagen generada en el módulo no difusivo y refinando al mismo tiempo cada paso en la cadena de Márkov. A continuación se brinda una breve descripción del estado del arte en los dos dominios médicos explorados en

²⁹ ZENG et al. “3D CVT-GAN: a 3D convolutional vision transformer-GAN for PET reconstruction”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pp. 516–526.

³⁰ ARMANIOUS et al. “MedGAN: Medical image translation using GANs”. In: *Computerized medical imaging and graphics* 79 (2020), p. 101684.

³¹ ÖZBEY et al. “Unsupervised medical image translation with adversarial diffusion models”. In: *IEEE Transactions on Medical Imaging* (2023).

este trabajo.

ACV isquémico. Particularmente en el dominio del ACV, Wang *et al.* emplearon una arquitectura U-Net para obtener imágenes de DWI a partir de imágenes de perfusión por TC (CTP) ³². Para esto, un extractor de características obtiene representaciones compactas a partir de secuencias de angiografía por TC. Posteriormente, estas representaciones son concatenadas a las imágenes de CTP e ingresan a la U-Net para generar las imágenes de DWI. También se utiliza una pérdida híbrida que pondera las regiones de la lesión isquémica con el fin de obtener una mayor fidelidad de las imágenes. Sin embargo, la baja calidad de los estudios de CTP impacta de manera negativa la representación de detalles de alta frecuencia, propios de las imágenes de DWI, en las imágenes sintéticas. Por su parte, Garzón *et al.* ³³ propusieron un método basado en una arquitectura codificador-decodificador, compuesta de bloques U-Net, y en una pérdida de consistencia cíclica para sintetizar imágenes de DWI a partir de imágenes de TC, con énfasis en la preservación de las regiones de lesión isquémica y de la estructura cerebral. No obstante, con esta estrategia no se garantiza la preservación de la forma, tamaño y localización de las lesiones en las imágenes sintéticas. Gutiérrez *et al.* incluyeron un proceso de atención guiada y una función de pérdida de consistencia del gradiente en una arquitectura CycleGAN para enfatizar las zonas de lesión isquémica y para preservar la morfología cerebral, respectivamente, en la síntesis de secuencias de FLAIR a partir de secuencias de NCCT. Sin embargo, en el caso de lesiones de gran tamaño o que presentan un bajo contraste con respecto al tejido cerebral, el modelo enfrentó dificultades

³² WANG et al. “Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks”. In: *Medical Image Analysis* 65 (2020), p. 101787.

³³ GARZÓN et al. “A deep CT to MRI unpaired translation that preserve ischemic stroke lesions”. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 2708–2711.

para representar las intensidades características de cada modalidad ³⁴.

Enfermedad de Parkinson. Particularmente, la generación de neuroimágenes fue evaluada por Gajjar *et al.* al emplear diversos esquemas generativos como GANs, autoencoders variacionales y modelos de difusión con el objetivo de generar imágenes de RM de alta calidad en el contexto de la EP y Alzheimer ³⁵. Mediante modelos de clasificación ajustados sobre datos propios del dominio destino, esta propuesta sugiere a los modelos de difusión y GANs como los esquemas más prometedores a la hora de codificar representaciones propias de la enfermedad empleando la tarea de síntesis de neuroimagen en modalidades complementarias. Por otra parte, estos esquemas generativos han evidenciado capacidades identificando patrones ocultos relacionados con poblaciones de estudio particulares mediante estrategias de aprendizaje autosupervisado y semisupervisado. Por ejemplo, Pinon *et al.* propone la detección de anomalías cerebrales relacionadas a la EP desde una caracterización del espacio latente de un auto-encoder ajustado a la reconstrucción de neuroimágenes de poblaciones control de referencia ³⁶. A pesar de estos avances, este enfoque no aborda la codificación de hallazgos radiológicos propios de la EP resultando en un campo aún inexplorado. En este sentido, Watanabe *et al.* explora la codificación de hallazgos radiológicos relacionados a la EP generando representaciones volumétricas 3D en secuencias SPECT a partir de una cantidad limitada de cortes superiores empleando bloques decodificadores de *transformers*

³⁴ GUTIERREZ et al. “Lesion-preserving unpaired image-to-image translation between MRI and CT from ischemic stroke patients”. In: *International Journal of Computer Assisted Radiology and Surgery* 18.5 (2023), pp. 827–836.

³⁵ GAJJAR et al. “An Empirical Analysis of Diffusion, Autoencoders, and Adversarial Deep Learning Models for Predicting Dementia Using High-Fidelity MRI”. in: *IEEE Access* (2024).

³⁶ PINON et al. “Brain subtle anomaly detection based on auto-encoders latent space analysis: application to de novo parkinson patients”. In: IEEE. 2023, pp. 1–5.

³⁷. A pesar de estos grandes avances, dada la naturaleza multifactorial de la EP, la adecuada codificación de los hallazgos radiológicos y su análisis a través de los diferentes estadios de la enfermedad sigue siendo un reto. Por otra parte, resultan escasos los trabajos en el estado del arte que exploren estos sistemas generativos como una alternativa a la obtención de modalidades complementarias que permitan una caracterización de la enfermedad y su relación a los diferentes estadios que la comprenden.

³⁷ WATANABE et al. “Generative image transformer (GIT): unsupervised continuous image generative and transformable model for [123 I] FP-CIT SPECT images”. In: *Annals of nuclear medicine* 35 (2021), pp. 1203–1213.

2. PROBLEMA DE INVESTIGACIÓN

Las técnicas de neuroimagenología desempeñan un rol esencial en el diagnóstico, tratamiento y el seguimiento de enfermedades como el ACV y la EP ³⁸. Idealmente, con el objetivo de soportar el diagnóstico, es necesario contar con múltiples secuencias diagnósticas para extraer información específica de cada modalidad y explotar datos complementarios. Sin embargo, factores como la accesibilidad, los costos y la necesidad de actuar con celeridad imponen restricciones significativas en la ejecución de este análisis multimodal ⁶. Ante estas limitaciones, han surgido diversas estrategias de aprendizaje profundo generativo dedicadas a la síntesis de secuencias diagnósticas cerebrales. Estas estrategias intentan superar las restricciones impuestas por la práctica clínica, facilitando la estimación de modalidades faltantes. Sin embargo, estos trabajos reportan limitaciones en cuanto al aprendizaje del mapeo entre modalidades debido a la relación no lineal entre las modalidades ⁹ y la variabilidad en la presentación de estas enfermedades. Estos desafíos frecuentemente provocan la pérdida de patrones de intensidad y de información crucial para el diagnóstico. En particular, aunque las imágenes sintéticas logran una traducción adecuada de las zonas de control, muestran una estimación inadecuada de los patrones visuales en áreas críticas para cada enfermedad. Específicamente, en el contexto del ACV, se observa una incapacidad para traducir correctamente la lesión isquémica ³⁴, mientras que en la EP, se evidencia la aparición/desaparición aleatoria de transportadores dopaminérgicos, además otros problemas asociados a la traducción de la estructura anatómica cerebral ³⁹. También, problemas relacionados con la convergencia prematura y el colapso del modo en arquitecturas generativas afectan la gen-

³⁸ E BERCOVICH and MC JAVITT. *Medical Imaging: From Roentgen to the Digital Revolution, and Beyond*. *Rambam Maimonides Medical Journal*, 9, e0034. 2018.

³⁹ COHEN et al. "Distribution matching losses can hallucinate features in medical image translation". In: *MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer. 2018, pp. 529–536.

eración de imágenes y la diversidad de las mismas ²⁵.

De esta manera, surge la siguiente pregunta de investigación:

¿Cómo implementar una estrategia de aprendizaje profundo generativo que permita la síntesis de secuencias diagnósticas complementarias en los contextos de pacientes con enfermedades cerebrovasculares y enfermedades neurodegenerativas?

3. OBJETIVOS

3.1. Objetivo general

Implementar una estrategia de aprendizaje profundo generativo para la síntesis de secuencias diagnósticas capturadas en pacientes con enfermedades cerebrovasculares y neurodegenerativas.

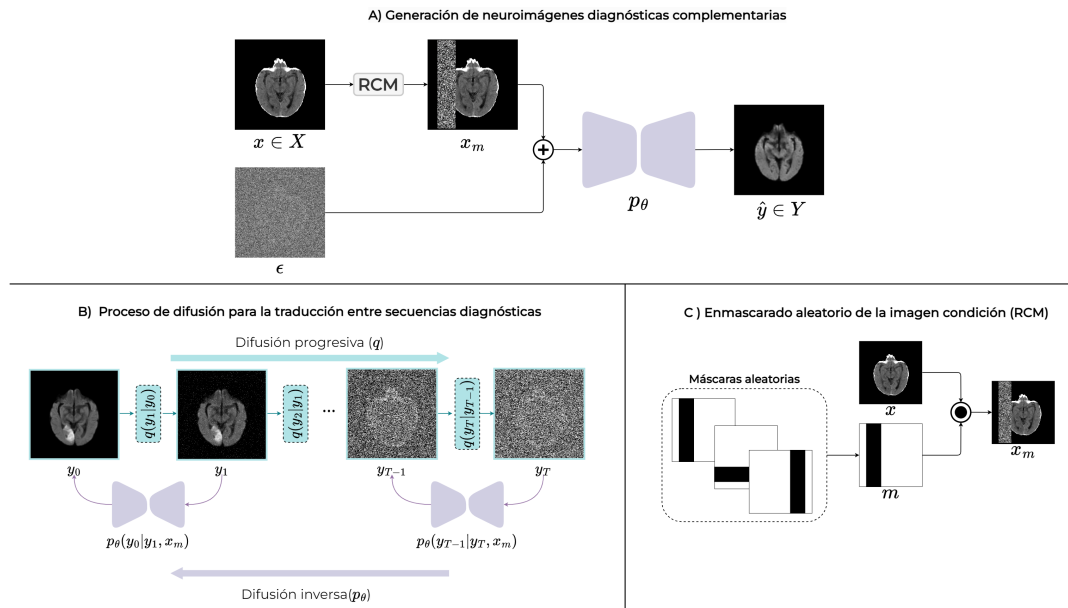
3.2. Objetivos específicos

- Seleccionar al menos dos conjuntos de datos de secuencias diagnósticas que comprendan modalidades complementarias, dentro de los contextos del accidente cerebrovascular y de la enfermedad de Parkinson.
- Implementar un modelo generativo para la síntesis entre dos secuencias diagnósticas clínicamente complementarias para la caracterización de la enfermedad de Parkinson.
- Implementar una arquitectura generativa para la traducción entre dos imágenes diagnósticas complementarias para soportar la caracterización de hallazgos radiológicos asociados al accidente cerebrovascular.
- Validar la capacidad de los modelos generativos implementados en la tarea de traducción entre secuencias diagnósticas complementarias en los contextos del accidente cerebrovascular y de la enfermedad de Parkinson.

4. MÉTODO PROPUESTO

En el presente trabajo se implementó una estrategia de aprendizaje profundo generativo para la síntesis de neuroimágenes diagnósticas en el contexto del ACV isquémico y la EP. Esta estrategia consiste en un modelo de difusión condicional, el cual realiza la traducción entre neuroimágenes clínicamente complementarias en los contextos de ACV y la EP (Ver Figura 8). Para esto, se entrenó una arquitectura autocodificador siguiendo un proceso iterativo de difusión, condicionado por neuroimágenes de un dominio de partida. Esta estrategia de difusión condicional consiste en dos procesos que permiten el aprendizaje gradual de los patrones visuales de modalidades de neuroimágenes que ofrecen mayor sensibilidad en el diagnóstico de cada enfermedad. En las siguientes subsecciones se describen detalladamente cada uno de estos procesos.

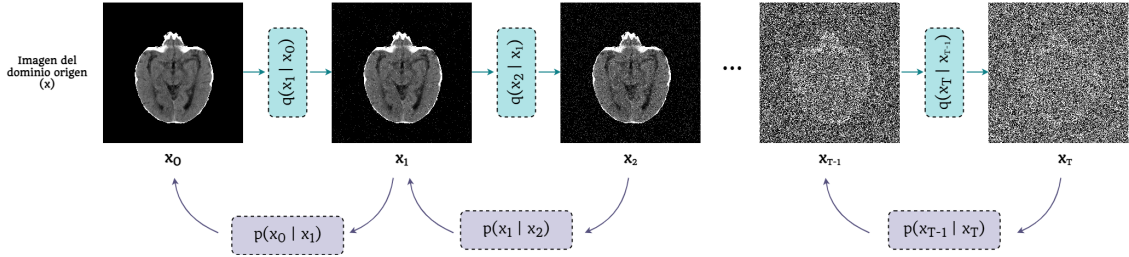
Figura 8. Esquema de la estrategia implementada. En **A** se representa el proceso de inferencia del modelo. **B** ilustra los procesos de difusión progresiva y difusión inversa durante el aprendizaje del mapeo entre modalidades. En **C** se muestra el mecanismo de enmascarado aplicado a la imagen condición en el proceso generativo.



4.1. Difusión progresiva Gaussiana

Los modelos basados en difusión consisten en una cadena de Márkov parametrizada que es entrenada para sintetizar datos a partir de ruido ⁴⁰. Posteriormente, se elimina el ruido agregado en cada paso de la cadena de Márkov utilizando una red neuronal hasta recuperar la imagen original. En la Figura 9 se ilustra este proceso básico de agregación de ruido y recuperación de la representación de la imagen.

Figura 9. Modelo de difusión típico, el cual inicia con una imagen de TC y la transforma progresivamente añadiendo ruido a través de una serie de pasos q . Al final de la secuencia, la imagen ha sido convertida completamente en ruido. Este proceso se invierte a través de una serie de predicciones, denotadas por p , restaurando la imagen original.



El entrenamiento de estos modelos se lleva a cabo por medio de dos procesos llamados difusión progresiva y difusión inversa. Particularmente, la difusión progresiva (q) es un proceso Markoviano que tiene como objetivo modelar una transición suave desde una imagen diagnóstica hacia una imagen que es completamente ruidosa y_T . Este proceso añade gradualmente ruido gaussiano a una neuroimagen del dominio de destino $y_0 \in \mathcal{Y}$ a lo largo de T iteraciones, hasta obtener una Gaussiana isotrópica y_T . Esta etapa produce una secuencia de versiones cada vez más ruidosas de y_0 , denotadas como y_t en cada uno de los pasos t ²⁵ (Este proceso implementado en la arquitectura implementada también se ilustra en la Figura 8B). Las distribuciones de las imágenes intermedias $q(y_t | y_{t-1})$ están descritas por:

⁴⁰ Jonathan HO; Ajay JAIN, and Pieter ABBEEL. “Denoising diffusion probabilistic models”. In: *Advances in neural information processing systems* 33 (2020), pp. 6840–6851.

$$q(y_t|y_{t-1}) = \mathcal{N}(y_t; \sqrt{\alpha_t}y_{t-1}, (1 - \alpha_t)I)$$

Donde α_t , un hiperparámetro dependiente del tiempo, regula la cantidad de ruido introducido en cada paso. A medida que t aumenta, la varianza del ruido incrementa, transformando progresivamente la imagen original en una versión cada vez más ruidosa. Este proceso garantiza que, al alcanzar el paso final T , la imagen resultante se asemeje a una Gaussiana isotrópica. La secuencia completa de imágenes ruidosas se expresa de la siguiente manera:

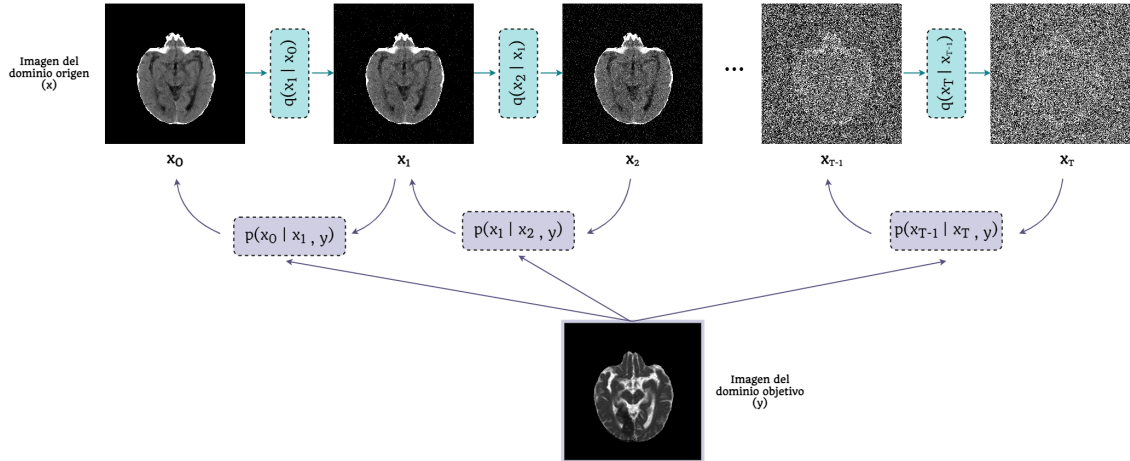
$$q(y_{1:T}|y_0) = \prod_{t=1}^T q(y_t|y_{t-1})$$

Siguiendo la formulación propuesta por Ho *et al.* ⁴⁰, la función de muestreo $q(y_t|y_0) = \mathcal{N}(y_t; \sqrt{\bar{\alpha}_t}y_0, (1 - \bar{\alpha}_t)I)$, donde $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$ corresponde al producto acumulativo de los parámetros del ruido hasta el paso t , elimina la necesidad de iterar sobre los pasos intermedios a partir de y_0 , lo que contribuye a reducir significativamente la complejidad computacional.

4.2. Traducción de neuroimágenes mediante Difusión Inversa Condicional

En un esquema condicional típico, se aplica un proceso de difusión inversa (p_θ) para estimar el ruido agregado en cada paso de q mediante un proceso de minimización. Sin embargo, en este trabajo se implementó un proceso condicional para guiar el proceso de síntesis. Particularmente, en este trabajo se implementó un proceso de difusión inversa condicional (p_θ) que permite capturar las relaciones complejas entre las modalidades de origen \mathcal{X} y destino \mathcal{Y} en neuroimágenes. Como se ilustra en la Figura 10, este proceso permite la traducción de patrones visuales entre los dominios, incorporando iterativamente características de y en la reconstrucción de x_0 , guiando efectivamente la traducción hacia el dominio objetivo.

Figura 10. Modelo de difusión condicional. En este, se realiza el proceso de *forward diffusion* convencional para una imagen de TC. Sin embargo, el proceso de *reverse diffusion* no solo está condicionado por el paso anterior, sino que también por una imagen objetivo, en este caso, una imagen de RM. Este proceso finaliza con la generación de una imagen sintética del dominio objetivo.



Particularmente en este trabajo, este proceso es el mecanismo generativo clave para recuperar la imagen original y_0 a partir de la versión ruidosa y_T producida por el proceso de difusión progresiva (ver adaptación en nuestro trabajo en la Figura 8B). Específicamente, se sigue un refinamiento iterativo ($y_{T-1}, y_{T-2}, \dots, y_0$) mediante distribuciones de transición condicionales aprendidas $p_\theta(y_{t-1}|y_t, x)$ que dependen de la imagen actual con ruido (y_t) y están condicionadas por una imagen de origen $x \in \mathcal{X}$ que proporciona información contextual crucial de estructuras anatómicas, distinciones entre tipos de tejido, y la presencia de lesiones o anomalías en el cerebro. De esta manera, el proceso de traducción se aproxima a la conservación de los detalles estructurales y anatómicos clave. Además, dado que el objetivo es aproximar la verdadera distribución condicional inversa $q(y_{t-1}|y_t, x)$ en cada paso t , se aprende una distribución parametrizada $p_\theta(y_{t-1}|y_t, x)$. Este proceso de difusión inversa en cada paso se describe como:

$$p_\theta(y_{t-1}|y_t, x) = \mathcal{N}(y_{t-1}; \mu_\theta(y_t, t, x_m), \Sigma_\theta(y_t, t)),$$

Donde μ_θ representa la media de la distribución Gaussiana en el paso t , predicha por una

arquitectura autocodificador f_θ , y x_m corresponde a una versión enmascarada aleatoriamente de la condición. Adicionalmente, la varianza Σ_θ es fija y se deriva de la calendarización del ruido α_t definida en q , lo que garantiza la consistencia entre los procesos de difusión progresiva e inversa.

4.2.1. Enmascarado aleatorio de la imagen condición Para aumentar la robustez del modelo frente a diferentes tipos de tejido (es decir, tejido sano y tejido afectado), se incorporó un mecanismo de enmascaramiento aleatorio, denotado como RCM, de la imagen condición x en el proceso de difusión inversa (Ver Figura 8C) ⁴¹. El propósito de estas máscaras es inducir al modelo a aprender cómo reconstruir las regiones enmascaradas basándose en el contexto circundante, al mismo tiempo que traduce la imagen desde la modalidad origen X hacia la modalidad objetivo Y . Más precisamente, durante la fase de entrenamiento, se genera una máscara de ruido que se aplica a la imagen condición x , ocultando una parte de la imagen en cada iteración. Estas máscaras varían en ubicación y orientación, asegurando que el modelo esté expuesto a un conjunto diverso de escenarios. La imagen condición enmascarada x_m se define como:

$$x_m = (1 - m) \odot x + m \odot \epsilon$$

Aquí, m es una máscara binaria donde los píxeles con valor 1 representan el área a enmascarar, y los píxeles con valor 0 representan el área no enmascarada. Adicionalmente, $\epsilon \sim \mathcal{N}(0, I)$.

4.2.2. Arquitectura U-Net Para aproximar la distribución condicional inversa, implementamos una arquitectura U-Net que predice la media de la distribución Gaussiana en cada paso t , permitiendo estimar imágenes menos ruidosas a través del proceso de difusión inversa.

⁴¹ Ziyun LIANG et al. “Modality cycles with masked conditional diffusion for unsupervised anomaly segmentation in mri”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 168–181.

Esta arquitectura está compuesta por bloques convolucionales residuales y recibe como entradas la imagen ruidosa $y_T \in \mathbb{R}^{1 \times H \times W}$ y la neuroimagen de condición $x \in \mathbb{R}^{1 \times H \times W}$. Estas imágenes se concatenan a lo largo de la dimensión de los canales, formando una única entrada representada como $[y_T; x] \in \mathbb{R}^{2 \times H \times W}$. A partir de esta concatenación, el codificador de la U-Net compacta progresivamente los patrones imagenológicos relacionados con enfermedades neurodegenerativas como la enfermedad de Parkinson (EP) y accidentes cerebrovasculares (ACV), específicos del dominio de neuroimágenes \mathcal{X} . Este proceso de codificación se lleva a cabo utilizando bloques que combinan capas convolucionales, normalización y *downsampling*. Adicionalmente, durante la fase de codificación, cada uno de los pasos t de la cadena de Márkov es codificado en embebidos que condicionan la red al paso actual del proceso de difusión inversa. Estos embebidos se integran en capas específicas de la red mediante la suma de los mismos con los mapas de características, lo que permite incorporar la información dependiente del tiempo de manera efectiva. A continuación, el decodificador utiliza las características extraídas por el codificador y, mediante bloques compuestos por capas convolucionales, de normalización y de *upsampling*, se recupera la información espacial perdida durante el proceso de codificación, a la vez que se estima la imagen $\hat{y} \in \mathbb{R}^{1 \times H \times W}$. Con el fin de recuperar detalles finos, el proceso de decodificación implementa conexiones de salto entre el codificador y el decodificador, lo que permite reintegrar la información que podría haberse perdido durante el proceso de codificación. El modelo f_θ se ajusta utilizando una función de pérdida basada en el error absoluto medio, definida como:

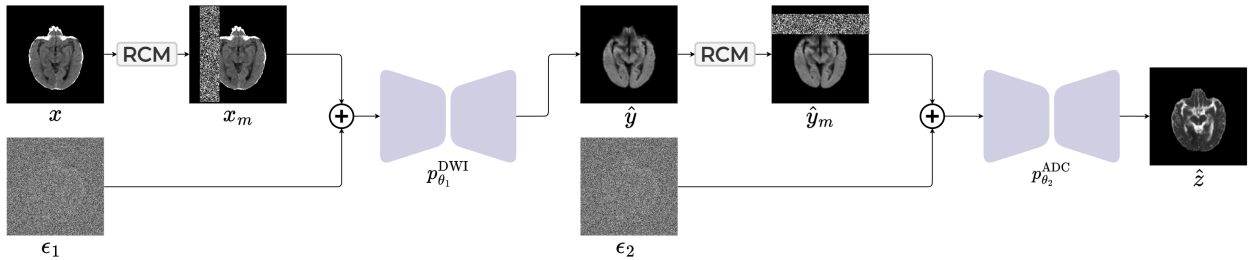
$$\mathcal{L} = \|f(y_t, x_m, \alpha_t) - y_0\|_1$$

4.3. Síntesis de secuencias diagnósticas para ACV isquémico

En este trabajo se utilizó como base de modelamiento el proceso típico de adquisición y el protocolo de análisis de neuroimágenes para la caracterización de lesiones isquémicas. En referencia a la resonancia magnética ponderada por difusión en ACV, primero se adquieren

las secuencias DWI y posteriormente se estiman los mapas de ADC. Entonces, en el método propuesto se implementó una estrategia secuencial de síntesis de neuroimágenes complementarias para ACV isquémico. Esta estrategia emula el orden natural de adquisición mediante una traducción secuencial de dos fases, basada en la estrategia de difusión implementada. Este proceso inicia con la estimación de secuencias DWI sintéticas $\hat{y} \in \mathcal{Y}$ a partir de secuencias NCCT de triaje, y posteriormente se sintetizan mapas de ADC $\hat{z} \in \mathcal{Z}$. En concreto, las traducciones se llevan a cabo mediante modelos independientes, $p_{\theta_1}^{\text{DWI}}(y_{t-1}|y_t)$ y $p_{\theta_2}^{\text{ADC}}(z_{t-1}|z_t, \hat{y}_m)$, los cuales realizan los mapeos NCCT \rightarrow DWI y DWI \rightarrow ADC, respectivamente. Este enfoque permite descomponer el mapeo complejo de NCCT hacia ADC en pasos más simples, aprovechando el alto contraste entre el tejido normal y el tejido de lesión en el mapeo de NCCT hacia DWI. Este proceso de síntesis en cascada se ilustra en la Figura 11, donde primero se genera la imagen DWI desde la NCCT y luego se estima el mapa paramétrico ADC.

Figura 11. Proceso generativo de secuencias diagnósticas del ACV isquémico. A partir de las imágenes de NCCT, un proceso $p_{\theta_1}^{\text{DWI}}$ realiza la síntesis de secuencias DWI, las cuales son empleadas posteriormente por un proceso $p_{\theta_2}^{\text{ADC}}$ para generar mapas de ADC sintéticos.

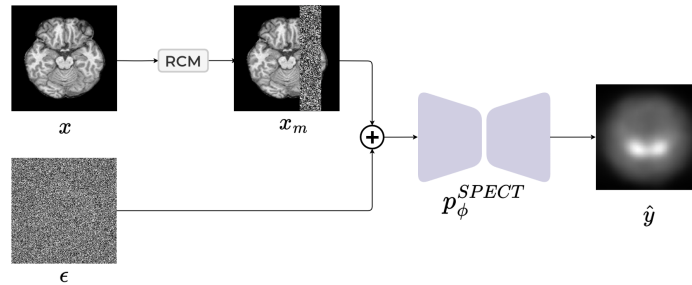


4.4. Síntesis de secuencias diagnósticas de la EP

En el contexto de la enfermedad de EP, donde las secuencias estructurales T1 ($x \in \mathcal{X}$) son más accesibles que las imágenes funcionales SPECT ($y \in \mathcal{Y}$), se implementó una estrategia para estimar imágenes SPECT sintéticas \hat{y} basada en la metodología de difusión implementada (ver Figura 12). La traducción entre estas modalidades se lleva a cabo mediante una función

parametrizada $p_\phi^{\text{SPECT}}(y_{t-1}|y_t, x_m)$, que captura las relaciones entre la información estructural y funcional del cerebro, aproximando la concentración de dopamina del paciente. Este enfoque reduce la necesidad de exponer a los pacientes a radiofármacos, utilizados durante la adquisición de imágenes SPECT, minimizando así los riesgos asociados a este procedimiento.

Figura 12. Proceso generativo de secuencias diagnósticas de la enfermedad de Parkinson. Usando neuroimágenes RM-T1, en el proceso p_ϕ^{SPECT} se sintetizan secuencias SPECT.



En la figura 12 se ilustra el proceso de difusión adaptado en el problema de la EP. En este caso se realiza una traducción desde hallazgos morfométricos hacia patrones dopaminérgicos, considerando cierta correlación espacial de las regiones cerebrales. En este proceso, la secuencia estructural RM-T1 se usa como condición para sintetizar una imagen funcional SPECT, mapeando las características propias de la modalidad condición y plasmándolas en las secuencias objetivo, la cual presenta una alta sensibilidad ante la EP.

5. DISEÑO EXPERIMENTAL

5.1. Conjuntos de datos

La estrategia desarrollada de síntesis por medio de un mecanismo de difusión, fue aquí validado en dos contextos donde se requieren neuroimagenes complementarias. A continuación, se describe el conjunto de datos utilizado tanto como para el contexto del ACV como para la EP. También se informa de los parámetros usados y los ajustes realizados a cada uno de los dominios, donde se desarrollo este trabajo.

Conjunto de datos para ACV isquémico El conjunto de datos utilizados en este trabajo fue recolectado en el marco de un proyecto de Minciencias en colaboración con el grupo de investigación BIVL²ab y la clínica FOSCAL. Particularmente el conjunto de datos acotado en este trabajo incluye estudios imagenológicos de 95 pacientes que presentaban síntomas de ACV, de los cuales 85 fueron diagnosticados con isquemia cerebral y 10 se clasificaron como pacientes control (sin evidencia de lesión isquémica). Para cada paciente se adquirió un estudio NCCT al momento de su ingreso hospitalario. Transcurrido un máximo de 24 horas, se procedió a la adquisición de imágenes de RM en la modalidad DWI y los correspondientes mapas de ADC. Las lesiones isquémicas fueron delineadas manualmente en los mapas de ADC por dos neurorradiólogos con más de cinco años de experiencia, observándose variabilidad en cuanto a la forma, tamaño y el número de cortes consecutivos en los que la lesión isquémica es visible. Los criterios de inclusión abarcan pacientes mayores de 18 años, sin signos de hemorragia cerebral, y que no hubieran recibido tratamiento alguno entre la adquisición de las imágenes NCCT y RM. Para la obtención de estas imágenes se emplearon dos escáneres de tomografía computarizada (tomógrafos) y dos escáneres de resonancia magnética (resonadores). La tabla 1 muestra la distribución demográfica del conjunto de datos.

Tabla 1. Información demográfica para el conjunto de datos para el ACV isquémico.

Variable	Entrenamiento	Validación
ACV / Control	54 / 6	31 / 4
Edad ($\mu \pm \sigma$)	70.20 \pm 12.80	73.30 \pm 15.20
Género (M/F)	32 / 28	16 / 19

En cuanto al preprocesamiento del conjunto de datos, el software SynthStrip fue aplicado para eliminar el cráneo de todas las secuencias imagenológicas, asegurando que solo el tejido cerebral permaneciera visible. En el caso de las imágenes de NCCT, para obtener una eliminación efectiva del cráneo, primero se realizó una reconstrucción aproximada T1 MP-RAGE con el fin de resaltar la estructura cerebral y facilitar una segmentación más precisa. Posteriormente, las imágenes de NCCT y DWI fueron alineadas con los mapas de ADC mediante un proceso de registro rígido utilizando Elastix. Este registro se enfocó en ajustar la posición y orientación de las imágenes, aplicando únicamente rotaciones y traslaciones. Se utilizó interpolación B-spline de tercer orden para asegurar la precisión de los valores de intensidad de los píxeles durante la transformación, garantizando una correcta alineación de las modalidades con las lesiones anotadas. Finalmente, se extrajeron los cortes (*slices*) de cada volumen de estudio, las intensidades de los píxeles se normalizaron al rango de -1 a 1, y cada corte se redimensionó a 256×256 píxeles. Estos cortes fueron agrupados como volúmenes y almacenados en formato NPY.

Conjunto de datos para la EP El enfoque propuesto también fue validado en el contexto de la enfermedad del Parkinson. La validación del método propuesto consistió en evaluar la capacidad del mismo para sintetizar la concentración dopaminérgica (altamente relacionada con el desarrollo de la enfermedad) típicamente estimada sobre secuencias SPECT pero a partir de secuencias convencionales RM-T1 (ampliamente disponibles en la rutina clínica). Para tal fin, se empleó la base de datos pública *Parkinson's Progression Markers Initiative* (PPMI) (www.ppmi-info.org/data). Esta base de datos, de naturaleza multicéntrica, es una iniciativa global que busca evaluar y validar la progresión de la enfermedad mediante el

análisis de neuroimágenes que incluyen las modalidades MRI, DTI, SPECT y PET, junto con información genética. De esta amplia base de datos, se seleccionó una cohorte de pacientes clínicamente diagnosticados con la EP y sujetos control de referencia que cumplieran con el criterio de adquisición para las modalidades de imagen RM-T1 y SPECT. En este sentido, el conjunto seleccionado incluye imágenes médicas de 150 pacientes, distribuidos equitativamente entre sujetos control de referencia y pacientes EP diagnosticados según la escala de medición Hoehn and Yahr. Para las poblaciones tenidas en cuenta en este estudio se reporta un promedio de 25 días transcurridos entre la toma inicial de RM-T1 y la secuencia SPECT asociada a cada paciente de la población control, mientras que para la población EP transcurrieron en promedio 255 días.

Variable	Entrenamiento	Validación
Parkinson/Control	60 / 60	15 / 15
Edad ($\mu \pm \sigma$)	64.01 \pm 9.89 / 60.24 \pm 11.83	66.07 \pm 8.91 / 61.47 \pm 10.84
H&Y escala (1/2/3/4)	8 / 49 / 1 / 2	1 / 7 / 5 / 2

Tabla 2. Características demográficas y clínicas de los pacientes con Enfermedad de Parkinson (EP) y controles, en los conjuntos de entrenamiento y prueba.

La Tabla 2 presenta la información demográfica y la distribución de los pacientes EP según la escala H&Y para los conjuntos de entrenamiento y test.

En cuanto al preprocesamiento de los datos recolectados y con el fin de garantizar tanto la uniformidad espacial de las secuencias radiológicas, como paso inicial, los voxeles de las modalidades RM-T1 y SPECT fueron ajustados a un espacio isométrico de 1 cm³. Particularmente, para las secuencias anatómicas T1-RM, los siguientes pasos de preprocesamiento fueron aplicados. Dada la irrelevancia clínica del tejido craneal a la hora del estudio de la EP, dicha información fue eliminada empleando la utilidad [synthstrip](#). Por otra parte, dada la naturaleza multicéntrica de la fuente de datos, las inhomogeneidades en las intensidades de los estudios imagenológicos fueron corregidas aplicando una corrección de campo de sesgo N4 fue aplicada usando [SimpleITK](#). De esta manera se mantienen los valores de intensidad para los tejidos de igual naturaleza. Con el fin de establecer un sistema común de coordenadas

que permita no solo estandarizar los estudios individuales, sino también tener una referencia anatómica, los casos T1-RM fueron registrados de manera rígida a un espacio anatómico de referencia MNI-152 mediante la herramienta [ANTsPy](#). Por otra parte, sobre las secuencias SPECT también se aplicó una corrección de campo de sesgo N4 posterior al ajuste isométrico. Debido a la naturaleza inherente del procedimiento, donde la aleatoriedad del componente radiactivo empleado y su posterior detección para determinar la presencia dopaminérgica, las secuencias SPECT resultan ser susceptibles de sufrir altos niveles de ruido en su señal. Para manejar esto, un filtro gaussiano isotrópico de 18 mm fue aplicado. Al finalizar el preprocesamiento para las dos modalidades, se tienen estudios con una resolución espacial de 256×256 píxeles. Para fines prácticos de la propuesta desarrollada, todos los análisis se desarrollan sobre la vista axial, donde para ambas modalidades, los cortes iniciales y finales de cada estudio no se tuvieron en cuenta dada la escasa información cerebral contenida y, por tanto, su irrelevancia clínica.

5.2. Configuración de la arquitectura desarrollada

La arquitectura desarrollada compartió parámetros de arquitectura y entrenamiento para los dos dominios evaluados (ACV, EP). A continuación se detallan los parámetros y configuraciones realizadas.

- En cuanto al proceso de difusión la agregación progresiva del ruido dentro del modelo de difusión propuesto fue ajustado a los contextos clínicos de interés, empleando un total de 1000 pasos con un α_t calendarizado de manera lineal en el rango definido $[0.0001, 0.02]$.
- En cuanto a la arquitectura generativa, del tipo U-Net, está construida por cuatro niveles de codificación y decodificación. Por una parte, el codificador inicia con 32 canales que son incrementados gradualmente a través de las capas utilizando una configuración de multiplicación de canales (1, 2, 4, 8). Cada bloque residual incorpora una normal-

ización de grupos seguida de activaciones *SiLU*. Además, se emplean mecanismos de atención en múltiples resoluciones (32×32 , 16×16 , 8×8) que incluyen normalización de grupos. El decodificador refleja la estructura del codificador para recuperar la resolución espacial original, empleando como activación final una *Tanh*.

- En cuanto a la etapa de entrenamiento se utilizó un optimizador *AdamW* con una tasa de aprendizaje de 0.0001. Por otra parte, el sobre ajuste del modelo fue mitigado empleando un dropout de 0.1. El entrenamiento se realizó a lo largo de 20000 iteraciones donde la función de ajuste del modelo fue multiplicada por un factor de 100.

5.3. Validación

Para los dos contextos clínicos manejados, los conjuntos de datos fueron validados mediante tareas de síntesis y de clasificación. Siendo la primera de ellas aplicada exclusivamente dentro del contexto de ACV y ambas para EP. Para la parte generativa, la validación incluye métricas de amplio uso reportadas por el estado del arte, las cuales permiten cuantificar la similitud y calidad de la traducción entre las imágenes obtenidas y de referencia. Las métricas usadas en este estudio se detallan a continuación:

Proporción máxima de señal-ruido (PSNR) El PSNR mide la calidad de la imagen generada en función de la relación del ruido existente entre ésta y la imagen de referencia. Su cálculo se da mediante la relación entre el error cuadrático medio (MSE) y el valor máximo de pixel (L) en una imagen, como se expresa a continuación:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{L^2}{\text{MSE}} \right), \quad (4)$$

Donde el MSE resulta del siguiente cálculo efectuado sobre la imagen generada I_{gen} y la imagen de referencia I_{ref} , de la siguiente manera:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (I_{\text{gen}}(i) - I_{\text{ref}}(i))^2, \quad (5)$$

Aquí, n es el número total de píxeles en la imagen. El PSNR se mide en decibelios (dB) en un rango de 0 a ∞ . Un PSNR alto indica una reconstrucción de alta similitud respecto a la imagen de referencia.

Índice de similitud estructural (SSIM) El SSIM evalúa la similitud entre dos imágenes, considerando cambios en la luminancia, el contraste y la estructura, y se define como:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (6)$$

Donde, la luminancia está dada por la comparación entre μ_x y μ_y , que representan las medias de las imágenes x e y ; el contraste está dado por σ_x^2 y σ_y^2 , que son sus varianzas; y σ_{xy} es la covarianza entre ambas imágenes, representando la comparación de la estructura. Por otra parte, C_1 y C_2 son constantes que estabilizan la división. Esta métrica toma valores entre 0 y 1, donde un valor de SSIM cercano a 1 indica un alto grado de similitud estructural entre las imágenes.

Validación de la síntesis enfocada en la región cerebral En un análisis más detallado, la validación del método generativo propuesto se enfocó de manera particular en la región cerebral. De esta manera, mediante un proceso de enmascarado, la región cerebral fue filtrada con el fin de eliminar información circundante clínicamente irrelevante (fondo). De esta manera, el cálculo de las métricas de similitud ofrece un análisis centrado en la región anatómicamente relevante, evitando cálculos sesgados por la influencia de las demás regiones. El proceso consistió en la generación de una máscara binaria $M(x)$ para cada imagen, definida como:

$$M(x) = \begin{cases} 1, & \text{si } x \in B \\ 0, & \text{si } x \notin B \end{cases}$$

donde B es el conjunto de píxeles pertenecientes al cerebro. La región de interés (ROI) se define entonces como el conjunto de todos los píxeles donde $M(x) = 1$. Así, las métricas PSNR y SSIM se calcularon únicamente en los píxeles dentro de la ROI:

$$\text{PSNR}_{\text{ROI}} = 10 \cdot \log_{10} \left(\frac{L^2}{\frac{1}{n_{\text{ROI}}} \sum_{i \in \text{ROI}} (I_{\text{gen}}(i) - I_{\text{ref}}(i))^2} \right), \quad (7)$$

donde n_{ROI} es el número de píxeles dentro de la región del cerebro definida por $M(x)$. De manera similar, el SSIM para la región cerebral se calcula restringido a la ROI. Esta evaluación específica permitió medir la calidad de las imágenes generadas únicamente en las regiones relevantes para el diagnóstico.

Validación enfocada en la preservación de características en la EP En una fase posterior a la fase generativa, la validación de las secuencias diagnósticas generadas con un sentido clínico relevante (SPECT) para el diagnóstico de la EP, fue llevada a cabo mediante una arquitectura convolucional del estado del arte ajustada al dominio de interés. Particularmente, dicha red toma como pesos de inicialización los ajustados a tareas de clasificación sobre imágenes del dominio natural (ImageNet). Posteriormente, sus capas intermedias y superiores fueron ajustadas al dominio de interés (SPECT) para lograr la diferenciación entre poblaciones, control de referencia y pacientes EP. Una vez ajustado el modelo al dominio de interés, las secuencias SPECT sintetizadas a partir de entradas RM-T1, fueron empleadas como un conjunto adicional de validación con el fin de determinar la presencia de los descriptores dopaminérgicos sintetizados a partir de las secuencias anatómicas. De esta manera, se puede estimar la preservación de las características en la EP.

Prueba estadística Con el fin de determinar la significancia estadística entre el método propuesto y las existentes arquitecturas generativas del estado del arte (U-Net, Pix2Pix y CycleGAN), la prueba de Mann-Whitney U fue empleada como técnica de estimación estadística. Esta prueba resulta adecuada para comparar métricas entre dos grupos independientes.

Particularmente, mediante esta prueba se contrastan las métricas de similitud SSIM y PSNR obtenidas por cada una de las arquitecturas generativas frente al método propuesto con el objetivo de identificar discrepancias significativas a la hora de realizar la síntesis bajo los dos contextos clínicos de uso.

6. EVALUACIÓN Y RESULTADOS

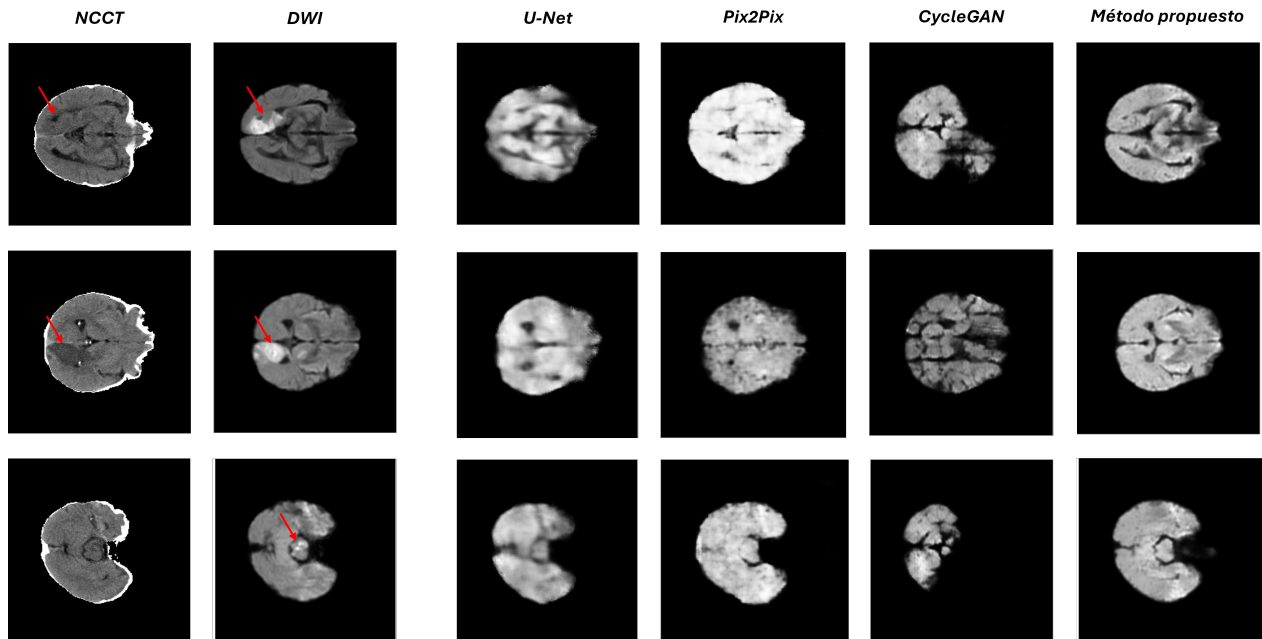
El enfoque propuesto fue validado bajo dos contextos de neuroimágenes con el fin de determinar el aporte del método a la hora de estimar secuencias radiológicas clínicamente complementarias. Particularmente, dentro del contexto de ACV y la EP, las estrategias generativas del estado del arte como U-Net, Pix2Pix y CycleGAN representan la línea base frente al método propuesto. A continuación se presentan los resultados obtenidos en la síntesis de secuencias diagnósticas dentro del contexto del ACV isquémico y de la EP.

6.1. Síntesis de secuencias DWI y mapas de ADC para el ACV isquémico

Para validar el enfoque generativo propuesto, se realizó una evaluación empleando diferentes arquitecturas generativas ampliamente abordadas en el estado del arte bajo diferentes contextos incluyendo el dominio médico. Esta validación dentro del contexto del ACV fue llevada a cabo con el conjunto de validación, compuesto por 35 pacientes (cerca de 878 cortes). La Figura 13 muestra los resultados de sintetizar secuencias DWI a partir de NCCT obtenidos por cada arquitectura generativa. El método propuesto evidencia una capacidad notable en cuanto a la preservación de detalles anatómicos cerebrales a comparación de los demás métodos, donde estas regiones tienden a ser suavizadas por las arquitecturas U-Net y Pix2Pix (ver columnas 3 y 4 de la Figura 13). Adicionalmente, la arquitectura CycleGAN tiende a eliminar regiones cerebrales que pueden conducir a neuroimágenes poco realistas, sin contexto anatómico y sin utilidad en el ambiente clínico. A pesar que el enfoque propuesto preserva de manera general la morfología cerebral, son evidentes las inconsistencias en cuanto al mapeo de las intensidades del tejido cerebral y por tanto de las lesiones isquémicas. Particularmente, en los casos de las filas 1 y 2, ninguna arquitectura logra reproducir la información de las lesiones en las imágenes sintéticas a pesar de que estas son visibles tanto en las secuencias NCCT como las DWI, sugiriendo que ajustes adicionales deben realizarse con el propósito

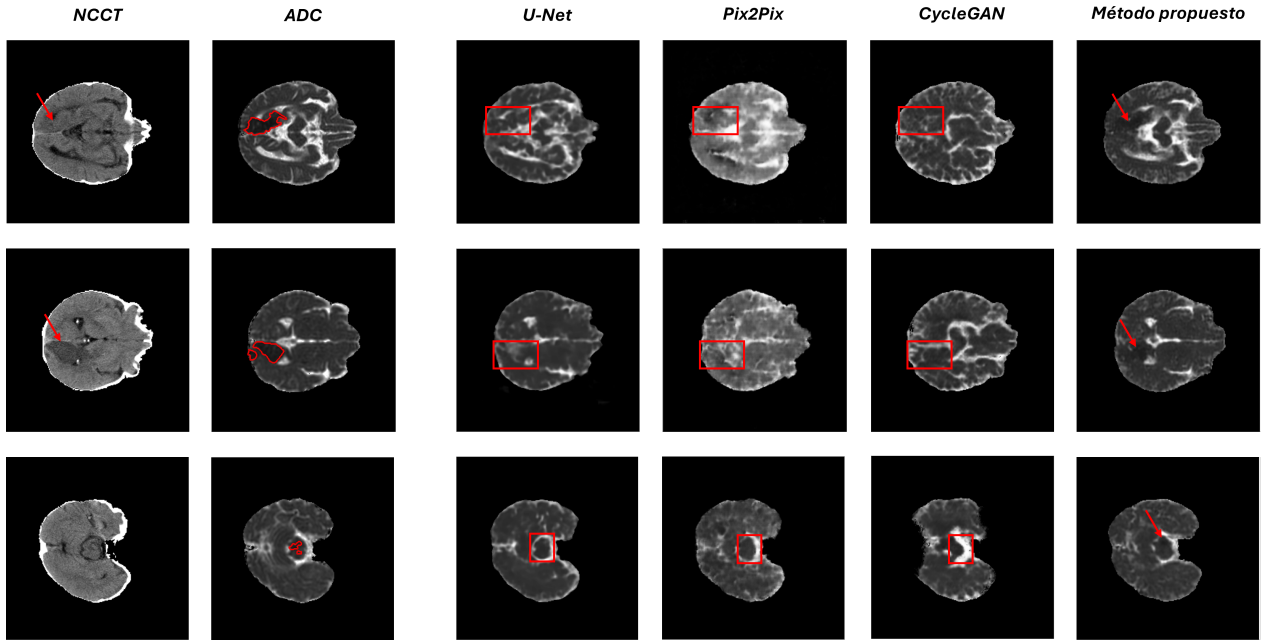
de preservar la lesión durante el proceso generativo.

Figura 13. Resultados cualitativos para la síntesis de secuencias DWI a partir de secuencias NCCT mediante distintas arquitecturas generativas. En las columnas 1 y 2 se presentan los estudios de NCCT, DWI reales, mientras que las columnas 3-6 muestran los DWI sintéticas por cada arquitectura. Cada fila representan estudios diferentes.



Por otra parte, la tarea generativa también fue evaluada con el fin de obtener mapas de ADC sintéticos a partir de las secuencias NCCT. La Figura 14 ilustra el desempeño logrado por las diferentes arquitecturas, donde el método propuesto es el único que logra preservar de manera visiblemente adecuada la información de las lesiones isquémicas en los mapas de ADC sintéticos, incluso en el caso donde la lesión es diminuta y no es visible en NCCT (fila 3). Adicionalmente, la coherencia estructural y anatómica del cerebro es preservada de mejor manera por el método propuesto a comparación de representaciones logradas por arquitecturas generativas base. Este resultado permite resaltar cómo el método propuesto es viable en un contexto clínico, dando aproximaciones coherentes del ADC.

Figura 14. Resultados cualitativos para la síntesis de mapas de ADC a partir de secuencias NCCT mediante distintas arquitecturas generativas. Las dos primeras columnas muestran las secuencias de referencia para NCCT y ADC, mientras que las columnas 3-6 muestran los resultados logrados por cada arquitectura generativa. Cada fila representa un estudio imagenológico diferente.



En términos cuantitativos, la Tabla 3 muestra los valores de SSIM y PSNR al realizar la síntesis de las secuencias DWI y de los mapas ADC. Particularmente, sobre la tarea de sintetizar estudios DWI a partir de secuencias NCCT, el método propuesto evidencia un desempeño superior respecto a las demás arquitecturas, logrando un SSIM de 0.55 y un PSNR de 18.8. Cabe resaltar que durante el cálculo de estas métricas se excluyó el fondo, el cual hace parte del contexto. La inclusión del fondo proporciona índices SSIM superiores al 90% dando información sobre la tarea general de reconstrucción, pero limitando un análisis más detallado de las estructuras cerebrales. Sin embargo, a pesar de que estos resultados sugieren una adecuada codificación de las características típicamente observables sobre las secuencias DWI, se hipotetiza que la mejora en dichas métricas puede lograrse llevando a cabo ajustes adicionales que focalicen la síntesis a nivel local (lesión isquémica).

En este mismo sentido, la síntesis de NCCT a ADC también fue valorada, donde el método

propuesto también evidencia resultados superiores respecto a los demás enfoques generativos, logrando valores de SSIM y PSNR de 0.45 y 16.94, respectivamente. Como era de esperarse, estos resultados son inferiores a los logrados para la síntesis de DWI debido a la dificultad del mapeo desde NCCT hacia ADC, que se hace particularmente notable aquí. Particularmente, las arquitecturas U-Net y Pix2Pix obtienen valores en las métricas SSIM/PSNR considerablemente bajos (0.10/7.83 y 0.27/12.02, respectivamente). Con lo anterior, se destaca la robustez del método propuesto ante las diferencias entre las modalidades imagenológicas, en este caso NCCT y ADC, en la representación de las intensidades del tejido cerebral y de lesión isquémica.

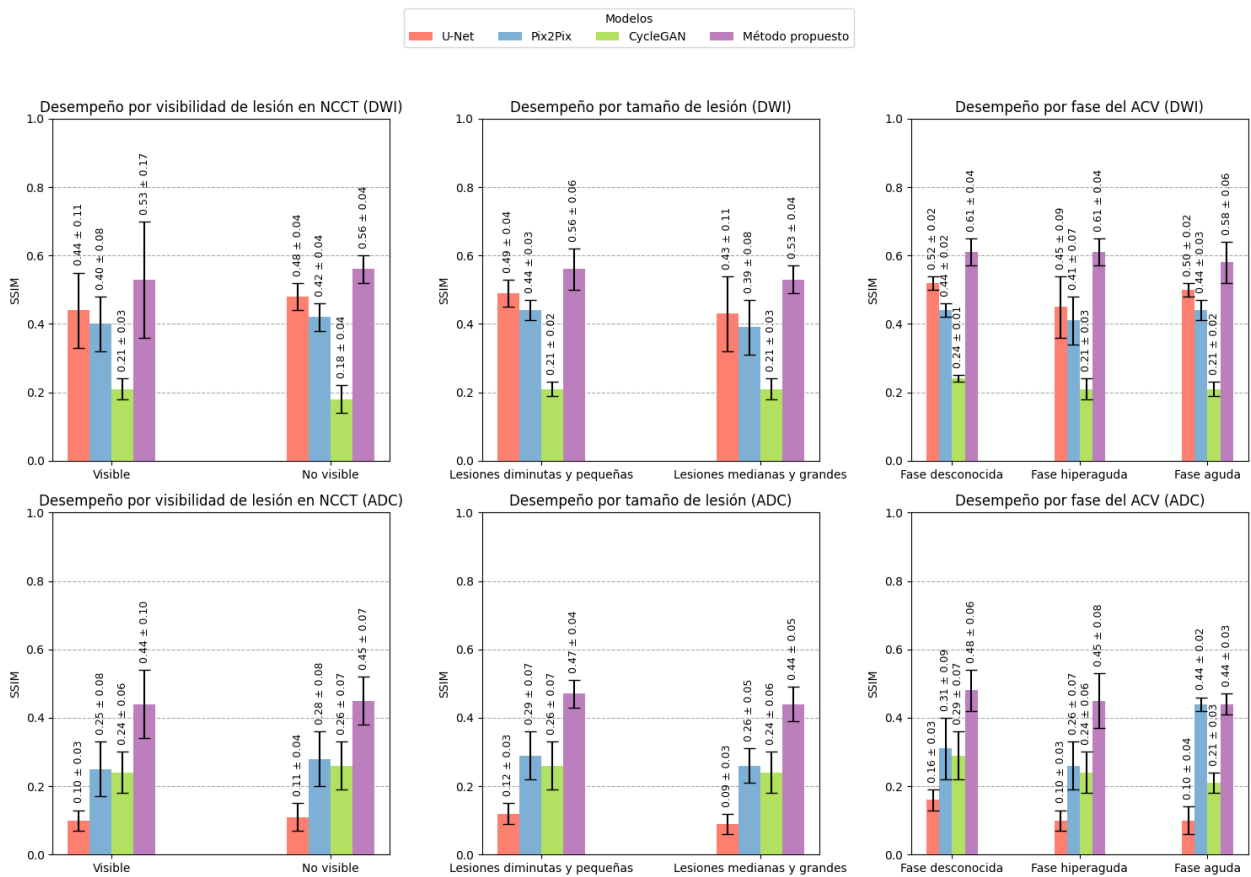
Tabla 3. Evaluación de las secuencias DWI y mapas ADC sintéticos generados a partir de distintas arquitecturas generativas.

Síntesis	Arquitectura	SSIM \uparrow	PSNR \uparrow
NCCT \rightarrow DWI	U-Net	0.46 \pm 0.08	12.72 \pm 1.20
	Pix2Pix	0.41 \pm 0.06	10.85 \pm 1.80
	CycleGAN	0.21 \pm 0.03	11.80 \pm 0.75
	Método propuesto	0.55 \pm 0.13	18.80 \pm 2.61
NCCT \rightarrow ADC	U-Net	0.10 \pm 0.03	7.83 \pm 0.46
	Pix2Pix	0.27 \pm 0.07	12.02 \pm 2.03
	CycleGAN	0.25 \pm 0.06	12.23 \pm 0.78
	Método propuesto	0.45 \pm 0.07	16.94 \pm 1.48

Con el objetivo de realizar una valoración del desempeño de las arquitecturas al momento de sintetizar las secuencias radiológicas complementarias, un análisis detallado se llevó a cabo siguiendo criterios clínicos como el tamaño de la lesión isquémica, la visibilidad de la misma en los estudios de NCCT y la fase del ACV. La Figura 15 muestra los valores de SSIM para cada criterio clínico. El método propuesto muestra resultados superiores de manera consistente con respecto a los demás modelos generativos en los tres criterios, tanto en SSIM como en PSNR. Particularmente, en cuanto al tamaño de lesión, se observa que el enfoque propuesto mantiene un rendimiento consistente tanto para las lesiones pequeñas como para las lesiones grandes, demostrando su eficacia para traducir detalles a baja escala. Este criterio

es relevante en el contexto del ACV, teniendo en cuenta que el tamaño (volumen) de la lesión determina el tipo de procedimiento que se debe llevar a cabo, pero además es un indicador determinante para el pronóstico del paciente.

Figura 15. Comparación de los valores obtenidos para la métrica del SSIM en las imágenes sintéticas de DWI y ADC según los criterios de visibilidad de la lesión en NCCT, el tamaño de la lesión y la fase del ACV. En la fila superior se muestra el desempeño de los modelos en la síntesis de secuencias DWI y en la fila inferior se presenta el desempeño de los mismos en la síntesis de los mapas de ADC.



Además, mediante la prueba estadística U de Mann-Whitney se determinaron diferencias estadísticamente significativas entre los resultados de los enfoques base y el propuesto. Dichas diferencias estadísticas resultan ser consistentes en la síntesis de secuencias DWI como de mapas de ADC. Sin embargo, al analizar el comportamiento del método propuesto en función de los criterios de la visibilidad de las lesiones en NCCT, el tamaño de la lesión y la fase

del ACV, no se encontraron diferencias significativas ($\rho > 0.05$), sugiriendo un rendimiento estable independiente de las características clínicas específicas.

Caracterización de la EP En este contexto de neuroimágenes, la síntesis de secuencias SPECT desde imágenes RM-T1 se realiza para tener una aproximación de los niveles dopaminérgicos, así una relación de la progresión de la enfermedad. En resumen, desde la aproximación SPECT se espera tener una clasificación de la EP. Entonces el primer paso consistió en validar la representación convolucional que logra una mayor discriminación de la EP desde las secuencias originales SPECT. Esta validación se llevó a cabo sobre un conjunto balanceado de secuencias SPECT captadas de 15 pacientes tanto para la población EP como control representando al rededor de 225 cortes para cada grupo, respectivamente. La tabla 4 muestra el desempeño de clasificación binaria para las diferentes arquitecturas. Interesantemente, la red VGG16 logra resultados superiores en todas las métricas a excepción de la sensibilidad (superada por un 1% por la MobileNet). Este desempeño puede deberse a los mecanismos clásicos de convolución empleados por la VGG16 donde las características primarias como bordes y contornos son capturados de mejor manera a comparación de esquemas compactos y ligeros que favorecen el desempeño de entrenamiento pero comprometen la capacidad de generalización. En este sentido, la arquitectura VGG16 fue seleccionada como el backbone de clasificación empleando como criterio de decisión la precisión y puntaje F1.

Tabla 4. Evaluación comparativa de cuatro arquitecturas: VGG16, ConvNeXt, MobileNet y ResNet50, en la tarea de clasificación de Parkinson vs. Control en el conjunto de validación.

Arquitectura	Exactitud \uparrow	Precisión \uparrow	Sensibilidad \uparrow	Puntaje F1 \uparrow
Vgg16	96%	97%	95%	96%
ConvNeXt	82%	76%	92%	83%
MobileNet	96%	96%	96%	96%
ResNet50	56%	76%	56%	44%

6.2. Síntesis de secuencias SPECT para la EP

Una vez se realizó este análisis, se procedió a realizar la síntesis de imágenes SPECT a partir de secuencias RM-T1, empleando arquitecturas generativas ampliamente empleadas en el estado del arte como referencia y comparándolas con el enfoque de difusión desarrollado en esta trabajo. Para esto, el subconjunto de validación compuesto por 15 sujetos control y 15 pacientes EP, de los cuales se contaba con las modalidades de interés fue seleccionado. De esta manera se tienen en promedio 225 cortes para la modalidad RM-T1 y 225 cortes para SPECT, para cada población respectivamente. La Figura 16 presenta los resultados cualitativos a la hora de estimar los patrones dopaminérgicos a partir de secuencias RM-T1 sobre sujetos EP a lo largo de los diferentes estadios y muestras control de referencia. Como era de esperarse, la codificación funcional del estado dopaminérgico cerebral a partir de información anatómica resulta ser una tarea difícil para los enfoques en general. A pesar de esto, el método propuesto es el que logra una codificación variante de la concentración dopaminérgica a lo largo de los diferentes estadios de la enfermedad, sin llegar ésta a ser lo suficientemente distintiva a pesar de los patrones asimétricos marcados propios de cada estadio de la enfermedad. Por otra parte, resulta interesante la capacidad de mantener la simetría en los casos control de referencia. Este comportamiento sugiere que el método propuesto ajustó la representación dopaminérgica a una media generalizada del conjunto de datos donde las características distintivas y las variaciones sutiles entre los diferentes casos clínicos requieren de ajustes adicionales. Por ejemplo, para las representaciones logradas por el método propuesto, como se observa en la fila 6 columna 4, muestran que los descriptores dopaminérgicos mapeados tienden a ser simétricos, independientemente de la población observada. De manera similar, métodos del estado del arte como U-Net, Pix2Pix y CycleGAN también presentan limitaciones incluso más significativas en la preservación de patrones específicos de la enfermedad de Parkinson (EP), como la asimetría en la concentración de dopamina entre hemisferios. Esto sugiere que, aunque estos modelos son capaces de aprender características generales, fallan en capturar detalles críticos necesarios para un diagnóstico

preciso.

Figura 16. Resultados cualitativos de secuencias de **SPECT** generadas a partir de neuroimágenes **RM-T1** en sujetos control y pacientes con EP en diferentes estadios de Hoehn y Yahr (**H&Y** 1-4). Las primeras filas de **RM-T1** y **SPECT** corresponden a las secuencias originales de los pacientes. Por otro lado, U-Net, Pix2Pix, CycleGAN y Método propuesto muestran secuencias **SPECT** sintéticas obtenidas mediante arquitecturas generativas a partir de las neuroimágenes **RM-T1**.

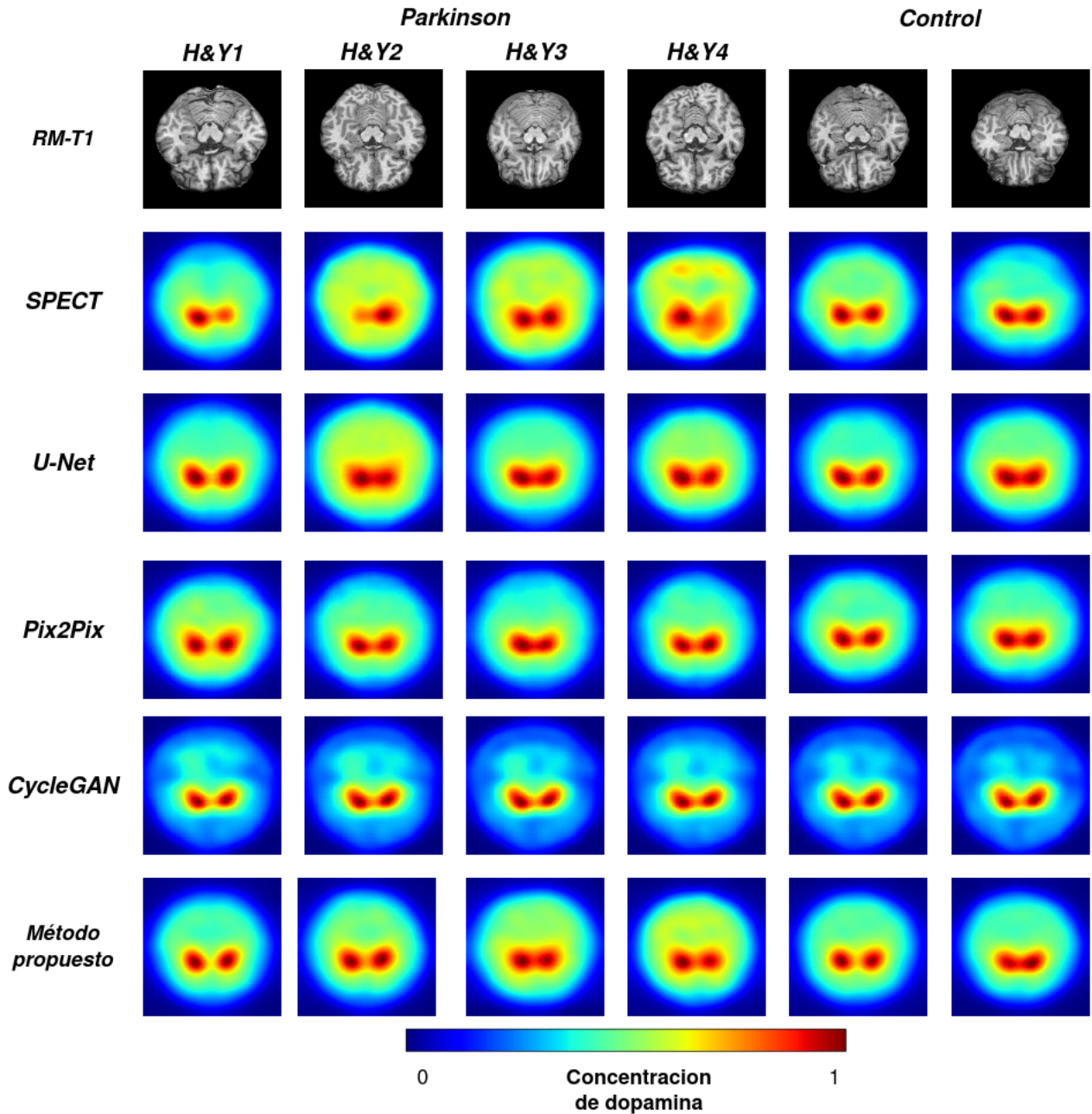


Tabla 5. Métricas de similitud sobre la generación de muestras SPECT sintéticas generadas a partir de distintas arquitecturas generativas.

Síntesis	Arquitectura	SSIM \uparrow	PSNR \uparrow
RM-T1 \rightarrow SPECT	U-Net	0.93 ± 0.03	19.38 ± 2.63
	Pix2Pix	0.93 ± 0.03	20.03 ± 2.50
	CycleGAN	0.87 ± 0.06	16.66 ± 2.70
	Método propuesto	0.94 ± 0.03	21.11 ± 3.00

Un análisis cuantitativo de la calidad de la síntesis lograda por cada una de las arquitecturas generativas se presenta en la Tabla 5. El método propuesto logra desempeños superiores sobre las métricas SSIM y PSNR superando a las demás arquitecturas por un 0.94 y un 21.11, respectivamente. Esto puede deberse a que el enfoque difusivo puede estar modelando de mejor manera los patrones globales propios de cada población pero sin lograr un nivel de representación detallado sobre los aspectos simétricos propios de cada estadio de la enfermedad. Esto guarda una coherencia con los resultados visuales donde al parecer la representación ajustada obedece a patrones dopaminérgicos promedio de las poblaciones presentes en el estudio.

Así mismo, las redes U-Net y Pix2Pix logran resultados SSIM de 0.93 en ambos casos, indicando un nivel alto de similitud estructural. Sin embargo, en términos de PSNR, Pix2Pix supera ligeramente a U-Net (20.03 vs 19.38), lo que sugiere una representación con menos presencia de ruido por parte de Pix2Pix. Esto puede explicarse principalmente por el diseño de su arquitectura, la cual, a diferencia de la red U-Net, tiene un discriminador dedicado a valorar la coherencia en la reconstrucción hacia el dominio interés de destino. Por otra parte, la red CycleGAN, logra un SSIM menor (0.87) y el valor más bajo de PSNR (16.66). Esto puede deberse a su enfoque no alineado de entrenamiento, donde los patrones dopaminérgicos son calculados sin tener en cuenta la dupla RM-T1 y SPECT para un mismos paciente, sugiriendo que dicha relación beneficia la complementariedad anatómico-funcional propia de los dominios donde los detalles finos propios de cada estadio de la enfermedad son relevantes en este contexto.

Tabla 6. Métricas de evaluación estructural (SSIM) y relación señal/ruido (PSNR) de secuencias SPECT generadas a partir de neuroimágenes RM-T1 en pacientes control y de diferentes estadios de la EP, enfocándose exclusivamente en las regiones correspondientes al cerebro.

Arquitectura	Métrica \uparrow	Control	H&Y 1	H&Y 2	H&Y 3	H&Y 4
U-Net	SSIM	0.93 ± 0.02	0.93 ± 0.01	0.91 ± 0.04	0.91 ± 0.00	0.92 ± 0.00
	PSNR	19.47 ± 2.55	20.67 ± 1.85	19.23 ± 2.99	18.55 ± 0.00	17.78 ± 0.00
Pix2Pix	SSIM	0.95 ± 0.02	0.94 ± 0.01	0.92 ± 0.04	0.91 ± 0.00	0.92 ± 0.00
	PSNR	20.85 ± 2.18	21.46 ± 1.46	18.95 ± 2.66	17.80 ± 0.00	17.89 ± 0.00
CycleGAN	SSIM	0.91 ± 0.02	0.86 ± 0.04	0.83 ± 0.06	0.81 ± 0.00	0.81 ± 0.00
	PSNR	18.53 ± 1.35	15.67 ± 1.90	14.67 ± 2.54	13.83 ± 0.00	13.49 ± 0.00
Método propuesto	SSIM	0.95 ± 0.01	0.93 ± 0.02	0.92 ± 0.03	0.91 ± 0.00	0.94 ± 0.00
	PSNR	23.11 ± 2.94	22.43 ± 2.51	20.73 ± 2.79	19.57 ± 0.00	22.74 ± 0.00

En un análisis detallado sobre la calidad de la síntesis en el área cerebral, la Tabla 6 presenta las métricas de evaluación estructural (SSIM) y relación señal/ruido (PSNR) de las secuencias SPECT generadas por cada una de las arquitecturas evaluadas a partir de neuroimágenes RM-T1. Este análisis se realizó considerando las poblaciones y en particular, los sujetos EP fueron valorados según sus estadios clínicamente confirmados según la escala Hoehn y Yahr. Notablemente, tanto el método propuesto como Pix2Pix logran un SSIM de 0.95 en el grupo control, destacándose como las mejores arquitecturas en términos de preservación de la estructura global. Sin embargo, el método propuesto muestra superioridad en PSNR (23.11 frente a 20.85), sugiriendo un mapeo con menos presencia de ruido en las señales de las características funcionales. En cuanto al análisis particular de cada una de las fases de la enfermedad, el método propuesto logra resultados competitivos de forma consistente a lo largo de los diferentes estadios. Sin embargo, respecto a los estadios 2 y 3, el método logra resultados similares a los logros por Pix2Pix. Estos resultados pueden deberse a que ambas arquitecturas preservan los componentes globales en la traducción pero el método propuesto logra codificaciones detalladas sobre los niveles dopaminérgicos particulares de cada población sin ser aún sensible a los diferentes estadios de la EP.

Dado los resultados anteriores, un análisis estadístico empleando el test de Mann-Whitney U sobre las métricas de similitud estructural SSIM y la relación señal/ruido PSNR. De

esta manera se determina que no existen diferencias estadísticamente significativas ($\rho > 0.05$) entre el método propuesto y la arquitectura generativa Pix2Pix en cuanto a la métrica SSIM. Por el contrario, en cuanto a PSNR, el método propuesto si presenta diferencias significativas ($\rho < 0.05$) en comparación con todas las demás arquitecturas, incluida Pix2Pix. Lo anterior sugiere comportamientos similares en las dos arquitecturas, utilizando el esquema experimental definido en este trabajo. Sin embargo como perspectivas del trabajo se hará necesario explorar escenarios con mayor numero de casos, para analizar la capacidad de generalización de las estrategias.

Adicionalmente, no se encontraron diferencias significativas ($\rho > 0.05$) en PSNR y SSIM al comparar los diferentes estadios de la EP generados por el método propuesto. La única excepción fue la métrica de SSIM, donde sí se encontraron diferencias significativas ($\rho < 0.05$) entre las secuencias SPECT obtenidas en pacientes del grupo control y aquellos en el estadio 2. Este hallazgo sugiere que, aunque el método propuesto mantiene una consistencia en la calidad de las imágenes a lo largo de los estadios clínicos, las diferencias estructurales entre los pacientes control y aquellos en el estadio 2 son lo suficientemente marcadas como para ser detectadas valoradas por SSIM.

Tabla 7. Comparación del rendimiento de diferentes modelos en la clasificación de secuencias SPECT generadas a partir de neuroimágenes RM-T1, evaluados en términos de exactitud, precisión, sensibilidad, y el puntaje F1 en sujetos control y pacientes con enfermedad de Parkinson (EP).

Modelo	Exactitud \uparrow	Precisión \uparrow	Sensibilidad \uparrow	Puntaje F1 \uparrow
RM-T1	56%	58%	56%	54%
U-Net	48%	47%	48%	46%
Pix2Pix	52%	51%	51%	51%
CycleGAN	52%	26%	50%	34%
Método Propuesto	54%	54%	53%	51%
SPECT	96%	97%	95%	96%

Dados los resultados anteriores, la Tabla 7 presentan los resultados de clasificación entre poblaciones EP y control de referencia empleando la modalidad RM-T1, sus diferentes esti-

maciones dopaminérgicas logradas por las diferentes arquitecturas y la secuencia SPECT de referencia. Los resultados validados en la tarea generativa resultan ser consistentes desde el punto de vista discriminativo por parte de la arquitectura propuesta donde los resultados de clasificación son superiores respecto a los de los demás enfoques generativos. A pesar de esto, los resultados logrados se encuentran muy por debajo de los SPECT de referencia, sugiriendo que aún existen limitaciones para realizar una discriminación de pacientes con EP.

7. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se implementó una estrategia generativa de aprendizaje profundo basada en un modelo de difusión condicional, con el objetivo de sintetizar secuencias diagnósticas complementarias en el contexto de la EP y el ACV. En particular, se emplearon configuraciones que priorizan neuroimágenes comúnmente disponibles en la práctica clínica para generar secuencias con mayor capacidad al soporte diagnóstico. Este enfoque fue entrenado y validado con dos conjuntos de datos de secuencias diagnósticas complementarias en el contexto de cada enfermedad. Particularmente, este método sigue un esquema de difusión y agregación de ruido, el cual se modela en el contexto de síntesis, siguiendo un proceso de difusión inversa condicional, permitiendo capturar y aprender relaciones complejas de traducción entre la neuroimagen de entrada y la de destino. Durante este proceso, un mecanismo de enmascaramiento aleatorio permitió forzar la arquitectura en aprender el contexto durante la tarea de traducción. El método de difusión fue comparado contra arquitecturas clásicas del estado del arte. La evaluación de las secuencias sintéticas generadas se realizó mediante métricas de similitud estructural y relación señal/ruido, enfocadas en la región cerebral.

Específicamente, para el ACV isquémico, se implementó una traducción secuencial NCCT \rightarrow DWI \rightarrow ADC, inspirados en el orden de captura de cada modalidad de imagen. Además, esta síntesis en cascada corresponde al estándar clínico, donde típicamente el mapa de ADC es generado desde el DWI. Los resultados obtenidos en la traducción de secuencias de ACV isquémico demostraron que la estrategia implementada tiene una notable capacidad para preservar detalles anatómicos y estructurales del cerebro. Esto se refleja en los valores de SSIM/PSNR de 0.55/18.80 para las imágenes DWI y 0.45/16.94 para las de ADC. Además, esta estrategia superó consistentemente a las arquitecturas clásicas en todas las métricas evaluadas para cada neuroimagen. Específicamente, se observó una diferencia promedio de 0.19 en SSIM y 7.01 en PSNR en la traducción hacia DWI, mientras que para ADC, las

diferencias fueron de 0.24 en SSIM y 6.25 en PSNR. Estas mejoras pueden atribuirse a la tendencia de los métodos clásicos a suavizar o perder detalles en regiones del cerebro, un problema identificado a través de evaluaciones visuales, mientras que el modelo implementado mostró una mayor robustez al representar estructuras cerebrales. Esta superioridad se mantuvo bajo diversos criterios, como la visibilidad y el tamaño de las lesiones en NCCT, así como en distintas fases del ACV, evidenciando la capacidad del modelo para preservar la anatomía cerebral en escenarios clínicos variados. Sin embargo, se observaron limitaciones en la preservación de las lesiones isquémicas en las secuencias sintéticas de DWI, donde no se logró reproducir con precisión los patrones radiológicos clave asociados al ACV, lo cual constituye un área de mejora en futuras investigaciones.

Con respecto a la EP, se implementó una traducción desde secuencias estructurales T1-RM hacia secuencias funcionales SPECT. En este contexto, se mostraron resultados destacables en el SSIM y el PSNR, lo cual podría atribuirse a la naturaleza funcional de la modalidad SPECT, que se dedica principalmente a resaltar el funcionamiento dinámico del cerebro en lugar de los detalles estructurales. La evaluación visual de las imágenes SPECT sintéticas reveló características relevantes, destacando la concentración de dopamina entre las diferentes poblaciones tenidas en cuenta durante el estudio e incluso presentando diferentes patrones en los hemisferios cerebrales según el estadio de la enfermedad, sin éstos llegar a ser lo suficientemente detallados. El modelo de difusión condicional implementado continuó superando a las arquitecturas clásicas del estado del arte, aunque con diferencias menos pronunciadas, observándose una diferencia promedio de 0.03 en SSIM y 2.42 en PSNR. Cabe destacar que, al comparar con Pix2Pix, no se encontraron diferencias estadísticas significativas. En cuanto a la evaluación estratificada por estadios de la enfermedad, según la escala de Hoehn y Yahr, los resultados indicaron que no hubo diferencias significativas entre las imágenes generadas para pacientes control y aquellos en diferentes estadios de la EP, lo que demuestra la consistencia del modelo en la traducción de neuroimágenes. Sin embargo, al evaluar la utilidad

de las imágenes SPECT sintéticas en una tarea de clasificación entre sujetos control y pacientes con Parkinson, se presentaron limitaciones, como lo muestra un bajo puntaje F1 de 0.51. Esto sugiere que, aunque el método de difusión condicional produce mejores imágenes que las otras arquitecturas, su utilidad en tareas de diagnóstico sigue siendo limitada y está lejos del valor clínico que aporta la modalidad SPECT original. Esta limitación puede estar asociada a la dificultad de aproximar una función que establezca relaciones entre información estructural y funcional del cerebro.

Como trabajo futuro, se podrán explorar y modelar mecanismos alternativos para lograr mayor robustez en las síntesis generadas. Por ejemplo, usando información clínica que permita orientar el proceso de síntesis en el contexto de ambas enfermedades. Se explorarán mecanismos de difusión que incorporen información imagenológica multimodal, con el objetivo de enriquecer la síntesis y mejorar la representación de los patrones radiológicos asociados a cada patología. Además, se investigarán mecanismos de traducción utilizando modelos de difusión para generar múltiples imágenes sin necesidad de pasos intermedios. Asimismo, se deben explorar estrategias que resalten los patrones imagenológicos de cada enfermedad para centrar el aprendizaje en zonas de interés. Finalmente, el estudio se beneficiaría de ampliar los grupos de estudio, con el fin de aumentar la representatividad, precisión y poder estadístico, al tiempo que reducir el sesgo y facilitar análisis más detallados.

BIBLIOGRAFÍA

ABU-SRHAN et al. “Paired-unpaired Unsupervised Attention Guided GAN with transfer learning for bidirectional brain MR-CT synthesis”. In: *Computers in Biology and Medicine* 136 (2021), p. 104763 (cit. on p. 29).

ARMANIOUS et al. “MedGAN: Medical image translation using GANs”. In: *Computerized medical imaging and graphics* 79 (2020), p. 101684 (cit. on p. 30).

AYAZ et al. “Automated methods for diagnosis of Parkinson’s disease and predicting severity level”. In: *Neural Computing and Applications* 35.20 (2023), pp. 14499–14534 (cit. on pp. 20, 21).

BERCOVICH, E and JAVITT, MC. *Medical Imaging: From Roentgen to the Digital Revolution, and Beyond*. *Rambam Maimonides Medical Journal*, 9, e0034. 2018 (cit. on p. 34).

BIDESI et al. “The role of neuroimaging in Parkinson’s disease”. In: *Journal of neurochemistry* 159.4 (2021), pp. 660–689 (cit. on p. 21).

CHOUGAR et al. “The role of magnetic resonance imaging for the diagnosis of atypical parkinsonism”. In: *Frontiers in Neurology* 11 (2020), p. 665 (cit. on p. 23).

COHEN et al. “Distribution matching losses can hallucinate features in medical image translation”. In: *MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer. 2018, pp. 529–536 (cit. on p. 34).

COHENPOUR, Mehrzad and GOLAN, Haim. “Nuclear neuroimaging of dopamine transporter in Parkinsonism—role in routine clinical practice”. In: *Harefuah* 146.9 (2007), pp. 698–702 (cit. on p. 23).

DAYARATHNA et al. “Deep learning based synthesis of MRI, CT and PET: Review and analysis”. In: *Medical Image Analysis* (2023), p. 103046 (cit. on pp. 27, 35, 38).

DORSEY et al. “The emerging evidence of the Parkinson pandemic”. In: *Journal of Parkinson’s disease* 8.s1 (2018), S3–S8 (cit. on p. 13).

FEIGIN et al. “World Stroke Organization (WSO): global stroke fact sheet 2022”. In: *International Journal of Stroke* 17.1 (2022), pp. 18–29 (cit. on p. 13).

GAJJAR et al. “An Empirical Analysis of Diffusion, Autoencoders, and Adversarial Deep Learning Models for Predicting Dementia Using High-Fidelity MRF”. In: *IEEE Access* (2024) (cit. on p. 32).

GARZÓN et al. “A deep CT to MRI unpaired translation that preserve ischemic stroke lesions”. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2022, pp. 2708–2711 (cit. on p. 31).

GONZÁLEZ et al. *Acute ischemic stroke*. Springer, 2011 (cit. on p. 17).

GUTIERREZ et al. “Lesion-preserving unpaired image-to-image translation between MRI and CT from ischemic stroke patients”. In: *International Journal of Computer Assisted Radiology and Surgery* 18.5 (2023), pp. 827–836 (cit. on pp. 32, 34).

HEIM et al. “Structural imaging in atypical parkinsonism”. In: *International Review of Neurobiology* 142 (2018), pp. 67–148 (cit. on pp. 21, 22).

HO, Jonathan; JAIN, Ajay, and ABBEEL, Pieter. “Denoising diffusion probabilistic models”. In: *Advances in neural information processing systems* 33 (2020), pp. 6840–6851 (cit. on pp. 38, 39).

KHACHNAOUI, Hajer; MABROUK, Rostom, and KHLIFA, Nawres. “Machine learning and deep learning for clinical data and PET/SPECT imaging in Parkinson’s disease: a review”. In: *IET Image Processing* 14.16 (2020), pp. 4013–4026 (cit. on p. 14).

LI et al. “Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy”. In: *Quantitative imaging in medicine and surgery* 10.6 (2020), p. 1223 (cit. on p. 29).

LIANG, Ziyun et al. “Modality cycles with masked conditional diffusion for unsupervised anomaly segmentation in mri”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 168–181 (cit. on p. 41).

LIN, Michelle P and LIEBESKIND, David S. “Imaging of ischemic stroke”. In: *Continuum: Lifelong Learning in Neurology* 22.5 (2016), p. 1399 (cit. on pp. 14, 15, 18, 19, 34).

MAKKAT et al. “Signs of acute stroke seen on fluid-attenuated inversion recovery MR imaging”. In: *American Journal of Roentgenology* 179.1 (2002), pp. 237–243 (cit. on p. 19).

MORTEZAZADEH et al. “Imaging modalities in differential diagnosis of Parkinson’s disease: opportunities and challenges”. In: *Egyptian journal of radiology and nuclear medicine* 52 (2021), pp. 1–12 (cit. on p. 24).

NIE et al. “Medical image synthesis with deep convolutional adversarial networks”. In: *IEEE Transactions on Biomedical Engineering* 65.12 (2018), pp. 2720–2730 (cit. on p. 24).

ÖZBEY et al. “Unsupervised medical image translation with adversarial diffusion models”. In: *IEEE Transactions on Medical Imaging* (2023) (cit. on p. 30).

PINON et al. “Brain subtle anomaly detection based on auto-encoders latent space analysis: application to de novo parkinson patients”. In: IEEE. 2023, pp. 1–5 (cit. on p. 32).

POEWE et al. “Parkinson disease”. In: *Nature reviews Disease primers* 3.1 (2017), pp. 1–21 (cit. on p. 21).

SAEED et al. “Imaging biomarkers in Parkinson’s disease and Parkinsonian syndromes: current and emerging concepts”. In: *Translational neurodegeneration* 6 (2017), pp. 1–25 (cit. on p. 23).

SCHULZ et al. “Magnetic resonance imaging–based volumetry differentiates idiopathic Parkinson’s syndrome from multiple system atrophy and progressive supranuclear palsy”. In: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 45.1 (1999), pp. 65–74 (cit. on p. 23).

SHARMA, Vijay K. “Cerebrovascular Disease”. In: *International Encyclopedia of Public Health (Second Edition)*. Ed. by Stella R. QUAH. Second Edition. Oxford: Academic Press, 2017, pp. 455–470. DOI: <https://doi.org/10.1016/B978-0-12-803678-5.00058-8> (cit. on p. 13).

SIVARANJINI, S and SUJATHA, CM. “Deep learning based diagnosis of Parkinson’s disease using convolutional neural network”. In: *Multimedia tools and applications* 79.21 (2020), pp. 15467–15479 (cit. on p. 14).

STEINMETZ et al. “Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021”. In: *The Lancet Neurology* (2024) (cit. on p. 13).

TOLOSA, Eduardo et al. “Challenges in the diagnosis of Parkinson’s disease”. In: *The Lancet Neurology* 20.5 (2021), pp. 385–397 (cit. on p. 21).

WANG et al. “Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks”. In: *Medical Image Analysis* 65 (2020), p. 101787 (cit. on p. 31).

— “Quantitative Cerebral Blood Volume Image Synthesis from Standard MRI Using Image-to-Image Translation for Brain Tumors”. In: *Radiology* 308.2 (2023), e222471 (cit. on p. 29).

— “SPECT molecular imaging in Parkinson’s disease”. In: *BioMed Research International* 2012 (2012) (cit. on p. 23).

WATANABE et al. “Generative image transformer (GIT): unsupervised continuous image generative and transformable model for [123 I] FP-CIT SPECT images”. In: *Annals of nuclear medicine* 35 (2021), pp. 1203–1213 (cit. on p. 33).

WINTERMARK et al. “Imaging recommendations for acute stroke and transient ischemic attack patients: a joint statement by the American Society of Neuroradiology, the American College of Radiology, and the Society of NeuroInterventional Surgery”. In: *American Journal of Neuroradiology* 34.11 (2013), E117–E127 (cit. on pp. 14, 17, 19).

YANG et al. “MRI cross-modality image-to-image translation”. In: *Scientific reports* 10.1 (2020), p. 3753 (cit. on pp. 15, 34).

YOO et al. “Image-to-image translation using a cross-domain auto-encoder and decoder”. In: *Applied Sciences* 9.22 (2019), p. 4780 (cit. on p. 24).

ZENG et al. “3D CVT-GAN: a 3D convolutional vision transformer-GAN for PET reconstruction”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pp. 516–526 (cit. on p. 30).