

# Estimación del mapa de anisotropía fraccional y difusividad media en materia blanca utilizando transformers

Daniel Bandala Álvarez, Jorge Perez Gonzalez

Universidad Nacional Autónoma de México,  
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas,  
México

ing.dbandala@gmail.com, jorge.perez@iimas.unam.mx

**Resumen.** Los algoritmos de visión computacional se han convertido en una tecnología clave en muchos campos, desde sistemas de asistencia al conductor para automóviles hasta interacciones de usuario en videojuegos. Así, un campo que ha encontrado aplicaciones importantes y que ha impulsado todo una nueva área en la ingeniería es el procesamiento de imágenes médicas, ya sea para simplemente limpiar las componentes de ruido en las muestras obtenidos o para extraer e inferir conocimiento nuevo a partir de estas, tal como en la detección temprana de cáncer de mama, el diagnóstico exacto de retinopatía diabética o detección y localización de tumores. Esto último respalda la opción de utilizar algoritmos de aprendizaje computacional y redes neuronales profundas para la asistencia e investigación médica. En el presente trabajo se propone utilizar una red neuronal tipo autoencoder basada en módulos Transformer para la estimación de los mapas de anisotropía fraccional y difusividad media. Se presentan los resultados obtenidos sobre un conjunto de sujetos control y se demuestra que el uso de arquitecturas basadas en autoatención son capaces de modelar estructuras altamente complejas como las fibras de axones en el cerebro.

**Palabras clave:** Tensor de difusión, materia blanca, anisotropía fraccional, difusividad media, transformers.

## Fractional Anisotropy and Mean Diffusivity Maps Estimation in White Matter Using Transformers

**Abstract.** Computer vision algorithms have become a key technology in many fields, from driver assistance systems for automotive to user interactions in video games. Among these fields, medical image processing has particularly benefited from the application of these techniques. By utilizing computer vision algorithms, it is possible to enhance medical images by removing noise components and extracting valuable information from them. This, in turn, enables early detection of breast cancer, precise diagnosis of diabetic retinopathy, and accurate detection and localization of tumors. Consequently, the integration of machine learning

algorithms and deep neural networks has become pivotal in both medical care and research. In this study, we propose a novel approach to medical image processing utilizing an autoencoder neural network. The key innovation lies in the incorporation of Transformer modules within the network architecture for the estimation of fractional anisotropy and mean diffusivity maps. The results obtained on a set of control subjects are presented and it is demonstrated that the implementation of self-attention based architectures are able to model highly complex structures such as axon fibers across the brain tissue.

**Keywords:** Diffusion tensor, white matter, fractional anisotropy, mean diffusivity, transformers.

## 1. Introducción

Las neuroimágenes son una herramienta estándar en el estudio clínico de pacientes con patologías neurodegenerativas o en el análisis exploratorio prequirúrgico para cirugías cerebrales. En el área de la investigación, se han realizado avances contundentes debido al desarrollo de técnicas de extracción y procesamiento de datos a partir de imágenes estructurales y funcionales del cerebro [19].

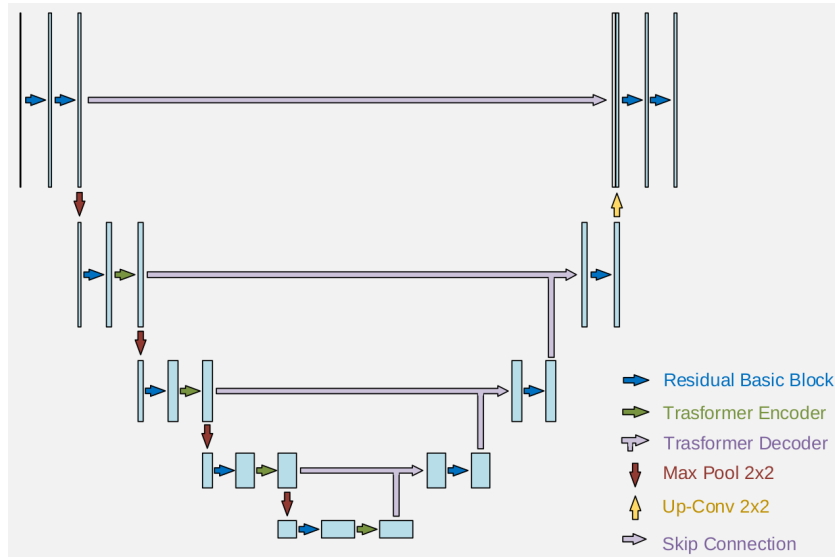
En concreto, las imágenes ponderadas por difusión (MRI-DWI) permiten estudiar la estructura del tejido de la materia blanca debido a su alta sensibilidad al desplazamiento de las moléculas de agua con una precisión del orden de micras. Debido a la energía térmica, estas moléculas se encuentran moviéndose constantemente en direcciones aleatorias en el tejido cerebral en un proceso denominado autodifusión [8, 9].

Y, debido a que las estructuras celulares pueden impedir el movimiento del agua a una escala microscópica, estas señales actúan como evidencia de la microestructura del tejido. Existen varias metodologías para modelar la difusión en sistemas biológicos, cada uno con distintas suposiciones y nivel de complejidad.

El modelo más simple es asumir una difusión libre, caracterizada por un único coeficiente de difusión. Sin embargo, debido a que las medidas de la difusividad claramente dependen de los parámetros experimentales, se incorporó el uso del coeficiente de difusión aparente (ADC) [5], en donde se calcula la difusividad relativa a una señal sin dirección de difusión del mismo escáner.

Por otra parte, la materia blanca cerebral se compone principalmente de axones, que son responsables de transmitir información eléctrica y química a través del cerebro. Son largas proyecciones celulares que se extienden desde las células nerviosas, o neuronas, y permiten que las señales eléctricas se propaguen entre terminales nerviosas, donde se comunican con otras neuronas o células del cuerpo.

Estos axones cerebrales están organizados en fascículos o haces, que a menudo se denominan tractos. Estos tractos están recubiertos por una capa de mielina, un material graso que actúa como un aislante eléctrico y acelera la velocidad de transmisión de los impulsos nerviosos. Además, se ha demostrado que la difusión en la materia blanca cerebral es dependiente de la dirección [15], lo que contiene información con la orientación de las fibras de axones dentro de las regiones de materia blanca en conjuntos de fibras coherentes.



**Fig.1.** Arquitectura UTNet implementada para la estimación de los mapas del tensor de difusión [4].

Esta información direccional de las señales de difusión presenta la única posibilidad de inferir las fibras de materia blanca dentro de un cerebro de un paciente vivo y de forma completamente no invasiva. Por tanto, el modelo del tensor de difusión se ha introducido como una extensión del modelo del coeficiente de difusión aparente con la capacidad de describir la difusión anisotrópica, es decir, la difusión que no es igual en todas las direcciones, y del cual se pueden extraer medidas como la difusividad media, la anisotropía fraccional y las direcciones principales de la difusión [11]. Gracias a la simplicidad de esta técnica, es utilizada como un estándar desde hace un par de décadas en especial en el área clínica.

### 1.1. Estado del arte

En general, los algoritmos para el seguimiento de los tractos o tractografía son simples, aunque sujetos a un número significativo de limitaciones, por lo tanto resulta interesante el desarrollo de nuevos modelos o técnicas computacionales que permitan aumentar las capacidades y reducir la complejidad de los métodos convencionales para el ajuste del tensor de difusión en la microestructura de tejidos biológicos [20].

En [18] se presenta una arquitectura que utiliza tres modelos de redes neuronales convolucionales para la estimación de la orientación principal de la difusión, la segmentación de los tractos de interés y la segmentación de los puntos de inicio y finalización del seguimiento de la tractografía.

En [7] se presenta un modelo similar basado en redes neuronales convolucionales que ajusta la relación no lineal entre las imágenes ponderadas por difusión y sus correspondientes mapas de difusión, utilizando únicamente 6 señales de difusión.

De manera similar, en [10] proponen una red neuronal tipo encoder-decoder llamada AGYnet para segmentar la materia blanca en el cerebro, esta toma como información de entrada imágenes estructurales ponderadas en tiempo T1 y los mapas de las direcciones principales de difusión. La arquitectura cuenta con compuertas de atención en la parte de decoder que permite combinar las características obtenidas de cada tipo de entrada.

Esto permite mantener las características originales y resaltar aquellas que realicen una mayor contribución de información a la estimación del modelo [13]. En consecuencia, gracias a estos mecanismos de auto atención ha surgido una alternativa a las redes neuronales convolucionales para el procesamiento de imágenes: los Transformers [16].

Así, se han presentado propuestas donde se combinan bloques convolucionales y bloques tipo Transformers para la segmentación de imágenes médicas [4, 1]. En ese contexto, Karimi et al. [6] han propuesto una arquitectura basada únicamente en Transformers para la estimación de los coeficientes del tensor de difusión obteniendo buenos resultados.

Y, de la misma forma, en el presente trabajo se propone un modelo de redes neuronales basado en Transformers y bloques convolucionales para la estimación de los mapas de anisotropía fraccional y difusividad media. A su vez, en comparación al método convencional, este modelo reduce significativamente el tiempo de procesamiento y el número de señales de difusión necesarias para la estimación de estos mapas.

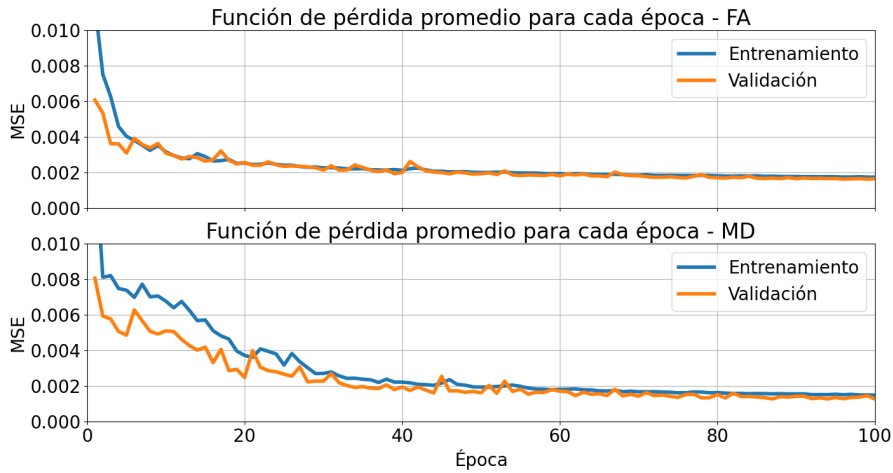
## **2. Materiales y métodos**

En esta sección se presenta la metodología de desarrollo del presente trabajo de investigación. Se describe la base de datos utilizada para la construcción del modelo y se muestran las características de la arquitectura implementada para la estimación de los mapas del tensor de difusión. Finalmente, se mencionan las métricas utilizadas para la evaluación de la arquitectura.

### **2.1. Base de datos**

La base de información de imágenes ponderadas por difusión se obtienen de dos proyectos incluidos en el Laboratorio de Neuroimágenes en la Universidad del Sur de California (USC): El Proyecto de Conectoma Humano (Human Connectome Project, HCP) y la Iniciativa de Neuroimágenes de la Enfermedad de Alzheimer (Alzheimer's Disease Neuroimaging Initiative, ADNI). La base de información de la Iniciativa de Neuroimágenes para la Enfermedad de Alzheimer cuenta con más de tres mil imágenes cerebrales de Resonancia Magnética, Tensor de Difusión, Tomografía Computarizada y por Emisión de Positrones.

De este conjunto de información se han recopilado muestras de 30 sujetos control con 38 señales de difusión cada una. Cada volumen contiene 80 imágenes de  $256 \times 256$  píxeles. Por otra parte, el Proyecto de Conectoma Humano (HCP) contiene muestras de 35 sujetos control en un rango de edad entre 20 y 89 años.



**Fig. 2.** Función de pérdida de las primeras 100 épocas durante el entrenamiento del modelo para la estimación de los mapas del tensor de difusión.

Cada volumen contiene 96 imágenes de  $140 \times 140$  píxeles, en donde las señales de difusión se han adquirido en cortes axiales oblicuos y con gradientes de difusión monopolares. Por tanto, en total se utilizan las muestras de 65 sujetos control para el entrenamiento, validación y evaluación del modelo.

## 2.2. Procesamiento con FSL

Para la construcción de la base de datos de referencia se utilizó el procesamiento estándar con la librería FSL (FMRIB Software Library) [3]. Este procedimiento consiste en extraer el conjunto de todas las imágenes que han sido obtenidas con un único valor de gradiente  $b$ , filtrar artificios generados por el propio escáner y, finalmente, ajustar el tensor de difusión mediante una método de mínimos cuadrados.

Además, se extrae todo el tejido no cerebral de las imágenes, mediante una máscara binaria, de modo que se ajusta el tensor de difusión para la región del cerebro únicamente. Así, a partir de cada imagen de tensor de difusión se ha extraído el mapa de fracción de anisotropía y difusividad media, los cuales servirán como referencia para el entrenamiento y validación del modelo propuesto.

## 2.3. Arquitectura de red neuronal

El diseño del modelo propuesto para la estimación de los mapas del tensor de difusión en imágenes cerebrales se basa en la arquitectura UTNet para segmentación binaria, presentada en [4], la cual se compone de bloques convolucionales, módulos Transformer, bloques de submuestreo y capas totalmente conectadas. Este modelo contiene una parte de contracción o compresión, como se muestra en la Figura 1, donde se reduce la dimensión espacial de las imágenes y se incrementa el número de canales o características; y una fase de expansión, empleada para la reconstrucción de las imágenes de referencia a sus dimensiones espaciales originales.

En la etapa de codificación de las imágenes se encuentran dos bloques convolucionales con conexión residual que extraen las primeras características locales de las señales de difusión y aumentan en un factor de 4 el número de canales de entrada de las imágenes. Se extraen estas últimas características con un submuestreo del máximo valor, utilizando una ventana de tamaño  $2 \times 2$ , y se introducen estos tensores de características en el primer codificador tipo Transformer.

Este toma las dimensiones espaciales de la imagen para proyectarlas sobre una dimensión  $x_{t_n} \in \mathbb{R}^{HW \times d}$ , donde  $HW$  representa el tamaño de la imagen y  $d$  es la dimensión del mapa de características, obteniendo así una secuencia de características de píxeles que capturan las principales propiedades de las imágenes.

Se repite este proceso desde el submuestreo hasta el codificador Transformer en 3 niveles y, en este punto, las dimensiones espaciales de las imágenes se ha reducido en un factor de 16, mientras que el número de canales o características ha aumentado en la misma proporción.

Entonces, a partir del espacio latente a la salida de la etapa de codificación del modelo, se comienza la reconstrucción de las imágenes de salida utilizando decodificadores tipo Transformer, el cual toma dos tensores de entrada, correspondientes a las características del espacio latente de la etapa de contracción y las características de mayor resolución mediante una conexión residual.

Así, en lugar de integrar el módulo de autoatención sobre los mapas de características que se han extraído de manera automática o manual como se plantea en la arquitectura ViT [2] y TransUNet [1], se aplica el módulo Transformer a cada nivel del codificador y decodificador para modelar la dependencia de largo alcance en múltiples escalas.

Además, dado que las imágenes son datos muy estructurados, la mayoría de los píxeles de los mapas de características de alta resolución dentro de una ventana local comparten características similares, excepto en el caso de las regiones limitantes. En consecuencia, el cálculo de la atención por pares entre todos los píxeles es ineficiente y redundante. Desde una perspectiva teórica, la autoatención es esencialmente de bajo rango para las secuencias largas [17], lo que indica que la mayor parte de la información se concentra en los valores singulares más grandes.

Por tanto, se utilizan los bloques de multiatención eficientes empleados en [14], en donde se aplican dos transformaciones para proyectar la clave y el valor del mapa de características  $K, V \in \mathbb{R}^{n \times d}$  en una codificación de menor dimensión  $\bar{K}, \bar{V} \in \mathbb{R}^{k \times d}$ , en donde  $k = hw \ll n$ ,  $h$  y  $w$  es el tamaño espacial reducido del mapa de características después del submuestreo. Entonces, la autoatención eficiente presentada en [4] se formula como:

$$\text{EfficientAttention}(Q, \bar{K}, \bar{V}) = \text{softmax} \left( \frac{Q\bar{K}}{\sqrt{d}} \right) \bar{V}. \quad (1)$$

De esta manera, la complejidad computacional del cálculo de los mapas de atención se reduce a  $O(nkd)$ . En particular, la proyección a una menor dimensión puede ser mediante cualquier operación de muestreo descendente. En este caso, se emplea una convolución con tamaño de kernel  $1 \times 1$  para reducir la muestra del mapa de características.

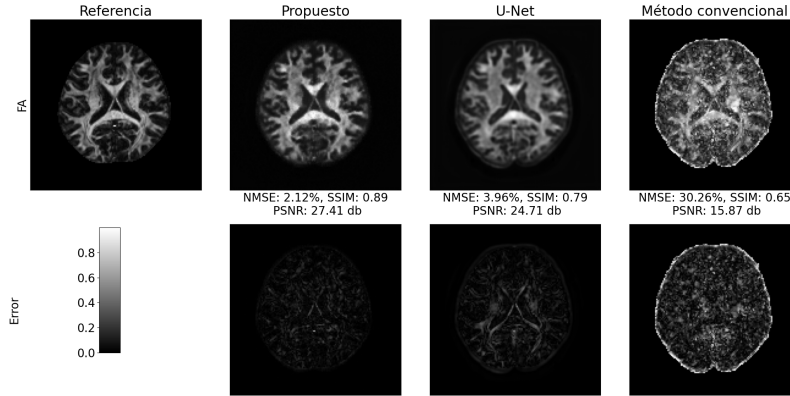


Fig. 3. Ejemplo de resultados obtenidos para mapa de anisotropía fraccional.

## 2.4. Validación

Para evaluar el desempeño del modelo se utilizan las métricas descritas en [7], que son el error cuadrático medio normalizado respecto al máximo valor de todos los errores para un mapa del tensor de difusión (NMSE); el índice de similitud estructural (SSIM), el cual se basa en comparar la luminancia, el contraste y la correlación estadística de los píxeles de una imagen con una imagen de referencia; y la relación de señal a ruido máxima (PSNR), utilizado para medir el desempeño en la limpieza de ruido de las imágenes de entrada del modelo.

## 3. Resultados y discusión

Todo el conjunto de imágenes de la base de datos que se ha recopilado se ha procesado utilizando la herramienta FSL. Entonces, para cada uno de los 65 casos control se han obtenido los volúmenes de anisotropía fraccional y difusividad media. Se ha utilizado el 80 % de los datos para el entrenamiento del modelo, el 10 % para validación y 10 % para una evaluación final.

En la implementación del modelo y la obtención de los resultados que se muestran a continuación, se ha utilizado el servicio de computación en la nube de Google Colabs, el cual proporciona un entorno Linux que cuenta con 32GB de memoria RAM y una tarjeta gráfica GPU Nvidia Tesla T4 con 16GB de memoria. Una ventaja de utilizar este servicio es la posibilidad de comunicar el entorno con cualquier repositorio público o privado, por lo que no es necesario copiar la base de datos utilizada para el entrenamiento al entorno del servicio.

Entonces, de acuerdo a la arquitectura del modelo que se ha implementado, es necesario realizar un acondicionamiento de la información original de la base de datos. Para empezar, se realiza un reordenamiento de las dimensiones de los volúmenes de las señales de difusión obtenidas por el escáner de resonancia magnética, los cuales consisten originalmente en un tensor de 4 dimensiones  $[h, w, d, c]$ , en donde  $h$  es la altura de las imágenes,  $w$  es el ancho de las imágenes,  $d$  es la profundidad del volumen y  $c$  es el número de señales de difusión.

**Tabla 1.** Métricas promedio obtenidas para la inferencia del conjunto de evaluación.

Mapa - Método	NMSE (%)	SSIM	PSNR (db)	Tiempo (min)
FA				
Propuesto	8.5307	0.8552	27.8626	2.74
UNet	10.8327	0.8112	26.6207	47.61
Método convencional	–	0.7708	17.2302	18.98
MD				
Propuesto	5.1540	0.9094	27.9955	3.02
UNet	7.1707	0.9166	29.0652	53.89
Método convencional	11.0139	0.8789	27.0625	18.98

Esta transformación se efectúa de manera que el tensor resultante tiene las dimensiones  $[d, c, h, w]$ . En cambio, el tensor de difusión tiene solo 6 grados de libertad, por lo que teóricamente es necesario únicamente 6 señales de difusión para calcular los coeficientes del tensor en cada punto o voxel del volumen. Por consiguiente, se toman un total de 7 señales de difusión para la estimación de los mapas del tensor de difusión: 6 señales de difusión en direcciones equitativamente distribuidas en la esfera y una imagen sin dirección de gradiente de magnetización.

El modelo se ha implementado utilizando la biblioteca de aprendizaje automático PyTorch, que se basa en la librería Torch de Linux Foundation umbrella. Para ajustar los parámetros del modelo se ha empleado un algoritmo de optimización estocástico basado en gradiente. Asimismo, se utiliza la desviación cuadrática media entre la salida del modelo y la referencia como función de pérdida a optimizar por el algoritmo antes mencionado. Así, la función de pérdida durante el entrenamiento se encuentra definida por la siguiente expresión:

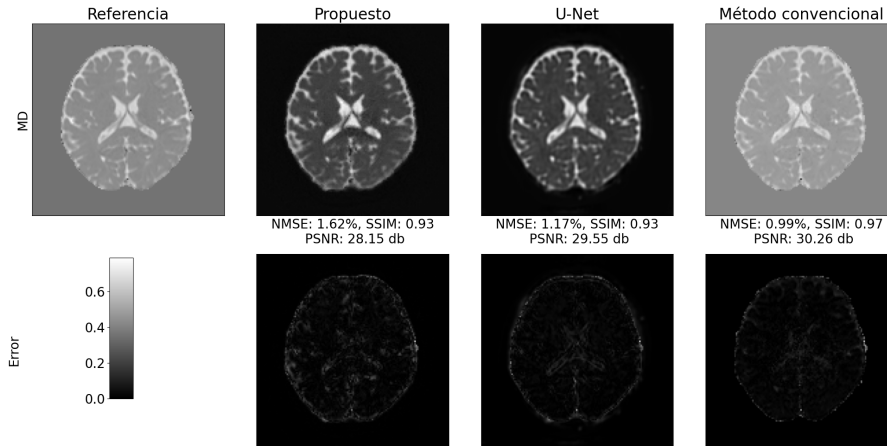
$$\text{MSE}(\Theta) = \frac{1}{n} \sum_{t=1}^n \|F(x^t; \Theta) - y^t\|^2, \quad (2)$$

donde  $F(\cdot)$  representa la función del modelo de la red neuronal y  $\Theta$  denota los parámetros de la red que se ajustan durante el entrenamiento. El modelo se entrena durante 140 épocas con un ritmo de aprendizaje inicial de 0.0002 y se reduce en un factor de 0.9 cada que la función de pérdida no disminuye durante una época. Asimismo, se ha realizado un aumento de datos durante el entrenamiento en donde se han rotado o reflejado horizontalmente y verticalmente las imágenes de manera aleatoria.

También, se ha entrenado el modelo individualmente para ajustar ambos mapas del tensor de difusión. En las gráficas de la Figura 2 se muestran la función de pérdida de las primeras 100 épocas durante el entrenamiento del modelo para el ajuste de los mapas de anisotropía fraccional y difusividad media.

Por otro lado, para evaluar el efecto de utilizar módulos Transformer en una arquitectura tipo autoencoder, se ha implementado una arquitectura UNet [12], la cual se compone únicamente de módulos convolucionales, para la estimación de los mapas del tensor de difusión. Y, además, se han obtenido estos mapas utilizando el método convencional pero con las mismas 7 señales de difusión que utiliza el modelo propuesto.





**Fig. 4.** Ejemplo de resultados obtenidos para mapa de difusividad media.

En la Figura 3 se presenta un ejemplo del mapa de anisotropía fraccional obtenido por el modelo propuesto, el modelo UNet y el método convencional, junto con los valores obtenidos de las métricas de desempeño y su correspondiente mapa de error absoluto para cada resultado. Es evidente, en este caso, que el modelo propuesto proporciona el mejor desempeño para la estimación de este mapa del tensor de difusión.

En la Figura 4 se muestra un ejemplo del mapa de difusividad media obtenido con el modelo propuesto y las dos metodologías de comparación mencionadas anteriormente, junto con sus correspondientes métricas de desempeño. En esta instancia, el modelo basado en módulos convolucionales ha logrado superar el desempeño del modelo propuesto y, aun más, el método convencional ha arrojado el mejor resultado para la estimación de este mapa, con un índice de similitud estructural de 0.97.

Esto último sugiere que la estimación del mapa de difusividad media es una tarea de menor complejidad que la estimación del mapa de anisotropía fraccional y no es necesaria la implementación de módulos basados en autoatención para mejorar la estimación de este mapa. De manera similar, en la Tabla 1 se presentan los valores promedio de las métricas de desempeño y el tiempo de procesamiento de un volumen completo obtenidos con los tres métodos a partir de todas las imágenes del conjunto de evaluación.

Entonces, para la estimación del mapa de anisotropía fraccional se ha obtenido un índice de similitud estructural promedio de 0.85, superando el modelo basado únicamente en bloques convolucionales y el método convencional, de los cuales se ha obtenido un índice de similitud estructural promedio de 0.81 y 0.65, respectivamente.

Además, el modelo propuesto es el que menor tiempo requiere para procesar un volumen completo de un paciente utilizando CPU, lo que abre la posibilidad de emplear este modelo en el área clínica. Por el contrario, para la estimación del mapa de difusividad media el modelo UNet ha generado el mejor desempeño, con un índice de similitud estructural promedio de 0.91, lo que sugiere que no es necesario el uso de bloques de autoatención para modelar imágenes con características altamente homogéneas, como es el caso de la segmentación binaria de imágenes.

En particular, la microestructura de las fibras de materia blanca forman una complicada red en todo el cerebro con características altamente estructuradas y, debido a esto, el modelar mapas de características a partir de imágenes de resonancia magnética ponderadas por difusión es todo un reto computacional. Asimismo, es importante destacar que la estimación del mapa de anisotropía fraccional y difusividad media es una tarea crucial en el procesamiento de imágenes de DTI, ya que estos mapas son una representación visual de la microestructura de los tejidos biológicos y proporcionan información importante sobre la integridad de las fibras nerviosas y la conectividad del cerebro.

Por tanto, el uso de redes neuronales profundas basadas en autoatención para la estimación de estos parámetros tiene varias ventajas en comparación con los métodos tradicionales de procesamiento de DTI, puesto que el modelo ha sido capaz de modelar características complejas y no lineales en la estructura de los mapas a partir de únicamente 6 señales de difusión, aún cuando la base de información utilizada para optimizar el modelo ha sido limitada. Esto representa un avance en el uso e implementación de modelos basados en Transformers para extracción de imágenes médicas en entornos clínicos, ya que la precisión y velocidad en la estimación de estos mapas es crucial para la detección temprana de enfermedades cerebrales y para el monitoreo de la progresión de estas mismas.

#### **4. Conclusiones**

En el presente trabajo se ha desarrollado la implementación de una variante del modelo U-Net para la estimación del mapa de anisotropía fraccional y la difusividad media en la materia blanca cerebral. Los resultados obtenidos demuestran que el modelo propuesto es una herramienta prometedora para la estimación de los mapas del tensor de difusión y, en general, para el procesamiento de imágenes médicas. Además, las arquitecturas compuestas por Transformers tienen la ventaja de ser altamente escalables y eficientes en términos de recursos computacionales. Esto resulta ideal para aplicaciones en tiempo real y el procesamiento de grandes conjuntos de datos.

**Agradecimientos.** Este trabajo ha sido apoyado por el programa UNAM-PAPIIT IA104622.

#### **Referencias**

1. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., Zhou, Y.: TransUNet: Transformers make strong encoders for medical image segmentation (2021) doi: 10.48550/ARXIV.2102.04306
2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16×16 words: Transformers for image recognition at scale (2021) doi: 10.48550/ARXIV.2010.11929

3. Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., Gorgolewski, K. J.: fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, vol. 16, no. 1, pp. 111–116 (2018) doi: 10.1038/s41592-018-0235-4
4. Gao, Y., Zhou, M., Metaxas, D.: Utnet: A hybrid transformer architecture for medical image segmentation (2021) doi: 10.48550/ARXIV.2107.00781
5. Jones, D. K., Knösche, T. R., Turner, R.: White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. *NeuroImage*, vol. 73, pp. 239–254 (2013) doi: 10.1016/j.neuroimage.2012.06.081
6. Karimi, D., Gholipour, A.: Diffusion tensor estimation with transformer neural networks. *arXiv:220105701v1*, pp. 102330 (2022) doi: 10.1016/j.artmed.2022.102330
7. Li, H., Liang, Z., Zhang, C., Liu, R., Li, J., Zhang, W., Liang, D., Shen, B., Zhang, X., Ge, Y., Zhang, J., Ying, L.: SuperDTI: Ultrafast DTI and fiber tractography with deep learning. *International Society for Magnetic Resonance in Medicine*, pp. 3334–3347 (2021) doi: 10.1002/mrm.28937
8. Mansouri, F. A., Koechlin, E., Rosa, M. G. P., Buckley, M. J.: Managing competing goals - a key role for the frontopolar cortex. *Nature Reviews Neuroscience*, vol. 18, no. 11, pp. 645–657 (2017) doi: 10.1038/nrn.2017.111
9. McLaren, D. G., Kosmatka, K. J., Oakes, T. R., Kroenke, C. D., Kohama, S. G., Matochik, J. A., Ingram, D. K., Johnson, S. C.: A population-average MRI-based atlas collection of the rhesus macaque. *NeuroImage*, vol. 45, no. 1, pp. 52–59 (2009) doi: 10.1016/j.neuroimage.2008.10.058
10. Nelkenbaum, I., Tsarfaty, G., Kiryati, N., Konen, E., Mayer, A.: Automatic segmentation of white matter tracts using multiple brain MRI sequences. In: *IEEE 17th International Symposium on Biomedical Imaging*, pp. 368–371 (2020) doi: 10.1109/ISBI45749.2020.9098454
11. Ozyurt, O., Dincer, A., Yildiz, M. E., Peker, S., Yilmaz, M., Sengoz, M., Ozturk, C.: Integration of arterial spin labeling into stereotactic radiosurgery planning of cerebral arteriovenous malformations. *Journal of Magnetic Resonance Imaging*, vol. 46, no. 6, pp. 1718–1727 (2017) doi: 10.1002/jmri.25690
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, pp. 234–241 (2015) doi: 10.48550/arXiv.1505.04597
13. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*, vol. 53, pp. 197–207 (2019) doi: 10.48550/arXiv.1808.08114
14. Shen, Z., Zhang, M., Zhao, H., Yi, S., Li, H.: Efficient attention: Attention with linear complexities (2020) doi: 10.48550/arXiv.1812.01243
15. Tournier, J.: Diffusion MRI in the brain – theory and concepts. *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 112–113, pp. 1–16 (2019) doi: 10.1016/j.pnmrs.2019.03.001
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000–6010 (2017) doi: 10.48550/arXiv.1706.03762
17. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-net: Efficient channel attention for deep convolutional neural networks. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020) doi: 10.1109/cvpr42600.2020.01155
18. Wasserthal, J., Neher, P., Hirjak, D., Maier-Hein, K.: Combined tract segmentation and orientation mapping for bundle-specific tractography. *Medical Image Analysis*, vol. 58, pp. 101559 (2019) doi: 10.1016/j.media.2019.101559

*Daniel Bandala Álvarez, Jorge Perez Gonzalez*

19. Young, P. N. E., Estarellas, M., Coomans, E., Srikrishna, M., Beaumont, H., Maass, A., Venkataraman, A. V., Lissaman, R., Jiménez, D., Betts, M. J., McGlinchey, E., Berron, D., O'Connor, A., Fox, N. C., Pereira, J. B., Jagust, W., Carter, S. F., Paterson, R. W., Schöll, M.: Imaging biomarkers in neurodegeneration: current and future practices. *Alzheimer's Research and Therapy*, vol. 12, no. 49 (2020) doi: 10.1186/s13195-020-00612-7
20. Zhao, H., Tsai, C. C., Zhou, M., Liu, Y., Chen, Y. L., Huang, F., Lin, Y. C., Wang, J. J.: Deep learning based diagnosis of parkinson's disease using diffusion magnetic resonance imaging. *Brain Imaging and Behavior*, vol. 16, no. 4, pp. 1749–1760 (2022) doi: 10.1007/s11682-022-00631-y