

DETECCIÓN Y SEGMENTACIÓN SEMÁNTICA DE DISPLASIAS CORTICALES FOCALES EN SUJETOS CON EPILEPSIA REFRACTARIA POR MEDIO DE IMÁGENES DE RESONANCIA MAGNÉTICA Y APRENDIZAJE PROFUNDO

David Jiménez Murillo

Programa

Maestría en Automatización y Control Industrial



Institución Universitaria

Instituto Tecnológico Metropolitano

Facultad de Ingenierías

Medellín, 2023

DETECCIÓN Y SEGMENTACIÓN SEMÁNTICA DE DISPLASIAS CORTICALES FOCALES EN SUJETOS CON EPILEPSIA REFRACTARIA POR MEDIO DE IMÁGENES DE RESONANCIA MAGNÉTICA Y APRENDIZAJE PROFUNDO

David Jiménez Murillo

Programa

Maestría en Automatización y Control Industrial

Directores: Andrés Eduardo Castro Ospina, MSc
Juan David Martínez Vargas, MSc, PhD

Grupos: Grupo de Máquinas Inteligentes y Reconocimiento de Patrones – MIRP

Laboratorios: Máquinas Inteligentes y Reconocimiento de Patrones – MIRP



Institución Universitaria

Instituto Tecnológico Metropolitano

Facultad de Ingenierías

Medellín, 2023

AGRADECIMIENTOS

Agradezco a los directores de tesis, los profesores Andrés Eduardo Castro Ospina y Juan David Martínez Vargas, por su orientación y por estar siempre dispuestos a brindar su conocimiento, en clases y asesorías, para llevar a buen puerto los objetivos de este proyecto. Quiero también expresar mi agradecimiento a los líderes y compañeros del grupo de investigación de Máquinas Inteligentes y Reconocimiento de Patrones (MIRP), en especial a Leonardo Duque, por su apoyo y por brindarme los recursos necesarios para realizar la maestría. También deseo agradecer a mi familia en especial a mi madre, Luz Amanda, y a mi tía, Olga Cecilia por su paciencia, dedicación y afecto. Este trabajo es dedicado a la memoria de Martha Lucía Murillo y Mónica Correa.

Finalmente, quiero dar un agradecimiento especial al Instituto de Alta Tecnología Médica (IATM), en especial a Jazmín Ximena Suarez y al Dr. Jorge Mario Vélez por su apoyo, al brindarnos las imágenes MRI de pacientes con lesiones de FCD y por el arduo trabajo de marcar cada una de ellas.

Este trabajo fue apoyado por el proyecto P20214 del Instituto Tecnológico Metropolitano: *"Metodología para evaluación cuantitativa de Imágenes de Resonancia Magnética estructural en sujetos con epilepsia no lesional"*.

RESUMEN

La epilepsia es una enfermedad que afecta aproximadamente a 50 millones de personas alrededor del mundo y se caracteriza por episodios convulsivos recurrentes que, cuando son resistentes al tratamiento médico, se le denomina epilepsia refractaria. Esta condición se presenta en aproximadamente el 30% de los casos de epilepsia y una de sus posibles causas es la presencia de malformaciones en la corteza cerebral conocidas como displasias corticales focales (FCD, por sus siglas en inglés). Estas pequeñas malformaciones suelen ser casi imperceptibles en las imágenes de resonancia magnética (MRI, por sus siglas en inglés) y escapan de la vista del radiólogo con considerable frecuencia. Por esta razón y dado que la correcta detección y segmentación de FCD es clave en el éxito de la cirugía, se hace necesario el desarrollo de técnicas computacionales para automatizar el proceso de diagnóstico. Uno de los métodos más efectivos para el reconocimiento de patrones en imágenes son las técnicas de aprendizaje profundo. En este trabajo se desarrolla una metodología basada en algoritmos de aprendizaje profundo para la detección y segmentación automática de FCD. Se cuenta con un conjunto de datos propio de volúmenes MRI que contiene pacientes con FCD delineados por un experto y controles. Sin embargo, debido al bajo número de estudios, los modelos de detección y segmentación propuestos son inicialmente entrenados con un conjunto de datos público de tumores cerebrales que contiene mayor cantidad de datos. Posteriormente, mediante transferencia de aprendizaje, los modelos son ajustados para detectar y delinear FCD. Como resultado, se entrenaron modelos que pueden detectar y segmentar la zona de la malformación de FCD, los cuales pueden ser usados como apoyo al diagnóstico por parte de expertos.

Palabras Claves: Deep Learning, Displasia cortical focal, Procesamiento de imágenes, Resonancia magnética, Segmentación semántica

ABSTRACT

Epilepsy is a disease that affects approximately 50 million people around the world and is characterized by recurrent seizure episodes. When these are resistant to medical treatment, they are called refractory epilepsy. This condition occurs in approximately 30% of epilepsy cases, and one of its possible causes is the presence of malformations in the cerebral cortex known as focal cortical dysplasias (FCD). These small malformations are usually almost imperceptible on magnetic resonance imaging (MRI) and escape the radiologist's view with considerable frequency. Given that the correct detection and segmentation of FCDs is paramount to the success of surgery, it is necessary to develop computational techniques to automate the diagnostic process. One of the most effective methods for pattern recognition in images is deep learning techniques. In this work, we develop a methodology based on deep learning algorithms for the automatic detection and segmentation of FCD. A dataset of MRI volumes containing expert-delineated FCD patients and controls is available. However, due to the low number of studies, the proposed detection and segmentation models are initially trained with a publicly available brain tumor dataset containing a larger amount of data. Subsequently, through transfer learning, the models are tuned to detect and delineate FCDs. As a result, models that can detect and segment the area of FCD malformation were trained, which can be used as diagnostic support by experts.

Keywords: Deep Learning, Focal Cortical Dysplasia, Image processing, Magnetic resonance imaging, Semantic Segmentation.

ÍNDICE GENERAL

Agradecimientos	III
Resumen	IV
Abstract	V
Índice General	VII
Lista de Acrónimos	VIII
Introducción	1
Justificación	1
Problema	2
Hipótesis	3
Objetivos	3
General	3
Específicos	3
Contribuciones	3
Referencias	4
1 Estado del Arte Sobre Detección y Segmentación de Displasias Corticales Focales en Imágenes de Resonancia Magnética	6
1.1 Métodos Visuales.	8
1.2 Métodos semiautomáticos	9
1.3 Métodos Automáticos	11
1.3.1 Métodos Matemáticos	11
1.3.2 Métodos Automáticos Basados en Morfometrías Volumétricas.	13
1.3.3 Métodos Automáticos Basados en Aprendizaje de Máquinas.	18
1.3.4 Métodos Automáticos Basados en Algoritmos de Aprendizaje Profundo.	24
Referencias	37
2 Marco Teórico	45
2.1 Imágenes de Resonancia Magnética.	46
2.1.1 Tiempos de relajación en la adquisición de MRI	47
2.1.2 Secuencia T1	47
2.1.3 Secuencia T2-FLAIR	48
2.2 Preprocesamiento	48
2.2.1 Herramienta de Extracción de Cráneo	48
2.3 Algoritmos de Aprendizaje y Optimización.	48
2.3.1 Red neuronal convolucional.	48
2.3.2 U-Net	49
2.3.3 Transferencia de aprendizaje	50
2.3.4 Estimación Adaptativa de los Momentos (Optimizador Adam)	50

2.3.5	Entropía Cruzada Binaria	51
2.4	Métodos de evaluación	52
2.4.1	Métricas basadas en matriz de confusión.	52
2.4.2	Coefficiente de Similitud Sørensen-Dice	53
	Referencias	55
3	Marco Experimental	56
3.1	Elección de bases de datos y preprocesamiento	57
3.1.1	Brain Tumor Segmentation 2020 dataset	58
3.1.2	Conjunto de datos Imágenes Ayudas Diagnósticas SURA	59
3.2	Detección automática de lesiones de FCD en MRI	62
3.2.1	Entrenamiento del clasificador de tumores cerebrales en MRI	63
3.2.2	Resultados de la CNN en detección de tumores	64
3.2.3	Entrenamiento del clasificador de lesiones de FCD en MRI.	65
3.2.4	Resultados de la CNN en detección de lesiones de FCD.	67
3.3	Segmentación semántica de lesiones de FCD en volúmenes MRI	69
3.3.1	Segmentación de tumores en MRI	69
3.3.2	Resultados de la U-Net en segmentación de tumores	71
3.3.3	Segmentación de lesiones FCD en imágenes MRI	73
3.3.4	Resultados de la U-Net en segmentación de lesiones de FCD.	76
	Referencias	83
4	Discusión y conclusiones	84
4.1	Discusión	84
4.2	Conclusiones	85
4.3	Trabajo Futuro	87
	Referencias	88

LISTA DE ACRÓNIMOS

AdaBoost	Adaptive Boost
ANN	Artificial Neural Network
AUC	Area Under the Curve
BET	Brain Extraction Tool
BCE	Binary Cross-Entropy
CNN	Convolutional Neural Networks
CSF	Cerebrospinal Fluid
DTI	Diffusion Tensor Imaging
ED	Peritumoral Edema
EEG	Electroencephalography
ET	Enhanced Tumor
FAST	FMRIB's Automated Segmentation Tool
FC	Fully Connected
FCD	Focal Cortical Dysplasia
FCN	Fully Convolutional Network
fKNN	Fuzzy K-Nearest Neighbours
FLAIR	Fluid Attenuated Inversion Recovery
fMRI	Functional MRI
FMRIB	Functional Magnetic Resonance Imaging of the Brain
FN	False Negative
FP	False Positive
FSL	FMRIB's Software Library
GAN	Generative Adversarial Network
GLM	Generalized Linear Model
GM	Gray Matter
GPML	Gaussian Processes for Machine Learning
GPU	Graphic Processing Unit
ILSN	Iterating Local Searches on Neighborhood
KNN	K-Nearest Neighbours
LCD	Local Cortical Deformation
LDA	Linear Discriminant Analysis
LDPO	Local Directional Probability Optimization
MDA	Mahalanobis Discriminant Analysis
MLP	Multi-layer Perceptron
MRI	Magnetic Resonance Image
NCR/NET	Necrotic and non-Enhancing Tumor Core
NIFTI	Neuroimaging Informatics Technology Initiative
NODDI	Neurite Orientation Dispersion and Density Imaging
OC-SVM	One Class - Support Vector Machine

PCA	Principal Components Analysis
PET	Positron Emission Tomography
PIB	Pattern Image Block
PNH	Periventricular Nodular Heterotopia
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QDA	Quadratic Discriminant Analysis
QPET	Quantitative Positron Emission Tomography
RAS	Right Anterior Superior
RF	Radio Frequency
ReLU	Rectified Linear Unit
ResPath	Residual Path
ROC	Receiver Operating Characteristic
RUSBoost	Random Undersampling Boost
sEEG	Stereo Electroencephalography
SGD	Stochastic Gradient Descent
SBM	Surface Based Morphometry
SMT	Spherical Mean Technique
SPM	Statistical Parametrical Mapping
SVM	Support Vector Machine
TLE	Temporal Lobe Epilepsy
TN	True Negative
TP	True Positive
UVMC	Unanimous Voting of Multiple Classifiers
VBM	Voxel Based Morphometry
VoIBM	Volume Based Morphometry
WM	White Matter
XGBoost	Extreme Gradient Boost

INTRODUCCIÓN

JUSTIFICACIÓN

La epilepsia es una enfermedad neurológica que aqueja la humanidad desde épocas remotas y aún hoy es el desorden neurológico más común afectando a individuos de todas las edades [1]. Se estima que la epilepsia afecta a alrededor de 50 millones de personas a nivel mundial, de los cuales el 80% provienen de países en vías de desarrollo, 5 millones de la región de Latinoamérica y el Caribe [2]. En un estudio reciente realizado en Colombia se encontró que la prevalencia de la enfermedad en el país es de 11,3 diagnósticos por cada 1000 habitantes y que esta cifra presenta poca variación entre regiones, a excepción de la región oriental, donde se encontró que la prevalencia alcanzó la cifra de 23 casos por cada 1000 habitantes [3]. La epilepsia representa una carga sustancial para el paciente diagnosticado debido a los efectos colaterales de los medicamentos, comorbilidades, disfunción cognitiva y física, mayor probabilidad de lesiones y mortalidad. Asimismo, se ha evidenciado que los pacientes de epilepsia gozan de una menor calidad de vida y representan un mayor impacto económico para el sistema de salud [4]. En un estudio demográfico realizado en la ciudad de Bogotá con una muestra de 107 pacientes diagnosticados con epilepsia se demostró que existe una correlación con mayor deserción escolar (tan solo el 10.2% tiene un nivel de educación superior), altas tasas de desempleo (76.7% estaban desempleados al momento del estudio) y dificultades para formar una familia (el 73.8% estaban solteros) [5].

Aproximadamente el 30% de los pacientes diagnosticados con epilepsia padece de epilepsia refractaria, es decir, no tienen respuesta al tratamiento médico [6]. Una causa común de epilepsia refractaria son las displasias corticales focales [7], estas son un tipo de malformaciones de la corteza cerebral que se dan desde la etapa de desarrollo del sistema nervioso central. Para los pacientes con epilepsia refractaria la única alternativa es la cirugía y para los prestadores de salud se ha demostrado que esta es una solución costo-efectiva comparado con el tratamiento tradicional con medicamentos [8].

Para detectar este tipo de malformaciones se ha recurrido a técnicas como MRI, electroencefalografía (EEG, por sus siglas en inglés) y tomografía por emisión de positrones (PET, por sus siglas en inglés). Las imágenes MRI son importantes porque se adquieren mediante una técnica no invasiva y permiten captar las diferentes estructuras cerebrales con una mayor resolución espacial, lo cual puede ser de gran ayuda en el diagnóstico de epilepsia refractaria. Según [9], los pacientes sometidos a cirugía obtienen mejores resultados posquirúrgicos, quedando libres de convulsiones, si se les ha detectado FCD en el análisis prequirúrgico con MRI. Asimismo, una completa resección de la región cerebral con lesión de FCD incrementa la probabilidad de que el paciente quede libre de convulsiones tras la cirugía [10].

Encontrar a simple vista aquellas regiones donde se hallan las displasias es bastante difícil dado que estas se pueden manifestar de forma muy sutil en la imagen de resonancia magnética y pasar desapercibidas para el radiólogo. Algunos estudios muestran que aproximadamente un 34% de imágenes MRI de pacientes con epilepsia refractaria y FCD comprobada histológicamente tienen diagnóstico

negativo [10]. Sin embargo, se ha clasificado cierto tipo de características presentes comúnmente en la imagen MRI de pacientes con FCD como son aumento del grosor cortical, borrosidad en la frontera entre materia gris (GM, por sus siglas en inglés) y materia blanca (WM, por sus siglas en inglés), entre otras [11]. Características que podrían ser detectadas automáticamente dentro de la imagen con ayuda de algoritmos de aprendizaje de máquina.

Los métodos automáticos de aprendizaje de máquina han venido ganando popularidad dada su capacidad para detectar malformaciones, pero su dependencia de la extracción manual de características los hace sujetos al error [12]. Los métodos computarizados de detección de objetos y segmentación automática de regiones que han mostrado mejor desempeño son aquellos que hacen uso de algoritmos de aprendizaje profundo como redes neuronales convolucionales (CNN, por sus siglas en inglés), redes neuronales recurrentes, grafos y otros métodos. En medicina, estos son aplicados sobre imágenes MRI, PET, entre otras, para la detección y segmentación de tejidos orgánicos [13], tumores cancerígenos [14] y otras patologías [15].

Por todo lo anterior, se hace importante el desarrollo de un método automático, que esté capacitado para realizar la tarea de detectar y segmentar semánticamente, con precisión, la región de las malformaciones FCD en imágenes MRI, con el fin de que este pueda usarse como apoyo en el proceso de diagnóstico, así como en el protocolo prequirúrgico de la enfermedad.

PROBLEMA

La detección y segmentación de FCD es un tema de suma importancia en el área médica ya que de su correcta implementación depende el éxito de un procedimiento quirúrgico en el que está en juego el bienestar físico y psicológico del paciente que debe ser sometido a este tratamiento. El éxito posquirúrgico trae beneficios, no solo para el paciente, sino también para las entidades y el sistema de salud en general ya que el tratamiento con medicamentos es bastante costoso. Asimismo, si no existe una buena segmentación de la malformación el paciente puede volver a tener episodios convulsivos posquirúrgicos. En este caso es necesario someter al paciente de nuevo al procedimiento con los costos que conlleva tanto económicos como físicos y psicológicos.

Sin embargo, a pesar de múltiples esfuerzos en materia de investigación, a la fecha no existe un método automático que sirva de ayuda en el diagnóstico de FCD y su correcta detección y segmentación en imágenes MRI, dado que el carácter sutil de las lesiones hace que estas sean difíciles de detectar tanto en el análisis visual por parte del personal neurológico, como por los métodos automatizados apoyados por computadora. En años recientes, con el auge de algoritmos de aprendizaje se han logrado resultados que se aproximan a un modelo eficiente, sin embargo, para tener resultados satisfactorios, estos algoritmos deben ser entrenados con una cantidad grande de datos, que en el área médica suelen ser escasos.

Es en este punto donde se plantea la siguiente pregunta de investigación: ¿es posible mejorar los resultados obtenidos en detección y segmentación automática de FCD utilizando imágenes MRI y técnicas de aprendizaje profundo y con esto obtener un avance en el diagnóstico de la epilepsia refractaria en pacientes con imágenes MRI negativa?

HIPÓTESIS

Mediante técnicas de aprendizaje profundo será posible detectar la presencia de displasias corticales focales y realizar una segmentación semántica de las mismas a partir de imágenes de resonancia magnética estructural del cerebro, logrando así construir una herramienta que realice una ubicación precisa de las malformaciones.

OBJETIVOS

GENERAL

Desarrollar una metodología basada en aprendizaje profundo para la detección y segmentación semántica de displasias corticales focales a partir de imágenes de resonancia magnética estructural del cerebro.

ESPECÍFICOS

- Desarrollar un método de detección de displasias corticales focales utilizando técnicas de aprendizaje profundo con el fin de identificar la presencia de malformaciones.
- Desarrollar una metodología, que integre el método de detección, para la segmentación semántica de displasias corticales focales basada en técnicas de aprendizaje profundo que permita la delineación anatómica de las malformaciones.
- Evaluar el desempeño de la metodología integral propuesta usando métricas de evaluación adecuadas para segmentación semántica con el fin establecer su utilidad como herramienta de diagnóstico.

CONTRIBUCIONES

- Artículo Publicado *Jiménez-Murillo, D., Castro-Ospina, A. E., Duque-Muñoz, L., Martínez-Vargas, J. D., Suárez-Revelo, J. X., Vélez-Arango, J. M., & de la Iglesia-Vayá, M. (2023). Automatic Detection of Focal Cortical Dysplasia Using MRI: A Systematic Review. Sensors, 23(16), 7072.*
- Jiménez-Murillo, D., Terraza-Arciniegas, D. F., Duque-Muñoz, L., Martínez-Vargas, J. D., & Castro-Ospina, A. E. *U-Net Based on MRI Image Patches for Skull Stripping*. SmartTech-IC 2021.
- Participación en el evento SIMDIA 2021 en modalidad de presentación de trabajos de Maestría; allí se presentaron y discutieron la hipótesis y los objetivos de la tesis, con el trabajo titulado *Detección y Segmentación Semántica de Displasias Corticales Focales en Sujetos con Epilepsia Refractaria por Medio de Imágenes de Resonancia Magnética y Aprendizaje Profundo*.
- Registro de software *PyTorch MRI (NIfTI) Dataset Generator*
- Registro de software *Delineador automático de displasias corticales focales sobre secuencias T1 axiales* (En proceso de registro)
- Dirección de trabajo de grado de ingeniería electrónica titulado *Segmentación de estructuras cerebrales usando aprendizaje profundo sobre imágenes de resonancia magnética estructural* del estudiante Stiven Alexander Henao Garcés
- Participación como estudiante vinculado del proyecto P20214 titulado: *"Metodología para evaluación cuantitativa de Imágenes de Resonancia Magnética estructural en sujetos con epilepsia no lesional"*.

REFERENCIAS

- [1] P. N. Banerjee, D. Filippi, and W. Allen Hauser, “The descriptive epidemiology of epilepsy—A review,” *Epilepsy Research*, vol. 85, no. 1, pp. 31–45, 2009.
- [2] J. Orozco-Hernández, J. Quintero-Moreno, D. Marín-Medina, J. Castaño-Montoya, P. Hernández-Coral, M. Pineda, J. Vélez, H. Villada, J. Martínez, and A. Lizcano, “Clinical and sociodemographic profile of epilepsy in adults from a reference centre in Colombia,” *Neurología (English Edition)*, vol. 34, no. 7, pp. 437–444, 2019.
- [3] A. Velez and J. Eslava-Cobos, “Epilepsy in Colombia: Epidemiologic Profile and Classification of Epileptic Seizures and Syndromes,” *Epilepsia*, vol. 47, no. 1, pp. 193–201, 2006.
- [4] N. Tian, M. Boring, R. Kobau, M. M. Zack, and J. B. Croft, “Active Epilepsy and Seizure Control in Adults — United States, 2013 and 2015,” *MMWR. Morbidity and Mortality Weekly Report*, vol. 67, no. 15, pp. 437–442, 2018.
- [5] C. Espinosa Jovel, C. Pardo, C. Moreno, J. Vergara, D. Hedmont, and F. Sobrino Mejía, “Perfil demográfico y social de la epilepsia en una población vulnerable y de bajos recursos económicos en Bogotá, Colombia,” *Neurología*, vol. 31, no. 8, pp. 528–534, 2016.
- [6] M. Gálvez M, G. Rojas C, J. Cordovez M, D. Ladrón de Guevara, M. Campos P, and I. López S, “Displasias corticales como causa de epilepsia y sus representaciones en las imágenes,” *Revista chilena de radiología*, vol. 15, 2009.
- [7] S. M. Sisodiya, “Surgery for malformations of cortical development causing epilepsy,” *Brain*, vol. 123, no. 6, pp. 1075–1091, 2000.
- [8] S. Wiebe, A. Gafni, W. T. Blume, and J. P. Girvin, “An economic evaluation of surgery for temporal lobe epilepsy,” *Journal of Epilepsy*, vol. 8, no. 3, pp. 227–235, 1995.
- [9] J. T. Lerner, N. Salamon, J. S. Hauptman, T. R. Velasco, M. Hemb, J. Y. Wu, R. Sankar, W. Donald Shields, J. Engel Jr, I. Fried, C. Cepeda, V. M. Andre, M. S. Levine, H. Miyata, W. H. Yong, H. V. Vinters, and G. W. Mathern, “Assessment and surgical outcomes for mild type I and severe type II cortical dysplasia: A critical review and the UCLA experience,” *Epilepsia*, vol. 50, no. 6, pp. 1310–1335, 2009.
- [10] L. Tassi, “Focal cortical dysplasia: neuropathological subtypes, EEG, neuroimaging and surgical outcome,” *Brain*, vol. 125, no. 8, pp. 1719–1732, 2002.
- [11] O. Colliot, T. Mansi, N. Bernasconi, V. Naessens, D. Klironomos, and A. Bernasconi, “Segmentation of focal cortical dysplasia lesions on MRI using level set evolution,” *NeuroImage*, vol. 32, no. 4, pp. 1621–1630, 2006.
- [12] E. Thomas, S. J. Pawan, S. Kumar, A. Horo, S. Niyas, S. Vinayagamani, C. Kesavadas, and J. Rajan, “Multi-Res-Attention UNet: A CNN Model for the Segmentation of Focal Cortical Dysplasia Lesions from Magnetic Resonance Images,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1724–1734, 2021.
- [13] S. Çimen Yetiş, A. Çapar, D. A. Ekinçi, U. E. Ayten, B. E. Kerman, and B. U. Töreyn, “Myelin detection in fluorescence microscopy images using machine learning,” *Journal of Neuroscience Methods*, vol. 346, p. 108946, 2020.

-
- [14] J. Zhou, Y. Zhang, K. Chang, K. E. Lee, O. Wang, J. Li, Y. Lin, Z. Pan, P. Chang, D. Chow, M. Wang, and M. Su, "Diagnosis of Benign and Malignant Breast Lesions on DCE-MRI by Using Radiomics and Deep Learning With Consideration of Peritumor Tissue," *Journal of Magnetic Resonance Imaging*, vol. 51, no. 3, pp. 798–809, 2019.
- [15] X. Bai, Y. Hu, G. Gong, Y. Yin, and Y. Xia, "A deep learning approach to segmentation of nasopharyngeal carcinoma using computed tomography," *Biomedical Signal Processing and Control*, vol. 64, p. 102246, 2021.

1

ESTADO DEL ARTE SOBRE DETECCIÓN Y SEGMENTACIÓN DE DISPLASIAS CORTICALES FOCALES EN IMÁGENES DE RESONANCIA MAGNÉTICA

Índice

1.1 Métodos Visuales	8
1.2 Métodos semiautomáticos	9
1.3 Métodos Automáticos	11
1.3.1 Métodos Matemáticos	11
1.3.2 Métodos Automáticos Basados en Morfometrías Volumétricas	13
1.3.3 Métodos Automáticos Basados en Aprendizaje de Máquinas	18
1.3.4 Métodos Automáticos Basados en Algoritmos de Aprendizaje Profundo	24
Referencias	37

SE han propuesto diversas técnicas para la detección y segmentación de FCD en MRI que se pueden agrupar de diferentes formas de acuerdo con el tipo de algoritmo elegido, el tipo de imágenes o datos empleado y la naturaleza del análisis – que puede ser matemático o estadístico, manual o automático, entre otros. En este trabajo se propone una división fundamental en métodos visuales, semiautomáticos y automáticos, y de esta división se desprenden otras subdivisiones, enfocándose en métodos automáticos, como se puede apreciar en el diagrama de flujo de la Figura 1.1. El desarrollo de estas técnicas ha dado lugar a una buena cantidad de documentación, lo cual hace necesario el uso de algún método ya establecido para recopilarla y revisarla sistemáticamente.

Dada la importancia de llevar una revisión de literatura actualizada y en vista de que tan solo se encuentra la revisión que en 2016 publican Wang and Alexopoulos [1] respecto al tema, se ha decidido llevar a cabo una revisión sistemática de la literatura científica, con una propuesta de agrupamiento de los artículos de acuerdo con el modo de detección de lesiones de FCD empleado por los diferentes

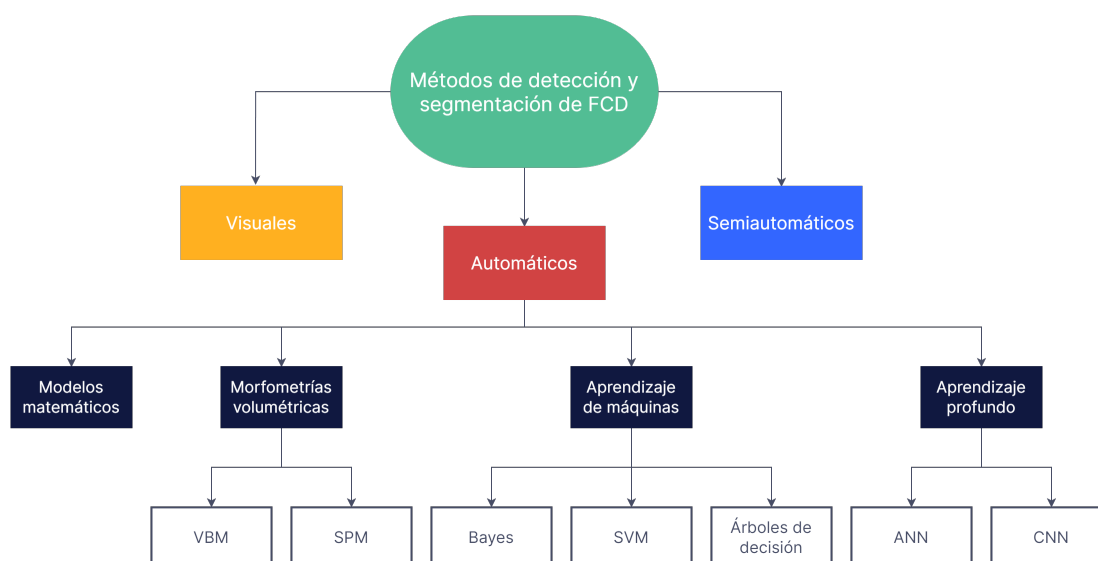


Figura 1.1: Diagrama de flujo donde se muestra la división propuesta de los métodos encontrados, en esta revisión de literatura, para la detección y segmentación de lesiones de FCD en imágenes MRI.

autores y siguiendo los lineamientos del protocolo de “Ítems preferidos en el reporte de revisiones sistemáticas y meta-análisis” (PRISMA, por sus siglas en inglés) para revisiones sistemáticas de literatura científica, enfocado originalmente en el área médica, pero aplicable en otras áreas [2].

La selección de artículos, según PRISMA, debe ser llevada a cabo por dos personas, en este caso por uno de los directores de esta disertación de Maestría y su autor principal. El primer paso fue definir las bases de datos de literatura científica donde se realizará la búsqueda de artículos y luego definir la ecuación de búsqueda apropiada para encontrar los artículos pertinentes a la investigación en estas bases de datos. Para la búsqueda de artículos se eligieron las bases de datos Scopus, Science Direct, Web of Science y PubMed, buscando abarcar un buen número de revistas, no solo enfocado en la parte técnica, sino buscando que aborden el problema desde un enfoque multidisciplinario, donde se tengan resultados de búsqueda tanto en el área médica como ingenieril. La ecuación de búsqueda definida, basada en los anteriores requerimientos, fue la siguiente:

((FCD) OR (Focal Cortical Dysplasia)) AND ((Automatic) OR (Automated)) AND ((MRI) OR (Magnetic Resonance Images))

Con esta ecuación, los motores de búsqueda de las bases de datos arrojaron 250 registros de interés, de los cuales 91 fueron encontrados en PubMed, 66 en Scopus, 70 en Web of Science y 23 en Science Direct, en la Figura 1.2 se muestra el diagrama de flujo empleado para seleccionar los artículos relevantes siguiendo el protocolo PRISMA. Luego de realizar una búsqueda manual, fueron removidos 110 artículos de los 250 registros encontrados inicialmente, por encontrarse duplicados entre las distintas bases de datos. Después de la depuración de registros repetidos quedaron 140 artículos. De estos se descartaron 70 artículos dado que, al revisar su título y resumen, se halló que no se ajustaban al tema de investigación; otro artículo fue descartado debido a que no estaba disponible para visualización en las bases de datos. Quedaron entonces 69 artículos para el siguiente criterio de selección, que consiste en leer el texto completo y, de acuerdo con el contenido de la lectura, se decide si el artículo

hará parte de la revisión de literatura. De la lectura del texto completo de los artículos encontramos cuatro artículos más que debían ser descartados: dos artículos descartados dado que utilizan otro tipo de imágenes o procedimientos médicos para encontrar las FCD, diferentes al enfoque de este artículo que es en MRI; un artículo fue descartado dado que trata otra patología diferente a FCD (esclerosis múltiple); y un artículo descartado porque su objetivo se aleja mucho de los objetivos propuestos en esta investigación.

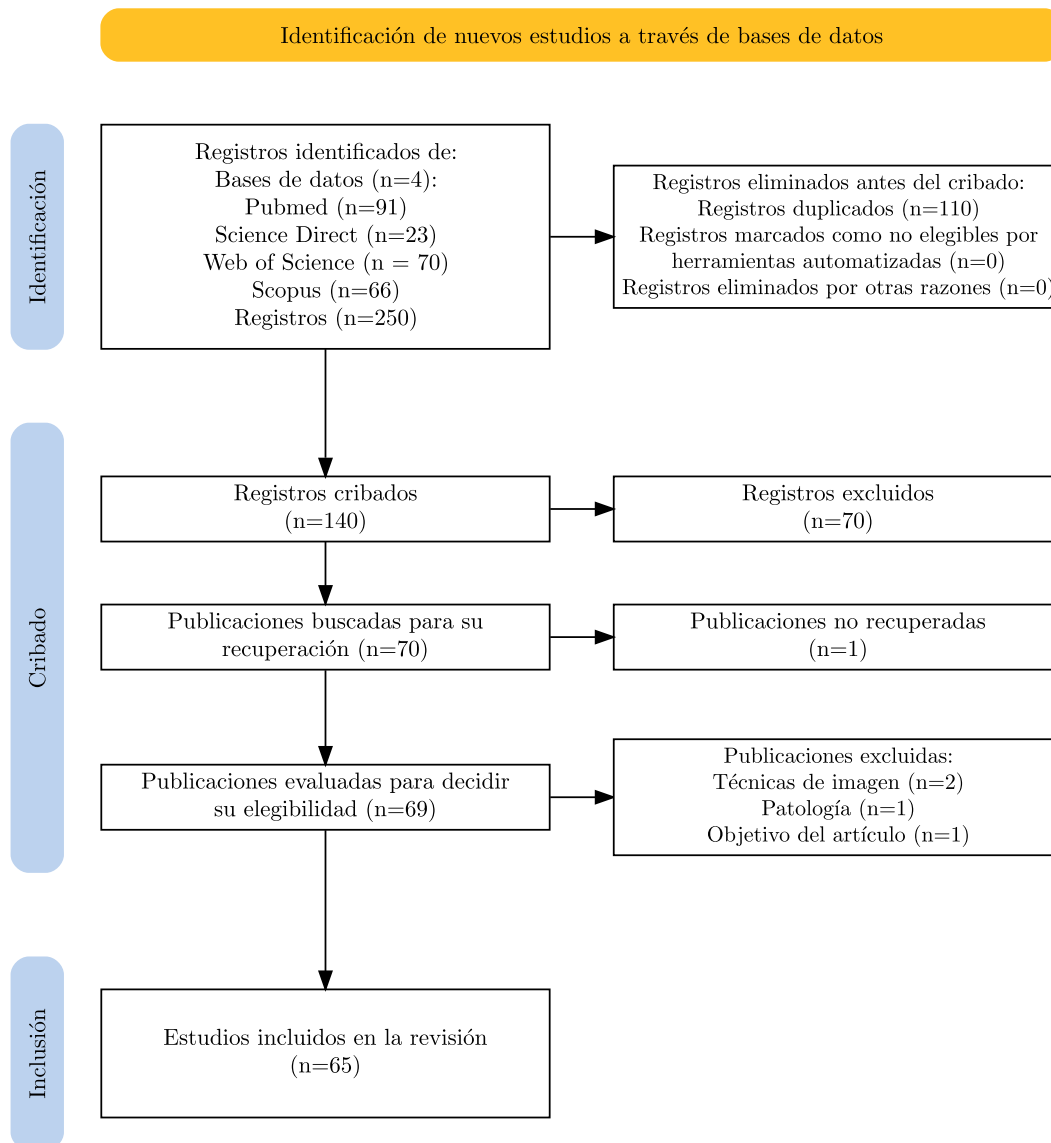


Figura 1.2: Diagrama de flujo PRISMA 2020 para revisiones sistemáticas de literatura científica empleado en este trabajo para seleccionar los registros relevantes en el área de detección automática de FCD.

1.1. MÉTODOS VISUALES

Si bien los primeros estudios se basaban en análisis histológicos y en actividad eléctrica obtenida mediante EEG, los avances en imagenología médica han hecho que estas tecnologías se hayan ido des-

plazando en favor de las imágenes médicas. La técnica de imagenología más común en la actualidad para la evaluación de displasias corticales son las imágenes MRI. Esta técnica implica un estudio completo del cerebro, en el que se realizan tomas de la secuencia T2 coronal y transversal, Recuperación de la inversión atenuada de fluido (FLAIR, por sus siglas en inglés) coronal y T1 coronal. Dentro de las características que han encontrado autores como Kabat and Król [3] para diferenciar las displasias en imágenes MRI se encuentran: grosor cortical, falta de definición de la unión entre materia gris y materia blanca y patrones anormales de los surcos o de los giros, entre otras.

Más recientemente se han empleado otras técnicas de imagenología con el fin de realizar la segmentación prequirúrgica. Una de ellas es la técnica de MRI de difusión, que al ser sensible a los axones en lugar de a la mielina permite obtener una “diferenciación más robusta de la corteza y la materia blanca y así mejorar la delineación de la malformación cortical prequirúrgica” (Lampinen, 2020). Otra secuencia MRI especializada e introducida recientemente para la delineación de FCD es la técnica 3D FLAIR la cual, al medir su desempeño para mostrar las características diferenciables de FCD, como grosor cortical, mostró “una diferencia estadística significativa (...) sobre las imágenes convencionales” Saini et al. [4].

La “difusión compleja” es un método basado en ecuaciones diferenciales parciales que puede ser aplicado en procesamiento de imágenes y que Rajan et al. [5] emplearon con el fin de mejorar la visibilidad de las lesiones de FCD en imágenes MRI del cerebro en secuencia T1. Al aplicarse difusión compleja a una imagen se obtienen los detalles de la imagen en un plano real y un plano complejo. En el plano real quedan las áreas “suaves” de la imagen, mientras que en el imaginario quedan los bordes. Contando con una base de datos de 48 pacientes de los cuales 25 fueron MRI-positivo, después de realizar preprocesamiento a las MRI y de segmentar los tejidos de GM, WM y fluido cerebroespinal (CSF, por sus siglas en inglés), los autores extraen el mapa de grosor cortical de las imágenes de acuerdo con el método de Laplace. El método de difusión compleja no lineal es aplicado a la imagen segmentada de GM para obtener la imagen de difusión compleja que consta de una parte real y otra imaginaria. Luego se calcula el mapa complejo, de acuerdo con la ecuación:

$$complex_map = \frac{grosor_cortical \times difusion_compleja_real}{difusion_compleja_imaginario}$$

Después de aplicar el método de difusión compleja se espera que aumente el contraste entre las áreas con y sin FCD. Finalmente, usando técnicas de morfometría basada en vóxeles (VBM, por sus siglas en inglés) para comparar los mapas obtenidos, los autores encontraron lesiones de FCD en el 76% (29/38) de los pacientes.

1.2. MÉTODOS SEMIAUTOMÁTICOS

En esta sección se agrupan aquellos estudios que se caracterizan por tener en su desarrollo dos componentes: una componente automática, que generalmente presenta la realización de cálculos en la imagen por medio de una computadora, y otra componente manual donde una mano experta supervisa o complementa el trabajo realizado por la máquina.

Antel et al. [6] buscan obtener un mayor contraste entre corteza con lesión y tejido sin lesión, partiendo del trabajo realizado por Bernasconi et al. [7], incorporan operadores de procesamiento de

imágenes para modelar tres de los atributos comunes de las lesiones de FCD en secuencia T1: incremento del grosor cortical, borrosidad de la frontera GM-WM e hiperintensidad de la señal relativa a la corteza normal. Todos los operadores fueron basados en *vóxel* y produjeron mapas tridimensionales de características que, a su vez, fueron combinados en un solo mapa de características compuesto. Luego, se calculó el contraste, definido como la media del *z-score* del tejido con lesión, tanto en los mapas originales como en el mapa compuesto de cada paciente y se calculó el porcentaje de cambio en contraste de los mapas originales al mapa compuesto. Evaluando el método en imágenes MRI de 14 pacientes, se encontró que el método mejora el contraste en 11 de ellas, además, en un paciente la lesión solo fue visible con la ayuda del mapa compuesto. Este resultado demuestra que incrementando el contraste entre tejido con lesión y tejido sin lesión se mejora la detección de lesiones a través del análisis visual de los mapas compuestos.

El principal objetivo en el estudio de Roca et al. [8] es realizar un análisis cuantitativo a nivel grupal de los patrones anormales en los surcos asociados a las lesiones de FCD en imágenes de pacientes MRI-negativo y MRI-positivo. El estudio está basado en: *i)* Un análisis cuantitativo automatizado de anomalías en los surcos usando un nuevo descriptor de los surcos llamado “energía del surco”; *ii)* Tres grupos de sujetos: MRI-positivo, MRI-negativo, sujetos de control; *iii)* Lesiones de FCD localizadas en la región central. El segundo objetivo del estudio consistió en evaluar la relevancia de los mapas *z-score* de la energía del surco para localizar la lesión epileptogénica a nivel individual. Para caracterizar la anatomía surco-giro se usó un procedimiento que constaba de cuatro pasos: 1. Extracción de los pliegues corticales en MRI secuencia T1 y su conversión a una representación basada en grafos de la corteza que contiene para cada pliegue una lista de descriptores morfológicos (área, profundidad, longitud, etc) y una organización espacial relativa a sus vecinos; 2. Reconocimiento automático de los surcos con un algoritmo basado en una congregación de redes neuronales con minimización de energía optimizada mediante recocido simulado bajo la tesis de que la energía mínima final es mayor en pacientes con FCD; 3. Generación del mapa de energía de los surcos calculado a partir de la energía mínima final del punto anterior; 4. Cálculo de los mapas *z-score* de energía de los surcos obtenidos a partir de la normalización de mapas de energía de los surcos, dividiendo la diferencia entre la energía de los surcos del paciente y la media de la energía de los surcos de los sujetos sanos por la desviación estándar de los controles. Se estudiaron las imágenes correspondientes a 29 pacientes, diestros, con lesiones FCD histológicamente confirmada en 25 de ellos y los 4 restantes siendo pacientes MRI-positivo y 25 sujetos de control, también diestros. A nivel grupal, se encontró que los pacientes presentaron mayor energía del surco central que los controles y que las FCD estaban asociadas con patrones anormales en surco central en comparación con los sujetos sanos. A nivel individual, se encontró que solamente la energía del surco no fue suficiente para detectar la lesión de FCD, sin embargo, se encontró una correspondencia entre los *z-score* máximos y el lugar donde se ubica la lesión de FCD.

También se han empleado métodos semiautomatizados asistidos por software en la detección de FCD en imágenes MRI como complemento a los métodos tradicionales y con resultados positivos. En un estudio realizado con 20 pacientes con FCD confirmada histológicamente, Sepúlveda et al. [9] contrastaron las segmentaciones realizadas por tres radiólogos expertos con aquellas realizadas por medio del software Freesurfer. La segmentación por software permitió encontrar dos casos más que los detectados manualmente (correspondiente a un 25 % más de casos), demostrando que es un método que ayuda en la detección de FCD invisibles con métodos convencionales.

Usando los avances a la fecha en software y hardware de MRI, Lorio et al. [10] buscan determinar

si los parámetros de difusión de modelos multi-sección demuestran cambios consistentes en probables lesiones FCD. Para ello se emplean nuevos modelos de difusión como “*Spherical Mean Technique*” (SMT, por sus siglas en inglés) y “*Neurite Orientation Dispersion and Density Imaging*” (NODDI, por sus siglas en inglés) que proveen medidas que potencialmente producen mapas más específicos de la microestructura del tejido anormal. El estudio fue evaluado usando las MRI de difusión de 33 pacientes pediátricos. Los mapas NODDI fueron calculados para cada *vóxel* del cerebro como tres secciones diferentes: intraneurita, extraneurita y sección libre de agua (CSF). Los modelos SMT fueron hallados estimando los mapas de tensor de difusión microscópica y de difusión microscópica multisección. Se obtuvieron perfiles de la lesión para cada mapa de difusión promediando los valores dentro de la máscara de la lesión y la región homóloga a lo largo de cada superficie de muestreo. Para evaluar los cambios en las lesiones FCD, los perfiles de los mapas de difusión fueron comparados estadísticamente entre lesión y la región homóloga de los 33 pacientes. Los mapas obtenidos con estos métodos demostraron tener mejor contraste que las imágenes en secuencia FLAIR en algunos casos y tuvieron cambios de señal consistentes específicos a FCD, sugiriendo que pueden mejorar los protocolos de imagenología en pacientes pediátricos con epilepsia y puede ser usados como características en la automatización de detección de lesiones.

1.3. MÉTODOS AUTOMÁTICOS

Los métodos automáticos para detección y segmentación de lesiones de FCD en MRI se han venido implementando con mayor frecuencia en los últimos años, dado el auge que han tomado los diferentes algoritmos de aprendizaje de máquina, supervisados o no supervisados, por sus buenos resultados soportados en diferentes estudios. El área de procesamiento de imágenes médicas no ha sido la excepción, beneficiándose de dichos algoritmos, bien sea en la detección de patologías en dichas imágenes que antes eran invisibles a los ojos del radiólogo o en la automatización de este proceso que realizado manualmente es costoso y demandante de tiempo. A continuación resumimos como se han implementado estos algoritmos en la detección o segmentación de lesiones de FCD en MRI y sus resultados.

1.3.1. MÉTODOS MATEMÁTICOS

Para la detección de FCD en MRI se ha recurrido a diferentes métodos basados en la solución de problemas de cálculo, como curvas de nivel o ecuaciones diferenciales, para el análisis de las estructuras cerebrales en MRI que permitan encontrar las regiones con malformación de forma analítica. A continuación se describen algunos artículos que han desarrollado este enfoque.

Debido a las deficiencias en segmentación de FCD en MRI que tienen los algoritmos existentes a la fecha del estudio y que fueron creados en su mayoría con el foco únicamente en detección de FCD, Colliot et al. [11] proponen un algoritmo basado en conjuntos de nivel para segmentación de FCD en MRI en secuencia T1. El modelo se apoya de un modelo deformable 3D basado en el método de conjuntos de nivel (las mismas curvas de nivel en 2D o superficies de nivel en 3D). En lugar de basar el modelo en los niveles de gris de la imagen, los autores proponen un modelo guiado por mapas de probabilidad derivados de características visuales de FCD como son: grosor cortical, borrosidad de la frontera GM-WM e hiperintensidad de la señal en la región displásica. Asimismo, el punto de partida para la evolución de los conjuntos de nivel se toma de un trabajo de clasificación previo de los autores [12]. La región de la lesión se forma con el método de competencia de regiones, se busca segmentar la imagen en varias regiones moviendo las interfaces que existen entre ellas. La evolución de las interfaces esta

dada por funciones, derivadas de las características de FCD, que indican la membresía a cada región. Así, la competencia de regiones ocurre en cada punto entre la clase lesión y la clase no-lesión más probable: GM, WM y CSF. El estudio contó con una base de datos de 24 pacientes MRI-positivos. El algoritmo de clasificación para iniciar el conjunto de nivel detectó 18/24 (75%) lesiones, por lo que el método fue ejecutado solo en estos casos. Las segmentaciones automáticas fueron comparadas con las segmentaciones manuales usando un índice de similitud, teniéndose que los conjuntos de nivel lograron una similitud de 62% respecto de las segmentaciones manuales.

En un estudio posterior, Colliot et al. [13] usan los mismos métodos que en Colliot et al. [11] pero esta vez apoyados en dos modelos 3D deformables sucesivos. El primer modelo, como en el trabajo anterior, está dado por los mapas de probabilidad obtenidos de características visuales de FCD en MRI en secuencia T1: grosor cortical, intensidad relativa y gradiente; con este primer modelo se pretende expandir la región de la lesión ubicada por el clasificador. El segundo modelo toma el resultado del modelo anterior para expandir la lesión, mediante gradiente de flujos vectoriales, desde la frontera GM-WM a otras regiones corticales subyacentes o suprayacentes. El algoritmo se evaluó sobre los mismos 18 pacientes del estudio anterior y obtuvo una similitud de 73% respecto de las segmentaciones manuales realizadas por un experto.

Despotovic et al. [14] proponen un algoritmo de cortes de grafos mejorado, para una segmentación más precisa de la corteza cerebral, integrando campos aleatorios de Markov basados en una función de energía. El método se basa en la segmentación de la corteza cerebral en tres etiquetas: GM, WM y CSF; de MRI en secuencia T1. El problema de minimización de energía es resuelto mediante el cálculo del costo mínimo en un grafo, igual a la energía total de la segmentación correspondiente. El método fue evaluado con imágenes simuladas y con MRI de ocho pacientes usando el coeficiente de similitud Sørensen-Dice. De acuerdo con el documento, se obtuvo una similitud de 95%, sin embargo, no es claro en el artículo si es un promedio de los ocho pacientes.

Snyder et al. [15] implementan un nuevo método para describir y detectar lesiones FCD. Creando primero un modelo, basado en una implementación de características locales 3D multiescala invariante a la rotación y a través de múltiples contrastes, que describe la variabilidad normal de la corteza de sujetos sanos, con las características de este modelo se crea una representación latente de los datos que permite la detección directa de casos atípicos en el espacio de características multivariado. En pacientes, muchas lesiones de FCD aparecen como casos atípicos pero con características subyacentes similares a regiones normalmente atípicas, por lo que se evalúa la similitud entre las características subyacentes a las FCD. La detección automática de FCD se basa en estos mapas de similitud, en los cuales la media del vector de características para cada parche cortical es proyectada sobre el vector unitario promedio de FCD. Con una base de datos de 15 pacientes y 30 sujetos de control, este estudio pudo detectar automáticamente 12/15 lesiones con un conjunto óptimo de hiperparámetros para una sensibilidad de 80% y un área bajo la curva de 0.96; de los 12 lesiones detectadas 11 pertenecían a pacientes MRI-positivo mientras que solo 1/3 pertenecía a pacientes MRI-negativo.

Investigando la arquitectura laminar neocortical en pacientes con FCD y heterotopía nodular periventricular (PNH, por sus siglas en inglés) para detección de anomalías corticales, Lotan et al. [16], por medio de regresión logística, comparan pacientes de FCD y PNH con sujetos sanos. Para ello realizaron el cálculo de las capas corticales en T1 hallando el mapa de probabilidad de GM; usando el mapa de GM, se crea un histograma de valores de GM de la secuencia T1 (*corticograma T1*), usado para

clasificar las capas T1. El *corticograma T1* se adecúa a un modelo de distribución-*t* resultando en seis distribuciones-*t* que presuntamente dan cuenta de las distribuciones de las seis capas corticales. Con la regla de Bayes se calcularon seis mapas de probabilidad cortical, uno para cada capa T1 que demostrara un patrón laminar. Para determinar la composición laminar de las diferentes áreas corticales, se usó un atlas de la 78 regiones corticales de interés que a su vez fueron divididas en 1000 sub-áreas de igual volumen. El mayor hallazgo de este estudio fue la observación de la extensión de la arquitectura cortical anormal en los pacientes comparado con los sujetos de control y que se extiende más allá de las zonas presuntamente epileptogénicas como se aprecia en el análisis MRI convencional.

1.3.2. MÉTODOS AUTOMÁTICOS BASADOS EN MORFOMETRÍAS VOLUMÉTRICAS

Los primeros métodos automáticos de detección y segmentación de FCD están basados en el cálculo asistido por computadora de diferentes modelos matemáticos y estadísticos que permiten caracterizar las estructuras del cerebro para crear mapas de cerebros sanos y realizar la comparación con el mapa cerebral de pacientes con FCD. En estos modelos se plantean algunos de los métodos de extracción de características de FCD en MRI, que en trabajos posteriores serán de uso común en el entrenamiento de algoritmos de aprendizaje de máquinas.

1.3.2.1. MORFOMETRÍA BASADA EN VÓXELES (VBM)

Como indican Ashburner and Friston [17], VBM implica realizar una comparación de la concentración local de GM entre dos grupos de sujetos. Para llevar a cabo este proceso se requiere normalizar espacialmente las imágenes de alta resolución de todos los sujetos en el estudio a un espacio estereotáctico equivalente.

En esta revisión de literatura, el primer estudio encontrado que implementa VBM buscando automatizar completamente la detección de FCD fue propuesto por Kassubek et al. [18] apoyados en el cálculo de los mapas de densidad de GM, donde cada *vóxel* tiene codificada la concentración de GM y su correspondiente posición en la MRI original. Se calcularon primero los mapas de densidad de GM de 30 sujetos sanos, obteniendo un solo mapa general calculando la media de cada uno de ellos; luego se calcularon los mapas de 7 pacientes con FCD. Para encontrar las lesiones en los mapas de densidad de GM de los 7 pacientes, se sustrajo el mapa de densidad media de GM de los sujetos de sanos a nivel de *vóxel*, buscando los máximos globales y locales de la diferencia resultante de las imágenes. Estos máximos fueron comparados con la desviación estándar de los valores de los mapas de densidad de los sujetos sanos en sus correspondientes locaciones. Con este método se encontraron lesiones en 6/7 pacientes, donde el máximo global, que representa la máxima diferencia en densidad de GM entre pacientes y sujetos sanos, correspondía con la ubicación de la lesión como fuera detectada por medio de análisis visual convencional.

Como extensión del trabajo realizado en [18] en el que se buscaba hacer más visibles las lesiones de FCD mediante mapas de densidad de materia gris, Huppertz et al. [19] proponen un nuevo método basado en VBM que, como novedad, busca hacer más visible una característica diferente de FCD como es la borrosidad de la unión GM-WM, con el fin de mejorar la delineación y reconocimiento de FCD. Para cada paciente, de una base de datos de 25 pacientes con FCD histológicamente comprobada, se calcularon dos mapas tridimensionales de características derivados de la secuencia T1 original; un primer mapa llamado por los autores “imagen de la juntura” que destaca las regiones cerebrales donde existe borrosidad en la unión GM-WM, y una “imagen de extensión” que es el mismo mapa de densidad de materia gris que fuera calculado en [18]. Asimismo, se calcularon estos dos mapas para cada uno de los 53 sujetos de control y, hallando la media de todos los 53 mapas, se generaron dos

mapas de sujetos sanos. Como en el trabajo anterior, los autores esperan que la diferencia máxima entre la “imagen de la juntura” y la T1 original mejore la visualización de la lesión, es decir, aumente el contraste entre las regiones normales y la regiones con FCD del cerebro. El método encontró lesiones en 18/25 pacientes aplicando el mapa de “imagen de la juntura”, demostrándose que la borrosidad en la frontera GM-WM es la característica de estas lesiones; combinando tanto el método que emplea los mapas de densidad de materia gris del artículo anterior con el método basado en la borrosidad en la frontera GM-WM de este trabajo, se detectaron 21/25 lesiones, de las cuales 4 no fueron reconocidas en análisis MRI convencional.

Buscando también mejorar la visualización de borrosidad en la frontera GM-WM, con el fin de automatizar la detección de FCD en pacientes cuyas MRI presentan esta característica y, ante la cantidad de falsos positivos (FP, por sus siglas en inglés) que introducen los métodos basados en gradiente (especialmente en regiones de donde la intensidad de la transición GM-WM es débil), Qu et al. [20] propusieron un nuevo algoritmo que llamaron “Iteración de búsquedas locales en el vecindario” (ILSN, por sus siglas en inglés). Este nuevo método mide el ancho de la región borrosa en lugar de la fuerza del gradiente. El estudio contenía 8 imágenes MRI en secuencia T1 de 8 pacientes cuya FCD se manifestó en la imagen como borrosidad en la frontera GM-WM. Luego de preprocesar las MRI (extracción de cerebro, corrección de *bias* e interpolación), los autores estiman la región de la frontera GM-WM extrayendo los mapas de GM y WM con la “Herramienta de Segmentación Automática de MRI Funcional del Cerebro” (FAST, por sus siglas en inglés), que pertenece a la “Librería de Software MRI Funcional del Cerebro (FMRIB, por sus siglas en inglés)” (FSL, por sus siglas en inglés); luego calculan la región de la frontera como aquellos *vóxeles* que tienen tanto información de GM como de WM. Luego, se modela el mapa de potencial de forma análoga a un campo eléctrico, donde la intensidad del campo en un punto dado, es decir, el valor del *vóxel*, equivale al negativo del gradiente de potencial en ese punto. Los resultados de la evaluación de este método indican que tiene mejor habilidad para identificar FCD caracterizadas por borrosidad en frontera GM-WM que los métodos basados en gradiente.

Colliot et al. [13] usan VBM para investigar cambios en la GM de los pacientes de manera individual. Esto, dado que los mismos autores en [6, 7] observaron que suelen descartarse lesiones de FCD que se manifiestan como hiperintensidades en la GM, en secuencia T1, al ser etiquetadas como parte de la WM. El estudio, que incluyó 27 pacientes y 39 controles, consistió en comparar los mapas de GM de cada sujeto con el mapa promedio de GM de todos los sujetos de control, obteniéndose un mapa *z-score* de cada individuo. Los mapas *z-score* resultantes mostraron, en 21/27 sujetos, que la lesión coincide con un aumento en la intensidad de GM-WM.

Dado que los mapas de características que resaltan la característica de borrosidad de la frontera WM-GM, para detección automática de FCD, dependen de pequeñas diferencias en la intensidad de la señal y que las secuencias T2 tienen una alta sensibilidad a diferencias sutiles en la intensidad de la señal. Esta es la motivación de House et al. [21] para realizar un estudio donde se evalúa cualitativa y cuantitativamente los resultados del desempeño del análisis morfométrico, para detección de FCD, de los mapas *z-score* obtenidos de comparar los mapas de la frontera GM-WM en secuencias T1 y T2, de 20 pacientes con FCD, contra los mapas promedio de los sujetos de control. Los resultados del análisis cualitativo arrojaron que en 16 de los 20 pacientes la característica de borrosidad en la frontera GM-WM fue visualizada con contraste más alto en la secuencia T2, en 3 pacientes no hubo una diferencia sustancial entre las dos secuencias y solo en un caso la secuencia T1 mostró un mejor contraste. Para la evaluación cuantitativa, los autores compararon la tasa del promedio de los *z-score* interno y externo a

la FCD de cada paciente para las secuencias T1 y T2. En 19 casos la tasa fue más alta para la secuencia T2 que para T1, confirmándose el resultado de la comparación cualitativa; el caso restante, donde la secuencia T1 mostró mejor contraste en la evaluación cuantitativa, corresponde con el mismo caso en la evaluación cualitativa. Se puede concluir entonces de este estudio que la secuencia T2 es más apropiada para evaluar o detectar FCD cuando esta se manifiesta como borrosidad en la frontera GM-WM.

Dado que con los mapas de la magnitud del gradiente, usados para modelar la borrosidad en frontera GM-WM, no se puede cuantificar el ancho de la unión GM-WM, Qu et al. [22] proponen un nuevo algoritmo llamado “Optimización de probabilidad direccional local” (LDPO, por sus siglas en inglés), con el cual pretenden detectar y cuantificar el ancho de la frontera GM-WM dentro de las áreas con lesión. Para ello, en primer lugar, se extrae la frontera GM-WM de la MRI por medio de un algoritmo de campos aleatorios de Markov. Luego, se generan las direcciones locales de los *vóxeles* en la frontera GM-WM que pasan entre la GM y la WM considerando la frontera GM-WM como un campo de potencial eléctrico. Por último, se estiman las direcciones locales óptimas para hallar el ancho de la frontera GM-WM, mediante la búsqueda iterativa en un vecindario. Los autores evaluaron la efectividad del método propuesto en volúmenes MRI del cerebro de 10 pacientes y 31 sujetos sanos. Los resultados indican que el ancho de la borrosidad en la frontera GM-WM en la región con lesión, calculado con el método LDPO, es mayor que en las regiones sin lesión. El método propuesto obtuvo puntajes *F-score* más altos en términos de la borrosidad GM-WM dentro de la región con lesión comparados con los puntajes obtenidos usando mapas de características.

Una variante de VBM, cuyos autores originales [23] denominan *MorphoBox* y que a su vez está basada en otro tipo de biomarcador conocido como “Morfometría basada en volumen” (VolBM, por sus siglas en inglés), es usada por Chen et al. [24] para estimar si las anomalías volumétricas (atrofia, hipertrofia) pueden ayudar en la detección automática de FCD. A diferencia de VBM, *MorphoBox* provee volúmenes de las diferentes estructuras cerebrales de interés y las compara con aquellas de sujetos sanos de una población similar en términos de edad y género; la técnica se concentra en una sola región de interés compuesta de múltiples *vóxeles*, donde se asume que el efecto acumulado de las anomalías de cada *vóxel* se hace más evidente que cuando se considera cada *vóxel* de manera individual, como es el caso de VBM, incrementando así la sensibilidad. Asimismo, *MorphoBox* considera tanto GM como WM lo cual puede mejorar la tasa de detección de FCD ya que las anomalías pueden no restringirse a un solo tejido. Evaluado en 16 pacientes con FCD histológicamente confirmada, se encontraron anomalías como atrofias e hipertrofias en todos los pacientes, zonas epileptogénicas en estructuras cerebrales con volumen anormal en 87.5% (14/16) de los pacientes; estas zonas se localizaron en regiones con anomalía atrófica en el 71.4% (10/14) de los pacientes. Estos hallazgos sugieren que es más probable que las lesiones de FCD se encuentren en regiones con volumen atrófico que en aquellas con volumen hipertrófico.

Qu et al. [25] proponen fusionar múltiples clasificadores mediante la optimización de sus parámetros con un algoritmo genético para tomar ventaja de la alta sensibilidad de los clasificadores reduciendo la cantidad de FP. Para ello, primero hallaron los mapas de 6 características de 10 pacientes con FCD: Grosor cortical, Gradiente, Intensidad relativa, Frontera GM-WM, Curvatura y Profundidad de los surcos. Los clasificadores seleccionados por los autores fueron Naive Bayes, Análisis de discriminantes lineales (LDA, por sus sigla en inglés), Análisis de discriminantes de Mahalanobis (MDA, por sus siglas en inglés) y Análisis de discriminantes cuadrático (QDA, por sus siglas en inglés). Cada mapa de características fue comparado con el correspondiente mapa promedio de 31 sujetos sanos para hallar

el correspondiente mapa *z-score*. Luego, cada mapa *z-score* se pasa por cada uno de los clasificadores y se calcula un *F-score* para cada clasificador. Por último, se calcula el *F-score* promedio por mapa para seleccionar los *P* mejores mapas. Los resultados experimentales del estudio muestran que el método fue capaz de detectar FCD en 9 de 10 pacientes y de clasificar correctamente los 31 sujetos de control para una especificidad de 100 %.

Feng et al. [26] proponen un método para detectar FCD automáticamente, en secuencias FLAIR negativas, basado en la medida de grosor cortical y su desviación estándar. Los autores primero extraen el mapa de grosor cortical promedio de un conjunto de imágenes MRI en secuencia T1 de 32 sujetos sanos y su desviación estándar. Luego, para las secuencias FLAIR de una base de datos de 6 pacientes con FCD histológicamente confirmada, se obtiene un mapa *z-score* restando el mapa de grosor cortical promedio con el mapa de grosor cortical de cada paciente y dividiendo por la desviación estándar del grosor cortical promedio de los pacientes sanos. Este método logró detectar 3/4 lesiones ubicadas en áreas no-temporales pero falló en detectar las lesiones en el lóbulo temporal de 3 pacientes.

1.3.2.2. MAPEO ESTADÍSTICO DE CARACTERÍSTICAS

El mapeo estadístico de características (SPM, por sus siglas en inglés) es un popular método de análisis de imágenes neurológicas, desarrollado por Friston et al. [27], que implementa una metodología basada en VBM. Como indican Yousaf et al. [28], la implementación del método implica la extracción de varios mapas de probabilidad de tejidos cerebrales como GM y WM mediante la aplicación del algoritmos de segmentación Bayesianos, suavizado mediante kernel Gaussiano y normalización. Finalmente, de acuerdo con los efectos de la concentración de GM relacionados con el caso de estudio, se realiza un análisis estadístico usando un “modelo lineal generalizado” (GLM, por sus siglas en inglés).

La metodología presentada por Srivastava et al. [29] aprovecha las características de intensidad de las lesiones de FCD en secuencia T1 y aporta como novedad el análisis estadístico de las mapas de características de FCD empleando SPM y GLM. A partir de los mapas de GM y WM, se construye un mapa de características específico de la lesión, calculando para cada *vóxel* de GM la tasa entre el grosor cortical y la intensidad del gradiente absoluto en la MRI. Los mapas de características, de una base de datos de 17 pacientes, fueron contrastados con mapas similares obtenidos de un grupo de control de 64 sujetos sanos. El modelo caracteriza las lesiones de FCD como áreas corticales que muestran engrosamiento en la corteza y borrosidad en la frontera GM-WM. El método detectó y localizó correctamente la lesión de FCD en 9/17 (53%) pacientes, usando un umbral que minimiza los FP, y 12/17 (71%) lesiones con un umbral que permite un mayor número de FP.

Focke et al. [30] ejecutaron análisis SPM sobre MRI en secuencia T2-FLAIR de 25 pacientes con 25 sujetos de control. La metodología consistió en tomar las MRI en secuencia T1 y registrarlas en el espacio de las MRI en secuencia T2-FLAIR, seguido de segmentación de tejidos en la secuencia T1. A continuación las imágenes FLAIR fueron normalizadas en intensidad de acuerdo con el valor “robusto” de intensidad media de la WM y se extrajo el cráneo usando el algoritmo de extracción de cráneo (BET, por sus siglas en inglés). Por último, se realizó convolución entre las imágenes FLAIR sin cráneo, un kernel Gaussiano y un GLM de SPM para comparar cada paciente contra el grupo de control. Con este algoritmo se encontraron lesiones en 22/25 (88%) con solo un falso positivo hallado en uno de los sujetos de control.

Al año siguiente, Focke et al. [31] utilizaron el mismo algoritmo de SPM que usaron en el anterior artículo [30] para detección automática de FCD pero esta vez con dos novedades. La primera es la ampliación a una base de datos de 70 pacientes que incluía MRI en secuencia T2-FLAIR y la misma cantidad (25) de sujetos de control. La segunda novedad es que, a diferencia de [30], donde las FCD fueron reportadas por dos expertos radiólogos, para este artículo los autores debieron apoyarse en telemetría de Video-Encefalografía ya que no fue posible para los expertos detectar las FCD. Los resultados obtenidos por el algoritmo fueron inferiores en este caso, encontrándose tan solo 10/70 lesiones en los pacientes para un 14.3%, resultado que los autores atribuyen a diversas causas como el hecho de que durante la toma de video telemetría algunos pacientes no presentaron convulsiones por lo que la FCD fue estimada de acuerdo con los hallazgos interictales y el historial clínico.

Con el objetivo de determinar la sensibilidad y especificidad del análisis de VBM y comparar sus resultados con otros métodos, Wong-Kisiel et al. [32] utilizan el método SPM-12 para comparar las características morfométricas de las MRI de una cohorte de 39 pacientes, que incluye pacientes 20 adultos y 19 pediátricos, con FCD histológicamente confirmada y un grupo de control compuesto por 64 adultos y 41 pediátricos, todos en secuencia T1. Luego de extraer las máscaras de GM, WM y CSF, se crea la imagen de extensión restando la media de GM de los sujetos de control de la máscara GM de cada paciente y dividiendo luego por la desviación estándar de la máscara GM de los controles, obteniéndose un mapa *z-score*. La imagen de unión se crea de forma similar pero, en lugar de la máscara de GM, se usa la máscara de la frontera GM-WM, y se obtiene el mapa *z-score*. Este método pudo detectar lesiones de FCD en 36/39 casos (92%); de los 36 casos detectados, la imagen de extensión encontró 34 y la imagen de unión 29 casos.

Mediante el análisis cuantitativo de imágenes PET en combinación con técnicas de SPM para identificar las anomalías presentes en las imágenes de una cohorte de pacientes MRI-negativo, Lin et al. [33] esperan comprobar su hipótesis de que las regiones coidentificadas en el análisis SPM y PET son regiones epileptogénicas. Las MRI en secuencia T1 de 104 pacientes fueron procesadas mediante la metodología SPM-12; luego se calculó la imagen de unión de cada paciente y su diferencia estadística con un base de datos de pacientes de control para hallar el mapa *z-score* y encontrar la FCD en forma de borrosidad en la frontera GM-WM. Las imágenes PET fueron evaluadas cuantitativamente para generar los mapas estadísticos conocidos como "PET cuantitativo" (QPET, por sus siglas en inglés). Solamente aquellas imágenes donde se detectó FCD mediante el análisis SPM-12 fueron contrastadas con su respectiva imagen QPET; la región se etiquetaba como positiva si y solo si en ambos casos coincidía, de lo contrario se etiquetaba como negativa. Con el análisis SPM-12 se detectaron 82/104 casos, de los cuales 77 fueron corroborados mediante el análisis QPET; la sensibilidad fue entonces de 74% con un umbral QPET $SD = -1$ elegido de acuerdo con la desviación estándar respecto de los sujetos de control.

Un estudio basado en SPM-12 con pacientes pediátricos fue propuesto por Wang et al. [34]. En total el estudio incluyó 78 MRI (40 en escáner de 3T y 38 en escáner de 1.5T) en secuencia T1 de pacientes MRI-negativo y 370 sujetos de control, en ambos casos con edades entre 5 y 21 años. Siguiendo los pasos comunes de la metodología de análisis morfométrico de MRI por SPM, se encontraron malformaciones en 44 de los 78 pacientes (56%), 7 de ellos con lesiones en múltiples regiones. De los 44 pacientes cuyas MRI fueron positivas, 21/40 (52%) correspondieron a muestras en MRI 3T y en 23/38 (60%) casos en 1.5T, los autores no interpretaron esto como una diferencia estadística significativa entre los dos grupos para sugerir que uno u otro escáner sea más apto para la detección automática de

FCD.

En la Tabla 1.1 se muestran los resultados obtenidos mediante SPM en la detección de lesiones de FCD en MRI. Se puede inferir de la tabla que, aunque los estudios [33] y [34] contaron con una cantidad mayor de pacientes, estos no fueron los que lograron mejores resultados. El mejor desempeño lo obtuvieron [32] con la misma versión del algoritmo y utilizando la misma secuencia MRI.

Tabla 1.1: Tabla de resultados del análisis morfométrico usando SPM

Estudio	Tecnica	Sensibilidad%	Secuencia	Pacientes	Controles	Grupo etario
Srivastava et al. [29]	SPM	53	T1	17	64	Adultos
Focke et al. [30]	SPM5	88	T2-FLAIR	25	25	Adultos
Wong-Kisiel et al. [32]	SPM-12	92	T1	39	105	Adultos/Pediátricos
Lin et al. [33]	SPM-12	74	T1	104	N/A	Adultos
Wang et al. [34]	SPM-12	56	T1	78	370	Pediátricos

1.3.3. MÉTODOS AUTOMÁTICOS BASADOS EN APRENDIZAJE DE MÁQUINAS

En esta sección se incluyen métodos tradicionales de aprendizaje de máquina que han demostrado tener un buen desempeño en el análisis de imágenes y por lo tanto han sido probados en diferentes estudios, tratando de medir su rendimiento en la detección automática de lesiones de FCD. En la revisión de literatura se encontró que los algoritmos más recurrentes son los clasificadores Naive Bayes y las máquinas de soporte vectorial (SVM, por sus siglas en inglés). En menor medida, otros algoritmos como los árboles de decisión y potenciación extrema de gradiente (XGBoost, por sus siglas en inglés) también son usados.

1.3.3.1. CLASIFICADOR BAYES

Quizás el intento más temprano por resolver el problema de clasificación de *vóxeles* como FCD o no FCD en MRI, usando técnicas de aprendizaje de máquinas, proviene del estudio donde Antel et al. [12] emplean una secuencia de clasificadores de Bayes en dos etapas entrenados en imágenes pertenecientes a una población de 18 pacientes con FCD confirmada histológicamente y 14 controles. El clasificador Bayes de la primera etapa fue entrenado a partir de características morfométricas de primer orden de la imagen. Primero segmentando la imagen en las clases GM, WM, y CSF. Luego, partir de estos mapas de las estructuras cerebrales, se extraen las características de la imagen para crear los modelos computacionales de grosor cortical, “borrosidad” de la frontera GM-WM e hiperintensidad en la señal T1. Este primer clasificador sirve para clasificar los *vóxeles* como lesionales ó no-lesionales. La segunda etapa del clasificador fue entrenada usando características de textura de segundo orden. Los autores seleccionaron tres características que fueron: segundo momento angular, diferencial de entropía y contraste. Este clasificador fue empleado para reclasificar los *vóxeles* clasificados como lesionales por el clasificador de la primera etapa. El clasificador combinado encontró 17 de las 20 lesiones etiquetadas manualmente.

En Yang et al. [35] usaron el algoritmo Naive Bayes para clasificación en un total de 21 sujetos con FCD. A diferencia de Antel et al. [12], no segmentaron la imagen en estructuras cerebrales, en su lugar seleccionando características que no dependieran de esta segmentación; y no clasificaron cada *vóxel* sino una vecindad de *vóxeles*. Las características calculadas fueron: grosor cortical, oblicuidad del grosor cortical, curtosis del grosor cortical, gradiente absoluto y varianza, oblicuidad y curtosis de las orientaciones de los vectores gradiente. Con esta aproximación, los autores encontraron lesiones en

21/21 pacientes.

En un estudio posterior, Yang et al. [36] usaron el mismo clasificador de Bayes empleado en [35] entrenado con las mismas características extraídas de la imagen y como etapa de entrada de un segundo clasificador, en este caso una SVM con un kernel polinomial de tercer orden, con el fin de reducir el número de FP. En esta metodología, a los grupos marcados como lesionales por el primer clasificador, se les calcula las mismas características del primer clasificador y luego se calcula la diferencia respecto de las características de este; estas diferencias constituyen las características de entrada del segundo clasificador, la SVM. Con este algoritmo, los autores obtuvieron una tasa de detección de 88 %.

En Strumia et al. [37] entrenan un clasificador de Bayes con características de intensidad, textura y forma extraídas a partir de MRI en secuencias T1 y FLAIR. La característica de intensidad se extrae a partir de secuencias T1 y FLAIR creando una distribución 2D para resaltar la hipointensidad de la señal T1 y la hiperintensidad de la señal FLAIR. La primera característica de textura es la magnitud del gradiente de la secuencia T1, que representa la borrosidad en la frontera GM-WM; la segunda resalta la baja definición que existe en la frontera GM-WM en las lesiones de FCD y es expresada como la varianza en la orientación espacial del gradiente calculado en la imagen T1. Se extraen dos características de forma a partir de la secuencia T1. Visualizando las FCD como una masa amorfa de tejido cerebral o como una región donde se fusionan los tejidos de GM y WM sin una frontera que los separe claramente, se extrae la primera característica de forma: “anisotropía fraccionaria”, esta se define como los eigenvalores de la matriz Hessiana en una ventana de 2 mm^3 . La segunda característica de forma es una medida de la asimetría del histograma de grosor cortical local. Estas características se calculan para todas las MRI en una base de datos de 11 pacientes y 20 sujetos de control. Se comparan las características calculadas para los pacientes con las características correspondientes en la base de datos de sujetos sanos a nivel de *vóxel*. Asumiendo una distribución Gaussiana para la probabilidad de que una característica f_i pertenezca al tejido normal, la probabilidad de este pertenezca a una lesión de FCD es su complemento. Se asume que las características son estadísticamente independientes y se realiza la clasificación con un clasificador Naive Bayes, obteniéndose un mapa de probabilidad espacial de FCD. Se obtuvieron los siguientes resultados promediando para todos los pacientes: precisión de 51 %, sensibilidad de 15 % y un coeficiente de similitud Sørensen-Dice de 13 % respecto de las etiquetas reales (*ground truth*).

Kulaseharan et al. [38] implementaron un clasificador Bayes de dos etapas basado en el clasificador implementado por Antel et al. [12], pero aplicado esta vez en volúmenes MRI en secuencias T1, T2 y FLAIR de 54 pacientes pediátricos y 13 sujetos de control. El clasificador de dos etapas es primero entrenado solo sobre características morfométricas; luego, los *vóxeles* clasificados como lesionales son reclasificados usando características de textura por el clasificador de la segunda etapa. Este método logró una sensibilidad de 97 % en pacientes MRI-positivos y de 70 % en pacientes MRI-negativos.

El estudio más reciente en utilizar un clasificador de Bayes fue desarrollado por Feng et al. [39] a partir de características extraídas de imágenes MRI en secuencia FLAIR de 7 pacientes MRI-negativos, para entrenar un clasificador Naive Bayes de tres etapas y generar un mapa 3D de la lesión. Este clasificador arrojó una sensibilidad de 87.5 %. En la Tabla 1.2 se relacionan los resultados obtenidos por los diferentes autores utilizando el algoritmo de Bayes.

Tabla 1.2: Tabla de resultados obtenidos con las diferentes implementaciones del clasificador de Bayes

Estudio	Técnica	Sensibilidad %	Pacientes	Grupo etario
Antel et al. [12]	Bayes de dos etapas	85	18	Adultos
Yang et al. [35]	Naive Bayes	62.49	21	Adultos
Yang et al. [36]	Naive Bayes + SVM	88	21	Adultos
Strumia et al. [37]	Naive Bayes	51	11	Pediátricos/Adultos
Kulaseharan et al. [38]	Bayes de dos etapas	70	54	Pediátricos
Feng et al. [39]	Bayes de tres etapas	87.5	7	Adultos

1.3.3.2. MÁQUINAS DE SOPORTE VECTORIAL

Otro algoritmo muy popular en aprendizaje de máquina usado frecuentemente para la detección automática de lesiones de FCD son las SVM. En Yang et al. [36] encontramos, en esta revisión de literatura, el primer uso de este algoritmo, en este caso para una segunda etapa de reclasificación de *vóxeles* clasificados como no-lesionales por un clasificador Bayes en una primera etapa arrojando una sensibilidad de 88 % en una base de datos compuesta por imágenes de 21 adultos.

En Hong et al. [40] usaron SVM para clasificar los tipos de lesiones presentes en imágenes MRI en secuencia T1 de una cohorte de 41 pacientes con FCD histológicamente confirmada (13 con FCD tipo I y 28 con FCD tipo II) y 41 controles. Mediante análisis de grosor cortical y de la complejidad de las plegaduras corticales (mapas de curvatura) de los pacientes a nivel individual y grupal, el algoritmo clasificó las lesiones en subtipos histopatológicos de FCD, esto es, como FCD tipo I ó FCD tipo II. Entrenada con estas características, la SVM pudo predecir el subtipo histológico de la FCD con una precisión promedio de 98 % (100 % en pacientes con FCD tipo I y 96 % en pacientes con FCD tipo II); en la lateralización del foco epileptogénico obtuvo un 92 % de precisión en FCD tipo I y 82 % en FCD tipo II; finalmente, predijo el resultado de la cirugía con una precisión de 92 % en pacientes con FCD tipo I y 82 % con FCD tipo II.

Una caso especial del algoritmo SVM llamado One-class SVM (OC-SVM, por sus siglas en inglés) fue empleado por El Azami et al. [41]. La OC-SVM es entrenada con ejemplos “negativos” de una base de datos de 77 sujetos sanos, para permitir la detección de anomalías a nivel de *vóxel* en un grupo de 11 pacientes con 13 lesiones (3 lesiones MRI-positivo). Para entrenar el algoritmo se calcularon dos mapas de características para todos los sujetos: un mapa de extensión de la GM en la WM, llamado “mapa de extensión”; y un mapa de la unión GM-WM, llamado “mapa de unión”. Para los pacientes de control se calcularon los mapas de media y desviación estándar de ambas características. Finalmente se obtuvo un mapa *z-score* restando el mapa de la media del mapa individual y dividiendo por el mapa de desviación estándar. El entrenamiento de la OC-SVM se hizo a nivel de *vóxel*, donde a cada *vóxel* se le asocia un clasificador OC-SVM y las características de entrada son el vector conformado por el valor del *vóxel* de la coordenada correspondiente en el “mapa extensión” y en el “mapa de unión”. Este algoritmo pudo detectar el 100 % de las lesiones en los casos MRI-positivos (3/13). En los casos MRI-negativos (10/13) detectó 7/10 lesiones para un 70 %.

Combinando características extraídas de imágenes MRI y PET de una cohorte de 28 pacientes con FCD histológicamente confirmada, Tan et al. [42] buscaron automatizar la detección de FCD sutiles o visualmente inidentificables. A partir de las imágenes, los autores extrajeron características morfológicas (grosor cortical y profundidad de los surcos) y características basadas en intensidad (intensidad

relativa de la materia gris y gradiente en la interfaz GM-WM). La clasificación fue realizada en dos etapas. El clasificador SVM de la primera etapa busca reconocer los vértices con lesión de FCD con la tasa de detección más alta. Las lesiones detectadas por el clasificador en la primera etapa sirven de entrada al clasificador basado en parches de la segunda, el cual, por medio de análisis estadístico de mapas tiene como objetivo eliminar los FP generados por el primer clasificador. De acuerdo con el análisis de sus experimentos, la máxima sensibilidad la obtuvieron incluyendo tanto las imágenes MRI y PET usando las dos etapas de clasificación, para lograr una sensibilidad de 93 %.

Finalmente, Wang et al. [43] proponen una comparativa de algoritmos de aprendizaje de máquina como los procesos Gaussianos para aprendizaje de máquinas (GPML, por sus siglas en inglés) y SVM. Aparte de que es el único estudio encontrado donde se tiene un algoritmo de GPML, la otra novedad que introduce es el entrenamiento con imágenes con tensor de difusión (DTI, por sus siglas en inglés) en conjunto con imágenes MRI en secuencia T2 de doce pacientes con FCD definida radiológicamente. Los resultados de la comparación del desempeño de ambos algoritmos indican que los clasificadores basados en GPML tienen mejor desempeño que el clasificador SVM, empleando la métrica de área bajo la curva (AUC, por sus siglas en inglés) se encontró que la media de AUC para el clasificador basado en GPML fue $AUC = 0,76$ y para la SVM fue de $AUC = 0,69$.

En la Tabla 1.3 se pueden observar los resultados obtenidos por los diferentes estudios que emplean SVM. Se excluye el estudio [43] ya que los autores emplean una métrica de evaluación diferente y por lo tanto no es posible su comparación con los demás. En la Tabla 1.3 se puede apreciar que Hong et al. [40] lograron un mejor desempeño, entrenando con 41 pacientes, para una sensibilidad de 98 %, seguido de Tan et al. [42] con 93% de sensibilidad y 28 pacientes. Los dos mejores resultados se deben probablemente a que cuentan con un mayor número de imágenes de entrenamiento y, como es de esperarse, se aprecia un descenso en el desempeño de los clasificadores a medida que el número de imágenes de entrenamiento disminuye.

Tabla 1.3: Tabla de resultados obtenidos con las diferentes implementaciones del clasificador SVM

Estudio	Técnica	Sensibilidad %	Secuencia	Pacientes	Grupo etario
Yang et al. [36]	Naive Bayes + SVM	88	T1	21	Adultos
Hong et al. [40]	SVM	98	T1	41	Adultos
El Azami et al. [41]	OC-SVM	77	T1	11 (13 lesiones)	Adultos
Tan et al. [42]	SVM-Mat	93	T1, PET	28	Adultos

1.3.3.3. ÁRBOLES DE DECISIÓN

Los árboles de decisión han sido utilizados también para resolver el problema de automatización de detección de FCD. Gill et al. [44], a través de características morfológicas y de intensidad basadas en superficie y provenientes de secuencias MRI T1, T2 y FLAIR junto con un mapeo de la relación FLAIR/T1 de 41 pacientes con FCD histológicamente confirmada y 38 controles, entrenan un clasificador de dos etapas. La primera etapa realiza clasificación de vértices para maximizar sensibilidad con el clasificador de potenciación de submuestreo aleatorio de gradiente (RUSBoost, por sus siglas en inglés) el cual, por medio de una combinación de sub-muestreo aleatorio y refuerzo adaptativo de aprendizaje (AdaBoost, por sus siglas en inglés), permite manejar el desbalance de clases en problemas de clasificación removiendo muestras de la clase que tiene el mayor número de ejemplos. La salida del

primer clasificador son agrupamientos de *vóxeles* etiquetados como lesionales y que sirven de entrada a la segunda etapa de clasificación basada en agrupamientos y que tiene como fin aumentar la especificidad del modelo removiendo los FP predichos en la primera etapa. Con este modelo los autores obtuvieron una sensibilidad de 83 % con 34/41 lesiones detectadas y una especificidad de 92 % siendo así que no se encontraron lesiones en 35/38 controles sanos.

Lin et al. [45] desarrollaron un clasificador entrenado con características corticales basadas en superficie para la localización precisa de lesiones en imágenes de 22 pacientes con FCD tipo IIb confirmada histopatológicamente. Integrando imágenes PET y MRI en secuencias T1 y FLAIR, extrajeron las siguientes características: grosor cortical, intensidad FLAIR, profundidad de los surcos, índice de *girificación* local, área de GM (mm^2), volumen de GM (mm^3), intensidad PET e intensidad del contraste GM-WM. El clasificador XGBoost seleccionado por los autores fue entrenado con solo características de MRI, solo PET y por último, con la combinación de PET y MRI. El clasificador que obtuvo mejor desempeño fue el entrenado con combinación de PET y MRI que arrojó 93 % en sensibilidad, 91 % de especificidad y la tasa de FP fue de 9 %.

En la Tabla 1.4 se muestran los resultados de los métodos que implementan árboles de decisión. Ambas versiones tienen resultados similares, sin embargo, Lin et al. [45] obtienen un resultado mayor usando casi la mitad de los pacientes que Gill et al. [44].

Tabla 1.4: Tabla de resultados obtenidos con las diferentes implementaciones del clasificador con árboles de decisión

Estudio	Técnica	Sensibilidad %	Secuencia	Pacientes
Gill et al. [44]	2 stage RUSBoost	92	T1, T2, FLAIR	41
Lin et al. [45]	XGBoost	93	T1, FLAIR, PET	22

1.3.3.4. OTROS MÉTODOS DE APRENDIZAJE DE MÁQUINA

Con el fin de automatizar la detección de FCD en pacientes con epilepsia extratemporal en pacientes MRI-negativo, Hong et al. [46] implementaron un clasificador LDA, también conocido como discriminante lineal de Fisher, el cual está en capacidad de realizar separación entre clases de acuerdo con la combinación lineal encontrada en los datos de entrada. El clasificador fue entrenado con características morfológicas (grosor cortical, profundidad de los surcos, curvatura) y de intensidad (intensidad relativa y gradiente) basadas en superficie y probado en una base de datos de 19 pacientes y 24 controles sanos y 11 controles con epilepsia en el lóbulo temporal. El algoritmo detectó lesiones en la misma ubicación que las etiquetas manuales en 14/19 pacientes para una sensibilidad de 74 % y 100 % de especificidad al no encontrarse lesiones en las MRI de los sujetos de control. Además, el clasificador fue evaluado en una segunda base de datos de 14 pacientes y 20 sujetos de control arrojando una sensibilidad de 71 % y especificidad de 95 %.

Ahmed et al. [47] implementan un conjunto de clasificadores de regresión logística para clasificar vértices lesionales y no-lesionales, entrenados con características morfométricas (grosor cortical, contraste GM-WM, profundidad de los surcos, curvatura y una medida de deformación cerebral global basada en las propiedades de los giros y los surcos), basadas en superficie, en pacientes MRI-negativos con FCD histológicamente verificada. Los autores emplean una estrategia innovadora ya que se entrenan clasificadores separados para regiones anormalmente gruesas y para regiones anormalmente finas; adicionalmente, se entrenan clasificadores separados para las paredes de los giros, los surcos y la corona para optimizar la detección de lesiones presentes en el fondo de los surcos. Este método

detectó lesiones en 6/7 pacientes MRI-positivos y 14/24 en pacientes MRI-negativos 58%.

Buscando reducir el número de FP predichos por los algoritmos de clasificación, Qu et al. [48] implementaron un clasificador por votación unánime de múltiples clasificadores (UVMC, por sus siglas en inglés), el cual fue desarrollado sobre una combinación de diferentes clasificadores como son: Naive Bayes, LDA, análisis discriminante de Mahalanobis y análisis discriminante cuadrático. Cada uno de estos clasificadores, de forma individual, es una componente del algoritmo UVMC y son alimentados con características extraídas de imágenes MRI de 10 pacientes con FCD y 31 controles para predecir si los píxeles pertenecen a la clase FCD-positivo ó FCD-negativo. De acuerdo con los autores, cuando todos los clasificadores etiquetan un *vóxel* como FCD-positivo, el método UVMC lo reclasifica como FCD-positivo, de lo contrario, lo reclasifica como FCD-negativo; de esta forma se reduce el número de FP y, aunque se reduce también el número de verdaderos negativos, se conserva la mayor parte de la región FCD-positiva. Si bien no se especifican los resultados numéricos obtenidos en la evaluación de este modelo, la evaluación de UVMC presenta una reducción de FP al ser comparado con otros algoritmos empleados en la literatura como, por ejemplo, el clasificador Bayes de dos etapas propuesto por [12] y lo mismo ocurre con cada uno de los clasificadores que lo componen por separado.

En un artículo posterior, Qu et al. [49] proponen un algoritmo UVMC similar empleando los mismos algoritmos de clasificación para decidir entre regiones lesionales y no-lesionales con algunas novedades. La primera novedad es que se extraen características en las imágenes con las que luego se forman varios grupos de características con las cuales se entrena cada uno de los clasificadores (Naive Bayes, LDA, análisis discriminante de Mahalanobis y análisis discriminante cuadrático) para escoger el mejor grupo de características de acuerdo con el desempeño de los clasificadores. Luego de seleccionar el mejor grupo de características, se clasifican los *vóxeles* como lesionales o no lesionales por el algoritmo de voto unánime de forma similar al trabajo anterior [48]. Por último, en otra novedad introducida al algoritmo, la clasificación es refinada mediante un análisis de regiones conectadas, *binarizando* la imagen predicha y aplicando una máscara para remover ruido de las regiones marcadas como lesión y luego removiendo aquellas regiones cuyo tamaño sea menor al de un umbral establecido. Este estudio también incluyó imágenes MRI en secuencia T1 de 10 pacientes con FCD y 31 sujetos de control, 8/10 pacientes fueron correctamente identificados como pacientes, mientras 30/31 sujetos de control fueron clasificados como sanos.

Lee et al. [50] proponen el único algoritmo de aprendizaje no supervisado que se encontró en la literatura. Se trata de un estudio ambicioso en el que se aplica *consensus clustering* (método que combina los resultados de múltiples algoritmos de agrupamiento) sobre características estructurales de FCD en MRI para identificar los grupos o clases que conforman una lesión; además se evalúa la relación entre las clases de FCD con su histopatología. La base de datos empleada en el estudio contiene 46 pacientes con FCD y un grupo de control de 35 sujetos. Para agrupar las regiones con lesión, se extraen características de superficie de todo el volumen MRI: grosor cortical, intensidad FLAIR normalizada, gradiente, relación T1/FLAIR y derivada funcionales en MRI funcional (fMRI, por sus siglas en inglés). Luego se realiza un análisis de estas características sobre la máscara de la región etiquetada como lesional por un experto. A las características de esta región se les aplica el algoritmo de *consensus clustering* para estimar la matriz de estabilidad que almacena la probabilidad de que las características intra-lesionales de FCD pertenezcan al mismo grupo o clase. Por último, los autores entrenan un algoritmo de clasificación XGBoost de dos etapas para relacionar los vértices etiquetados como lesionales a una clase específica de FCD, usando las mismas características del algoritmo de agrupamiento junto

con los valores de sus medias en vértices vecinos. La primera etapa del clasificador buscaba maximizar la sensibilidad y consta de cuatro clasificadores cada uno sintonizado para alguna de las cuatro clases descubiertas por el algoritmo de agrupamiento y cuya salida alimenta un meta-clasificador encargado de realizar la predicción final. El propósito de la segunda etapa es remover los FP detectados en la primera, usando los mismos cuatro clasificadores seguidos de un meta-clasificador. Como resultado se tiene que las lesiones FCD tipo II fueron parceladas en cuatro clases con distintos perfiles estructurales, histopatológicos e impacto funcional por el algoritmo de agrupamiento. En cuanto a detección de FCD, el clasificador entrenado con información de la clase lesión detectó un número mayor de lesiones que el clasificador entrenado sin esta información ($77 \pm 3\%$ vs. $73 \pm 3\%$).

1.3.4. MÉTODOS AUTOMÁTICOS BASADOS EN ALGORITMOS DE APRENDIZAJE PROFUNDO

Los métodos basados en morfometría, a pesar de ser prácticos y robustos en la implementación, tienen sus propias limitaciones. En primer lugar, dicen los autores, “no contienen una relación espacial a través de la superficie cortical, por lo que cualquier error en el registro puede resultar en que las lesiones más sutiles se pierdan de la imagen. Segundo, se requiere que un radiólogo experto realice la lectura del análisis morfométrico arrojado por el método” [51]. Estas limitaciones han volcado la atención hacia los nuevos métodos de aprendizaje de máquinas y aprendizaje profundo. Alrededor de comienzos de la década de 2000, los métodos de aprendizaje de máquina crecieron en popularidad dado el aumento de la capacidad de cómputo de los computadores personales y debido a su buen desempeño en tareas de análisis digital de imágenes. Sin embargo, en épocas recientes se han visto relegados a un segundo plano por los métodos de aprendizaje profundo, cuyo rendimiento ha demostrado ser superior.

En esta sección se presentan métodos de detección automática de FCD basados en dos tipos de algoritmos de aprendizaje profundo: redes neuronales artificiales (ANN, por sus siglas en inglés) y CNN. Aunque los dos métodos se pueden categorizar como aprendizaje profundo, se decidió presentarlos en dos subsecciones diferentes dado que estos tienen una gran diferencia: para el entrenamiento de las ANN es necesario realizar el proceso de extracción de características de la imagen, mientras que las CNN extraen automáticamente sus propias características de alto nivel.

1.3.4.1. REDES NEURONALES ARTIFICIALES

Utilizando morfometría y una congregación de redes neuronales para extraer automáticamente surcos cerebrales en MRI, Besson et al. [52] buscan cuantificar la relación especial que existe entre las lesiones de FCD y los surcos. Para la extracción de los surcos, los autores primero segmentan la imagen en mapas de GM, WM y CSF y reconstruyen las superficies de las interfaces GM-WM y GM-CSF, finalmente, extraen los surcos segmentando la interfaz GM-CSF después de un proceso de “esquele-tonización”. Los surcos son etiquetados usando una congregación de redes neuronales entrenada con etiquetas delineadas manualmente. Para cada sujeto de control se identificaron los surcos que coincidían con aquel relacionado a la FCD de cada paciente e inspeccionando su morfología y etiquetado, se aseguró la mejor equivalencia posible de los surcos entre sujetos. La profundidad de un surco dado fue calculada obteniendo la línea del “fondo surcal”, que está definida por el borde interior de este. La profundidad de cada punto a lo largo de esta línea se determinó calculando la distancia de Chamfer más corta a la superficie cortical exterior. Se evaluó la media y la máxima profundidad del surco para cada sujeto de una base de datos de 43 pacientes y 21 sujetos sanos, así como la profundidad promedio de la porción de fondo del surco ubicada en el vecindario de la lesión. Las lesiones fueron clasificadas entre pequeñas y grandes de acuerdo con un umbral definido por una medida de entropía. Con un umbral $T = 3093mm^3$, el método clasificó 21/43 casos como lesiones FCD pequeñas, entre ellas, 17/21 (81%)

habían sido ignoradas en la evaluación radiológica inicial; asimismo, 18/21 (86%) lesiones fueron localizadas en el fondo del surco cerebral, dos estaban relacionadas a las paredes de los surcos y una estaba ubicada en la corona del giro.

Profundizando el alcance de su método de clasificación de lesiones FCD entre lesiones pequeñas y grandes [52], Besson et al. [53] proponen un nuevo método para detectar lesiones de FCD pequeñas en MRI secuencia T1, apoyándose en características basadas en superficie que modelan la morfometría y textura de las lesiones FCD. La detección automática fue realizada por un clasificador por vértices basado en redes neuronales entrenado con etiquetas manuales. La clasificación por vértices estuvo compuesta por una ANN *feed-forward* de cuatro capas con 6 neuronas en la primera capa, 4 en la segunda, 4 en la tercera y una neurona en la capa de salida. De una base de datos que incluía 41 pacientes, se determinó que 19 de ellos tenían lesión pequeña de FCD de acuerdo con el método expuesto en [52], estos 19 pacientes fueron seleccionados para el estudio junto con 45 sujetos sanos. El clasificador detectó lesiones en 17/19 (89%) pacientes seleccionando un umbral probabilístico $T = 0,93$ que ofrecía el mejor balance entre FP y tasa de detección.

Posteriormente, Besson et al. [54] complementan su metodología para detección de lesiones pequeñas de FCD, tomando como base el clasificador empleado en su trabajo anterior [53] y añadiendo un nuevo clasificador con el fin de reducir el número de FP. La detección automática fue entonces realizada por un clasificador de dos pasos, donde, al clasificador por vértices basado en redes neuronales y entrenado con etiquetas manuales, de la primera etapa, se le adicionó un segundo clasificador por agrupamiento. En la primera etapa se utilizó el mismo algoritmo de ANN salvo la modificación de algunos hiperparámetros. Los grupos generados por este clasificador fueron separados entre FP o lesión de FCD, por el segundo clasificador, basándose en características globales. En el segundo clasificador se empleó un algoritmo de K-vecinos más cercanos difuso (fKNN, por sus siglas en inglés) para determinar la membresía a la clase lesión FCD. La base de datos contó con los mismos 41 pacientes, de los cuales en 19 su lesión había sido clasificada como lesión pequeña de FCD de acuerdo con el método expuesto en [52]; estos 19 pacientes fueron seleccionados para el estudio, esta vez junto con una base de datos de sujetos sanos incrementada a 48 sujetos. El primer clasificador detectó lesiones en 18/19 (95%) pacientes, con un nuevo umbral probabilístico $T = 0,87$; el clasificador fKNN, con $k = 20$, mantuvo 13/19 (68%) lesiones y redujo eficientemente la cantidad de FP.

Para desarrollar una herramienta para la detección automática de FCD que aborde el desafío de pacientes pediátricos, Adler et al. [55] buscan optimizar la habilidad para encontrar y cuantificar cada área de la corteza en términos de como difiere de una corteza sana. Para ello se calcula medidas estructurales y se aplican métodos de pos-procesamiento para cuantificar el número de identificadores radiológicos de las lesiones de FCD. Primero, se aborda el problema que tienen marcadores estructurales establecidos de FCD (grosor cortical, intensidad, etc), que pueden aparecer como regiones normales y ocultar valores locales anormales dentro de una lesión FCD, normalizando las medidas entre sujetos, calculando las asimetrías interhemisféricas entre esta medidas y normalizando los valores para cada vértice relativo a los sujetos de control. Los marcadores estructurales comunes a las regiones con lesión de FCD, como grosor cortical, contraste GM-WM e intensidades en la señal FLAIR, fueron cuantificados con un método denominado “*doughnut*”, el cual calcula la diferencia entre un área de la corteza y su anillo circundante en cada vértice, resaltando dónde estas diferencias son mayores. Adicionalmente, los autores desarrollaron una medida de deformación cortical local, basada en la magnitud de la curvatura intrínseca que rodea cada vértice, como una medida más robusta de la forma cortical.

Estos marcadores estructurales y métodos de pos-procesamiento (deformación cortical local, asimetría interhemisférica y “*doughnuts*”) fueron combinados con características de superficie comunes en detección de FCD (grosor cortical, contraste en la frontera GM-WM, intensidad de la señal FLAIR, curvatura y profundidad de los surcos) para entrenar una ANN para clasificar regiones corticales entre vértices con lesión y sin lesión. La ANN estuvo compuesta por una sola capa oculta y el número de nodos fue determinado mediante el análisis de componentes principales (PCA, por sus siglas en inglés) de las características de entrada, usando el número de componentes que representara más del 99% de la varianza. La ANN se entrenó con un total de 28 características basadas en los procedimientos anteriormente descritos; adicionalmente, los autores entrenaron ANN separadas usando características individuales y subconjuntos de datos para evaluar su valor discriminatorio. El estudio, basado en pacientes pediátricos, incluyó 22 pacientes con diagnóstico radiológico de FCD y 28 sujetos de control. Individualmente, de las características estándar de FCD, se encontró que la que aparentemente tiene mejor valor discriminatorio es la intensidad FLAIR (AUC = 0.83), seguida del contraste GM-WM (AUC = 0.80) y grosor cortical (AUC = 0.63). De las nuevas características propuestas en el artículo, la asimetría en la intensidad FLAIR mostró el mejor valor discriminatorio (AUC = 0.87). Empleando tanto características nuevas como características establecidas en el análisis de FCD, el modelo fue capaz de detectar 16/22 (73%) lesiones de FCD. Usando solo características establecidas de FCD, el modelo detectó 13/22 (53%) lesiones de FCD, lo cual es evidencia de que la inclusión de las nuevas características fue de ayuda en la detección de las FCD.

Jin et al. [51] proponen un algoritmo para la detección automática de FCD mediante una ANN, con el fin de evaluar el valor de diagnóstico que tienen las características basadas en morfometría de superficie (SBM, por sus siglas en inglés) en pacientes con FCD tipo II histológicamente confirmada. Para entrenar la ANN, los autores adquirieron seis características de cada vértice de la reconstrucción cortical 3D: grosor cortical, contraste de la intensidad GM-WM, curvatura, profundidad de los surcos, mapas *doughnut* y deformación cortical local (LCD, por sus siglas en inglés). Para las características de grosor cortical, el contraste de la intensidad GM-WM y LCD, se calculó también su asimetría interhemisférica restando los valores de los vértices del hemisferio derecho de los valores del hemisferio izquierdo para crear el mapa de asimetría del hemisferio izquierdo; y viceversa para crear el mapa de asimetría del hemisferio derecho. Adicionalmente, se entrenaron ANNs separadas para cada característica con el fin de evaluar su valor discriminatorio. Para evaluar el desempeño de la ANN se usó una estrategia de validación cruzada con $k = 5$, donde los vértices del 5% superior fueron identificados y conectados con los grupos vecinos. Adicionalmente, se estableció un umbral, mediante análisis de la curva característica operativa del receptor (ROC, por sus siglas en inglés), de forma que los vértices cuyos valores están por encima del umbral fueron etiquetados como lesión y por debajo como no-lesión. Finalmente, se tiene que la detección tuvo éxito si la región detectada se superpone con la etiqueta manual. Este estudio incluyó bases de datos de tres centros radiológicos diferentes, con población de sujetos tanto pediátricos como adultos, para un total de 61 pacientes, 17 de ellos MRI-negativo, y 120 sujetos de control; adicionalmente, se incluyó una base de datos de prueba con 15 pacientes y 35 sujetos sanos. Se evaluaron tres factores para determinar su efecto en el desempeño del clasificador: 1. Número de casos de entrenamiento; 2. Tamaño de la base de datos de sujetos de control; 3. Tipo de escáner. En cada factor, definiendo grupos de entrenamiento separados para cada una de las tres bases de datos. El algoritmo mostró un mejor desempeño con un umbral $T = 0,9$, con una sensibilidad de 73.7% y especificidad de 90% (para los datos de prueba, 91.4% de especificidad en sujetos sanos y 86.7% en pacientes). En cuanto al análisis del efecto del número de pacientes, se detectó que entre menor era el número de pacientes incluidos en el entrenamiento, menor era la sensibilidad y la es-

pecificidad del modelo; un resultado similar se encontró al reducir el número de sujetos sanos en el entrenamiento. En cuanto al efecto del tipo de escáner, se encontró que esto no tiene un efecto significativo en el desempeño del algoritmo.

Combinando características cuantitativas y multimodales basadas en superficie con aprendizaje de máquina, Mo et al. [56] construyen un modelo para detección automática de FCD con el fin de evaluar su valor clínico. El trabajo incluyó imágenes estructurales y funcionales en secuencias T1, T2-FLAIR y PET de 40 pacientes con MRI-positivo con FCD tipo II histológicamente confirmada y de 33 sujetos de control, con esclerosis hipocampal o quistes epidérmicos, dado que los autores no pudieron reunir una base de datos de sujetos sanos que incluyera imágenes PET. Después del pre-procesamiento y normalización de las imágenes, los autores realizaron reconstrucción cortical donde, entre otros procesos, se extrajo el tejido no cerebral, se segmentaron las estructura de GM y WM y se hizo teselado de la frontera GM-WM. Luego, se extrajeron características morfológicas (intensidad del contraste GM-WM, deformación cortical local, grosor cortical, curvatura media, profundidad de los surcos, *doughnut* de intensidad, *doughnut* de grosor), características de intensidad (intensidad FLAIR a un nivel diferente de la profundidad cortical, *doughnut* de FLAIR) y metabólicas (PET hipointensa, asimetría PET). La detección automática fue realizada mediante un clasificador ANN entrenado con las anteriores características tomadas en los vértices etiquetados como lesión y en vértices seleccionados etiquetados como no-lesión. Se aplicó PCA para reducir la dimensionalidad de entrada y acelerar el entrenamiento y se dividió la base de datos en 70% para entrenamiento, 15% para validación y 15% para prueba. El entrenamiento se evaluó mediante validación cruzada con $k = 5$ durante 100 iteraciones, encontrando lesiones cuya ubicación correspondía con la etiqueta manual en 31/40 pacientes para una tasa de detección de 77.5% y una sensibilidad de 70.5%.

Wagstyl et al. [57] buscan evaluar la factibilidad de incorporar algoritmos de aprendizaje profundo, entrenados con MRI, en el planeamiento de implantación quirúrgica de cátodos de estereo electroencefalografía (sEEG, por sus siglas en inglés) en pacientes pediátricos con FCD, analizando el grado de correspondencia entre la localización de la lesión por el algoritmo automático y la zona de convulsión detectada mediante sEEG. El estudio contó con imágenes en secuencia T1 y FLAIR de una cohorte de 34 pacientes MRI-positivo y una cohorte de 34 pacientes previamente intervenidos con sEEG, ambas bases de datos compuestas por sujetos entre 3 y 18 años; adicionalmente, se incluyó un grupo de 20 sujetos de control sin ningún historial neurológico. Para entrenar el algoritmo de aprendizaje profundo, los autores extrajeron las siguientes características morfológicas y de intensidad, calculadas por vértice a través de la superficie cortical: grosor cortical, intensidad del contraste GM-WM, curvatura, profundidad de los surcos, curvatura intrínseca, intensidad de la señal FLAIR muestreada al 25%, 50%, y 75% del grosor cortical y de la frontera GM-WM y, finalmente, fueron calculados mapas de asimetría interhemisférica. La arquitectura de la ANN contó con una capa oculta, cuyo número de nodos, calculado usando PCA sobre las características de entrada, fue el mínimo número de componentes que representó al menos el 99% de la información. El algoritmo pudo encontrar lesiones de FCD en 25/34 pacientes MRI-positivo para una sensibilidad de 74% y no se encontró lesión en ninguno de los 20 sujetos de control (especificidad 100%). De los 34 pacientes intervenidos con sEEG, se identificó lesión de FCD en 21 de ellos. Comparado con los pacientes que fueron intervenidos con sEEG, la lesión predicha automáticamente correspondió con la determinada por sEEG en 13/21 pacientes para una correspondencia de 62%, indicando que el algoritmo automático puede llegar a ser útil para evitar el implante quirúrgico de electrodos.

Con el fin de mejorar los métodos de detección automática de FCD existentes, Ganji et al. [58] diseñaron e implementaron un sistema de diagnóstico computarizado para identificar lesiones de FCD, asumiendo que por medio de métodos morfométricos y de aprendizaje se puede detectar y localizar con precisión las lesiones de FCD. El estudio se realizó con imágenes estructurales en secuencia T1 y T2-FLAIR de 30 pacientes con FCD tipo II confirmada y 28 sujetos de control para evaluar especificidad. Las imágenes pasaron por un proceso de reconstrucción cortical para luego extraer las siguientes características basadas en morfología e intensidad, parcelando la corteza cerebral en 34 regiones por hemisferio y calculando para cada una de ellas: grosor cortical promedio, curvatura Gaussiana, curvatura media, índice de curvatura intrínseco e índice de pliegues; se incluyeron también índices estadísticos como media, mínimo, máximo, rango, desviación estándar y relación señal a ruido para la intensidad del contraste. La clasificación sirvió dos propósitos: primero, diagnosticar la región como lesión o no-lesión; segundo, detectar el área de la lesión. El clasificador ANN fue entrenado durante 30 épocas, con el 70% de los datos para entrenamiento y el resto se dividió en iguales proporciones para validación y prueba. La ubicación de la lesión se determinó en dos pasos: primero, se determinó el hemisferio de la lesión (hemisferio derecho o izquierdo), y luego el lóbulo cerebral (lóbulo frontal, temporal, parietal u occipital). Después de entrenar el algoritmo por 30 épocas, los valores promedio de sensibilidad, especificidad y precisión fueron 96.7, 100 y 98.6%, respectivamente. Evaluado el método en 10 pacientes MRI-negativo, se tuvo una precisión de 91.3%. Asimismo, la precisión del clasificador fue de 84.2% para detectar el hemisferio de la lesión y 77.3% para la detección del lóbulo cerebral.

Buscando obtener una clasificación robusta y automatizada de FCD, David et al. [59], buscan desarrollar un clasificador, integrando la información de tres tipos de mapas morfométricos con los resultados de la segmentación de tejidos de una ANN, y probar su generalización y desempeño en un conjunto de datos grande e independiente de pacientes con FCD histológicamente confirmada. Primero, los autores realizaron análisis morfométrico de las imágenes para extraer las imágenes o mapas de extensión, unión y grosor. Luego, con una base de datos compuesta por MRI en secuencia T1 de 113 pacientes (pediátricos y adultos) y 362 sujetos sanos (adultos), se extrajeron las siguientes mapas de características para entrenar la ANN: imagen T1 normalizada, mapas morfométricos (extensión, unión y grosor), mapas morfométricos suavizados, mapas de tejidos (GM, WM y CSF), mapas de tejidos suavizados, mapa del cerebro y mapa de la frontera GM-WM. La ANN fue diseñada con una arquitectura de dos capas con cinco neuronas *sigmoide* en la capa oculta y dos neuronas *softmax* en la capa de salida, con el fin de distinguir entre dos clases: *vóxel* displásico y no-displásico. Para la clasificación de un *vóxel* específico, los valores de los 15 mapas de características fueron leídos en el *vóxel* correspondiente y usados como entrada a la ANN. El desempeño del entrenamiento fue evaluado mediante validación cruzada con $k = 10$ iteraciones. La base de datos de validación incluyó 60 pacientes con FCD (pediátricos y adultos) y 76 sujetos sanos (adultos). En la evaluación con los datos de entrenamiento, la ANN detectó en promedio 98.8/113 FCDs para una sensibilidad media de 87.4% con una especificidad promedio de 85.4%, lo que corresponde a una precisión de 85.9%. El desempeño del modelo también fue evaluado con una base de datos independiente compuesta por 58 pacientes con FCD y 70 sujetos sanos. Con estos datos independientes, el método detectó 47/58 FCDs alcanzando una sensibilidad de 81% y, en cuanto a los sujetos sanos, detectó lesiones tan solo en 11/70 sujetos para una especificidad de 84.3% en la base de datos independiente.

El objetivo de De Freitas Oliveira Baffa et al. [60] es desarrollar un método de clasificación, basado en extracción de características morfológicas, en el cual un algoritmo de aprendizaje de máquinas pueda reconocer patrones y asistir al personal sanitario en la detección de lesiones de FCD. En este trabajo

se presentó un método de clasificación basado en *vóxel* por medio de un *perceptrón multicapa* (MLP, por sus siglas en inglés). Adicionalmente, los autores entrenaron múltiples clasificadores tradicionales para compararlos y seleccionar el método de mejor desempeño, además con las siguientes novedades: (a) Desarrollo de un vector de características extendido, (b) Desarrollo y evaluación de un ensamble de métodos de aprendizaje para detección de FCD, (c) Evaluación de la clasificación a nivel de paciente. Después de pasar por una etapa de preprocesamiento y extracción de cerebro, se extraen un total de 20 características morfológicas y de textura para alimentar los algoritmos de clasificación. El vector descriptor se compone de ocho características de textura de Haralick: energía, entropía, correlación, contraste, tono del agrupamiento, homogeneidad, prominencia del agrupamiento, correlación de Haralick; junto con el nivel de intensidad y grosor cortical que es una característica morfológica. Estas características se comparan luego con una plantilla o mapa que representa una corteza cerebral sana y se calcula el *z-score* de cada característica, resultando en un vector de características de tamaño 20. Adicionalmente, se generaron 10 nuevas características basadas en la suma de las 10 características de mayor impacto con la 10 de menor impacto, midiendo el impacto de cada una por medio del algoritmo *SelectKBest* de *Scikit-Learn*, para finalmente resultar con un vector de características de tamaño 30. Los clasificadores, implementados con hiperparámetros obtenidos empíricamente, fueron: (a) XGBoost, (b) K-vecinos más cercanos (KNN, por sus siglas en inglés), (c) árboles de decisión, (d) bosques aleatorios, (e) MLP, (f) Naive Bayes, (g) AdaBoost y (h) SVM. El clasificador MLP propuesto fue conformado por una capa escondida con 100 neuronas con función de activación lineal rectificadas (ReLU, por sus siglas en inglés), optimizador Adam y entrenado por 5000 épocas con una tasa de aprendizaje de 0.001. Se utilizó el procedimiento de validación cruzada con $k = 7$ para tener al menos dos pacientes para prueba. El estudio contó con MRI de 15 pacientes. A nivel de *vóxel*, MLP logró el mejor desempeño con una exactitud promedio de 96.81 %, precisión de 97.12 % y sensibilidad 96.47 %. A nivel paciente, la SVM obtuvo la exactitud más alta con 90.64 %, seguida del clasificador AdaBoost con 90.22 % y el MLP con 89.3 %, lográndose buenos resultados tanto a nivel de *vóxel* como a nivel paciente.

1.3.4.2. REDES NEURONALES CONVOLUCIONALES

Aprovechando la capacidad que tienen las CNN de extraer características automáticamente, Gill et al. [61] proponen un algoritmo con un preprocesamiento de datos mínimo y usando para entrenamiento imágenes MRI en secuencia T1 y T2, validando el algoritmo, por primera vez según los autores, en una base de datos proveniente de siete diferentes centros radiológicos alrededor del mundo. Los autores proponen entrenar el algoritmo con parches extraídos de las MRI con el fin de eliminar el desbalance de clases inherente a las lesiones de FCD, donde existe un número significativamente mayor de *vóxeles* de tejido sano que de *vóxeles* con malformación, creando una base de datos de entrenamiento compuesta por un número igual de ejemplos de ambas clases y tomando solo aquellos *vóxeles* donde la hiperintensidad en la secuencia FLAIR fuera superior a un umbral establecido. El clasificador diseñado por los autores consistió de dos CNN idénticas, en cascada, con sus pesos optimizados independientemente. La primera CNN fue entrenada con el objetivo de maximizar el hallazgo de *vóxeles* con lesión, mientras la segunda CNN fue entrenada para disminuir el número de FP manteniendo una óptima sensibilidad. Cada CNN se compuso de tres pilas de capas convolucionales y de *max-pooling*, cada una con 48, 96 y 2 filtros, respectivamente. Se usó activación ReLU en las dos primeras capas, mientras que en la capa de salida se usó activación *softmax* para obtener una distribución binomial de las clases estimadas. Para entrenar el algoritmo se dividió la base de datos, compuesta por una cohorte de 40 sujetos con FCD histológicamente confirmada, en 75 % para entrenamiento y 25 % para validación, los pesos fueron inicializados aleatoriamente y la función de costo empleada fue entropía cruzada binaria (BCE, por sus siglas en inglés). Durante el entrenamiento, el modelo fue validado me-

dianter validación cruzada con $k = 5$ durante cinco épocas. La evaluación del modelo se realizó sobre una cohorte independiente, compuesta por 67 sujetos con FCD histológicamente confirmada (adultos y pediátricos) de seis centros radiológicos diferentes, 38 sujetos de control sanos y 63 sujetos de control con epilepsia en el lóbulo temporal (TLE, por sus siglas en inglés). La validación cruzada en entrenamiento de las CNN resultó en una sensibilidad de $87 \pm 4\%$, detectando en promedio 35/40 lesiones, y una especificidad de 95% en sujetos sanos y de 90% en sujetos con TLE. Evaluando el modelo en los sujetos de prueba de los siete diferentes centros, la sensibilidad fue de 91% detectando 61/67 lesiones, observándose 3 ± 2 FP en 47/67 casos.

En Dev et al. [62] los autores proponen el primer algoritmo de CNN, para detección y segmentación de FCD, entrenado exclusivamente usando MRI en secuencia FLAIR, ya que en esta secuencia son más evidentes las regiones hiperintensas por presencia de FCD. La metodología desarrollada por los investigadores incluyó los pasos de preprocesamiento (reducción de ruido y extracción de cráneo), segmentación de FCD (arquitectura U-Net) y postprocesamiento (filtrado). La arquitectura del consiste de una red U-Net que toma como entrada imágenes de tamaño 256×256 . En el camino de contracción, cada bloque consiste de dos capas convolucionales, con filtros de 3×3 , seguidos de una función de activación ReLU y de una capa de normalización del *batch* y finalizando con una operación de *max-pooling*. Para compensar la pérdida de información espacial debido a la operación de *max-pooling*, se dobla el número de filtros en la capa convolucional de los bloques subsiguientes. En el camino de expansión, se incrementan las dimensiones espaciales de los mapas de características por medio de una operación convolucional transpuesta y, luego, se concatenan los mapas de características de alta resolución del camino de contracción con la imagen sobremuestreada resultante. Después de cada concatenación, se aplica una capa de *dropout* para reducir el efecto del sobreentrenamiento. Finalmente, se aplica una capa convolucional de 1×1 con activación sigmoide cuya salida es una imagen con las mismas dimensiones de la original, y que luego es umbralizada para obtener la máscara binaria de segmentación. El algoritmo fue entrenado durante cien épocas, como función de costo se empleó una combinación de BCE y Sørensen-Dice, y para actualizar los pesos se empleó el optimizador Adam con una tasa de aprendizaje de 0.001, reducida por un factor de 0.4 cuando el costo en validación no mejora su desempeño. Durante el entrenamiento del modelo se aplicaron técnicas de aumento de datos (rotación, zoom y *giramiento* horizontal) en tiempo real. Finalmente, con el fin de reducir el número de falsos positivos, se aplica un filtro promedio con *stride* = 1 y tamaño 5×5 . Para el experimento los autores contaron con una base de datos de 43 sujetos con epilepsia refractaria debido a FCD, con 17 MRIs tomadas en escáner de 1.5T con dimensiones 256×224 y número de *slices* entre 144 y 192; y 26 MRI tomadas en escáner de 3T con dimensiones 512×512 con 320 *slices*. Se entrenó el algoritmo con *slices* 2D de cada volumen y se dividieron las imágenes en 80% para entrenamiento y 20% para prueba, dando lugar a 5680 imágenes de entrenamiento, 1419 para validación y 1780 para prueba. Este modelo logró una tasa de detección de 82.5% (33/40), con una precisión de 80.69% y un coeficiente Sørensen-Dice de 52.47%.

Wang et al. [63] proponen una técnica para detección automática de FCD usando MRI y aprendizaje profundo por medio de una arquitectura de CNN entrenada con parches. El estudio incluyó imágenes MRI en secuencia T1 de una cohorte de 10 pacientes con FCD confirmada, 10 sujetos sanos y 10 sujetos con TLE como grupo de control. Estas imágenes pasaron por una etapa de preprocesamiento que incluye, entre otros, extracción de cerebro y corrección de intensidades y normalización. Luego, se extrajeron parches rectangulares de la corteza cerebral en el eje axial de la MRI, primero extrayendo parches máscara del mapa de probabilidad de GM y luego aplicando la máscara a la MRI

preprocesadas del cerebro. En total se extrajeron un total de 129780 parches (4326 parches/imagen), 2302 de ellos etiquetados como FCD 127478 como no-FCD. Para solucionar este desbalance de clases, se aplica aumento de datos usando el método de sobremuestreo para incrementar el número de parches con etiqueta FCD, resultando en un total de 36786 parches de la clase FCD, luego se selecciona aleatoriamente el mismo número de parches etiquetados como no-FCD para balancear los datos de entrenamiento. Todos los parches se dividen en 70% para entrenamiento, 15% para validación y 15% para prueba. La arquitectura de la CNN propuesta consiste de cinco capas convolucionales, una capa de *max-pooling* y dos capas totalmente conectadas (FC, por sus siglas en inglés). Cada capa convolucional consiste de tres operaciones: convolución, normalización de *batch* y activación ReLU. Luego de la capa convolucional se pasa por la capa de *pooling*, donde por medio de submuestreo *max-pooling* con un filtro de 2×2 , se reducen los parámetros a entrenar, luego se reducen las dimensiones por medio de una operación de *flatten* para terminar con vectores de características de 8448×1 . Por último, los vectores de características pasan por los 256 nodos de la primera capa FC cuya salida va a la segunda capa FC compuesta por dos nodos. La activación final es una capa *softmax* que entrega la probabilidad de cada clase. La función de costo para el entrenamiento es la función de entropía cruzada para clasificación multiclase optimizada mediante el descenso de gradiente estocástico (SGD, por sus siglas en inglés) con momentum. Con el método propuesto, los autores obtienen una exactitud de 0.941 y un área bajo la curva ROC de 0.985 en entrenamiento, indicando un buen desempeño del algoritmo. Para validación se utiliza el método de validación cruzada y se realiza el cálculo de sensibilidad, especificidad, exactitud y coeficiente Sørensen-Dice. Se detectan 9/10 lesiones de FCD en la base de datos de sujetos con lesión para una sensibilidad de 90%, evaluada en los sujetos de control, la especificidad fue de 85%, la exactitud de 88% y 0.78 fue la media del coeficiente Sørensen-Dice.

Usando una CNN, Feng et al. [64] tienen el objetivo de clasificar y localizar FCD en MRI en pacientes con epilepsia a partir de simulación de FCD. Para entrenamiento, los autores obtienen 636 MRI de 89 sujetos sanos (pediátricos y adultos) a las cuales añaden, localmente, una región brillante para simular el área de la lesión FCD. Se crearon tres conjuntos de datos para entrenamiento: el primero, con una magnificación de intensidad de 1.5 veces en el área donde se simula la lesión; el segundo, magnificado por 1.2; y el tercero magnificado por 1.1. Cada conjunto de datos quedó compuesto por 1272 imágenes, de las cuales 636 correspondieron a muestras positivas y las otras 636 a muestras negativas. La base de datos de prueba consistió de 53 volúmenes MRI, de los cuales 18 volúmenes pertenecían a 10 pacientes (pediátricos y adultos) con FCD y 35 volúmenes a 6 sujetos sanos en edad adulta. La arquitectura de la CNN consistió de cuatro capas convolucionales con activación sigmoide, acompañadas cada una de una operación de *average pooling*; la última capa, es una capa totalmente conectada con las funciones de activación de las neuronas de la capa anterior y cuya activación también es la función sigmoide. La función de costo es el error cuadrático medio optimizada mediante el método del descenso de gradiente, con una tasa de aprendizaje $\alpha = 1$ durante 2000 épocas. Después de entrenados los modelos para cada una de las tres bases de datos generadas, se obtuvo para la primera base de datos de entrenamiento (con magnificación de 1.5), un 100% en clasificación y 65% en localización, y en prueba 67.92% y 16.67% en clasificación y localización respectivamente. Para la segunda base de datos de entrenamiento (magnificación de 1.2), 100% en clasificación y 89% en localización, y en prueba 79.25% y 71.43% en clasificación y localización, respectivamente. Los mejores resultados fueron obtenidos con la tercera base de datos (magnificación de 1.1), con 100% en clasificación y 92% en localización en entrenamiento y en prueba 92.45% en clasificación y 92.86% en localización.

En un estudio posterior, Feng et al. [65] proponen un nuevo método, basado en CNN, con el pro-

pósito de detectar y segmentar lesiones de FCD en imágenes FLAIR de pacientes MRI-negativo con FCD histológicamente confirmada. Los autores denominan el método “Maximización de activación y localización convolucional” y consta de cinco pasos: (a) preprocesamiento de la imagen, (b) diseño y entrenamiento de una CNN, (c) obtención de una serie de bloques de patrones en la imagen (PIB, por sus siglas en inglés) usando la CNN entrenada, (d) usar el PIB promedio con convolución para localizar y segmentar el área de la lesión, y (e) evaluación cuantitativa del desempeño del método. La fase de preprocesamiento incluyó los pasos de extracción de cerebro de la imagen, redimensionamiento a 256×256 píxeles y normalización; luego se extraen, para entrenamiento, bloques positivos de 28×28 píxeles del área de la lesión y bloques negativos, de igual dimensión, del área normal; en total se obtuvieron 126684 bloques negativos y 148428 positivos para el entrenamiento de la CNN. La fase de entrenamiento de CNN, esta compuesta por una CNN denominada Net-Pos, la cual consiste de dos capas convolucionales con activación ReLU, dos capas de *max-pooling* y dos capas totalmente conectadas. Los autores seleccionaron, como función de costo, la función de verosimilitud logarítmica optimizada mediante SGD con momentum y con una tasa de aprendizaje de 0.01; los datos fueron divididos en 80% para entrenamiento y 20% para prueba. Para obtener los PIB de la tercera fase del método, se usa el concepto de maximización de activación mediante descenso de gradiente, primero, alimentando el modelo entrenado con ruido o una imagen aleatoria, luego, calculando los gradientes de una neurona específica de una capa específica con respecto a la imagen ruidosa para maximizar la neurona y, por último, actualizando iterativamente cada píxel en la imagen ruidosa para obtener la imagen patrón o PIB. En la siguiente fase se calcula el PIB promedio y se emplea para realizar una operación convolucional de emparejamiento de patrones. En la última fase, la evaluación de desempeño es llevada a cabo mediante los métodos de exhaustividad, especificidad y exactitud. La técnica fue evaluada usando 18 imágenes de 12 pacientes FLAIR-negativo, arrojando una sensibilidad a nivel de sujeto de 83.33% (15/18 lesiones detectadas), y el desempeño de la segmentación, medido con el coeficiente Sørensen-Dice, fue de 52.68%.

Mediante mejoras a los recientes métodos de aprendizaje profundo, en particular al método propuesto en [63], Aliev et al. [66] intentaron resolver el problema de detección automática de FCD en MRI. La base de datos de entrenamiento incluyó MRI etiquetadas de 15 sujetos con FCD, 15 MRI no etiquetadas y 17 sujetos sanos de control. Para este estudio, las etiquetas fueron marcadas por un radiólogo experto mediante cajas de delimitación 2D por vista (axial, coronal y sagital). Los autores tomaron [63] como modelo base, el cual consiste de cuatro pasos: (a) preprocesamiento, (b) extracción de parches de cerebro, (c) clasificación mediante aprendizaje profundo, y (d) posprocesamiento. El preprocesamiento incluyó, entre otros, extracción de cerebro, normalización y obtención de regiones *ground truth* formando paralelepípedos (3D) a partir de las cajas de delimitación 2D. La extracción de parches es una parte fundamental del algoritmo de acuerdo con los autores, por lo que evalúan los modelos de acuerdo con las probabilidades asignadas a los parches, para ello introducen el método de puntaje *top-k* como métrica de evaluación del modelo. La métrica consiste en tomar los k mejores parches por probabilidad y estimar si al menos uno de ellos interseca la región *ground truth*, en ese caso la detección es exitosa. El modelo base tuvo un puntaje *top-k* de 0.2, sin embargo, este fue mejorado por los autores implementando algunas mejoras: (1) los paralelepípedos con las máscaras fueron reemplazados por elipses 3D. (2) se implementó un etiquetado “suave”, donde, en lugar de solo ceros y unos, de acuerdo con la cantidad de solapamiento entre la elipse y la región *ground truth*, se asignan valores cercanos a 1 cuando hay mucho solapamiento y cercanos a 0.5 cuando este es de menor tamaño. (3) Se modificó el tamaño de los parches de 16×32 , ya que su tamaño no es suficiente para cubrir todas las regiones cerebrales, por el nuevo tamaño de 24×40 . (4) Se incluyeron *slices* de los

ejes coronal y sagital, usando aquellos *slices* que se cruzan en el centro con su respectivo *slice* en el eje axial, así como sus vecinos reflejados. (5) Dado que en 8/15 pacientes la lesión de FCD está ubicada en giro temporal y en 7/15 en la región no-temporal, y dada la diferente naturaleza de estas dos regiones, se entrenan modelos diferentes para cada uno de los dos grupos de sujetos. (6) Para sacar ventaja de los MRI de 15 sujetos no etiquetados, se preentrena un *autoencoder* con la arquitectura *3 heads - 3 tails*, donde cada *slice* de cada eje tiene su propia convolución y los resultados se concatenan seguidos de una operación lineal. El modelo que obtuvo el mejor puntaje *top-k*, es decir de mejor desempeño, pudo detectar 11/15 sujetos para una sensibilidad de 73.3%.

En un intento por minimizar las desventajas de la red U-Net implementada en [62], Thomas et al. [67] proponen una red U-Net con compuertas de atención óptimamente localizadas y con la elección correcta de los mapas de características para la señal de activación de las compuertas, así como con la ubicación óptima de las rutas residuales (ResPath, por sus siglas en inglés) dentro de la arquitectura propuesta para segmentación de FCD. Los autores consideraron que con la adición de las siguientes características de diseño, se podrían eliminar las desventajas de la arquitectura U-Net base [62]: (a) Atención a las características salientes (para prevenir la pérdida de las características más finas entre las capas intermedias y las capas finales, permitiendo enfocar el modelo en regiones específicas de la imagen); (b) Robustez a la escala y reducción de la brecha semántica (dado que las imágenes para las tareas de segmentación de tumores, lesiones, órganos, etc., contienen escalas variantes de las ROI; asimismo, la brecha semántica aumenta la discrepancia en el aprendizaje y es perjudicial para el desempeño del modelo). Para resolver el problema de invariabilidad a la escala, los autores incorporaron bloques multiresiduales los cuales, primero, reducen los altos requerimientos de memoria de los *kernel* de 5×5 y 7×7 usando cadenas de *kernel* de 3×3 en serie y, segundo, el bloque incrementa el número de filtros de las tres capas sucesivas en pasos, previniendo que la sobrecarga de memoria de las capas anteriores se propague a las capas siguientes. Para disminuir los efectos de la brecha semántica se adopta el procedimiento de ResPath, bajo la suposición de que un modelo muy profundo no es deseable, dado el pequeño tamaño de las lesiones de FCD, se propone una arquitectura con profundidad-3, donde las características extraídas por el bloque codificador en la profundidad-0 son de bajo nivel y, al mismo tiempo, la capa decodificadora reflejada, en la profundidad-0, recibe mapas de características de alto nivel, esta discrepancia es más pronunciada en la profundidad-0 y es allí donde los autores confinan el uso de ResPath, introduciendo no linealidad en los mapas de características para reducir la brecha semántica. El bloque de atención consiste de una señal de activación de atención que se toma de bloque multiresidual más profundo, contiene las características más finas e informa al bloque de atención sobre cuales son las características más relevantes para enfocarse en ellas, resultando en un mapa de características que da atención a las características salientes de la imagen de entrada. El algoritmo fue entrenado usando MRI en secuencia FLAIR de 26 pacientes, la base de datos fue dividida en 80% para entrenamiento (de aquí se usa 15% para validación) y 20% para prueba. Como técnica de validación se usó validación cruzada con $k = 5$, y para reducir el sobreentrenamiento se usó aumento de datos en tiempo real. La función de costo de entrenamiento representó como la suma del costo Sørensen-Dice y de BCE usando el optimizador de Adam con tasa de aprendizaje 0.001. El método propuesto logró una tasa de detección de 92% a nivel de paciente. Asimismo, con coeficiente Sørensen-Dice de 76.62%, el modelo propuesto constituye una mejora en el coeficiente Sørensen-Dice, a nivel de región, de 5.09%, 6.71% y 6.40% respecto del modelo base [62] y de los modelos propuestos en [68, 69]. De otro lado, los autores concluyen que el modelo se queda corto en términos de identificar la verdadera extensión de la lesión de FCD.

Gill et al. [70] emplean una CNN para validar la hipótesis de que un algoritmo de aprendizaje profundo, validado en una base de datos multi-centro, puede detectar de lesiones de FCD en pacientes MRI-negativo. Este estudio incluyó cohortes retrospectivas de nueve centros terciarios de epilepsia alrededor del mundo con lesiones de FCD confirmadas histológicamente con imágenes MRI en secuencias T1 y FLAIR. El centro primario fue conformado por 62 pacientes con FCD y un grupo de control con 42 sujetos sanos y 89 sujetos con TLE (49 pacientes de esclerosis hipocampal visible en MRI y 40 no visible); de los ocho centros restantes, la cohorte comprendió 86 pacientes con FCD. Las MRI en secuencia FLAIR de cada individuo fueron normalizadas en intensidad mediante *z-score* descartando las intensidades menores al percentil 10, esto resultó en una máscara con los *vóxeles* pertenecientes a la interfaz GM-WM. De esta máscara se extrajeron luego parches 3D de las MRI corregistradas en secuencia T1 y FLAIR, los cuales sirvieron de entrada a la CNN. Los autores diseñaron un sistema de clasificación en cascada, con dos CNN, en el cual la primera CNN (CNN-1) sirvió de entrada a la segunda CNN (CNN-2). La CNN-1 tenía como objetivo maximizar la detección de *vóxeles* con lesión, mientras que el objetivo de la CNN-2 era reducir el número de FP manteniendo la sensibilidad. Los algoritmos fueron entrenados con los parches 3D de las MRI en secuencias T1 y FLAIR de los centros radiológicos incluidos en el estudio usando validación cruzada con el método de dejar-un-centro-afuera; adicionalmente, se creó una base de datos de prueba, totalmente independiente del entrenamiento, con 23 pacientes de FCD, 70% MRI negativo, tomados de los centros 1 y 2. El desempeño del modelo fue evaluado tanto a nivel de paciente como a nivel de *vóxel*. A nivel paciente, la sensibilidad en el entrenamiento fue de 93% (detectando 137 de 148 lesiones de FCD) con un promedio de 6 FP por paciente; en la cohorte independiente, la sensibilidad fue de 83% (19/23 lesiones de FCD detectadas) con un promedio de 5 FP por paciente. La especificidad fue de 90% en sujetos sanos y 89% en sujetos con TLE. A nivel de *vóxel*, el área bajo la curva ROC fue de 0.83, lo cual indica que el modelo obtuvo una alta sensibilidad y especificidad.

House et al. [71] desarrollaron una nueva CNN 3D para la detección y segmentación automática de FCD, regularizada por *autoencoder* y validada de manera prospectiva en MRIs rutinarias. El estudio contó con una base de datos de entrenamiento compuesta por 201 MRIs en secuencias T1 y FLAIR y 172 MRIs en secuencia T2; adicionalmente, 100 MRIs de sujetos normales fueron incluidas junto con 50 MRIs de pacientes con diferentes patologías neurológicas. La base de datos de prueba consistió de 100 MRIs consecutivas obtenidas de forma rutinaria en pacientes con FCD. Todas las bases de datos fueron conformadas con MRIs de sujetos tanto adultos como pediátricos. El preprocesamiento de las imágenes consistió de tres etapas: segmentación de tejidos cerebrales (GM, WM y CSF), análisis morfométrico (creación de imágenes de juntura y de extensión) y preparación final para entrada a la CNN (recorte, *padding* y corrección de *bias*). Los autores propusieron una CNN que combina varias técnicas, la primera es una variación de la arquitectura codificador-decodificador, luego, aplica un *autoencoder variacional* como una rama separada del decodificador y como método de regularización; adicionalmente, los autores probaron otras técnicas como transferencia de aprendizaje, conexiones densas, entre otras. sin conseguir mejorar los resultados. El elemento principal de la arquitectura es el bloque residual, este contiene dos convoluciones seguidas de la función de activación ReLU y normalización grupal con el mismo número de grupos que de canales, seguido de un salto de conexión (*skip connection*) al final del bloque. El codificador tiene tres niveles, cada nivel tiene dos bloques residuales (excepto el primero con solo un bloque residual) con el primero incrementando el número de canales en un factor de dos; como capas de submuestreo se usan convoluciones con *stride*; y capas de normalización posicional para capturar la información estructural de la imagen. El decodificador usa solo un bloque residual en cada nivel y después del sobremuestreo con interpolación trilineal, su

salida se añade parcialmente al (*skip connection*) del nivel correspondiente en el codificador; luego, se usa (*moment shortcut*) como método para revertir la normalización posicional, aplicándose, al final, la función de activación sigmoide para obtener un mapa de probabilidades entre 0 y 1. El decodificador *variacional* sirve como mecanismo de regularización del modelo y no tiene influencia en la segmentación, este usa tres convoluciones y dos capas lineales de sobremuestreo. Las 201 MRI en secuencia T1 y FLAIR fueron divididas en subconjuntos de 80% para entrenamiento y para validación 20% y se usó validación cruzada con $k = 5$ y se usó el optimizador Adam iniciando con una tasa de aprendizaje de 0.0001 por un lapso de 80 épocas. Adicionalmente, se implementaron técnicas de aumento de datos (rotación, zoom, etc.). La función de costo para el entrenamiento del algoritmo consistió de una combinación del costo Sørensen-Dice, el error cuadrático medio y la divergencia Kullback-Leibler, y para medir el desempeño del algoritmo, se usó el coeficiente Sørensen-Dice, precisión, sensibilidad y *F-score*. Los mejores resultados de entrenamiento los arrojó el algoritmo que incluyó 100 MRI de pacientes con otras patologías diferentes a FCD, así, se obtuvo una sensibilidad de 70.1%, una precisión de 54.3% en la detección de FCD y un coeficiente Sørensen-Dice de 34.1% en segmentación. Entre tanto, en la evaluación prospectiva se obtuvo una sensibilidad de 77.8% y 5.5% de especificidad.

Una comparación del desempeño de dos tipos de modelo para clasificar y segmentar lesiones de FCD en pacientes pediátricos con epilepsia refractaria es la propuesta de Aminpour et al. [72]. El primer modelo es un tipo especial de CNN llamado “Red totalmente convolucional” (FCN, por sus siglas en inglés) y, el segundo modelo, es una FCN multisequencia. El modelo basado en FCN fue entrenado con *slices* de MRI en secuencias T1, T2 y FLAIR de forma separada, mientras que el modelo multisequencia fue entrenado con la información combinada de *slices* MRI en secuencias T1, T2 y FLAIR simultáneamente. La base de datos consistió de MRI en secuencias T1, T2 y FLAIR de 80 (56 MRI-positivo, 24 MRI-negativo) pacientes pediátricos y 15 sujetos sanos de control. El preprocesamiento de las imágenes incluyó extracción de cerebro, delineación de máscaras manuales y redimensionamiento de la imagen mediante adición o remoción de *vóxeles* de fondo. La arquitectura de la FCN consistió de cuatro bloques de dos capas de convoluciones y *pooling*, luego, las capas totalmente conectadas se reemplazan por tres bloques de sobremuestreo con una capa convolucional y una capa de sobremuestreo. A cada capa le sigue una función de activación ReLU y se utilizó salto de capas, para mejorar la salida de las capas de sobremuestreo, añadiendo el mapa de características submuestreado respectivo de la etapa codificadora. Adicionalmente, se implementaron capas de normalización de *batch* y *dropout* para evitar sobreentrenamiento. Por su parte, la arquitectura de la FCN multisequencia consistió de dos redes trabajadas juntas para extraer características de todas las secuencias, combinarlas y predecir un mapa de segmentación. La red diseñada para extraer características fue entrenada con diferentes secuencias y etiquetas a nivel de sujeto, luego, la salida de la última capa convolucional fue usada como entrada de la segunda red. La segunda red fue una FCN que toma los tres mapas de características de las secuencias T1, T2 y FLAIR, los concatena y sobremuestra la entrada para llegar al tamaño de *slice* original y generar el mapa de segmentación. Al igual que la arquitectura de la primera FCN, esta incluyó cuatro bloques de tres capas convolucionales. En cada bloque, las primera dos convoluciones tenían filtros de 3×3 con *stride* 1 y *padding* 0, la última capa convolucional tenía filtro de tamaño 2×2 con *stride* 2. Los bloques fueron seguidos de una capa convolucional de 1×1 . Luego, siguieron tres capas totalmente conectadas, con *dropout* y normalización de *batch*, para clasificar las características extraídas entre las clases “lesión” y “no-lesión”. Todas las capas usaron función de activación ReLU, excepto la última que usó la función de activación sigmoide. La función de costo empleada fue BCE optimizada mediante SGD. La FCN de sobremuestreo consistió de dos capas convolucionales con 128 filtros de 3×3 , *stride* 1 y *padding* 0, seguidas de tres bloques de sobremuestreo. Los dos primeros

bloques de sobremuestreo fueron una capa de sobremuestreo a 2x y una capa convolucional de 3×3 seguida de normalización de *batch*. El último bloque consistió de una capa de sobremuestreo de 4x seguida de una convolución de 1×1 , la función de activación fue ReLU en todas las capas excepto la última donde se usó la función sigmoide. Para entrenar y evaluar el modelo se utilizó la técnica de *leave-one-out*. Después de obtener las predicciones del modelo, para determinar si el sujeto pertenece a la etiqueta lesión o no-lesión, los autores dividieron el número de píxeles predichos como lesión sobre los píxeles predichos como sanos; esta relación se *umbraliza* con un valor de umbral obtenido mediante la curva ROC para dar la etiqueta al píxel. El mejor desempeño del algoritmo FCN se obtuvo en la secuencia T1 con 91 % de detección (52/56 MRI-positivo y 21/24 MRI-negativo) y coeficiente Sørensen-Dice de 56%, seguido de la secuencia T2. Sin embargo, el modelo multisequencia fue el que mostró el mejor desempeño, detectando 96% de los casos (55/56 MRI-positivo y 22/24 MRI-negativo) y con un coeficiente Sørensen-Dice de 57%. Los autores destinaron una cohorte de 6 pacientes para prueba, en esta cohorte, el algoritmo multisequencia detectó 4/6 casos para una sensibilidad de 67%, con un coeficiente Sørensen-Dice de 58%.

Tabla 1.5: Tabla de resultados obtenidos empleando diferentes arquitecturas de CNN

Estudio	Sensibilidad %		Secuencia	Pacientes	Controles	Grupo etario
	Train	Test				
Gill et al. [61]	87	91	T1, T2-FLAIR	107	101	Pediátricos/ Adultos
Wang et al. [63]	-	90	T1	10	20	-
Dev et al. [62]	82.5	-	FLAIR	43	-	-
Feng et al. [64]	100	92.45	T1	10	95	Pediátricos/ Adultos
Feng et al. [65]	-	83.33	FLAIR	19	-	Pediátricos/ Adultos
Aliev et al. [66]	-	73.3	-	30	17	-
Thomas et al. [67]	-	92	FLAIR	26	-	-
Gill et al. [70]	93	83	T1, FLAIR	171	131	Pediátricos/ Adultos
House et al. [71]	70.1	77.8	T1, FLAIR	201	150	Pediátricos/ Adultos
De Freitas Oliveira Baffa et al. [60]	-	96.47	-	15	-	-
Aminpour et al. [72]	96	67	T1, T2, FLAIR	80	15	Pediátricos

En la Tabla 1.5 se muestran los resultados obtenidos con los diferentes algoritmos de aprendizaje profundo. Algunos estudios logran tasas de detección muy altas por encima del 90%, sin embargo, en algunos casos esto puede indicar sesgos en el aprendizaje, especialmente en los estudios donde los autores disponían de pocos volúmenes para entrenar el algoritmo. El algoritmo de Feng et al. [64] que emplea una base de datos aumentada volúmenes con lesiones de FCD simulados, obtiene un buen resultado con una sensibilidad de 92.45%, lo cual es prometedor para estudios futuros donde se aumenten los datos por medio de simulación de FCD empleando otras técnicas como redes adversarias generativas (GAN, por sus siglas en inglés). Finalmente, si bien no se muestra en la tabla los resultados en coeficiente Sørensen-Dice ya que pocos emplean la métrica, se puede concluir que el

mejor desempeño se logra en el estudio [67], dado que aparte de tener una tasa de detección de 92 %, también publican un resultado de coeficiente Sørensen-Dice de 76.62 % que es el valor más alto en esta métrica de segmentación conseguido por ningún algoritmo a la fecha de revisión de esta literatura.

REFERENCIAS

- [1] Irene Wang and Andreas Alexopoulos. MRI postprocessing in presurgical evaluation. *Current Opinion in Neurology*, 29(2):168–174, 4 2016. doi: 10.1097/wco.0000000000000305. URL <http://dx.doi.org/10.1097/wco.0000000000000305>.
- [2] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, page n71, 3 2021. doi: 10.1136/bmj.n71. URL <http://dx.doi.org/10.1136/bmj.n71>.
- [3] Joanna Kabat and Przemysław Król. Focal cortical dysplasia—review. *Polish journal of radiology*, 77(2):35, 2012.
- [4] Jitender Saini, Atampreet Singh, Chandrasekharan Kesavadas, Bejoy Thomas, Chathurbhuj Rathore, Biji Bahuleyan, Ashalatha Radhakrishnan, and Kurupath Radhakrishnan. Role of three-dimensional fluid-attenuated inversion recovery (3D FLAIR) and proton density magnetic resonance imaging for the detection and evaluation of lesion extent of focal cortical dysplasia in patients with refractory epilepsy. *Acta Radiologica*, 51(2):218–225, 3 2010. doi: 10.3109/02841850903433805. URL <https://doi.org/10.3109/02841850903433805>.
- [5] Jeny Rajan, K. Kannan, C. Kesavadas, and Bejoy Thomas. Focal Cortical Dysplasia (FCD) lesion analysis with complex diffusion approach. *Computerized Medical Imaging and Graphics*, 33(7): 553–558, 10 2009. doi: 10.1016/j.compmedimag.2009.05.004. URL <http://dx.doi.org/10.1016/j.compmedimag.2009.05.004>.
- [6] Samson B Antel, Andrea Bernasconi, Neda Bernasconi, D Louis Collins, Robert E Kearney, Rajjan Shinghal, and Douglas L Arnold. Computational Models of MRI Characteristics of Focal Cortical Dysplasia Improve Lesion Detection. *NeuroImage*, 17(4):1755–1760, 12 2002. doi: 10.1006/nimg.2002.1312. URL <http://dx.doi.org/10.1006/nimg.2002.1312>.
- [7] Andrea Bernasconi, Samson B Antel, Donald L Collins, Neda Bernasconi, Andre Olivier, François Dubeau, Gilbert B Pike, Frederick Andermann, and Douglas L Arnold. Texture analysis and morphological processing of magnetic resonance imaging assist detection of focal cortical dysplasia in extra-temporal partial epilepsy. *Annals of Neurology*, 49(6):770–775, 2001. doi: 10.1002/ana.1013. URL <http://dx.doi.org/10.1002/ana.1013>.
- [8] Pauline Roca, Charles Mellerio, Francine Chassoux, Denis Rivière, Arnaud Cachia, Sylvain Charon, Stéphanie Lion, Jean-François Mangin, Bertrand Devaux, Jean-François Meder, and Catherine Oppenheim. Sulcus-Based MR Analysis of Focal Cortical Dysplasia Located in the Central Region. *PLOS ONE*, 10(3):e0122252, 3 2015. doi: 10.1371/journal.pone.0122252. URL <http://dx.doi.org/10.1371/journal.pone.0122252>.

- [9] M. Magdalena Sepúlveda, Gonzalo M. Rojas, Evelyng Faure, Claudio R. Pardo, Facundo las Heras, Cecilia Okuma, Jorge Cordovez, María de la Iglesia-Vayá, José Molina-Mateo, and Marcelo Gálvez. Visual analysis of automated segmentation in the diagnosis of focal cortical dysplasias with magnetic resonance imaging. *Epilepsy & Behavior*, 102:106684, 2020. doi: 10.1016/j.yebeh.2019.106684.
- [10] Sara Lorio, Sophie Adler, Roxana Gunny, Felice D’Arco, Enrico Kaden, Konrad Wagstyl, Thomas S. Jacques, Chris A. Clark, Judith Helen Cross, Torsten Baldeweg, and David W. Carmichael. MRI profiling of focal cortical dysplasia using multi-compartment diffusion models. *Epilepsia*, 61(3): 433–444, 3 2020. doi: 10.1111/epi.16451. URL <http://dx.doi.org/10.1111/epi.16451>.
- [11] O. Colliot, T. Mansi, N. Bernasconi, V. Naessens, D. Klironomos, and A. Bernasconi. Segmentation of Focal Cortical Dysplasia Lesions Using a Feature-Based Level Set. *Lecture Notes in Computer Science*, pages 375–382, 2005. doi: 10.1007/11566465\47. URL http://dx.doi.org/10.1007/11566465_47.
- [12] Samson B Antel, D Louis Collins, Neda Bernasconi, Frederick Andermann, Rajjan Shinghal, Robert E Kearney, Douglas L Arnold, and Andrea Bernasconi. Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis. *NeuroImage*, 19(4):1748–1759, 8 2003. doi: 10.1016/s1053-8119(03)00226-x. URL [http://dx.doi.org/10.1016/s1053-8119\(03\)00226-x](http://dx.doi.org/10.1016/s1053-8119(03)00226-x).
- [13] O. Colliot, T. Mansi, N. Bernasconi, V. Naessens, D. Klironomos, and A. Bernasconi. Segmentation of focal cortical dysplasia lesions on MRI using level set evolution. *NeuroImage*, 32(4):1621–1630, 10 2006. doi: 10.1016/j.neuroimage.2006.04.225. URL <http://dx.doi.org/10.1016/j.neuroimage.2006.04.225>.
- [14] I. Despotovic, I. Segers, L. Platasa, E. Vansteenkiste, A. Pizurica, K. Deblaere, and W. Philips. Automatic 3D graph cuts for brain cortex segmentation in patients with focal cortical dysplasia. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 8 2011. doi: 10.1109/iembs.2011.6091968. URL <http://dx.doi.org/10.1109/iembs.2011.6091968>.
- [15] Kathryn Snyder, Emily P. Whitehead, William H. Theodore, Kareem A. Zaghloul, Souheil J. Inati, and Sara K. Inati. Distinguishing type II focal cortical dysplasias from normal cortex: A novel normative modeling approach. *NeuroImage: Clinical*, 30:102565, 2021. doi: 10.1016/j.nicl.2021.102565. URL <http://dx.doi.org/10.1016/j.nicl.2021.102565>.
- [16] Eyal Lotan, Omri Tomer, Ido Tavor, Ilan Blatt, Hadassah Goldberg-Stern, Chen Hoffmann, Galia Tsarfaty, David Tanne, and Yaniv Assaf. Widespread cortical dyslamination in epilepsy patients with malformations of cortical development. *Neuroradiology*, 63(2):225–234, 9 2020. doi: 10.1007/s00234-020-02561-2. URL <http://dx.doi.org/10.1007/s00234-020-02561-2>.
- [17] John Ashburner and Karl J. Friston. Voxel-Based Morphometry—The Methods. *NeuroImage*, 11(6):805–821, 6 2000. doi: 10.1006/nimg.2000.0582. URL <https://doi.org/10.1006/nimg.2000.0582>.
- [18] Jan Kassubek, Hans-Jürgen Huppertz, Joachim Spreer, and Andreas Schulze-Bonhage. Detection and Localization of Focal Cortical Dysplasia by Voxel-based 3-D MRI Analysis. *Epilepsia*, 43(6): 596–602, 6 2002. doi: 10.1046/j.1528-1157.2002.41401.x. URL <http://dx.doi.org/10.1046/j.1528-1157.2002.41401.x>.

- [19] Hans-Jürgen Huppertz, Christina Grimm, Susanne Fauser, Jan Kassubek, Irina Mader, Albrecht Hochmuth, Joachim Spreer, and Andreas Schulze-Bonhage. Enhanced visualization of blurred gray–white matter junctions in focal cortical dysplasia by voxel-based 3D MRI analysis. *Epilepsy Research*, 67(1-2):35–50, 10 2005. doi: 10.1016/j.eplesyres.2005.07.009. URL <http://dx.doi.org/10.1016/j.eplesyres.2005.07.009>.
- [20] Xiaoxia Qu, Ljiljana Platisa, Ivana Despotovic, Asli Kumcu, Tingzhu Bai, Karel Deblaere, and Wilfried Philips. Estimating blur at the brain gray-white matter boundary for FCD detection in MRI. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 8 2014. doi: 10.1109/embc.2014.6944333. URL <http://dx.doi.org/10.1109/embc.2014.6944333>.
- [21] Patrick M. House, Michael Lanz, Brigitte Holst, Tobias Martens, Stefan Stodieck, and Hans-Jürgen Huppertz. Comparison of morphometric analysis based on T1- and T2-weighted MRI data for visualization of focal cortical dysplasia. *Epilepsy Research*, 106(3):403–409, 10 2013. doi: 10.1016/j.eplesyres.2013.06.016. URL <http://dx.doi.org/10.1016/j.eplesyres.2013.06.016>.
- [22] Xiaoxia Qu, Jian Yang, Danni Ai, Hong Song, Luosha Zhang, Yongtian Wang, Tingzhu Bai, and Wilfried Philips. Local Directional Probability Optimization for Quantification of Blurred Gray/White Matter Junction in Magnetic Resonance Image. *Frontiers in Computational Neuroscience*, 11, 9 2017. doi: 10.3389/fncom.2017.00083. URL <http://dx.doi.org/10.3389/fncom.2017.00083>.
- [23] Daniel Schmitter, Alexis Roche, Bénédicte Maréchal, Delphine Ribes, Ahmed Abdulkadir, Meritxell Bach-Cuadra, Alessandro Daducci, Cristina Granziera, Stefan Klöppel, Philippe Maeder, Reto Meuli, and Gunnar Krueger. An evaluation of volume-based morphometry for prediction of mild cognitive impairment and Alzheimer’s disease. *NeuroImage: Clinical*, 7:7–17, 2015. doi: 10.1016/j.nicl.2014.11.001. URL <http://dx.doi.org/10.1016/j.nicl.2014.11.001>.
- [24] Xin Chen, Tianyi Qian, Bénédicte Maréchal, Guojun Zhang, Tao Yu, Zhiwei Ren, Duanyu Ni, Chang Liu, Yongjuan Fu, Nan Chen, and Kuncheng Li. Quantitative volume-based morphometry in focal cortical dysplasia: A pilot study for lesion localization at the individual level. *European Journal of Radiology*, 105:240–245, 8 2018. doi: 10.1016/j.ejrad.2018.06.019. URL <http://dx.doi.org/10.1016/j.ejrad.2018.06.019>.
- [25] Xiaoxia Qu, Karel Deblaere, Wilfried Philips, Jian Yang, Ljiljana Platis, Asli Kumcu, Danni Ai, Bart Goossens, Tingzhu Bai, Yongtian Wang, and Jing Sui. Multiple Classifier Fusion and Optimization for Automatic Focal Cortical Dysplasia Detection on Magnetic Resonance Images. *IEEE Access*, 6:73786–73801, 2018. doi: 10.1109/access.2018.2883583. URL <http://dx.doi.org/10.1109/access.2018.2883583>.
- [26] Cuixia Feng, Hulin Zhao, Maoyu Tian, Miaomiao Lu, and Junhai Wen. Detecting focal cortical dysplasia lesions from FLAIR-negative images based on cortical thickness. *BioMedical Engineering OnLine*, 19(1), 2 2020. doi: 10.1186/s12938-020-0757-8. URL <http://dx.doi.org/10.1186/s12938-020-0757-8>.
- [27] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210, 1994. doi: 10.1002/hbm.460020402. URL <http://dx.doi.org/10.1002/hbm.460020402>.

- [28] Tayyabah Yousaf, George Dervenoulas, and Marios Politis. Advances in MRI Methodology. *International Review of Neurobiology*, pages 31–76, 2018. doi: 10.1016/bs.irn.2018.08.008. URL <http://dx.doi.org/10.1016/bs.irn.2018.08.008>.
- [29] Siddharth Srivastava, Frederik Maes, Dirk Vandermeulen, Wim Van Paesschen, Patrick Dupont, and Paul Suetens. Feature-based statistical analysis of structural MR data for automatic detection of focal cortical dysplastic lesions. *NeuroImage*, 27(2):253–266, 8 2005. doi: 10.1016/j.neuroimage.2005.03.045. URL <http://dx.doi.org/10.1016/j.neuroimage.2005.03.045>.
- [30] Niels K. Focke, Mark R. Symms, Jane L. Burdett, and John S. Duncan. Voxel-based analysis of whole brain FLAIR at 3T detects focal cortical dysplasia. *Epilepsia*, 49(5):786–793, 5 2008. doi: 10.1111/j.1528-1167.2007.01474.x. URL <http://dx.doi.org/10.1111/j.1528-1167.2007.01474.x>.
- [31] Niels K. Focke, Silvia B. Bonelli, Mahinda Yogarajah, Catherine Scott, Mark R. Symms, and John S. Duncan. Automated normalized FLAIR imaging in MRI-negative patients with refractory focal epilepsy. *Epilepsia*, 50(6):1484–1490, 6 2009. doi: 10.1111/j.1528-1167.2009.02022.x. URL <http://dx.doi.org/10.1111/j.1528-1167.2009.02022.x>.
- [32] Lily C. Wong-Kisiel, Diego F. Tovar Quiroga, Daniel L. Kenney-Jung, Robert J. Witte, Alexandra Santana-Almansa, Gregory A. Worrell, Jeffrey Britton, and Benjamin H. Brinkmann. Morphometric analysis on T1-weighted MRI complements visual MRI review in focal cortical dysplasia. *Epilepsy Research*, 140:184–191, 2 2018. doi: 10.1016/j.eplesyres.2018.01.018. URL <http://dx.doi.org/10.1016/j.eplesyres.2018.01.018>.
- [33] Yicong Lin, Yu-Hua Dean Fang, Guiyun Wu, Stephen E. Jones, Richard A. Prayson, Ahsan N. V. Moosa, Margit Overmyer, James Bena, Mykol Larvie, William Bingaman, Jorge A. Gonzalez-Martinez, Imad M. Najm, Andreas V. Alexopoulos, and Z. Irene Wang. Quantitative positron emission tomography-guided magnetic resonance imaging postprocessing in magnetic resonance imaging-negative epilepsies. *Epilepsia*, 59(8):1583–1594, 6 2018. doi: 10.1111/epi.14474. URL <http://dx.doi.org/10.1111/epi.14474>.
- [34] W. Wang, Y. Lin, S. Wang, S. Jones, R. Prayson, A. N. V. Moosa, A. McBride, J. Gonzalez-Martinez, W. Bingaman, I. Najm, A. Alexopoulos, and Z. I. Wang. Voxel-based morphometric magnetic resonance imaging postprocessing in non-lesional pediatric epilepsy patients using pediatric normal databases. *European Journal of Neurology*, 3 2019. doi: 10.1111/ene.13916. URL <http://dx.doi.org/10.1111/ene.13916>.
- [35] Chin-Ann Yang, Mostafa Kaveh, and Bradley J Erickson. Automated detection of Focal Cortical Dysplasia lesions on T1-weighted MRI using volume-based distributional features. *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 3 2011. doi: 10.1109/isbi.2011.5872541. URL <http://dx.doi.org/10.1109/isbi.2011.5872541>.
- [36] Chin-Ann Yang, Mostafa Kaveh, and Bradley Erickson. Cluster-based differential features to improve detection accuracy of focal cortical dysplasia. *SPIE Proceedings*, 2 2012. doi: 10.1117/12.905313. URL <http://dx.doi.org/10.1117/12.905313>.
- [37] Maddalena Strumia, Georgia Ramantani, Irina Mader, Jürgen Henning, Li Bai, and Stathis Hadjidemetriou. Analysis of Structural MRI Data for the Localisation of Focal Cortical Dysplasia in Epilepsy. *Clinical Image-Based Procedures. From Planning to Intervention*, pages 25–32, 2013. doi: 10.1007/978-3-642-38079-2_4. URL http://dx.doi.org/10.1007/978-3-642-38079-2_4.

- [38] Sancgeetha Kulaseharan, Azad Aminpour, Mehran Ebrahimi, and Elysa Widjaja. Identifying lesions in paediatric epilepsy using morphometric and textural analysis of magnetic resonance images. *NeuroImage: Clinical*, 21:101663, 2019. doi: 10.1016/j.nicl.2019.101663. URL <http://dx.doi.org/10.1016/j.nicl.2019.101663>.
- [39] Cuixia Feng, Hulin Zhao, Yueer Li, Zhibiao Cheng, and Junhai Wen. Improved detection of focal cortical dysplasia in normal-appearing FLAIR images using a Bayesian classifier. *Medical Physics*, 48(2):912–925, 12 2020. doi: 10.1002/mp.14646. URL <http://dx.doi.org/10.1002/mp.14646>.
- [40] Seok-Jun Hong, Boris C Bernhardt, Dewi S Schrader, Neda Bernasconi, and Andrea Bernasconi. Whole-brain MRI phenotyping in dysplasia-related frontal lobe epilepsy. *Neurology*, 86(7):643–650, 1 2016. doi: 10.1212/wnl.0000000000002374. URL <http://dx.doi.org/10.1212/wnl.0000000000002374>.
- [41] Meriem El Azami, Alexander Hammers, Julien Jung, Nicolas Costes, Romain Bouet, and Carole Lartzien. Detection of Lesions Underlying Intractable Epilepsy on T1-Weighted MRI as an Outlier Detection Problem. *PLOS ONE*, 11(9):e0161498, 9 2016. doi: 10.1371/journal.pone.0161498. URL <http://dx.doi.org/10.1371/journal.pone.0161498>.
- [42] Yee-Leng Tan, Hosung Kim, Seunghyun Lee, Tarik Tihan, Lawrence Ver Hoef, Susanne G Mueller, Anthony James Barkovich, Duan Xu, and Robert Knowlton. Quantitative surface analysis of combined MRI and PET enhances detection of focal cortical dysplasias. *NeuroImage*, 166:10–18, 2 2018. doi: 10.1016/j.neuroimage.2017.10.065. URL <http://dx.doi.org/10.1016/j.neuroimage.2017.10.065>.
- [43] Yanming Wang, Yawen Zhou, Huijuan Wang, Jin Cui, Benedictor Alexander Nguchu, Xufei Zhang, Bensheng Qiu, Xiaoxiao Wang, and Mingwang Zhu. Voxel-based automated detection of focal cortical dysplasia lesions using diffusion tensor imaging and T2-weighted MRI data. *Epilepsy Behavior*, 84:127–134, 7 2018. doi: 10.1016/j.yebeh.2018.04.005. URL <http://dx.doi.org/10.1016/j.yebeh.2018.04.005>.
- [44] Ravnoor S Gill, Seok-Jun Hong, Fatemeh Fadaie, Benoit Caldairou, Boris Bernhardt, Neda Bernasconi, and Andrea Bernasconi. Automated Detection of Epileptogenic Cortical Malformations Using Multimodal MRI. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 349–356, 2017. doi: 10.1007/978-3-319-67558-9_40. URL http://dx.doi.org/10.1007/978-3-319-67558-9_40.
- [45] Yaoyun Lin, Jiajie Mo, Huiwen Jin, Xueliang Cao, Yang Zhao, Changjun Wu, Kai Zhang, Wenhan Hu, and Zhiguo Lin. Automatic analysis of integrated magnetic resonance and positron emission tomography images improves the accuracy of detection of focal cortical dysplasia type IIb lesions. *European Journal of Neuroscience*, 53(9):3231–3241, 3 2021. doi: 10.1111/ejn.15185. URL <http://dx.doi.org/10.1111/ejn.15185>.
- [46] Seok-Jun Hong, Hosung Kim, Dewi Schrader, Neda Bernasconi, Boris C Bernhardt, and Andrea Bernasconi. Automated detection of cortical dysplasia type II in MRI-negative epilepsy. *Neurology*, 83(1):48–55, 6 2014. doi: 10.1212/wnl.0000000000000543. URL <http://dx.doi.org/10.1212/wnl.0000000000000543>.
- [47] Bilal Ahmed, Carla E Brodley, Karen E Blackmon, Ruben Kuzniecky, Gilad Barash, Chad Carlson, Brian T Quinn, Werner Doyle, Jacqueline French, Orrin Devinsky, et al. Cortical feature analysis

- and machine learning improves detection of “MRI-negative” focal cortical dysplasia. *Epilepsy Behavior*, 48:21–28, 7 2015. doi: 10.1016/j.yebeh.2015.04.055. URL <http://dx.doi.org/10.1016/j.yebeh.2015.04.055>.
- [48] Xiaoxia Qu, Jian Yang, Shaodong Ma, Yitian Zhao, and Tingzhu Bai. An unanimous voting of the multiple classifiers method for detecting focal cortical dysplasia on brain magnetic resonance image. *2015 IET International Conference on Biomedical Image and Signal Processing (ICBISP 2015)*, 2015. doi: 10.1049/cp.2015.0767. URL <http://dx.doi.org/10.1049/cp.2015.0767>.
- [49] Xiaoxia Qu, Jian Yang, Shaodong Ma, Tingzhu Bai, and Wilfried Philips. Positive Unanimous Voting Algorithm for Focal Cortical Dysplasia Detection on Magnetic Resonance Image. *Frontiers in Computational Neuroscience*, 10, 3 2016. doi: 10.3389/fncom.2016.00025. URL <http://dx.doi.org/10.3389/fncom.2016.00025>.
- [50] Hyo M Lee, Ravnoor S Gill, Fatemeh Fadaie, Kyoo H Cho, Marie C Guiot, Seok-Jun Hong, Neda Bernasconi, and Andrea Bernasconi. Unsupervised machine learning reveals lesional variability in focal cortical dysplasia at mesoscopic scale. *NeuroImage: Clinical*, 28:102438, 2020. doi: 10.1016/j.nicl.2020.102438. URL <http://dx.doi.org/10.1016/j.nicl.2020.102438>.
- [51] Bo Jin, Balu Krishnan, Sophie Adler, Konrad Wagstyl, Wenhan Hu, Stephen Jones, Imad Najm, Andreas Alexopoulos, Kai Zhang, Jianguo Zhang, Meiping Ding, Shuang Wang, and Zhong Irene Wang. Automated detection of focal cortical dysplasia type II with surface-based magnetic resonance imaging postprocessing and machine learning. *Epilepsia*, 59(5):982–992, 4 2018. doi: 10.1111/epi.14064. URL <http://dx.doi.org/10.1111/epi.14064>.
- [52] Pierre Besson, Frederick Andermann, Francois Dubeau, and Andrea Bernasconi. Small focal cortical dysplasia lesions are located at the bottom of a deep sulcus. *Brain*, 131(12):3246–3255, 9 2008. doi: 10.1093/brain/awn224. URL <http://dx.doi.org/10.1093/brain/awn224>.
- [53] Pierre Besson, Olivier Colliot, Alan Evans, and Andrea Bernasconi. Automatic detection of subtle focal cortical dysplasia using surface-based features on MRI. *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 5 2008. doi: 10.1109/isbi.2008.4541326. URL <http://dx.doi.org/10.1109/isbi.2008.4541326>.
- [54] Pierre Besson, Neda Bernasconi, Olivier Colliot, Alan Evans, and Andrea Bernasconi. Surface-Based Texture and Morphological Analysis Detects Subtle Cortical Dysplasia. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, pages 645–652, 2008. doi: 10.1007/978-3-540-85988-8_77. URL http://dx.doi.org/10.1007/978-3-540-85988-8_77.
- [55] Sophie Adler, Konrad Wagstyl, Roxana Gunny, Lisa Ronan, David Carmichael, J Helen Cross, Paul C. Fletcher, and Torsten Baldeweg. Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy. *NeuroImage: Clinical*, 14:18–27, 2017. doi: 10.1016/j.nicl.2016.12.030. URL <http://dx.doi.org/10.1016/j.nicl.2016.12.030>.
- [56] Jia-Jie Mo, Jian-Guo Zhang, Wen-Ling Li, Chao Chen, Na-Jing Zhou, Wen-Han Hu, Chao Zhang, Yao Wang, Xiu Wang, Chang Liu, Bao-Tian Zhao, Jun-Jian Zhou, and Kai Zhang. Clinical Value of Machine Learning in the Automated Detection of Focal Cortical Dysplasia Using Quantitative Multimodal Surface-Based Features. *Frontiers in Neuroscience*, 12, 1 2019. doi: 10.3389/fnins.2018.01008. URL <http://dx.doi.org/10.3389/fnins.2018.01008>.

- [57] Konrad Wagstyl, Sophie Adler, Birgit Pimpel, Aswin Chari, Kiran Seunarine, Sara Lorio, Rachel Thornton, Torsten Baldeweg, and Martin Tisdall. Planning stereoelectroencephalography using automated lesion detection: Retrospective feasibility study. *Epilepsia*, 61(7):1406–1416, 6 2020. doi: 10.1111/epi.16574. URL <http://dx.doi.org/10.1111/epi.16574>.
- [58] Zohreh Ganji, Mohsen Aghaee Hakak, Seyed Amir Zamanpour, and Hoda Zare. Automatic Detection of Focal Cortical Dysplasia Type II in MRI: Is the Application of Surface-Based Morphometry and Machine Learning Promising? *Frontiers in Human Neuroscience*, 15, 2 2021. doi: 10.3389/fnhum.2021.608285. URL <http://dx.doi.org/10.3389/fnhum.2021.608285>.
- [59] Bastian David, Judith Kröll-Seger, Fabiane Schuch, Jan Wagner, Jörg Wellmer, Friedrich Woermann, Bernhard Oehl, Wim Van Paesschen, Tobias Breyer, Albert Becker, Hartmut Vatter, Elke Hattingen, Horst Urbach, Bernd Weber, Rainer Surges, Christian Erich Elger, Hans-Jürgen Huppertz, and Theodor Rüber. External validation of automated focal cortical dysplasia detection using morphometric analysis. *Epilepsia*, 62(4):1005–1021, 2 2021. doi: 10.1111/epi.16853. URL <http://dx.doi.org/10.1111/epi.16853>.
- [60] Matheus De Freitas Oliveira Baffa, João Pereira, Fabrício Henrique Simozo, Luiz Otavio Murta, Junior, and Joaquim Cezar Felipe. Focal cortical dysplasia classification for refractory epilepsy detection using artificial neural network. *Computer methods in biomechanics and biomedical engineering. Imaging visualization*, pages 1–5, 2 2022. doi: 10.1080/21681163.2022.2043780.
- [61] Ravnoor S. Gill, Seok-Jun Hong, Fatemeh Fadaie, Benoit Caldaïrou, Boris C. Bernhardt, Carmen Barba, Armin Brandt, Vanessa C. Coelho, Ludovico d’Incerti, Matteo Lenge, Mira Semmelroch, Fabrice Bartolomei, Fernando Cendes, Francesco Deleo, Renzo Guerrini, Maxime Guye, Graeme Jackson, Andreas Schulze-Bonhage, Tommaso Mansi, Neda Bernasconi, and Andrea Bernasconi. Deep Convolutional Networks for Automated Detection of Epileptogenic Brain Malformations. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 490–497, 2018. doi: 10.1007/978-3-030-00931-1_56. URL http://dx.doi.org/10.1007/978-3-030-00931-1_56.
- [62] KM Bijay Dev, Pawan S Jogi, S Niyas, S Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan. Automatic detection and localization of focal cortical dysplasia lesions in mri using fully convolutional neural network. *Biomedical Signal Processing and Control*, 52:218–225, 2019. doi: 10.1016/j.bspc.2019.04.024. URL <https://www.sciencedirect.com/science/article/pii/S1746809419301211>.
- [63] Huiquan Wang, S Nizam Ahmed, and Mrinal Mandal. Automated detection of focal cortical dysplasia using a deep convolutional neural network. *Computerized Medical Imaging and Graphics*, 79:101662, 2020. doi: 10.1016/j.compmedimag.2019.101662. URL <https://doi.org/10.1016/j.compmedimag.2019.101662>.
- [64] Cuixia Feng, Hulin Zhao, Jun Zhang, Zhibiao Cheng, and Junhai Wen. Automated localization of Epileptic Focus Using Convolutional Neural Network. *International Conference Big Data Engineering and Technology*, 1 2020. doi: 10.1145/3378904.3378928. URL <https://doi.org/10.1145/3378904.3378928>.
- [65] Cuixia Feng, Hulin Zhao, Yueer Li, and Junhai Wen. Automatic localization and segmentation of focal cortical dysplasia in FLAIR-negative patients using a convolutional neural network. *Journal*

- of *Applied Clinical Medical Physics*, 21(9):215–226, 8 2020. doi: 10.1002/acm2.12985. URL <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/acm2.12985>.
- [66] Ruslan Aliev, Ekaterina Kondrateva, Maxim Sharaev, Oleg Bronov, Alexey Marinets, Sergey Subbotin, Alexander Bernstein, and Evgeny Burnaev. Convolutional Neural Networks for Automatic Detection of Focal Cortical Dysplasia. *Advances in intelligent systems and computing*, pages 582–588, 10 2020. doi: 10.1007/978-3-030-71637-0_67.
- [67] Edwin Thomas, S. J. Pawan, Shushant Kumar, Anmol Horo, S. Niyas, S. Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan. Multi-Res-Attention UNet: A CNN Model for the Segmentation of Focal Cortical Dysplasia Lesions from Magnetic Resonance Images. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1724–1734, 5 2021. doi: 10.1109/jbhi.2020.3024188.
- [68] Nabil Ibtehaz and Md. Mustafizur Rahman. MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121:74–87, 1 2020. doi: 10.1016/j.neunet.2019.08.025.
- [69] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias P. Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*, 53:197–207, 2 2019. doi: 10.1016/j.media.2019.01.012. URL <https://doi.org/10.1016/j.media.2019.01.012>.
- [70] Ravnoor S. Gill, Hyo Suk Lee, Benoit Caldairou, Hyo Geun Choi, Carmen Barba, Francesco Deleo, Ludovico D’Incerti, Vanessa Félix Nascimento Coelho, Matteo Lenge, Mira Semmelroch, Dewi V. Schrader, Fabrice Bartolomei, Maxime Guye, Andreas Schulze-Bonhage, Horst Urbach, Kyoo Ho Cho, Fernando Cendes, Renzo Guerrini, Graeme D. Jackson, R. Edward Hogan, and Andrea Bernasconi. Multicenter Validation of a Deep Learning Detection Algorithm for Focal Cortical Dysplasia. *Neurology*, 97(16):e1571–e1582, 9 2021. doi: 10.1212/wnl.0000000000012698.
- [71] Patrick M. House, Mikhail Kopelyan, Natalia Braniewska, Bartosz Silski, Agata Chudzinska, Brigitte Holst, Thomas Sauvigny, Tobias Martens, Stefan Stodieck, and Sirko Pelzl. Automated detection and segmentation of focal cortical dysplasias (FCDs) with artificial intelligence: Presentation of a novel convolutional neural network and its prospective clinical validation. *Epilepsy Research*, 172: 106594, 2 2021. doi: 10.1016/j.epilepsyres.2021.106594.
- [72] Azad Aminpour, Mehran Ebrahimi, and Elysa Widjaja. Lesion segmentation in paediatric epilepsy utilizing deep learning approaches. 2022. doi: 10.54364/AAIML.2021.1128.

2

MARCO TEÓRICO

Índice

2.1 Imágenes de Resonancia Magnética	46
2.1.1 Tiempos de relajación en la adquisición de MRI	47
2.1.2 Secuencia T1	47
2.1.3 Secuencia T2-FLAIR	48
2.2 Preprocesamiento	48
2.2.1 Herramienta de Extracción de Cráneo	48
2.3 Algoritmos de Aprendizaje y Optimización	48
2.3.1 Red neuronal convolucional	48
2.3.2 U-Net	49
2.3.3 Transferencia de aprendizaje.	50
2.3.4 Estimación Adaptativa de los Momentos (Optimizador Adam).	50
2.3.5 Entropía Cruzada Binaria	51
2.4 Métodos de evaluación	52
2.4.1 Métricas basadas en matriz de confusión	52
2.4.2 Coeficiente de Similitud Sørensen-Dice	53
Referencias	55

EN este capítulo se exponen los conceptos teóricos utilizados para abordar este trabajo, correspondientes a las etapas de adquisición de imágenes de resonancia magnética y los tipos de secuencias existentes, la etapa de preprocesamiento de las bases de datos donde, entre otros, se reorientan y registran todos los volúmenes MRI al espacio radiológico común “Right-Anterior-Superior” (RAS, por sus siglas en inglés) con las mismas dimensiones y tamaño de vóxel, y se realiza un proceso de extracción de cráneo. Posteriormente, se abordan las secciones teóricas de entrenamiento de los algoritmos de aprendizaje profundo y transferencia de aprendizaje, su función de costo, algoritmo de optimización y métodos de evaluación.

2.1. IMÁGENES DE RESONANCIA MAGNÉTICA.

Las imágenes de resonancia magnética son producto del comportamiento de los núcleos atómicos en presencia de un campo magnético, en particular, de la orientación y el movimiento del momento nuclear magnético dentro del campo. Para tomar una MRI, los momentos nucleares magnéticos de los átomos que componen los tejidos, originalmente alineados con un campo magnético estático, son puestos a girar en un plano perpendicular al plano del campo estático aplicando una señal que contiene una frecuencia de resonancia (básicamente una señal de radio frecuencia, RF por sus siglas en inglés) lo cual provoca que el vector de magnetización sea desplazado de su posición original en equilibrio. Al retirarse la señal de RF, la magnetización retorna a su valor original después de cierto tiempo. A este último fenómeno se le conoce como tiempo de relajación y puede ser manipulado para obtener diferentes características en la imagen.

Las imágenes de MRI permiten que se empleen diferentes mecanismos para lograr cambios en el contraste y así lograr que resalten otras características de los tejidos en la imagen. Adicionalmente, tienen la ventaja de que para la adquisición de la imagen no es necesario someter al paciente a niveles altos de radiación energética como en el caso de los Rayos-X y PET, sino que la imagen es adquirida mediante la interacción entre un campo magnético de alta intensidad y una señal de RF lo cual, en principio, no es potencialmente peligroso para el paciente.

Un volumen MRI en formato NIfTI (Neuroimaging Informatics Technology Initiative, por sus siglas en inglés) es la asociación de tres cosas:

- Un arreglo con los datos de la imagen que puede estar dado en 3D o 4D de acuerdo con las características del escáner o de la secuencia en que se toma la imagen.
- Una matriz de afinidad que indica la posición de los datos del arreglo que contiene la imagen en un espacio de referencia dado.
- Metadatos que describen las características de la imagen.

Las coordenadas del arreglo de la imagen vienen dadas en *vóxeles*, sin embargo, estos no dan información acerca de la posición real, en este caso del cerebro del paciente, dentro del escáner. Puede crearse un espacio de referencia tridimensional definido por un conjunto ordenado conformado por tres ejes independientes, para lo que se necesita escoger un origen para los ejes, su dirección y sus unidades. El origen de los ejes está ubicado en el isocentro del electroimán del escáner, este tiene coordenadas $(0, 0, 0)$ y todos los tres ejes pasan por el isocentro. Las unidades vienen dadas en milímetros (mm). La dirección depende de la posición del observador y puede ser izquierda-derecha, anterior-posterior o inferior-superior y viceversa: derecha-izquierda, posterior-anterior, superior-inferior.

Para pasar de un espacio de referencia a otro se necesita una función de transformación $f(\cdot)$ que lleve de las coordenadas (i, j, k) de un espacio de referencia a las coordenadas (x, y, z) del otro espacio de referencia como se muestra en la ecuación (2.1).

$$(x, y, z) = f(i, j, k) \tag{2.1}$$

Dado que los escáner toman los datos en un rejilla ordenada, la relación entre uno y otro espacio de referencia es lineal y puede ser codificado por medio transformaciones lineales que incluyen traslaciones, rotaciones y *zoom* para acercar o alejar la imagen. Se puede escribir f como una combinación

de multiplicación matricial por una matriz M de rotación y *zoom* seguido por la suma de un vector de traslación (a, b, c) , como se muestra en la ecuación (2.2).

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = M \begin{bmatrix} i \\ j \\ k \end{bmatrix} + \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad (2.2)$$

donde M puede ser una matriz de rotación de los *vóxeles* en el eje x , de la forma:

$$M = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.3)$$

Se tiene entonces que si $m_{i,j}$ es el valor de la fila i columna j de la matriz M , la matriz de afinidad A viene dada por:

$$A = \begin{bmatrix} m_{1,1} & m_{1,2} & m_{1,3} & a \\ m_{2,1} & m_{2,2} & m_{2,3} & b \\ m_{3,1} & m_{3,2} & m_{3,3} & c \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.4)$$

Añadir la fila extra $[0, 0, 0, 1]$ al final de A , permite integrar en una misma matriz las partes de traslación y rotación y *zoom*. Finalmente, la ecuación para pasar de un espacio de referencia a otro queda expresado en la ecuación (2.5):

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} m_{1,1} & m_{1,2} & m_{1,3} & a \\ m_{2,1} & m_{2,2} & m_{2,3} & b \\ m_{3,1} & m_{3,2} & m_{3,3} & c \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i \\ j \\ k \\ 1 \end{bmatrix} = A \begin{bmatrix} i \\ j \\ k \\ 1 \end{bmatrix} \quad (2.5)$$

Este tratamiento matemático fue adaptado de la [documentación de nibabel](#).

2.1.1. TIEMPOS DE RELAJACIÓN EN LA ADQUISICIÓN DE MRI

Cuánticamente, la excitación por medio de una señal de RF resonante hace que el *spin* de los núcleos pase de un estado de baja energía a otro de más alta energía. Una vez retirado el pulso de excitación, los *spin* vuelven al equilibrio en su estado de baja energía inicial después de transcurrido un instante de tiempo, a este instante de tiempo se le conoce como tiempo de relajación.

Físicamente se dan dos fenómenos en el proceso de relajación a diferentes escalas de tiempo. En el primero, la componente longitudinal de la magnetización se relaja exponencialmente a una componente de tiempo constante T_1 debida a la relajación *spin-lattice*. De otro lado, la componente transversal se relaja a una escala diferente de tiempo T_2 que debe su origen a otro fenómeno conocido como relajación *spin-spin*.

2.1.2. SECUENCIA T1

La secuencia T1, también conocida como secuencia *spin-lattice*, es una de las secuencias pulsadas básicas en MRI y muestra las diferencias en los tiempos de relajación de los diferentes tejidos expuestos a la radiación. La secuencia T1 se apoya en la relajación longitudinal del vector de magnetización neto de un tejido. Los espines alineados en un campo externo B se ubican en el plano transversal por

medio de pulso de RF. Luego, se desplazan de vuelta hacia el estado de equilibrio alineado con B . Dado que no todos los tejidos regresan al estado de equilibrio al mismo tiempo, la secuencia T1 refleja la cantidad de tiempo que tardan los espines de los protones del tejido en volver al estado de relajación. Por ejemplo, los tejidos grasos reordenan de forma rápida su magnetización longitudinal con B , por lo tanto aparecen como zonas brillantes en una MRI en secuencia T1. Por el contrario, el agua tarda más en reordenarse después del pulso RF, por lo tanto su señal es más tenue y aparece oscura en la imagen [1].

2.1.3. SECUENCIA T2-FLAIR

Otra de las secuencias pulsadas básicas en MRI es la secuencia T2, la cual se basa en la relajación transversal, también conocida como relajación *spin-spin* del vector de magnetización neta. Después de excitar el tejido con un pulso RF, se relajan los espines desde el plano transversal hacia el vector de magnetización longitudinal de B , esto es la secuencia T1. Al mismo tiempo, los espines decaen de su precesión alineada en el plano transversal, las diferencias en este decaimiento forman la imagen en secuencia T2 [2].

2.2. PREPROCESAMIENTO

La etapa de preprocesamiento incluyó todos aquellos procedimientos aplicados a las MRI de la base de datos con el fin de acondicionarlas para la posterior etapa de entrenamiento de los algoritmos de aprendizaje, garantizando que sean entrenados con datos de entrada óptimos. A continuación, se describen los métodos de preprocesamiento implementados.

2.2.1. HERRAMIENTA DE EXTRACCIÓN DE CRÁNEO

La herramienta de software BET elimina automáticamente el tejido no cerebral de una imagen entera de la cabeza y está incluida dentro de la librería FSL para análisis de MRI. También permite estimar las superficies interior y exterior del cráneo, así como el cuero cabelludo si se dispone de imágenes T1 y T2 de buena calidad. El método se basa en un modelo deformable que evoluciona para ajustarse a la superficie cerebral aplicando un conjunto de modelos adaptables localmente. Adicionalmente, la implementación de BET no requiere que la imagen sea prerregistrada ni de otros pasos de preprocesamiento [3].

2.3. ALGORITMOS DE APRENDIZAJE Y OPTIMIZACIÓN

2.3.1. RED NEURONAL CONVOLUCIONAL

Una red neuronal es un algoritmo de aprendizaje de máquina que además constituye el núcleo de los algoritmos de aprendizaje profundo. Estas se componen de nodos organizados en una capa de entrada, una o más capas ocultas y una capa de salida. Cada nodo de cada capa está conectado con la siguiente y tiene unos pesos y umbral asociados. Si la salida de un nodo de una capa está por encima del valor de umbral, este nodo se activa, enviando la información a los nodos conectados en la siguiente capa de la red neuronal.

Las redes neuronales convolucionales se utilizan frecuentemente en visión por computadora para desempeñar tareas de clasificación y segmentación. Sin embargo, estas requieren grandes volúmenes de datos para su entrenamiento y son computacionalmente costosas, requiriendo el uso de unidades de procesamiento de gráficos (GPU, por sus siglas en inglés) para reducir los tiempos de cómputo. Sin embargo, las redes neuronales convolucionales se destacan por su desempeño superior en tareas de visión por computadora y reconocimiento de lenguaje en señales de audio.

Las redes neuronales convolucionales se componen principalmente por tres tipos de capas:

- Capa convolucional,
- Capa de reducción de muestreo (*Pooling*),
- Capa totalmente conectada

2.3.1.1. CAPA CONVOLUCIONAL

La capa convolucional es la pieza fundamental de una CNN, en esta es donde ocurren la mayor parte de los cálculos. Requieren de unos datos de entrada y filtros que generan mapas de características. De acuerdo con el problema a resolver, los datos de entrada pueden ser una imagen o un archivo de audio, entre otros. Cada filtro o kernel, actúa como un detector de características, moviéndose a través de la imagen en busca de presencia de alguna característica en un proceso que se conoce como convolución. Después de cada convolución se suele aplicar una capa de activación ReLU para añadir no-linealidad al sistema. Las dimensiones de salida de cada convolución vienen dadas por la ecuación (2.6)

$$Dim_{out} = \frac{Dim_{in} + (2 \times padding) - (dilation \times (kernelsize - 1)) - 1}{stride} \quad (2.6)$$

2.3.1.2. CAPA DE REDUCCIÓN DE MUESTREO

Las capas de reducción de muestreo (*pooling*), reducen las dimensiones de la imagen, reduciendo a su vez el número de parámetros necesarios para el aprendizaje. Como en la capa convolucional, la capa de *pooling* realiza un barrido de los datos de entrada por medio de un filtro aunque esta vez sin calcular ningún peso. Pueden seleccionarse dos tipos de *pooling* que son: *Max pooling* que selecciona el píxel de mayor valor dentro del filtro, y *pooling* promedio que calcula y selecciona el valor promedio de los píxeles dentro del filtro. A pesar de la pérdida de información asociada a la reducción de las dimensiones de la imagen, las capas de *pooling* ayudan a mejorar la eficiencia de la red, reducen su complejidad y limitan el riesgo de sobre-entrenamiento. Las dimensiones de salida de la capa de reducción de muestreo vienen dadas también por la Ecuación (2.6).

2.3.1.3. CAPA TOTALMENTE CONECTADA

Como su nombre lo indica, cada nodo de la capa totalmente conectada se encuentra conectado directamente a los nodos de la capa previa. Esta capa realiza la tarea de aprendizaje de acuerdo con la información presente en las características extraídas por las capas previas. A diferencia de las capas convolucionales que suelen estar seguidas de una capa ReLU, las capas FC se conectan a su salida con una capa de activación *Softmax* que clasifica las entradas y entrega un valor entre 0 y 1. La capa totalmente conectada tiene dimensiones de salida equivalentes al número de parámetros de entrada de la siguiente etapa, si esta seguida de otra capa totalmente conectada, en caso contrario las dimensiones de salida son iguales al número de clases del modelo.

2.3.2. U-NET

La red U-Net es una CNN desarrollada por Ronneberger et al. [4] con el propósito de poder entrenar eficientemente un algoritmo de CNN con solo unas pocas muestras anotadas disponibles y apoyándose fuertemente en técnicas de aumento de datos. La arquitectura de la red U-Net puede ser dividida en dos etapas. La primera, conocida como el camino de contracción o codificador, tiene la tarea de extraer características de alto nivel en la imagen. La segunda etapa, conocida como camino simétrico de

expansión o decodificador, permite a la red localizar la información de clasificación. Esta arquitectura le da a la U-Net su forma de “U” característica.

De acuerdo con la descripción de Siddique et al. [5], el camino de contracción de la red U-Net consiste de arreglos repetidos de dos convoluciones sucesivas de tamaño 3×3 seguidas de una capa de activación ReLU y una capa de *max-pooling*. En el camino de expansión, los mapas de características de la imagen son sobre muestreados mediante convoluciones transpuestas de tamaño 2×2 por etapa y luego concatenadas con el mapa de características de la capa correspondiente en el camino de contracción. A continuación se realizan dos convoluciones consecutivas de 3×3 y una activación ReLU. La imagen segmentada se obtiene aplicando una convolución de 1×1 para reducir el mapa de características al número de canales deseados.

2.3.3. TRANSFERENCIA DE APRENDIZAJE

La transferencia de aprendizaje es una metodología de aprendizaje profundo, donde se toma un modelo previamente entrenado, generalmente con una cantidad grande de datos, para utilizarlo, bien sea como punto de partida para entrenar una segunda red o como un potente extractor de características genéricas [6]. Este modelo previo permite el aprendizaje para realizar una nueva tarea, teniendo en cuenta que las primeras capas de una red neuronal entrenada con imágenes aprenden características generales (que no son específicas a la tarea o base de datos) de la imagen y estas van volviéndose más específicas a medida que se van analizando las capas más profundas.

El método de transferencia de aprendizaje consiste en cargar el modelo preentrenado en el nuevo algoritmo de entrenamiento y “congelar” todas o algunas capas del modelo preentrenado. En el caso de utilizar el modelo preentrenado como extractor de características, se toma la salida de la capa totalmente conectada para que sirva como características de entrada a un nuevo modelo. Si lo que desea es utilizar los pesos del modelo preentrenado como punto de partida para un nuevo modelo, se congelan las capas anteriores a la capa deseada y se crean nuevas capas con parámetros “entrenables” a partir de la última capa congelada, ajustándose al número de clases del nuevo problema a resolver.

2.3.4. ESTIMACIÓN ADAPTATIVA DE LOS MOMENTOS (OPTIMIZADOR ADAM)

Parafraseando a Kingma and Ba [7], Adam es un “algoritmo para optimización de primer orden basada en gradiente de funciones objetivo estocásticas, basado en estimaciones adaptables de momentos de bajo orden”. A diferencia del SGD, en el cual la tasa de aprendizaje es constante para todos los pesos durante todo el entrenamiento, en el optimizador Adam la tasa de aprendizaje es adaptativa y se calcula para cada parámetro.

El algoritmo Adam está basado en dos variaciones del algoritmo SGD: el Algoritmo de Gradiente Adaptativo (AdaGrad) [8], el cual calcula una tasa de aprendizaje para cada parámetro; y la Propagación de Raíz Cuadrática Media [9] que también tiene una tasa de aprendizaje por parámetro y se adapta a la razón de cambio de los gradientes. Adam usa los primeros y segundos momentos del SGD, de ahí su nombre. El enésimo momento de una variable aleatoria es el valor esperado de esa variable a la potencia n :

$$m_n = E[X^n] \quad (2.7)$$

donde m_n es el enésimo momento y X es la variable aleatoria. El primer momento es la media y el segundo momento es la varianza no centrada. Para estimar estos momentos Adam usa los promedios

móviles:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (2.8)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2.9)$$

donde m_t y v_t son los estimados del primer y segundo momento; g es el gradiente de la función de costo para la actual época de entrenamiento; β_1, β_2 son hiperparámetros cuyos valores por defecto son $\beta_1 = 0,9$ y $\beta_2 = 0,999$, convencionalmente se conservan estos valores.

Ya que m_t y v_t son inicializados como vectores de ceros, sus valores tienden también a cero, especialmente durante las primeras iteraciones de entrenamiento. Para esto se usan las ecuaciones con sesgo corregido de los promedios móviles:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (2.10)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (2.11)$$

La tasa de aprendizaje adaptativa de cada parámetro se calcula de la anterior ecuación. Y por lo tanto, la regla de actualización del optimizador Adam está dada por (2.12):

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (2.12)$$

donde θ es el vector de pesos del modelo, η es la tasa de aprendizaje y ϵ es un escalar usado para evitar la división por cero, usualmente $\epsilon = 1 \times 10^{-8}$.

2.3.5. ENTROPÍA CRUZADA BINARIA

BCE es una función de costo comúnmente usada en aprendizaje de máquinas cuya expresión matemática se deriva del concepto de entropía, que es una medida de la incertidumbre asociada a una distribución de probabilidad $q(y)$. En este orden de ideas, si todos los puntos de la distribución pertenecen a una misma clase, la incertidumbre es igual a 0. Y en el peor de los casos, cuando la mitad de los datos pertenecen a una clase y la otra mitad pertenecen a la otra clase, la entropía viene dada por la expresión (2.13):

$$H(q) = \log(2) \quad (2.13)$$

Para todos los otros casos la entropía viene dada por (2.14):

$$H(q) = - \sum_{c=1}^C q(y_c) \cdot \log(q(y_c)) \quad (2.14)$$

Sin embargo, en el caso del aprendizaje de máquinas no se conoce la distribución $q(y)$ a priori, sino que tratamos de aproximarnos a ella mediante otra distribución $p(y)$. Es decir, se asume que los datos vienen de $p(y)$ pero se sabe que en realidad obedecen la distribución $q(y)$. Esto se conoce como la entropía cruzada entre las dos distribuciones y se puede expresar como en la ecuación (2.15):

$$H_p(q) = - \sum_{c=1}^C q(y_c) \cdot \log(p(y_c)) \quad (2.15)$$

Idealmente, $H(q) = H_p(q)$, pero para que esto se de las dos distribuciones deben ser iguales también $q(y) = p(y)$. Esto es poco probable, teniéndose que en la mayoría de los casos la entropía cruzada $H_p(q)$ es mayor a la entropía $H(q)$.

$$H_p(q) - H(q) \geq 0 \quad (2.16)$$

Esta diferencia se conoce como la divergencia de Kullback-Leibler [10] y sirve para medir la similitud entre dos distribuciones dadas:

$$D_{KL}(q||p) = H_p(q) - H(q) = \sum_{c=1}^C q(y_c) \cdot [\log(q(y_c)) - \log(p(y_c))] \quad (2.17)$$

La anterior expresión indica que entre más cerca esté $p(y)$ de $q(y)$, menor será el valor de la divergencia y, por lo tanto, menor la entropía cruzada. La tarea de un clasificador es entonces encontrar la mejor distribución $p(y)$ que minimice la entropía cruzada. Durante el entrenamiento, el clasificador usa los N datos de entrenamiento para calcular la función de costo y encontrar la distribución $p(y)$ óptima:

$$H_p(q) = -\frac{1}{N} \sum_{c=1}^C \log(p(y_i)) \quad (2.18)$$

Con un poco de manipulación matemática se tiene que para cualquier dato:

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \quad (2.19)$$

donde y es el vector de etiquetas y $p(y)$ la probabilidad de que dicha etiqueta pertenezca a una clase dada, y N es el número total de datos de entrenamiento.

2.4. MÉTODOS DE EVALUACIÓN

Se implementaron varios tipos de métricas para evaluar el desempeño de los algoritmos dada la diferente naturaleza de los dos modelos a implementar para la consecución de los objetivos de detección y segmentación. Para el algoritmo de detección se implementaron métricas basadas en matriz de confusión, mientras que el modelo de segmentación se evaluó con una combinación de técnicas que miden la similitud entre dos muestras y de métricas basadas en matriz de confusión.

2.4.1. MÉTRICAS BASADAS EN MATRIZ DE CONFUSIÓN

A la matriz de confusión también se le conoce como matriz de error, esta permite visualizar el desempeño de un clasificador binario y, a partir de ahí realizar el cálculo de diferentes métricas cuantitativas como exactitud, precisión, sensibilidad y especificidad.

2.4.1.1. EXACTITUD (ACCURACY)

La exactitud mide qué tan cerca están un grupo de medidas u observaciones de su valor verdadero o esperado. Se calcula mediante la ecuación (2.20):

$$Acc = \frac{1}{N} \sum_i^N 1(y_i = \hat{y}_i) \quad (2.20)$$

donde N es el número de muestras, y_i representa las etiquetas y \hat{y}_i las predicciones. Esta ecuación en clasificación binaria se expresa en la ecuación (2.21) como la razón entre las clasificaciones correctas y la cantidad total de datos:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.21)$$

donde TP y FP son los verdaderos positivos (TP, por sus siglas en inglés) y FP respectivamente, y TN y FN los verdaderos negativos (TN, por sus siglas en inglés) y falsos negativos (FN, por sus siglas en inglés), respectivamente.

2.4.1.2. PRECISIÓN

La precisión, también llamada valor predictivo positivo, mide qué tan cerca están las mediciones unas de otras, es decir, la tasa de verdaderos positivos sobre todos los valores estimados como positivos. Se calcula para problemas binarios mediante la ecuación (2.22).

$$Precisión = \frac{TP}{TP + FP} \quad (2.22)$$

2.4.1.3. SENSIBILIDAD (RECALL)

La sensibilidad mide la probabilidad de que el resultado de una prueba sea positivo, dado que la muestra es positiva. En el caso de un diagnóstico médico, es la probabilidad de diagnosticar una enfermedad en un paciente que se encuentra enfermo. La sensibilidad se calcula mediante la ecuación (2.23).

$$Sensibilidad = \frac{TP}{TP + FN} \quad (2.23)$$

2.4.1.4. ESPECIFICIDAD

La especificidad se calcula mediante la ecuación (2.24). Esta mide la capacidad del clasificador de dar resultados negativos correctamente, es decir, volviendo al ejemplo del paciente, la capacidad de diagnosticar como sano un paciente que se encuentra verdaderamente sano.

$$Especificidad = \frac{TN}{TN + FP} \quad (2.24)$$

2.4.2. COEFICIENTE DE SIMILITUD SØRENSEN-DICE

El coeficiente de similitud Sørensen-Dice [11, 12] sirve para medir qué tan similares son dos muestras en una escala de cero a uno, siendo cero una total diferencia entre las dos muestras, y uno, dos muestras completamente iguales. Para calcular su valor se emplea la ecuación (2.25):

$$Dice = \frac{2C}{|A| + |B|} = \frac{2|A \cap B|}{|A| + |B|} \quad (2.25)$$

donde A es la máscara predicha, B es la máscara manualmente segmentada y C es la cantidad de *vóxeles* o píxeles que están en la intersección de A y B , es decir, $C = |A \cap B|$.

Esta medida puede adaptarse para usarse también como función de costo en los algoritmos de aprendizaje automático para segmentación semántica, como propuso Milletari et al. [13]. Especialmente en problemas donde existe desbalance de clases, como suele ocurrir en el caso del procesamiento de imágenes médicas, cuya región de interés suele ocupar tan solo una pequeña porción de la totalidad del volumen MRI. En este caso, la ecuación viene dada de la siguiente forma:

$$L_{Dice} = 1 - \frac{2C}{|A| + |B|} = 1 - \frac{2|A \cap B|}{|A| + |B|} = 1 - Dice \quad (2.26)$$

REFERENCIAS

- [1] Yahya Baba and Jeremy Jones. T1 weighted image. *Radiopaedia.org*, 3 2009. doi: 10.53347/rid-5852. URL <https://doi.org/10.53347/rid-5852>.
- [2] A Haouimi and J Jones. T2 weighted image. *Radiopaedia.org*, 2005. doi: 10.53347/rID-6345. URL <https://doi.org/10.53347/rID-6345>.
- [3] Stephen M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3): 143–155, 11 2002. doi: 10.1002/hbm.10062. URL <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/hbm.10062>.
- [4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [5] Nahian Siddique, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access*, 9:82031–82057, 2021. doi: 10.1109/access.2021.3086020.
- [6] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. 6 2014. doi: 10.1109/cvprw.2014.131. URL <https://doi.org/10.1109/cvprw.2014.131>.
- [7] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv (Cornell University)*, 1 2015. doi: 10.48550/arxiv.1412.6980. URL <http://arxiv.org/pdf/1412.6980>.
- [8] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.
- [9] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop, coursera: Neural networks for machine learning. *University of Toronto, Technical Report*, 2012.
- [10] Seymour Sherman. Solomon kullback, information theory and statistics. *Bulletin of the American Mathematical Society*, 66(6):472–472, 1960.
- [11] Thorvald Julius Sørensen. *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons*. I kommission hos E. Munksgaard, 1948.
- [12] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3): 297–302, 1945.
- [13] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 10 2016. doi: 10.1109/3dv.2016.79. URL <https://doi.org/10.1109/3dv.2016.79>.

3

MARCO EXPERIMENTAL

Índice

3.1 Elección de bases de datos y preprocesamiento	57
3.1.1 Brain Tumor Segmentation 2020 dataset	58
3.1.2 Conjunto de datos Imágenes Ayudas Diagnósticas SURA	59
3.2 Detección automática de lesiones de FCD en MRI	62
3.2.1 Entrenamiento del clasificador de tumores cerebrales en MRI	63
3.2.2 Resultados de la CNN en detección de tumores	64
3.2.3 Entrenamiento del clasificador de lesiones de FCD en MRI	65
3.2.4 Resultados de la CNN en detección de lesiones de FCD	67
3.3 Segmentación semántica de lesiones de FCD en volúmenes MRI	69
3.3.1 Segmentación de tumores en MRI	69
3.3.2 Resultados de la U-Net en segmentación de tumores	71
3.3.3 Segmentación de lesiones FCD en imágenes MRI	73
3.3.4 Resultados de la U-Net en segmentación de lesiones de FCD	76
Referencias	83

EN este capítulo se aborda la metodología empleada para llevar a cabo los experimentos que sustentan este trabajo. Esta ha sido dividida en secciones o etapas, mediante el desarrollo de las cuales se tiene previsto alcanzar cada uno de los objetivos específicos planteados y con la realización de todos ellos, lograr el objetivo general del proyecto. Se detallará el proceso que se llevó a cabo para lograr los objetivos de detección y segmentación de lesiones de FCD y la integración de ambos, mientras que la evaluación de los modelos se realizará y documentará de manera transversal a cada objetivo conseguido.

El diagrama de la Figura 3.1 muestra cómo será el desarrollo de cada una de estas etapas. En la primera, el objetivo es obtener un modelo de detección de lesiones de FCD, a partir de un modelo entrenado para la detección de tumores por medio de transferencia de aprendizaje. El objetivo de la segunda etapa es obtener un modelo de segmentación de lesiones de FCD, también vía transferencia de aprendizaje, a partir de un modelo entrenado para la segmentación de tumores. En la tercera etapa

el objetivo es formar una herramienta, donde se integren los modelos de las etapas anteriores, que sirva de ayuda en el diagnóstico preoperatorio de lesiones de FCD. La evaluación de los modelos se realizará de forma transversal a lo largo del desarrollo de la metodología.

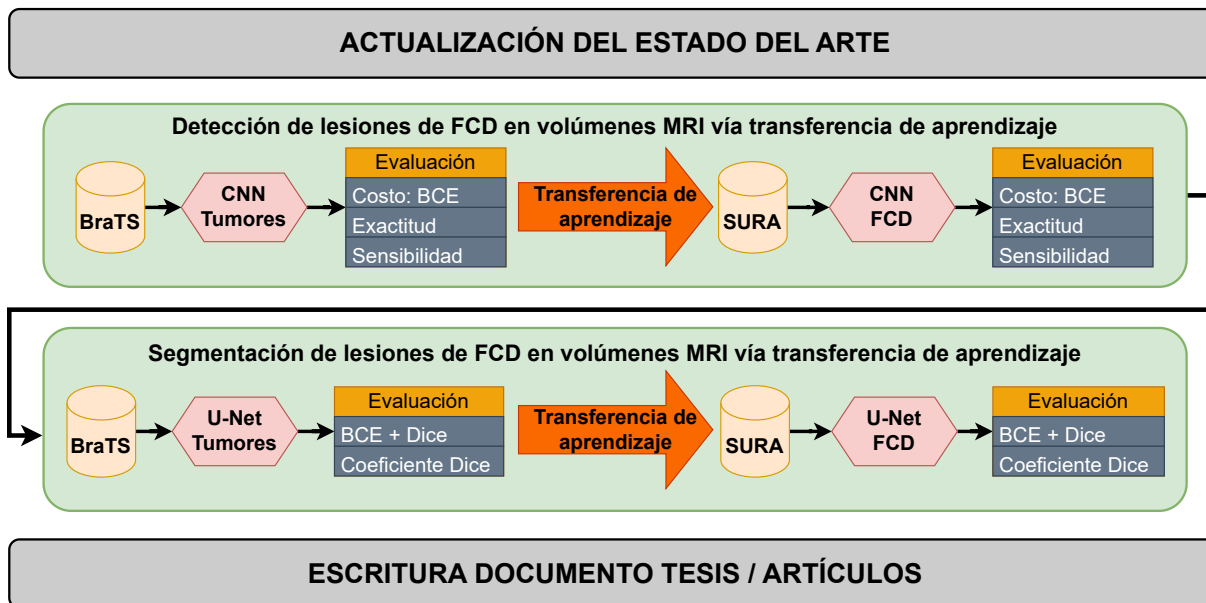


Figura 3.1: Diagrama de flujo que describe la metodología empleada en la realización de este trabajo.

3.1. ELECCIÓN DE BASES DE DATOS Y PREPROCESAMIENTO

Para seleccionar las bases de datos necesarias para lograr los objetivos de detección y segmentación de lesiones de FCD, se tuvo en cuenta que la metodología de transferencia de aprendizaje a implementar implicaba resolver dos problemas por objetivo: (a) detección y segmentación de tumores, y (b) detección y segmentación de FCD. Además, se debía contar con bases de datos abiertas y balanceadas. En detección de tumores existían varias opciones para elegir base de datos, sin embargo, para pacientes con lesiones de FCD no se encontró ninguna base de datos pública que pudiese tomar como referencia. Adicionalmente, a pesar de haber hecho varias solicitudes de intercambio de datos con los autores de artículos de investigación previos a este trabajo relacionados con detección y segmentación de lesiones de FCD, todas ellas fueron denegadas aduciendo conflictos éticos o institucionales.

Se optó entonces por elegir para las etapas de detección y segmentación de tumores, la base de datos del "Programa de segmentación de Tumores en el Cerebro" (BraTS, por sus siglas en inglés) de 2020, compuesta por 295 volúmenes y que describimos en la Subsección 3.1.1. Mientras que para entrenar los algoritmos de detección y segmentación de FCD, se contó con una base de datos suministrada por el equipo coejecutor, conformada por 48 volúmenes MRI pertenecientes a 27 pacientes con lesiones de FCD, etiquetada por un neurólogo experto, depurada luego para contar con 23 volúmenes para entrenamiento. Si bien es escasa la cantidad de volúmenes de pacientes con lesiones de FCD, el enfoque metodológico, donde se emplea transferencia de aprendizaje para resolver los problemas de detección y segmentación de FCD, fue diseñado para compensar esta falencia. En efecto, los parámetros aprendidos por los modelos diseñados para resolver la tarea de detección y segmentación de tumores, que

Modelos de tumores y FCD: cantidad de volúmenes y slices para entrenamiento, validación y prueba									
Caso	Etapa	Entrenamiento		Validación		Prueba		Total	
		Volúmenes	Slices	Volúmenes	Slices	Volúmenes	Slices	Volúmenes	Slices
Tumores		100	6400	10	640	10	640	120	7680
FCD		19	1216	2	128	2	128	23	1472

disponen de bases de datos públicas con una mayor cantidad de datos, se transfieren en el entrenamiento a los modelos de detección y segmentación de lesiones de FCD, para los cuales se contó con una menor cantidad de datos. Asimismo, con el enfoque de entrenamiento de los modelos por *slices*, en lugar de volúmenes completos, como estrategia de aumento de datos, junto con la transferencia de aprendizaje, se busca impulsar el entrenamiento de modelos mejor ajustados. A continuación se describen las dos bases de datos con las que se contó para realizar los experimentos.

3.1.1. BRAIN TUMOR SEGMENTATION 2020 DATASET

La base de datos elegida para entrenamiento del modelo de segmentación de tumores cerebrales fue el dataset BraTS [1–3]. Esta base de datos utiliza volúmenes MRI de múltiples instituciones y se enfoca en cuatro tareas de segmentación, con sus respectivas etiquetas: Tarea 1: segmentación de tumores cerebrales o gliomas. Tarea 2: predicción general de supervivencia de pacientes. Tarea 3: distinción entre pseudo-progresión y recurrencia real de tumor por medio de análisis integrador de características *radiómicas* y algoritmos de aprendizaje de máquina. Tarea 4: evaluación de la incertidumbre algorítmica en segmentación de tumores.

Todas las imágenes en la base de datos están disponibles en archivos en formato NIfTI en secuencias T1, T1 postcontraste, T2 y T2-FLAIR; adquiridas usando diferentes protocolos y escáneres de 19 instituciones diferentes. Todas las imágenes fueron segmentadas manualmente por uno a cuatro evaluadores, siguiendo el mismo protocolo, y aprobados por radiólogos experimentados. Las etiquetas manualmente delineadas corresponden a los siguientes tejidos:

- 1. Tumor GD-mejorado (ET, por sus siglas en inglés. Etiqueta: 4)
- 2. Edema peritumoral (ED, por sus siglas en inglés. Etiqueta: 2)
- 3. Núcleo necroso no-mejorado (NCR/NET, por sus siglas en inglés. Etiqueta: 1)

Todos los volúmenes fueron publicados después de ser preprocesados, corregistrados a la misma plantilla anatómica, interpolados a la misma resolución de *vóxel* (1mm^3) y tras pasar por un proceso de extracción de cerebro. La base de datos reúne un total de 369 volúmenes, con su respectiva segmentación manual, y con resolución de $240 \times 240 \times 155$ *vóxeles*.

Para el algoritmo de detección de tumores se toma de esta base de datos el volumen en la secuencia T1. La etapa de preprocesamiento realizado a esta base de datos, como preparación para su entrada al algoritmo, consistió de los siguientes pasos:

1. Reescalamiento de la imagen al tamaño de las dimensiones de entrada de la CNN, $128 \times 128 \times 64$, para lo cual se reescaló primero la matriz de afinidad teniendo en cuenta las nuevas dimensiones de entrada y definiendo visualmente un nuevo tamaño de *vóxel* de $2 \times 2 \times 2\text{mm}$ donde quepan las nuevas dimensiones. Luego, se redimensionó el arreglo de la imagen, y por último, se generó un nuevo volumen en formato NIfTI.

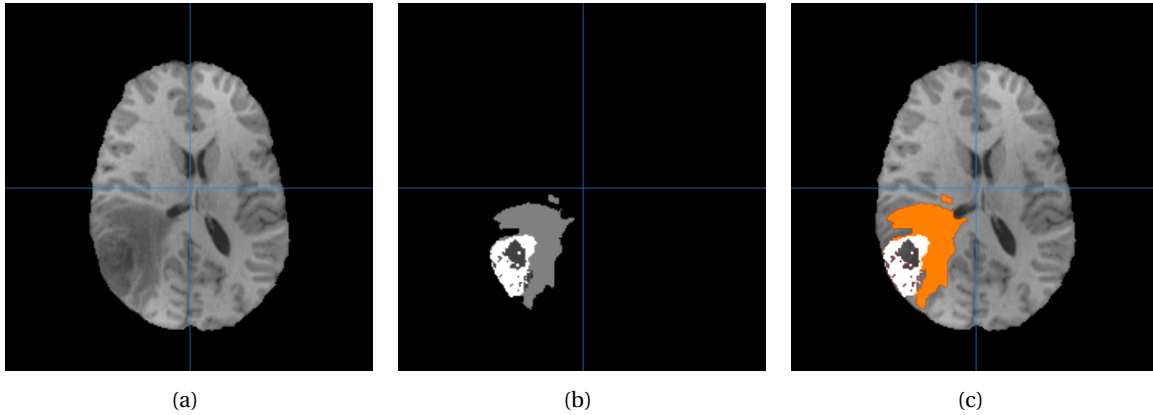


Figura 3.2: Ejemplo de *slice* de una imagen de la base de datos BraTS donde se puede apreciar el tumor segmentado manualmente proyectado sobre la imagen T1. (a) Secuencia T1. (b) Etiquetas manuales ó *ground truth*. (c) Proyección de las etiquetas sobre la imagen en secuencia T1.

2. Se normalizaron los valores de intensidad de los *vóxeles* del volumen a un rango entre [0, 1] de acuerdo con la ecuación (3.1):

$$vol = \frac{vol - min}{max - min} \quad (3.1)$$

donde *vol* es el valor de la intensidad del *vóxel*, *min* es el valor mínimo de intensidad de los *vóxeles* y *max* es el valor máximo.

3.1.2. CONJUNTO DE DATOS IMÁGENES AYUDAS DIAGNÓSTICAS SURA

La base de datos de resonancia magnética estructural de cerebro seleccionada y depurada por la entidad coejecutora, está compuesta por estudios de resonancia magnética (MRI) de cráneo, estos incluyen secuencias volumétricas T1, T2 y T2-FLAIR, de pacientes que consultaron en ayudas diagnósticas SURA. A partir de lo reportado en la lectura radiológica de las imágenes, se contó con 46 MRI pertenecientes a 27 sujetos con diagnóstico de displasia cortical focal (15 mujeres, edad: 28.07 ± 16.19 años), y 63 sujetos sanos (23 mujeres, edad = 33.13 ± 8.78 años). Los estudios seleccionados fueron adquiridos en diferentes centros del servicio de ayudas diagnósticas, por lo que provienen de diferentes resonadores (Siemens, Philips y GE), así como diferentes intensidades de campo magnético (1.5 y 3 Tesla).

Esta base de datos se depuró partiendo de que el entrenamiento se realizará solo con MRI en secuencia T1, descartándose los pacientes que no tienen imagen en esta secuencia. Se decidió entrenar solamente con pacientes que tuvieran una única lesión de FCD en la marcación, descartándose los estudios con múltiples lesiones, para evitar sesgos en el entrenamiento hacia estos pacientes. Después de este proceso de filtrado, la base de datos quedó compuesta por 23 volúmenes MRI correspondientes a 23 pacientes, cada MRI con su respectiva marcación manual delineada por el experto.

Las marcaciones manuales se realizaron bien sea sobre la secuencia T1, la secuencia T2 o la secuencia T2-FLAIR, está configuración de estudios de paciente se da dependiendo de en cuál secuencia la lesión fuera evidenciada por el experto con mayor facilidad. De acuerdo con esto, se presentaron las siguientes conjuntos de imagen MRI con marcación:

1. MRI en secuencia T1 con marcación en esa secuencia (9 casos).
2. MRI en secuencia T1 y T2 con marcación en secuencia T2 (7 casos).
3. MRI en secuencia T1 y T2-FLAIR con marcación en secuencia T2-FLAIR (7 casos).
4. MRI en secuencia T1 y T2-FLAIR con marcación en secuencia T1 y T2-FLAIR (2 casos).
5. MRI en secuencia T2, marcación en secuencia T2 (1 caso).
6. MRI en secuencia T1, T2 y T2-FLAIR con múltiples marcaciones en secuencia no especificada (1 caso).

Adicional a esto, los volúmenes suministrados en la base de datos, en su mayoría, tenían diferente resolución, orientación y tamaño de *vóxel*. En la Figura 3.3 se aprecia que la MRI en secuencia T1 tiene menos resolución. Por otro lado, la marcación, al ser proyectada sobre la secuencia T1, queda ubicada en un *slice* diferente del de la secuencia T2-FLAIR.

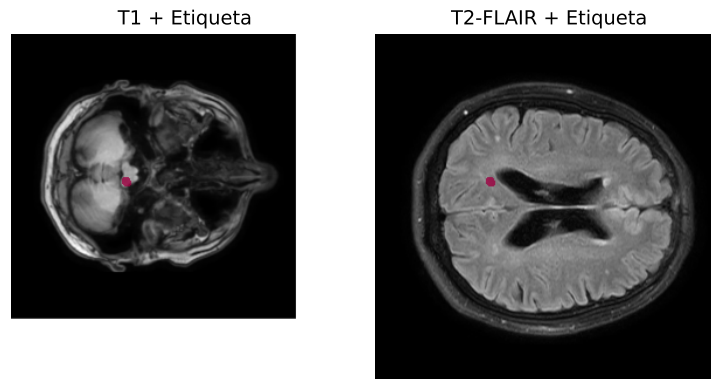


Figura 3.3: Diferencias entre tipos de secuencia en el conjunto de datos imágenes ayudas diagnósticas SURA.

Esto implicó la realización de un preprocesamiento particular casi con cada imagen, por lo que se elaboró un *pipeline*, que se muestra en la Figura 3.4, que depende de la secuencia MRI donde estaba incluida la marcación. Asimismo, se estableció que las dimensiones de entrada de los modelos de aprendizaje tendrían dimensión $128 \times 128 \times 64$, orientación en el espacio RAS y tamaño de *vóxel* $2 \times 2 \times 2mm$. Todos los volúmenes debieron ser llevados a este espacio común para así estandarizar dimensiones, orientación y tamaño de *vóxel*.

A continuación se explica brevemente cómo se llevaron a cabo los pasos de registro, orientación y remuestreo de la imagen.

3.1.2.1. REGISTRO

Por esta etapa pasan las imágenes que no tienen la marcación de lesión de FCD en la secuencia T1 sino que fueron marcadas en la secuencia T2 o T2-FLAIR. La etapa de registro consiste en llevar una MRI objetivo al espacio de referencia de una segunda MRI de referencia. Para esto el algoritmo debe calcular una nueva matriz de afinidad para llegar al espacio de la imagen de referencia, teniendo en cuenta las nuevas dimensiones y el tamaño de *vóxel*.

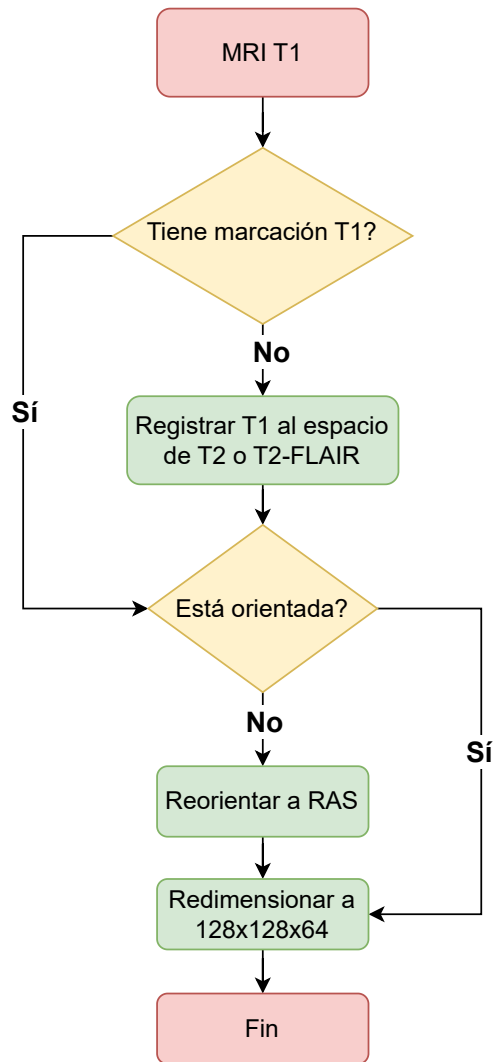


Figura 3.4: Diagrama de flujo que muestra las etapas de preprocesamiento de acuerdo con la configuración de estudios realizados al sujeto. Teniendo como entrada la MRI en secuencia T1, se procede según la secuencia en la que haya sido realizada la marcación de la lesión de FCD.

3.1.2.2. ORIENTACIÓN

Luego de realizar el registro en los casos que fuese necesario, se debió efectuar el proceso de reorientación de la imagen cuando la marcación, al ser sobrepuesta sobre la imagen T1 no se presenta en el mismo espacio de la imagen en el que se encuentra la lesión de FCD.

Este proceso se lleva a cabo leyendo los metadatos de la MRI, donde se encuentra la información de cómo está orientada la imagen, y verificando el espacio en que se encuentra. Los casos que se encontraron fueron los siguientes, tomando como referencia las coordenadas X, Y, Z del plano cartesiano:

1. MRI con orientación X, Y, Z (3 casos).
2. MRI con orientación $-X, -Y, Z$ (5 casos).
3. MRI con orientación $-X, -Y, -Z$ (2 casos).

4. MRI con orientación $-Y, -Z, -X$ (13 casos).

Dada la pluralidad de espacios de referencia en la base de datos, se decidió tomar el espacio de referencia estándar RAS y llevar tanto la MRI como la marcación a este espacio. Para ello se transponen los ejes que están marcados con signo negativo. Por su parte, los ejes que no están en la posición X, Y, Z correspondiente fueron llevados a esta posición mediante cambios de posición de eje.

3.1.2.3. REMUESTREO

El último paso de preprocesamiento consiste en el remuestreo de las MRI y las marcaciones para llevarlas a una resolución y tamaño de *vóxel* estándar, que además son las dimensiones de entrada de los algoritmos de aprendizaje. Esto se efectúa calculando una nueva matriz de afinidad usando la ecuación (2.4) para cada volumen y aplicando la función de transformación en la ecuación (2.5).

Tras efectuar todas las etapas de preprocesamiento en los volúmenes que así lo requirieron, se tiene el resultado mostrado en la Figura 3.5, donde tanto el volumen MRI como la marcación están registrados en el mismo espacio de referencia RAS, tienen las mismas dimensiones y tamaño de *vóxel* y, adicionalmente, la marcación de lesión de FCD se encuentra con la orientación correspondiente a su posición en el volumen.

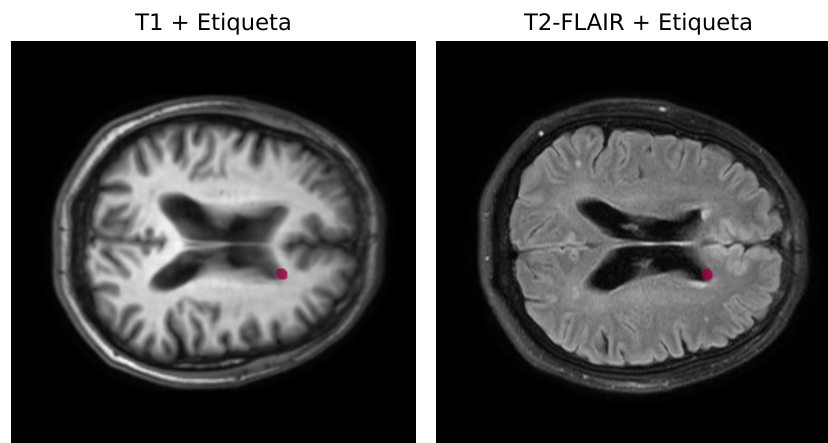


Figura 3.5: Volumen y marcación de FCD registradas, remuestreadas y orientadas al mismo espacio de referencia, listas para servir de entrada al algoritmo de aprendizaje.

3.2. DETECCIÓN AUTOMÁTICA DE LESIONES DE FCD EN MRI

Este paso consiste en establecer, mediante la implementación de un modelo de clasificación, si en el volumen de entrada hay presencia o no de lesión de FCD. Esto es un problema de índole binario que puede ser resuelto mediante un algoritmo de clasificación. Este algoritmo de clasificación fue implementado en dos etapas: en la primera, debido a la poca disponibilidad de volúmenes MRI con lesiones de FCD, se decidió entrenar un modelo de detección de tumores cerebrales en volúmenes MRI, usando los volúmenes de la base de datos abierta BraTS descrita en la Subsección 3.1.1. En la segunda etapa, se usó el modelo de detección de tumores entrenado en la primera etapa para inicializar los pesos del modelo de detección de lesiones de FCD mediante el método de transferencia de aprendizaje, descrito en la Subsección 2.3.3. Para entrenar este modelo se usó la base de datos de imágenes ayudas diagnósticas SURA, descrita en la Subsección 3.1.2.

Los pasos ejecutados en cada etapa fueron: (a) selección y preprocesamiento de base de datos de tumores cerebrales, (b) entrenamiento y evaluación de modelo de detección de tumores en MRI, (c) preprocesamiento de base de datos de pacientes con FCD, (d) entrenamiento, por medio de transferencia de aprendizaje, del modelo de detección de FCD en MRI, y (e) evaluación del modelo de detección de FCD. Al final de esta sección el algoritmo debe estar en capacidad de decidir si en el volumen de entrada existe o no lesión de FCD.

3.2.1. ENTRENAMIENTO DEL CLASIFICADOR DE TUMORES CEREBRALES EN MRI

El objetivo de este clasificador binario es detectar si existen tumores cerebrales en el volumen de entrada para luego usar este modelo como punto de partida en el entrenamiento del clasificador de lesiones de FCD por medio de transferencia de aprendizaje. Para entrenarlo se eligió implementar una CNN 2D compuesta por 4 capas convolucionales, seguidas de dos capas totalmente conectadas y finalizando con una activación sigmoide. Cada capa convolucional contiene una convolución 2D con kernel de 3×3 , *padding* 0 y *stride* 1, el número de filtros de la primera capa convolucional es de 16 y a partir de la tercera convolución se duplica esta cifra para terminar con 64 filtros en la última convolución. A cada convolución le sigue un activación ReLU y una capa de *max pooling* con kernel de 2×2 . En la Figura 3.6 se muestra cómo están dispuestas las capas dentro de la arquitectura de la CNN.

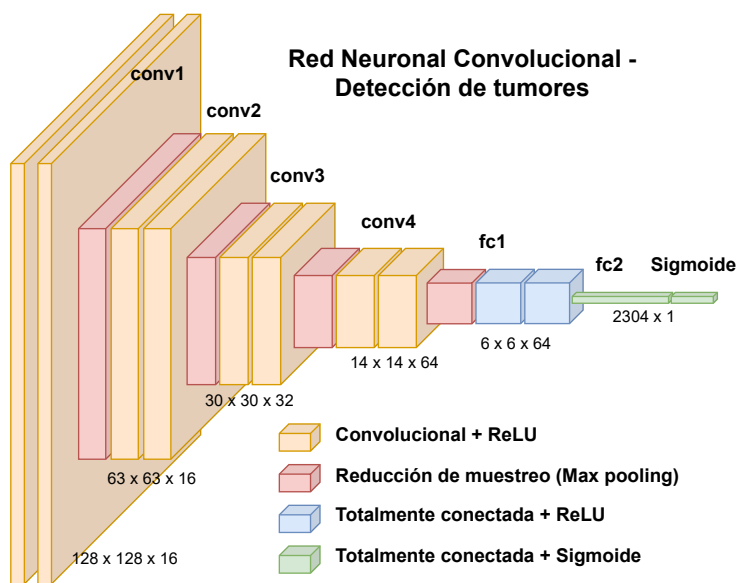


Figura 3.6: Arquitectura de la red neuronal convolucional implementada para detección de tumores a nivel de *slice* en volúmenes MRI. La red consiste de cuatro capas convolucionales seguidas de activación ReLU y reducción de muestreo, posteriormente dos capas totalmente conectadas, la primera con activación ReLU y la segunda con activación sigmoide.

Para entrenar esta red se tomaron 120 volúmenes MRI de la base de datos BraTS y se dividieron en 100 volúmenes para entrenamiento, 10 para validación y 10 para prueba. Las convoluciones son 2D por lo que la red es entrenada con la totalidad de *slices* de cada volumen (64), por lo tanto se cuenta con un total de 6400 *slices* para entrenamiento, 640 para validación y 640 para prueba. En entrenamiento, el número total de *slices* positivos (con tumor) fue de 3317 y de negativos (sin tumor) fue de 3083. En validación 313 positivos, 327 negativos. Esta combinación de número de capas convolucionales y dadas las dimensiones de entrada dan como resultado una cantidad de 5,338,641 parámetros entrenables por la CNN.

Los valores para los hiperparámetros con los cuales se configuró la inicialización de la red CNN para detección de tumores, seleccionados luego de varias sesiones de entrenamiento con una cantidad reducida de volúmenes, fueron los siguientes:

- Dimensiones de entrada del modelo: (128, 128, 1)
- Tasa de aprendizaje: 0.0001
- Numero de épocas: 20
- Tamaño de *batch*: 4
- Umbral de decisión: 0.5

La tasa de aprendizaje se inicializó en 0.0001 y actualiza su valor cada 5 épocas por medio de una función que multiplica el valor actual por un factor de 0.1. El optimizador elegido fue el optimizador Adam descrito en la Subsección 2.3.4 y la función de costo BCE con *logits*, descrita en la Subsección 2.3.5.

3.2.2. RESULTADOS DE LA CNN EN DETECCIÓN DE TUMORES

Para evaluar el desempeño del clasificador se tuvieron en cuenta las métricas exactitud, precisión, especificidad y sensibilidad, basadas en matriz de confusión, referidas en la Subsección 2.4.1. Estas medidas se tomaron en la etapa de entrenamiento, validación y prueba. Adicionalmente, en entrenamiento se monitorizó el descenso de la función de costo.

3.2.2.1. ENTRENAMIENTO

Adicional a la función de costo en entrenamiento, se tomaron medidas de exactitud, precisión, especificidad y sensibilidad que fueron llevadas a la Figura 3.7. En esta gráfica se puede ver que las métricas, en especial la sensibilidad, se comportan de forma ligeramente ascendente. El mejor desempeño en entrenamiento se da con las métricas de exactitud y especificidad, lo que indica que el algoritmo se inclina a realizar buenas predicciones de la condición del paciente.

3.2.2.2. VALIDACIÓN

Para validación se contó con un total de 10 volúmenes de pacientes con tumores cerebrales. Al final de cada época de entrenamiento se pasó cada volumen por el modelo y se tomaron medidas de exactitud, precisión, especificidad y sensibilidad. Los valores quedaron consignados en la Figura 3.8 donde se puede apreciar que las métricas permanecen constantes en esta etapa, con una exactitud, en promedio, de 0.7 aproximadamente.

3.2.2.3. PRUEBA

Para prueba se contó con 10 volúmenes de pacientes con tumores que no fueron usados en las etapas de entrenamiento y validación. A diferencia de validación, en esta sección se toma el modelo seleccionado por la mejor época de entrenamiento, de acuerdo con la lectura más alta de exactitud en la etapa de validación, que en este caso fue la época número 20. En la Tabla 3.1 se consignan los resultados en prueba del algoritmo de detección de tumores, la exactitud muestra un buen desempeño con valor promedio de 0.885 mientras que la especificidad muestra un descenso comparada con los valores que venía mostrando en validación.

Los códigos para este algoritmo pueden encontrarse en su repositorio de GitHub [repositorio de GitHub](#).

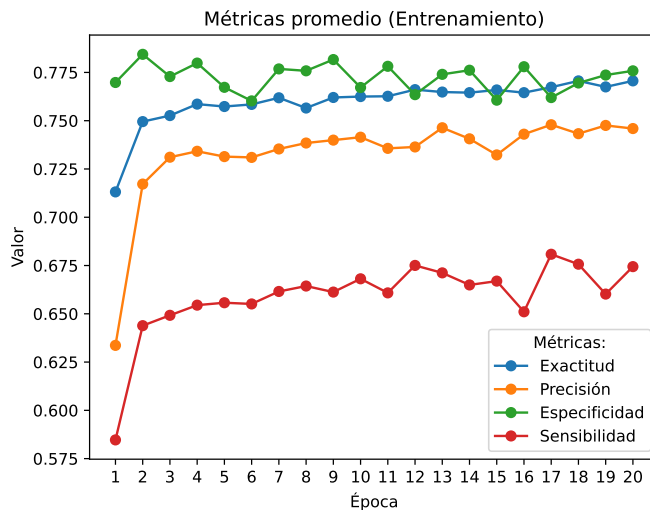


Figura 3.7: Representación gráfica de los valores promedio de exactitud, precisión, especificidad y sensibilidad tomados en cada época de entrenamiento. Se destaca un buen desempeño en exactitud y especificidad.

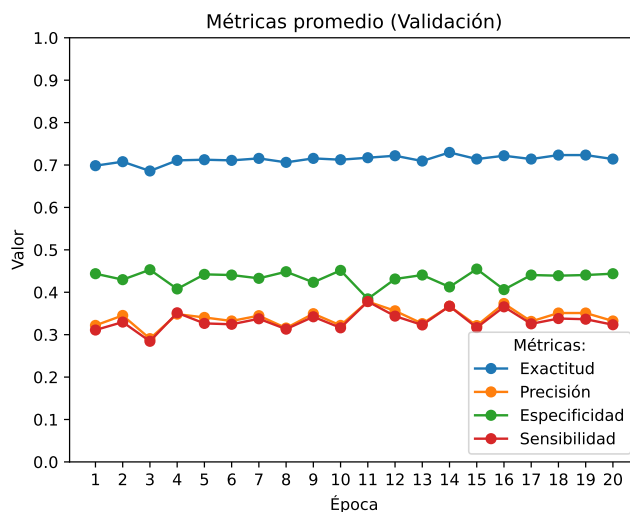


Figura 3.8: Las métricas de exactitud, precisión, especificidad y sensibilidad en la etapa de validación permanecen en un valor constante durante el entrenamiento.

Tabla 3.1: Tabla de resultados en prueba para detección de tumores donde el valor alcanzado de 0.855 exactitud sugiere una buena tasa de detección.

Exactitud	Precisión	Especificidad	Sensibilidad
0.855	0.521	0.407	0.513

3.2.3. ENTRENAMIENTO DEL CLASIFICADOR DE LESIONES DE FCD EN MRI

En esta etapa se toma como insumo el modelo entrenado en la Subsección 3.2.1 para inicializar los pesos del clasificador de lesiones de FCD, cuyo objetivo es detectar si existen o no lesiones de FCD

en el volumen de entrada. Para entrenarlo se tomó la misma arquitectura CNN 2D de dicha sección, compuesta por 4 capas convolucionales, seguidas de dos capas totalmente conectadas y finalizando con una activación sigmoide.

Para implementar el proceso de transferencia de aprendizaje, se instanció el modelo de CNN 2D y luego se procedió a congelar todas las capas de la red. Las capas que no se deseaba reentrenar se dejaron congeladas, mientras que para las capas entrenables se actualizaron con los mismos parámetros para ser sintonizados en la etapa de entrenamiento. Específicamente, se dejaron congeladas las cuatro primeras capas convolucionales. Las dos últimas capas, totalmente conectadas, fueron entrenadas de nuevo. Finalmente, se carga el modelo preentrenado para detección de tumores como pesos iniciales de los parámetros de esta nueva red. En total, se reentrenaron 5,313,025 de 5,338,641 parámetros. En la Figura 3.9 se muestra cómo están distribuidas dentro de la arquitectura de la CNN 2D las capas, tanto congeladas como entrenables.

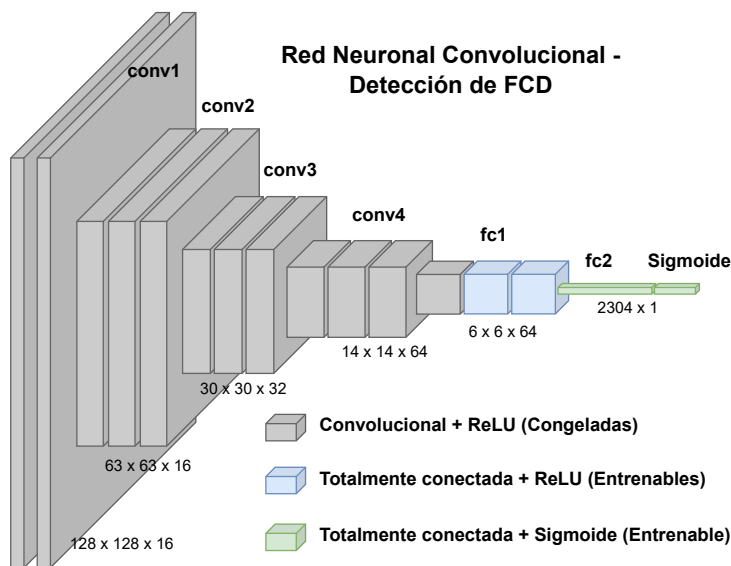


Figura 3.9: Arquitectura de la red neuronal convolucional implementada para detección de lesiones de FCD a nivel de *slice* en volúmenes MRI por medio de transferencia de aprendizaje. Las cuatro primeras capas de la CNN fueron congeladas, entrenándose únicamente las últimas dos capas totalmente conectadas.

Los hiperparámetros de entrenamiento con los cuales se inicializó la red CNN para transferencia de aprendizaje y detección de lesiones de FCD, seleccionados luego de monitorizar el desempeño de la red en varias sesiones de entrenamiento con una cantidad reducida de volúmenes, fueron los siguientes:

- Dimensiones del modelo: (128, 128, 1)
- Tasa de aprendizaje: 0.0005
- Número de épocas: 20
- Tamaño de *batch*: 4
- Umbral de decisión: 0.5

- Número de capas entrenables: 2

La tasa de aprendizaje se inicializó en 0.0005 y actualiza su valor cada 3 épocas por medio de una función que multiplica el valor actual por un factor de 0.1. El optimizador elegido fue el optimizador Adam descrito en la Subsección 2.3.4 y la función de costo BCE con *logits*, descrita en la Subsección 2.3.5.

Esta red se entrenó con los volúmenes de la base de datos de pacientes con lesiones de FCD, descrita en la Subsección 3.1.2. Se cargó cada volumen y se extrajeron los 64 *slices* que lo conforman después de pasar por la etapa de preprocesamiento, seleccionándose para entrenamiento solamente los *slices* que tuvieran marcación del experto, es decir, que contienen FCD. Adicionalmente, se añadieron tres *slices* sin lesión de FCD por cada volumen. El número total de *slices* para entrenamiento fue de 167, 113 positivos y 54 negativos. Si bien existe un desbalance en los datos de entrenamiento, este fue un escenario deseable ya que se buscó compensar la baja cantidad de *slices* con lesión y el pequeño tamaño del área de lesión de FCD en la imagen.

3.2.4. RESULTADOS DE LA CNN EN DETECCIÓN DE LESIONES DE FCD

El primer indicador del desempeño del modelo de detección de lesiones de FCD en MRI fue la curva de descenso de la función de costo durante el entrenamiento del algoritmo. Adicionalmente, se usaron las métricas basadas en matriz de confusión, referidas en la Subsección 2.4.1, tomadas en la etapa de entrenamiento, validación y prueba.

3.2.4.1. ENTRENAMIENTO

En la Figura 3.10 puede verse que los valores promedio por época de la función de costo descienden mostrando una curva pendiente, lo que es señal de una buena optimización en el entrenamiento. Sin embargo, en magnitud la diferencia entre el costo final y el costo inicial no es muy significativa.

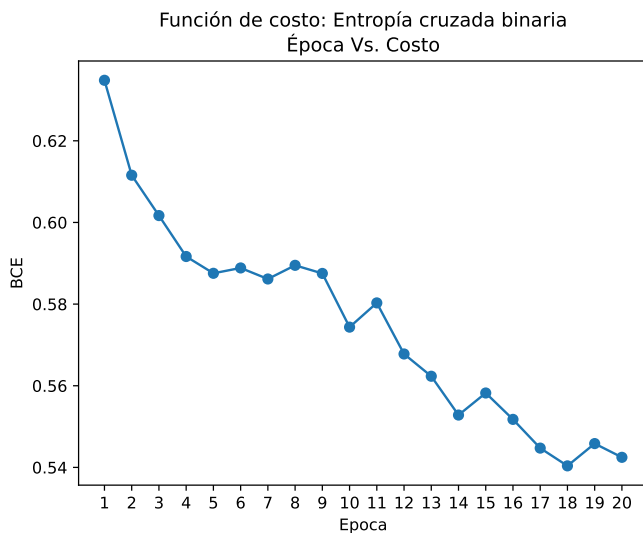


Figura 3.10: Gráfica del valor promedio de la función de costo por época de entrenamiento. La pendiente de la curva es empinada y descendente, sin embargo, en magnitud no tiene un descenso significativo.

Las métricas de exactitud, precisión, sensibilidad y especificidad, en entrenamiento, se tomaron

por cada *batch* y se promediaron luego para cada época, en la Figura 3.11 se recogen los valores de cada una.

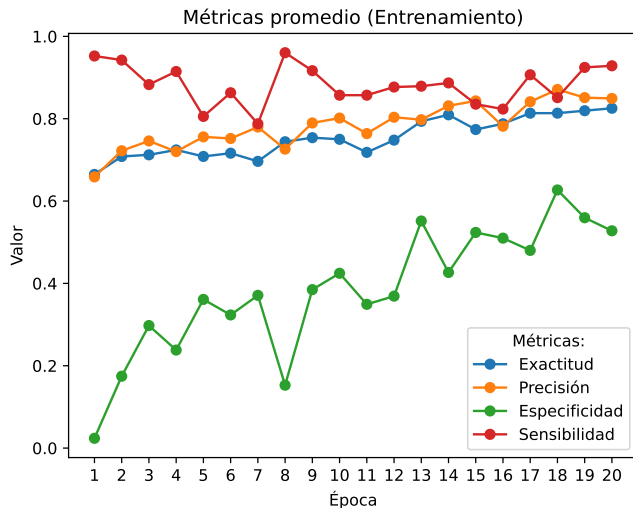


Figura 3.11: Representación gráfica de los valores promedio de exactitud, precisión, especificidad y sensibilidad tomados en cada época de entrenamiento. La especificidad, precisión y exactitud muestran una tendencia al crecimiento por época de entrenamiento, mientras que la sensibilidad permanece casi constante en valores cercanos a uno.

La especificidad es la métrica que exhibe una curva ascendente que es más acorde con la etapa de entrenamiento, si bien la exactitud y precisión también muestran un leve ascenso. La sensibilidad se comporta de manera casi estable a lo largo del entrenamiento y en valores cercanos a uno, lo que indica que el modelo está predispuesto a detectar solo los casos positivos, probablemente por desbalance en los datos de entrenamiento. Sin embargo, la curva en ascenso de especificidad demuestra una tendencia a revertir este sesgo, logrando también precisión en los casos negativos.

3.2.4.2. VALIDACIÓN

Los promedios de las métricas de exactitud, precisión, sensibilidad y especificidad en validación, por época, se muestran en la Figura 3.12. La especificidad muestra una curva ascendente mientras que las demás métricas permanecen en niveles relativamente constantes, donde la exactitud y la sensibilidad muestran los valores más altos.

3.2.4.3. PRUEBA

El modelo de detección de lesiones de FCD fue evaluado usando una base de datos de prueba, conformado por dos volúmenes de la base de datos presentada en la Subsección 3.1.2, que en ningún momento formaron parte de las etapas de entrenamiento y validación. Los resultados se muestran en la Tabla 3.2. La exactitud, quedándose en un valor de 0.477, no tuvo el comportamiento esperado teniendo en cuenta que el modelo de detección de tumores de la Subsección 3.2.1 tuvo un valor de 0.855 en la etapa de prueba.

Los códigos para este algoritmo pueden encontrarse en su repositorio de GitHub [repositorio de GitHub](#).

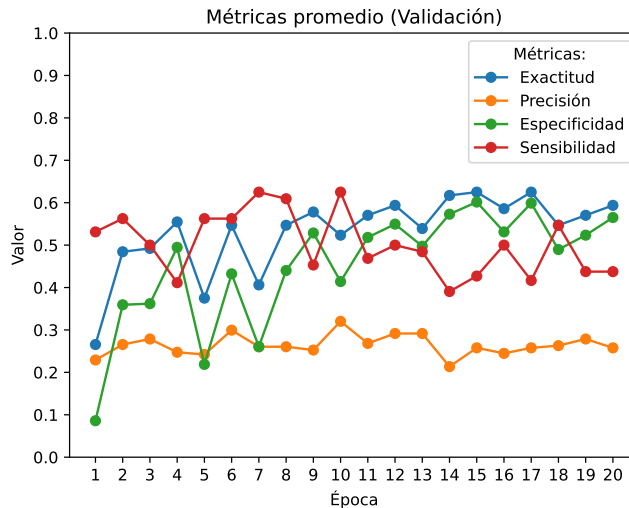


Figura 3.12: Desempeño del modelo de detección de lesiones de FCD en la etapa de validación. Se presenta inestabilidad en las métricas, principalmente de especificidad y sensibilidad, con una exactitud que en las últimas etapas de entrenamiento oscila entre 0.5 y 0.6.

Tabla 3.2: Tabla de resultados en prueba para detección de lesiones de FCD. La exactitud desciende en valor comparada con el modelo de detección de tumores y se observa una caída en general de las demás métricas.

Exactitud	Precisión	Especificidad	Sensibilidad
0.477	0.148	0.328	0.188

3.3. SEGMENTACIÓN SEMÁNTICA DE LESIONES DE FCD EN VOLÚMENES MRI

En esta sección se describen los pasos realizados para alcanzar el objetivo de segmentación de lesiones de FCD en MRI. Esto es, obtener un modelo que esté en capacidad de delinear automáticamente el contorno de la lesión de FCD. Para esto hacemos uso de la CNN conocida como U-Net que se describió en la Subsección 2.3.2, combinada con la metodología de transferencia de aprendizaje de la Subsección 2.3.3 para compensar la escasa cantidad de datos. Para conseguir este objetivo, se entrena un primer clasificador de segmentación de tumores en el cerebro y luego, por medio de transferencia de aprendizaje, se cargan los pesos aprendidos en esta etapa para inicializar el entrenamiento de un segundo modelo cuyo propósito es demarcar los *vóxeles* de la región con lesión de FCD.

Similar a la metodología implementada en la Sección 3.2, los pasos ejecutados en cada etapa fueron: (a) selección y preprocesamiento de base de datos de tumores cerebrales, (b) entrenamiento y evaluación de U-Net para segmentación de tumores en MRI, (c) preprocesamiento de base de datos de pacientes SURA con FCD, (d) entrenamiento, por medio de transferencia de aprendizaje, de U-Net para segmentación de lesiones de FCD en MRI, y (e) evaluación del modelo de detección de lesiones FCD. A continuación se describen en detalle y se dan los resultados de las etapas de esta sección.

3.3.1. SEGMENTACIÓN DE TUMORES EN MRI

La primera etapa consistió en el entrenamiento de una red U-Net con el objetivo de delinear el tumor cerebral presente en la MRI. Para lograrlo, se utilizaron los volúmenes de la base de datos de

tumores BraTS, presentado en la Subsección 3.1.1 como entrada para entrenamiento de la red U-Net de la Subsección 2.3.2. En la Figura 3.13 se muestra la arquitectura de la U-Net. Los dos caminos, contracción y expansión, fueron diseñados de la siguiente forma para este experimento:

El camino de contracción se compuso de cinco bloques convolucionales, cada bloque con 64, 128, 256, 512 y 1024 canales, respectivamente. Cada bloque convolucional estuvo conformado por dos capas convolucionales con $kernel$ de 3×3 , $stride$ 1 y $padding$ 1. A cada capa convolucional, exceptuando la última, le sigue una función de activación ReLU y una capa de max -pooling con $kernel$ de 2×2 .

Por su parte, el camino de expansión se conformó con cuatro bloques convolucionales, cada bloque con una capa convolucional hacia arriba con $kernel$ de 2×2 , $stride$ 2. Cada convolución hacia arriba es seguida de dos capas convolucionales con $kernel$ de 3×3 , $stride$ 1 y $padding$ 1 y función de activación ReLU. Cada bloque convolucional con número de canales 512, 256, 128 y 64. Finalmente, se cierra la U-Net con una convolución con $kernel$ de 1×1 y activación sigmoide.

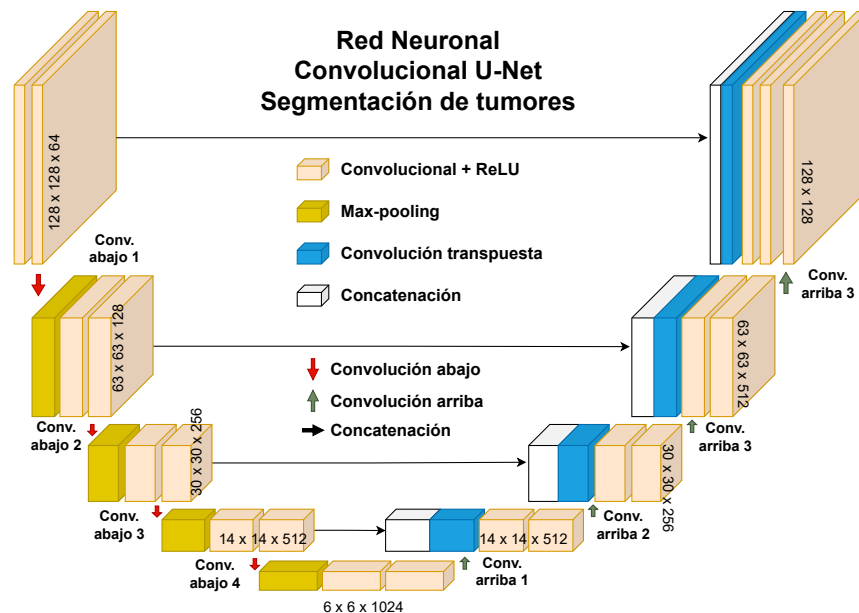


Figura 3.13: Arquitectura de la red U-Net implementada para segmentación de tumores en volúmenes MRI. La red se compone de un camino de contracción con cinco bloques convolucionales y activación ReLU, y un camino de expansión con cuatro bloques convolucionales y activación sigmoide.

Los datos de entrada fueron 120 volúmenes MRI de la base de datos BraTS divididos en 100 volúmenes para entrenamiento, 10 para validación y 10 para prueba. La red fue entrenada con la totalidad de *slices* de cada volumen (64), por lo tanto se cuenta con un total de 6400 *slices* para entrenamiento, 640 para validación y 640 para prueba. La marcación de tumores en esta base de datos está compuesta por varias etiquetas, según la zona tumoral demarcada, como se describe en Subsección 3.1.1, en este experimento se unificaron todas las etiquetas con el valor 1, con el fin de tomar la totalidad del tumor. El número total de *slices* positivos (con tumor) fue de 3317 y de negativos (sin tumor) fue de 3083. En validación 313 positivos, 327 negativos. El número total de parámetros de entrenamiento quedó en 31,030,593 dadas las dimensiones de la imagen y la arquitectura de la red.

Después de realizadas varias pruebas probando con diferentes valores, el entrenamiento de la red U-Net para segmentación de tumores se llevó a cabo con la siguiente configuración de hiperparámetros:

- Dimensiones del modelo: (128, 128, 1)
- Tasa de aprendizaje: 0.0001
- Numero de épocas: 20
- Tamaño de *batch*: 1
- Umbral de decisión: 0.1

La tasa de aprendizaje se inicializó en 0.0001 y actualiza su valor cada 5 épocas por medio de una función que multiplica el valor actual por un factor de 0.1. El optimizador elegido fue el optimizador Adam descrito en la Subsección 2.3.4. Como función de costo se eligió implementar una combinación de BCE, descrita en la Subsección 2.3.5, con la función de costo basada en el coeficiente de similitud Sørensen-Dice descrito en la Subsección 2.4.2 y dada por la ecuación (2.26), siendo el costo total la suma del valor de ambas funciones como se muestra en la ecuación (3.2). El tamaño de *batch* fue establecido en un valor de 1, ya que el coeficiente Sørensen-Dice no mostró buen comportamiento cuando se cargaron lotes más grandes de datos, dado el desbalance de clases.

$$L = BCE + L_{Dice} \quad (3.2)$$

3.3.2. RESULTADOS DE LA U-NET EN SEGMENTACIÓN DE TUMORES

El algoritmo de segmentación de tumores fue evaluado mediante el coeficiente de similitud Sørensen-Dice de la Subsección 2.4.2, implementado como métrica, donde un valor de cero es una total diferencia entre la marcación manual y la predicción, mientras que un valor de uno es una total correspondencia. Estas medidas se tomaron en la etapa de entrenamiento, validación y prueba. Adicionalmente, en entrenamiento se monitorizó el descenso de la función de costo.

3.3.2.1. ENTRENAMIENTO

El valor de la función de costo BCE+Dice se tomó en cada *batch* y se promedió por época de entrenamiento. Como puede verse en la Figura 3.14, la curva que describe es descendente para las primeras seis épocas, sin embargo, se estabiliza aproximadamente a partir de la sexta época de entrenamiento, oscilando en valores cercano a 0.4.

En entrenamiento también se tomó medida del coeficiente de similitud Sørensen-Dice cuyos resultados mostramos en la Figura 3.15. Como se esperaba, esta curva se comporta de forma similar, aunque de forma inversa a la curva del costo, mostrando una curva ascendente pero estabilizándose aproximadamente en la época cinco.

3.3.2.2. VALIDACIÓN

En validación, etapa que se ejecutó siempre al final de cada época de aprendizaje, usando dos volúmenes independientes de pacientes con lesión de FCD, se tomó medida del coeficiente de similitud Sørensen-Dice. Los resultados se muestran en la Figura 3.16, la curva es levemente ascendente, iniciando cerca de 0.2 y estabilizándose en valores cercanos a 0.6, indicando un 60% de similitud entra la máscara segmentada por un experto y la predicción del modelo.

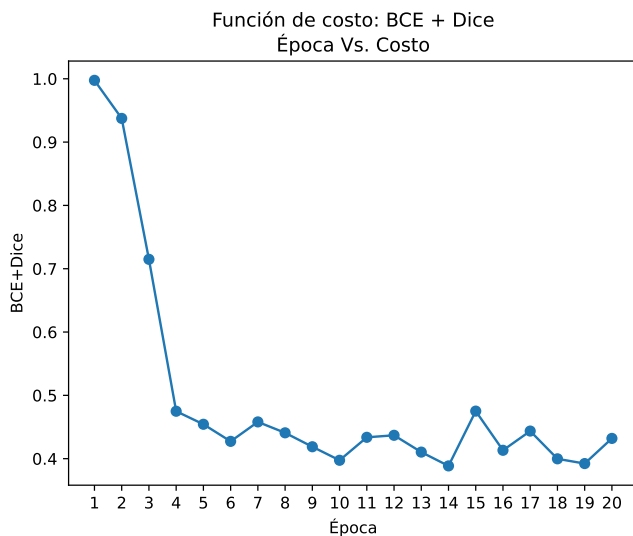


Figura 3.14: Valor promedio por época de la función de costo en entrenamiento. La pendiente de la curva es descendente en las primeras épocas, pero se estabiliza a partir de la quinta con valores entre 0.5 y 0.4.

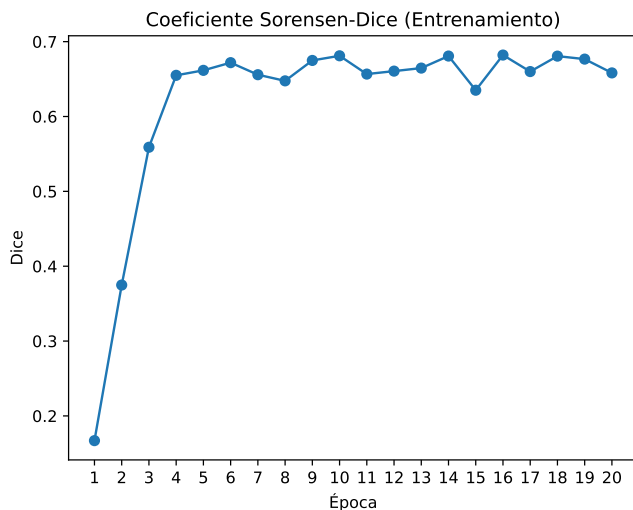


Figura 3.15: Valor promedio por época del coeficiente de similitud Sørensen-Dice en entrenamiento. Se aprecia un comportamiento ascendente de la métrica de evaluación en las primeras cinco épocas, iniciando en valores menores a 0.2 y estabilizándose en 0.65 en promedio, aproximadamente.

3.3.2.3. PRUEBA

Finalmente, cargando el modelo de segmentación de tumores entrenado por la red U-Net, se realizaron las predicciones sobre los datos de prueba y se obtuvo un coeficiente de similitud Sørensen-Dice promedio de 0.413, evaluado sobre los dos volúmenes específicamente dedicados a esta etapa de evaluación. En la Figura 3.17 se muestran los resultados de segmentación en la base de datos de prueba. El algoritmo logra acercarse a la forma y área de la marcación original e incluso, en los casos donde se presenta más de una región tumoral, puede segmentar ambas regiones.

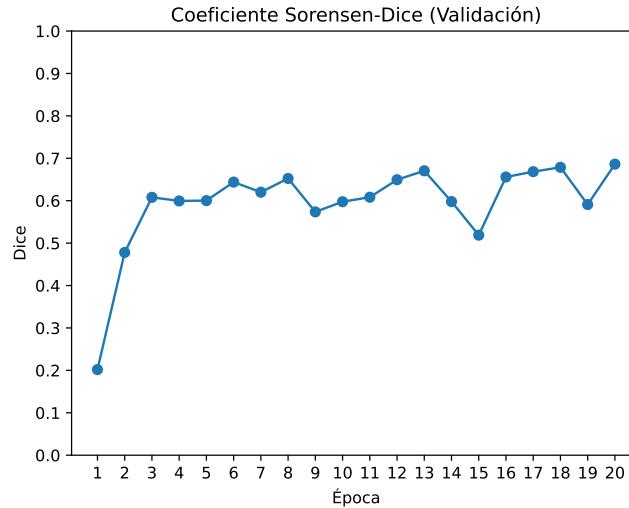


Figura 3.16: Valor promedio por época del coeficiente de similitud Sørensen-Dice en validación. La pendiente de la curva es empinada y ascendente en las primeras épocas, pero se estabiliza en valores cercanos a 0.6.

En los resultados también se encontraron casos de segmentación que no fueron satisfactorios, estos se muestran en la Figura 3.18. Por ejemplo, pudimos observar predicciones en las que el algoritmo no pudo encontrar la región tumoral y por lo tanto no devuelve ninguna marcación. Se presentó también el caso contrario, en que el volumen no tiene ningún tumor según la marcación del experto pero la red sí predice región tumoral, o donde solo existe un tumor pero la red encuentra varios de ellos. Los casos de falsos positivos, pueden servir al personal de diagnóstico para tener en cuenta patologías que pueden ser invisibles a simple vista.

Los códigos para este algoritmo pueden encontrarse en su repositorio de GitHub [repositorio de GitHub](#).

3.3.3. SEGMENTACIÓN DE LESIONES FCD EN IMÁGENES MRI

En esta segunda etapa, se carga el modelo de segmentación de tumores, entrenado en la Subsección 3.3.1, para inicializar los pesos de las capas de una red U-Net, con idéntica arquitectura: camino de contracción compuesto por cinco bloques convolucionales y el camino de expansión de cuatro bloques convolucionales con activación sigmoide.

Luego de instanciar la red U-Net con la misma arquitectura de la Subsección 3.3.1, se congelaron todas sus capas y se cargaron los pesos del modelo de segmentación de tumores. Como capas entrenables, quedaron solo la última convolución hacia arriba, seguida de dos convoluciones con *kernel* de 3×3 y activación ReLU y la convolución final con *kernel* de 1×1 y activación sigmoide. El número total de parámetros de entrenamiento de la red U-Net fue de 31,030,593, de los cuales 143,617 quedaron como parámetros entrenables. La arquitectura de la red se muestra en la Figura 3.19.

Para determinar el número de capas entrenables, se efectuaron varias sesiones de entrenamiento, con una muestra menor de volúmenes MRI, monitorizando la función de costo, el coeficiente de similitud Sørensen-Dice y haciendo un análisis visual de las predicciones. Primero se entrenó solo con

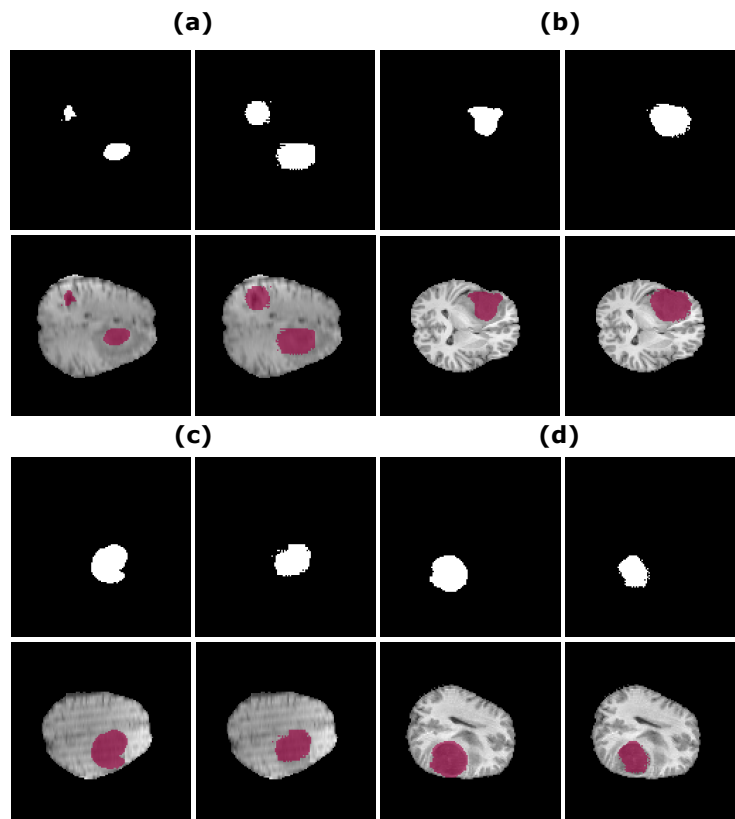


Figura 3.17: Resultados de la red U-Net en segmentación de tumores para cuatro sujetos en la base de datos de prueba. La columna izquierda de las imágenes contiene la marcación *ground-truth*, la columna de la derecha contiene las predicciones, arriba se muestran las máscaras y abajo la superposición de la máscara sobre la MRI original. (a) El algoritmo logra segmentar las dos regiones con tumor. (b), (c) y (d) El algoritmo logra acercarse a la forma y área de la marcación original.

la última capa convolucional, dejando congeladas el resto de capas y luego se incrementó paulatinamente el número de capas entrenables hasta llegar a una cantidad de siete. El mejor desempeño se dio entrenando las cuatro últimas capas, a partir de ahí, si se aumentaba el número de capas, el desempeño disminuía, aumentando el número de falsos positivos.

El resto de los hiperparámetros de entrenamiento de la red U-Net para detección de lesiones de FCD fueron los siguientes:

- Dimensiones del modelo: (128, 128, 1)
- Tasa de aprendizaje: 0.0005
- Número de épocas: 20
- Tamaño de *batch*: 4
- Umbral de decisión: 0.1
- Número de capas entrenables: 4

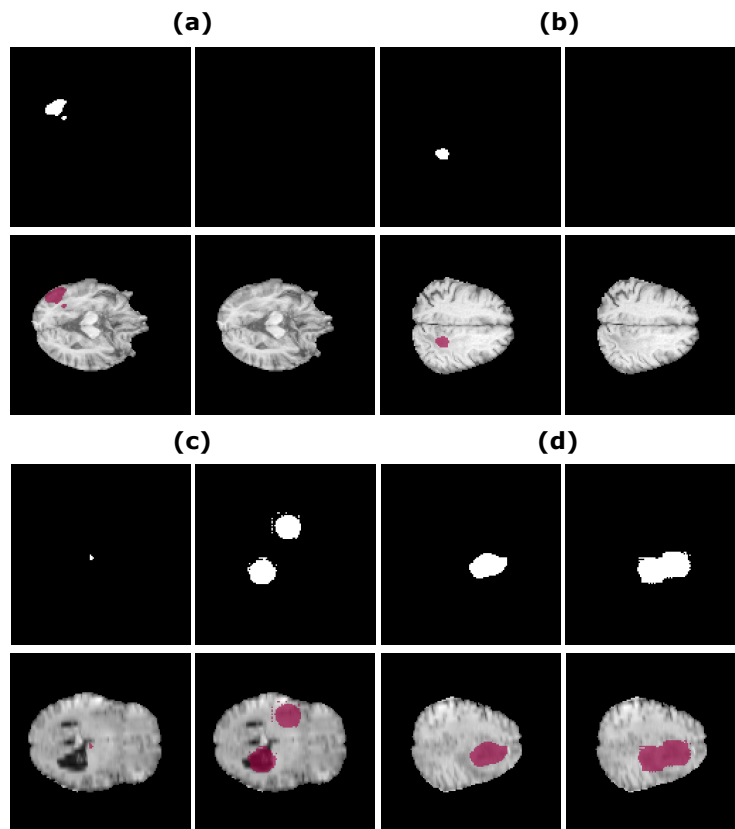


Figura 3.18: Resultados no satisfactorios, para cuatro diferentes sujetos, de la red U-Net en segmentación de tumores en la base de datos de prueba. La columna izquierda de las imágenes contiene la marcación *ground-truth*, la columna de la derecha contiene las predicciones. Tanto en (a) como en (b) el modelo no puede encontrar la región tumoral por lo tanto la predicción es solo fondo. En (c) el modelo detecta dos regiones con tumor a pesar de que solo existe una pequeña zona demarcada como tumor. Y en (d) el modelo predice una zona tumoral más grande que la marcación original.

Se usó una tasa de aprendizaje constante que se inicializó en 0.001, valor elegido por tener mejor desempeño luego de hacer pruebas con valores de diferentes que iban desde 0.005 hasta 0.0001. Como función de costo se implementó la modificación del coeficiente de similitud *Sørensen-Dice* de la Subsección 2.4.2, con la ecuación (2.26). Finalmente, se usó el optimizador Adam descrito en la Subsección 2.3.4 para calcular los gradientes de la función de costo.

Los volúmenes para entrenar este modelo fueron tomados de la base de datos de pacientes con lesiones de FCD de SURA descrita en la Subsección 3.1.2. Los 23 volúmenes fueron divididos en una base de datos de entrenamiento conformada por 19 volúmenes, otra base de datos de validación con 2 volúmenes y la base de datos de prueba de 2 volúmenes. Para entrenamiento, se tomaron solo los *slices* que tuviesen lesiones de FCD ya que su área es tan pequeña que entrenar con *slices* donde no existía lesión solo tendía a sesgar el algoritmo, aún más, a predecir todos los píxeles en el *slice* como fondo. Por lo tanto, para entrenamiento se contó con 113 *slices*, todos con presencia de lesión de FCD. Para validación y prueba se contó con la totalidad de *slices* de los volúmenes de estas bases de datos para un total 128 *slices* cada una.

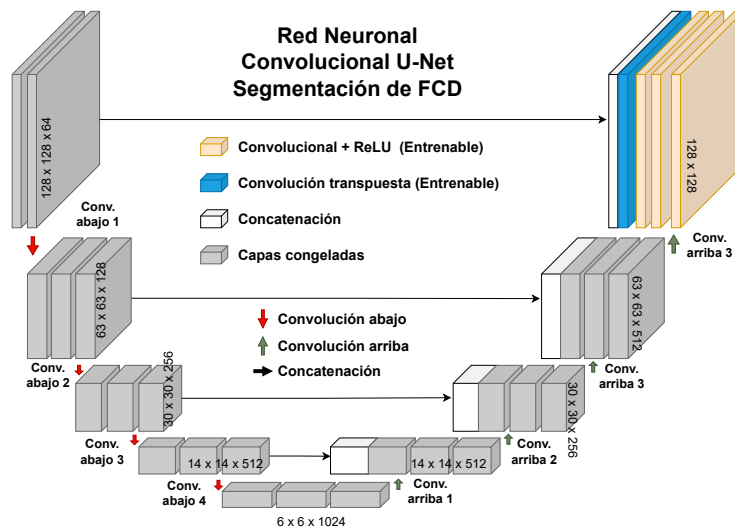


Figura 3.19: Arquitectura de la red U-Net implementada para segmentación de lesiones FCD a nivel de *slice* en volúmenes MRI. Solo las cuatro últimas capas convolucionales del camino de expansión fueron entrenadas, las demás fueron inicializadas con los pesos del modelo de segmentación de tumores.

3.3.4. RESULTADOS DE LA U-NET EN SEGMENTACIÓN DE LESIONES DE FCD

Los resultados del algoritmo de segmentación de lesiones de FCD fueron evaluados usando el coeficiente de similitud Sørensen-Dice de la Subsección 2.4.2, implementado como métrica, en las etapas de entrenamiento, validación y prueba. También se tomaron medidas de exactitud, sensibilidad y precisión. Asimismo, se realizó una inspección visual para comprobar que los resultados numéricos si se corresponden con los objetivos que se esperaba obtener en la sección.

3.3.4.1. ENTRENAMIENTO

En la etapa de entrenamiento se usaron 113 *slices* con lesión de FCD de los 19 volúmenes del dataset SURA de la Subsección 3.1.2. Solo se entrenó con *slices* donde existiera lesión ya que luego de realizar pruebas con la totalidad de *slices* e incluso tomando solo una pequeña muestra de *slices* sin lesión, los resultados fueron inferiores que entrenando solo con *slices* que tuvieran lesión, lo cual puede atribuirse al alto desbalance de clases. Para cada *batch* se tomó el valor del coeficiente Sørensen-Dice como función de costo y se promedió por época de entrenamiento. Como puede verse en la Figura 3.20, la función de costo describe una curva que es descendente durante todas las épocas de entrenamiento, sin embargo, la magnitud del descenso fue de 0.128, medida como la diferencia entre el costo final y el costo inicial.

En entrenamiento también se tomó medida del coeficiente de similitud Sørensen-Dice cuyos resultados mostramos en la Figura 3.21. En este caso el modelo no se comporta como se espera ya que la métrica se mantiene en valores cercanos a cero mientras que visualmente se observa que el modelo ya segmenta lesiones de FCD, aunque con presencia de algunos falsos positivos. La baja magnitud del coeficiente Sørensen-Dice puede deberse a que es un problema con un alto desbalance de clases, donde la mayoría de píxeles constituyen el fondo y solo una pequeña cantidad corresponde a la lesión. Adicional al coeficiente Sørensen-Dice se tomaron medidas de exactitud, sensibilidad y precisión.

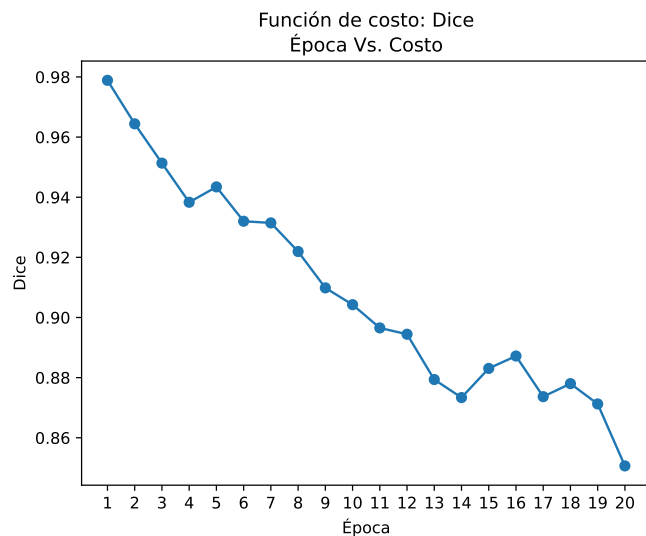


Figura 3.20: Valor promedio por época del coeficiente de similitud Sørensen-Dice, como función de costo, en entrenamiento. La pendiente de la curva es descendente a lo largo de todas las épocas de entrenamiento.

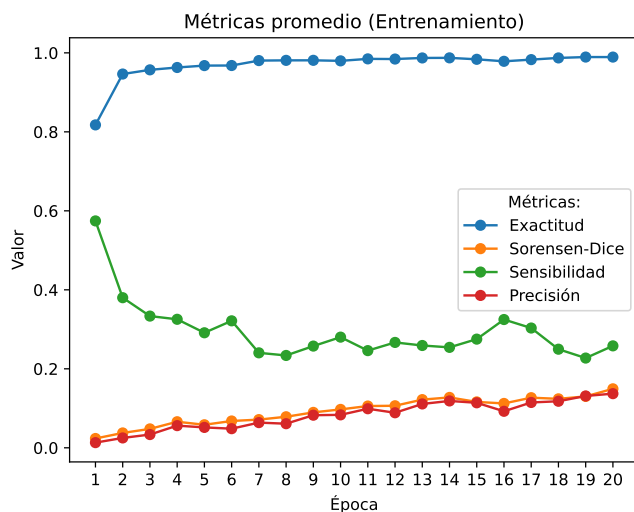


Figura 3.21: Valor promedio por época del coeficiente de similitud Sørensen-Dice, exactitud, sensibilidad y precisión en entrenamiento. Todas las métricas, excepto sensibilidad, muestran una leve pendiente ascendente lo que sugiere un aprendizaje aunque la magnitud sea baja.

3.3.4.2. VALIDACIÓN

Al final de cada época de aprendizaje se ejecutó la etapa de validación, usando dos volúmenes de pacientes con lesión de FCD. Se tomó medida del coeficiente de similitud Sørensen-Dice, exactitud, sensibilidad y precisión con resultados similares al entrenamiento. De nuevo el desbalance de clases implica unos valores muy bajos en la magnitud de las métricas, excepto por la exactitud, que muestra valores debido a que el área del fondo es mucho mayor que el área de la lesión de FCD. Sin embargo, la curva descrita por estas métricas no parece que sirva como un estimativo del nivel de aprendizaje del algoritmo, ya que se puede apreciar la aparición de la lesión de FCD en la inspección visual. Los

resultados se muestran en la Figura 3.22.

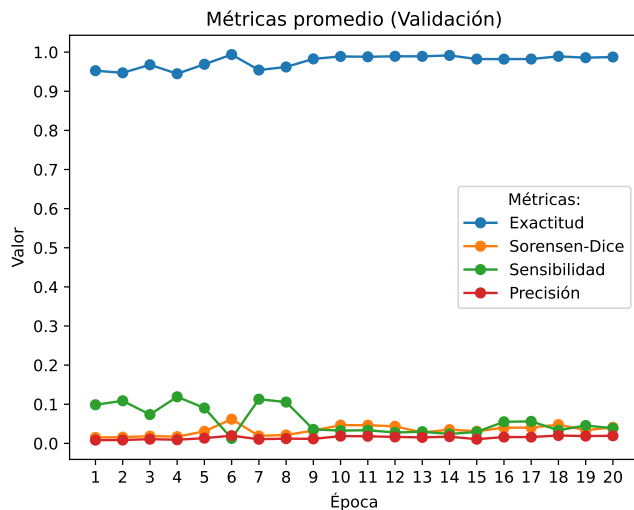


Figura 3.22: Valor promedio por época del coeficiente de similitud Sørensen-Dice, exactitud, sensibilidad y precisión en la etapa de validación. Como en el entrenamiento, las métricas muestran una baja magnitud, excepto la exactitud que muestra valores altos dada la preponderancia que tiene el área de fondo sobre el área de la lesión.

En la Figura 3.23 se muestran los resultados de segmentación de la lesión de FCD para el sujeto 1 de la base de datos de validación. En ella se ve que la lesión de FCD aparece en la mayoría de *slices*, llegando a aproximarse al área total de la marcación original en la imagen (d) y (e). También es de resaltar que el modelo detecta la lesión de FCD en *slices* consecutivos del volumen, aunque erosionada en los *slices* iniciales y finales donde la lesión se hace más pequeña. También puede verse que existe una cantidad alta de falsos positivos, que tienden a desaparecer con un umbral de decisión más alto como se muestra en las imágenes sobrepuestas.

Los resultados para el sujeto 2 de los volúmenes de validación se muestran en la Figura 3.24. Similar a los resultados del modelo en el sujeto 1, la lesión de FCD aparece en la mayoría de *slices*, con un área que tiende a aproximarse al área de la marcación original, especialmente en los *slices* donde se ubica el centro de la lesión de FCD (figuras (d) y (e) en la imagen). También aparece la lesión de FCD en *slices* consecutivos del volumen y presencia de falsos positivos, con un poco más de erosión en los *slices* iniciales y finales.

3.3.4.3. PRUEBA

El modelo de segmentación de lesiones de FCD entrenado por la red U-Net se evaluó en la base de datos de prueba, conformada por dos volúmenes en secuencia T1 de dos sujetos con lesión de FCD. Se obtuvo un coeficiente de similitud Sørensen-Dice promedio de 0.079 evaluado sobre cada *slice* de ambos volúmenes, lo cual, como se ha venido señalando, es una causa de la desproporción entre píxeles de fondo y píxeles con lesión de FCD.

En la Figura 3.25 se muestran *slices* consecutivos de los resultados de segmentación del volumen del sujeto 1 de la base de datos de prueba. En este caso el modelo detecta la lesión de FCD y continua

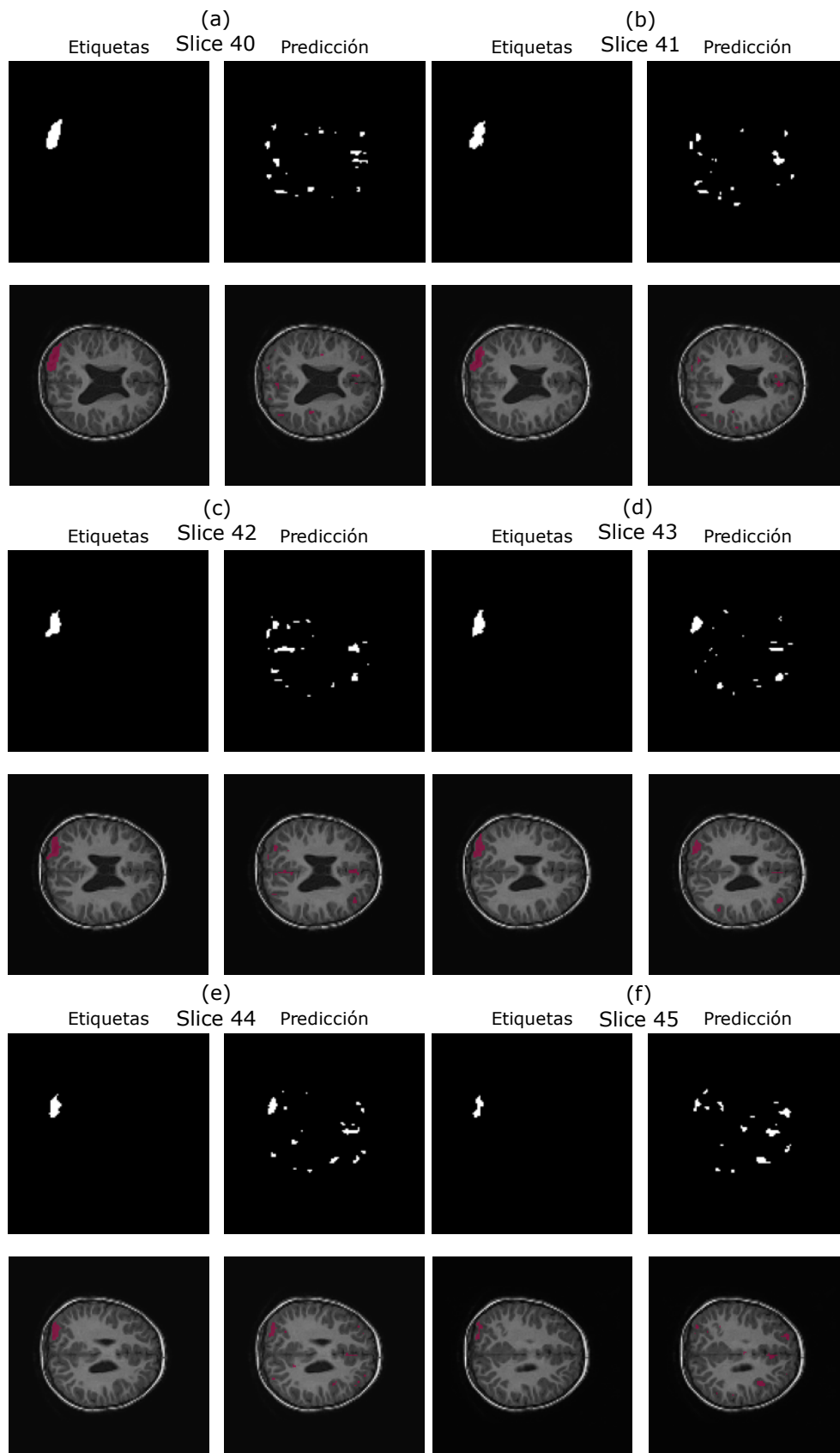


Figura 3.23: Las figuras (a)-(f) contienen, en la fila superior, las etiquetas y predicciones del modelo, y en la fila inferior la superposición de las etiquetas y de las predicciones, respectivamente, sobre la MRI original del primer sujeto de prueba con lesión de FCD.

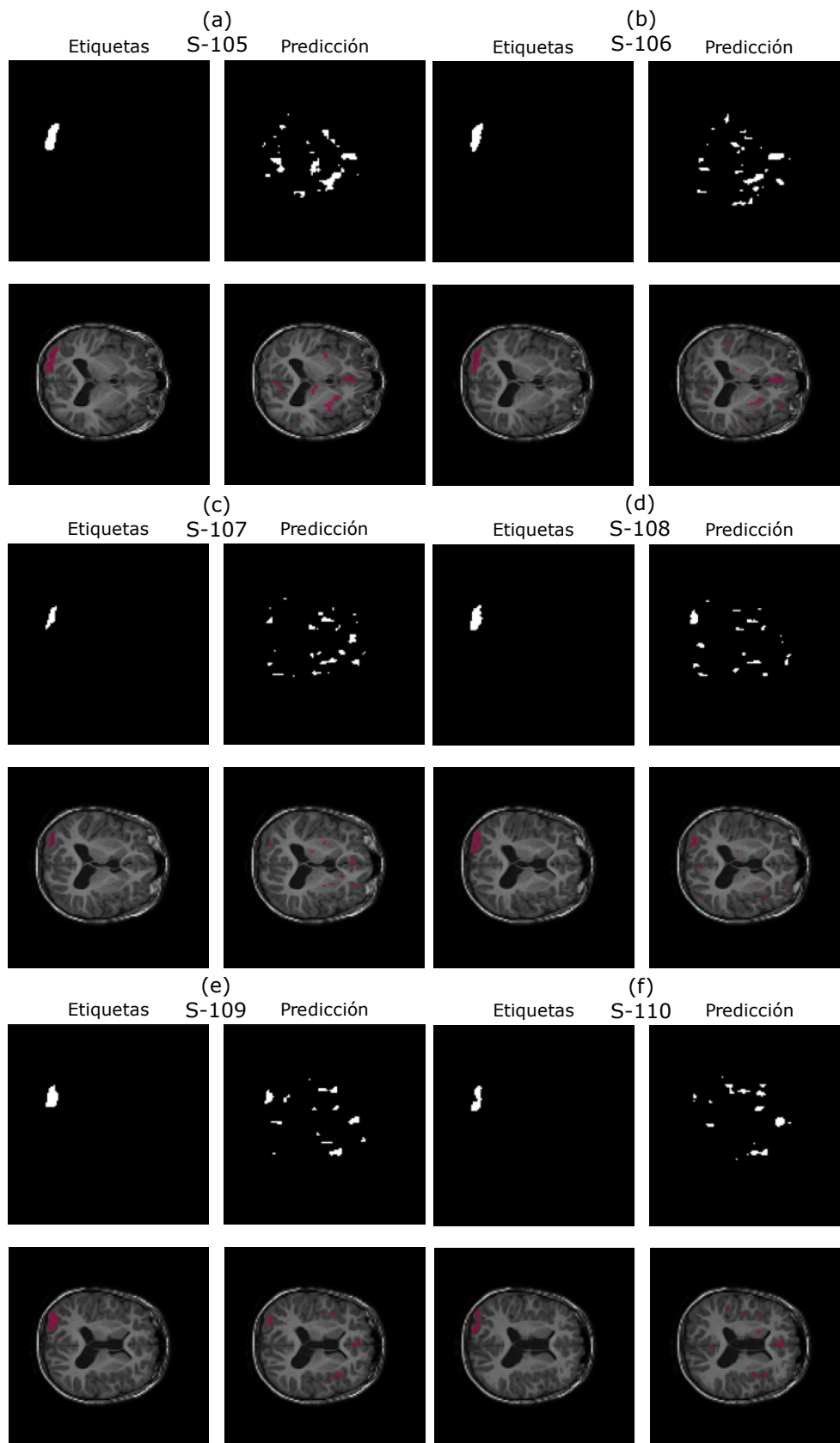


Figura 3.24: Las figuras (a)-(f) contienen, en la fila superior, las etiquetas y predicciones del modelo, y en la fila inferior la superposición de las etiquetas y de las predicciones, respectivamente, sobre la MRI original del segundo sujeto de prueba con lesión de FCD.

mostrándola en *slices* consecutivos lo cual es una buena característica del modelo entrenado. Sin embargo, la lesión aparece en una proporción menor de *slices* consecutivos que en la etapa de validación y en algunos casos, como en la imagen (c), parece confundirse con una región de falsos positivos, los cuales aparecen con un área aparentemente mayor.

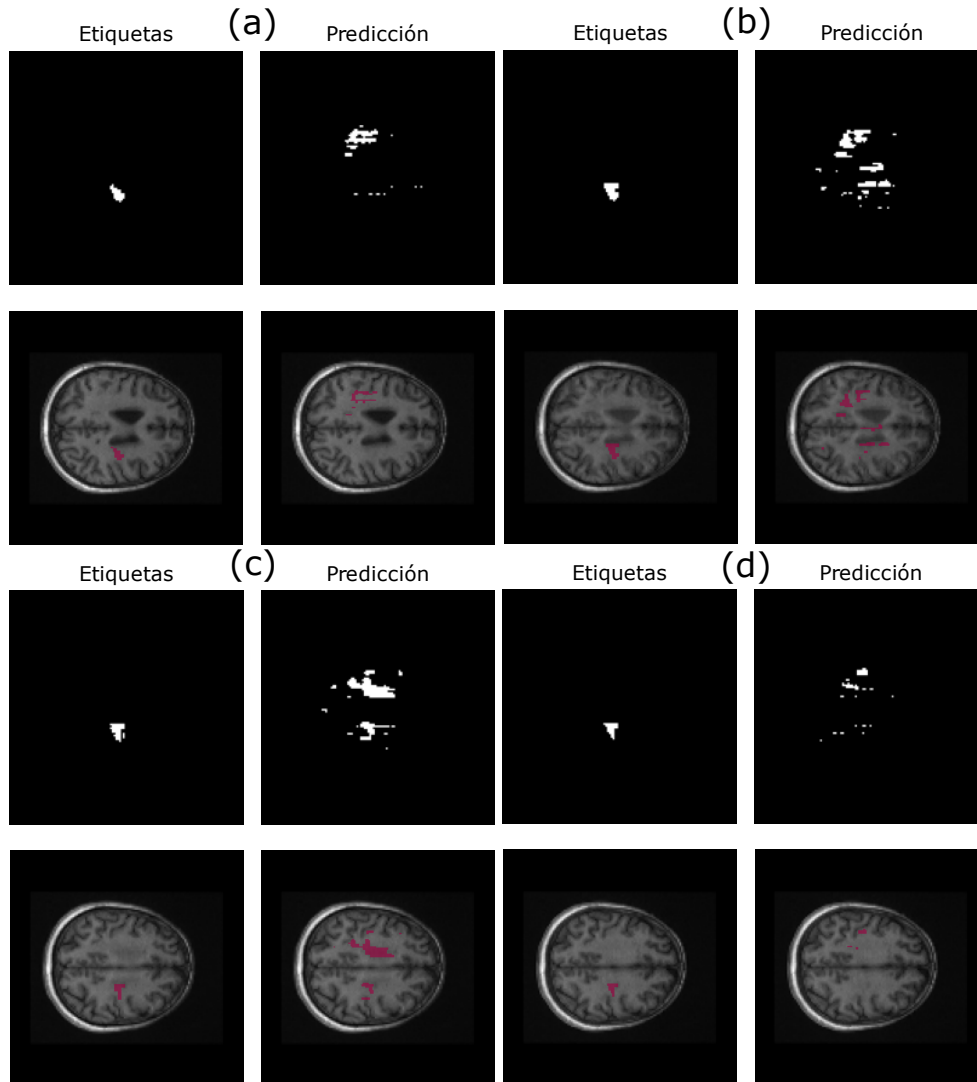


Figura 3.25: Predicciones del modelo para el sujeto 1 de la base de datos de prueba. En las imágenes centrales (c) y (d) aparece la lesión de FCD. Los *slices* donde se encuentran las regiones externas de la FCD, (a) y (d) las predicciones del modelo no corresponden con la marcación del experto.

En el segundo sujeto la lesión de FCD aparece con más frecuencia, y en *slices* consecutivos, en las predicciones del modelo, como se muestra en la Figura 3.26. Las imágenes (b) y (c) muestran fluctuaciones en el área delineada, sin embargo la ubicación de la lesión en el volumen es la esperada y se muestra una tendencia ascendente en el tamaño del área. En este caso se destaca que el modelo no entrega falsos positivos, lo que es un buen indicativo de que en el modelo sí existió entrenamiento, aunque las métricas de evaluación lo contradigan.

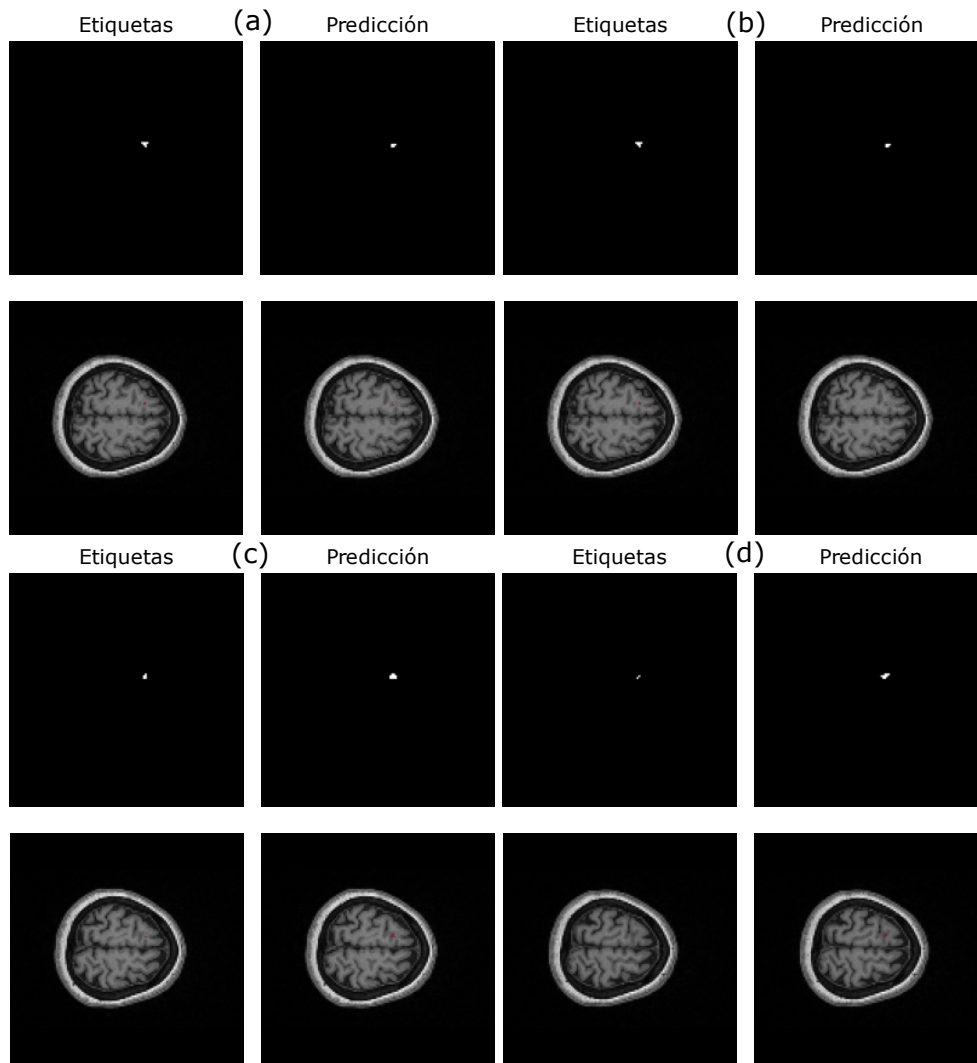


Figura 3.26: Predicciones del modelo para el sujeto 2 de la base de datos de prueba. En este caso la lesión de FCD aparece en todas las predicciones en la ubicación donde se espera, si bien en algunos casos, como en (c) y (d), con área mayor a la marcación del experto. En estas predicciones resalta la ausencia de falsos positivos en la imagen, en especial tratándose de una lesión particularmente pequeña.

Los códigos para este algoritmo pueden encontrarse en su repositorio de GitHub [repositorio de GitHub](#).

REFERENCIAS

- [1] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, N Porz, Johannes Slotboom, Roland Wiest, Levente Lenczi, Elizabeth R. Gerstner, Marc-André Weber, Tal Arbel, Brian B. Avants, Nicholas Ayache, Patricia Buendia, D. Louis Collins, Nicolas Cordier, Jason J. Corso, Antonio Criminisi, Tilak Das, Hervé Delingette, Çağatay Demiralp, Christopher R. Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan M. Iftekharuddin, Raj Jena, Nigel W. John, Ender Konukoglu, Danial Lashkari, Jose Mario Mariz, Raphael P. H. Meier, Sergio Luiz Pereira, Doina Precup, Stephen J. Price, Tammy Riklin Raviv, Syed Azer Reza, Michael J. Ryan, Duygu Sarikaya, Lawrence B. Schwartz, Hoo-Chang Shin, Jamie Shotton, Carlos A. Silva, Nuno Sousa, Nagesh K. Subbanna, Gábor Székely, Thomas N. Taylor, Owen R.T. Thomas, Nicholas J. Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes, and Koen Van Leemput. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 10 2015. doi: 10.1109/tmi.2014.2377694. URL <https://doi.org/10.1109/tmi.2014.2377694>.
- [2] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4(1), 9 2017. doi: 10.1038/sdata.2017.117. URL <https://www.nature.com/articles/sdata2017117.pdf>.
- [3] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018.

4

DISCUSIÓN Y CONCLUSIONES

Índice

4.1 Discusión	84
4.2 Conclusiones	85
4.3 Trabajo Futuro	87
Referencias	88

4.1. DISCUSIÓN

En esta investigación se pretendió evaluar la posibilidad de utilizar técnicas de aprendizaje profundo para el diagnóstico de pacientes con epilepsia refractaria utilizando volúmenes MRI. Este es un problema que, ante la falta de bases de datos abiertas con volúmenes MRI de pacientes con lesiones de FCD delineadas por expertos, ha obligado a buscar alternativas que puedan suplir esta ausencia y crear modelos que se ajusten a la tarea. Dentro de estos métodos alternativos se encontró el entrenamiento por parches, entrenamiento por medio de *slices* 2D y simulación de las lesiones de FCD en MRI y, de manera transversal a estos, el uso de métodos de visión por computadora como son rotaciones y filtros, entre otros.

En el entrenamiento de algoritmos con parches para detección y segmentación de FCD, se divide bien sea el volumen para crear parches 3D [1–5] o cada *slice* para crear parches 2D [6–10] del volumen en segmentos de imagen más pequeños. Para citar un ejemplo, usando un enfoque por parches, partiendo de 10 volúmenes de pacientes con FCD y 20 volúmenes de pacientes de control, Wang et al. [7] logran obtener 2302 parches con lesión y 127478 sin lesión y adicionalmente, usando técnicas de visión por computadora para aumento de datos, logran balancear la diferencia entre parches con y sin lesión, obteniendo un coeficiente Sørensen-Dice de 0.78 y exactitud de 0.88.

Aunque resulta poco habitual, dada la poca cantidad de volúmenes con que se suele contar, [11] utilizan el volumen completo en el entrenamiento de los modelos, sin embargo, en su mayoría este enfoque fue implementado en métodos como VBM y en modelos de aprendizaje de máquina como SVM que conllevan menor gasto en memoria ya que, en el primer caso se carga cada volumen individualmente y en el segundo solo se carga el volumen para extraer características. Dado que aparte

de reducir el consumo de memoria, sirve como método de aumento de datos, es más común el entrenamiento de algoritmos de aprendizaje profundo mediante *slices* 2D del volumen MRI en artículos como [6–8, 12, 13], entre otros. Los *slices* son tomados tras seleccionar un eje del volumen y extraer la imagen que queda conformada por los píxeles de los otros dos ejes, pasando de entrenar con un solo volumen a entrenar con una cantidad de *slices* cuyo valor depende del eje seleccionado. Estos factores fueron fundamentales para la elección de entrenamiento de los algoritmos de esta investigación ya que mediante *slices* 2D se pudo observar que, en efecto, los algoritmos se ven favorecidos por este enfoque.

Una posible solución a la ausencia de bases de datos, que hasta ahora ha sido poco explorada, es la simulación de volúmenes con lesiones de FCD. En el único estudio encontrado en la literatura, Feng et al. [12] parten de una de las manifestaciones más comunes de FCD en MRI como es la hiperintensidad en la frontera GM-WM y, usando métodos de visión por computadora, seleccionan regiones en el volumen a las que incrementan su intensidad en píxel por un factor multiplicativo para generar nuevos volúmenes. Los autores del estudio publican una tasa de 100% de clasificación que sugiere que hubo sobre-entrenamiento. Sin embargo, esta fue una solución considerada para este trabajo, empleando redes generativas para producir volúmenes MRI sintéticos con lesión de FCD y que fue descartada ante la necesidad de contar con una mano experta que realizara la evaluación de la simulación.

Los resultados obtenidos por los trabajos usando los enfoques que se acaban de exponer, motivaron la elección de entrenamiento por *slices* y de entrenamiento por transferencia de aprendizaje como forma de revertir el problema de ausencia de datos y como aporte novedoso a la solución del problema de detección y segmentación de FCD. Si bien no es un método de aumento de datos, se buscó remediar la ausencia de estos resolviendo un problema similar que contara con una base de datos apropiada, para posteriormente transferir los parámetros aprendidos al problema en cuestión.

Es difícil realizar una comparación directa de los resultados obtenidos por los modelos anteriormente mencionados con los obtenidos en este trabajo, dado que cada trabajo cuenta con su propio conjunto de datos fruto de su propio protocolo de adquisición de MRI y la diversidad de métricas usadas como exactitud, sensibilidad, coeficiente Sørensen-Dice, entre otras. Adicionalmente, se observó diferencias en la definición de algunas métricas como la exactitud. En el caso puntual del coeficiente de similitud Sørensen-Dice, no es claro en la literatura cuáles son valores aceptables de esta métrica, ya que no existe una medida del desempeño humano en términos de coeficiente Sørensen-Dice. Sin embargo, la evaluación visual de las imágenes segmentadas y la exactitud del modelo de detección, sugieren que el enfoque usado en este trabajo puede, en combinación con otros métodos de aumento de datos, llegar a dar una solución al problema de diagnóstico prequirúrgico de lesiones de FCD en volúmenes MRI.

4.2. CONCLUSIONES

A partir de técnicas de aprendizaje profundo se desarrolló una metodología que permite la detección y segmentación semántica de displasias corticales focales. Se logró crear una herramienta que integra los modelos de detección y segmentación y cuyo propósito es que sea un apoyo para el personal radiológico en el diagnóstico prequirúrgico de lesiones de FCD en imágenes MRI. La herramienta está en capacidad detectar y delinear la zona de las lesiones de FCD en volúmenes MRI en secuencia T1, secuencia que hace parte del protocolo estándar en el tratamiento de dichos pacientes y por lo tanto no requiere de gastos adicionales.

Dado que para los algoritmos de detección y de segmentación de FCD se empleó una base de datos retrospectiva, es decir, que recopila volúmenes MRI de tomas realizadas en fechas anteriores al planteamiento y desarrollo de este trabajo, no hubo uniformidad en los datos de entrada al modelo, ya que estos fueron tomados en tiempos diferentes, con distintos escáneres y siguiendo protocolos diferentes. Lo anterior implicó largos procesos de análisis de las imágenes y la elaboración de una compleja etapa de preprocesamiento que se puede evitar, bien creando un protocolo para la toma de datos a futuro o un sistema de preprocesamiento estandarizado.

Para el desarrollo de la metodología planteada se aplicó la técnica de transferencia de aprendizaje que, mediante transferencia de parámetros aprendidos en la solución de un problema similar al que se desea resolver, permite inicializar los pesos del nuevo modelo y así darle un impulso al aprendizaje de la nueva tarea. Se buscó que los modelos preentrenados estuvieran dirigidos a solucionar un problema similar a la detección/segmentación de FCD, con una mayor cantidad de datos disponibles, en bases de datos abiertas, para el entrenamiento de los modelos. En este caso, el problema similar elegido como punto de partida, fue la detección y segmentación de tumores cerebrales. Estos modelos preentrenados posteriormente fueron sintonizados para resolver la tarea de detección y segmentación de FCD, respectivamente.

La integración de los modelos de detección y segmentación de FCD en un software, fue pensado para que sea una herramienta de apoyo en el diagnóstico de lesiones de FCD y en ningún momento su intención es reemplazar al personal capacitado para esta compleja tarea. El diseño fue pensado para que las componentes del software pudieran ser empleadas de manera flexible, es decir, el usuario puede ejecutar ambos modelos en cascada, primero detectando si hay lesión o no en el volumen y, si la hay, realizar la segmentación de la lesión. También se puede usar cada modelo de forma individual, cuando solo se desea saber si en el volumen hay displasia, o si únicamente interesa delinear la región que el modelo prediga como lesión de FCD.

En este orden de ideas, era muy importante que cada modelo fuera tan preciso como fuera posible, por lo que se toma la decisión de realizar la evaluación de cada modelo individualmente. Esto permitió aprovechar al máximo los datos de FCD que se tenían disponibles, ya que en el modelo integrado la etapa de detección ignora los *slices* sin malformación, dejando solo una pequeña cantidad a la siguiente etapa que es la de segmentación. Asimismo permitió una mejor monitorización de cada etapa de entrenamiento y evaluación de tal manera que la transferencia de aprendizaje se ejecutara de forma concisa y específica.

Una gran limitante para el entrenamiento de algoritmos de aprendizaje cuando las clases se encuentran muy desbalanceadas, es encontrar una función de costo adecuada, donde no se penalice la clase con menor cantidad de muestras y que no permita que el modelo se “relaje” y se concentre solo en la clase más recurrente. Esta falencia se notó, especialmente, en el entrenamiento de los algoritmos de detección y segmentación de FCD, en los que se evidenció que el modelo tiene un sesgo hacia la clase mayoritaria (fondo) y, aunque las métricas muestran un valor alto y sugieren un buen entrenamiento, en realidad lo que ocurre es que la clase que tiene menos ejemplos, como es el caso de las lesiones de FCD y otras patologías, se vuelve irrelevante para el modelo ya que tiene menos peso numéricamente.

4.3. TRABAJO FUTURO

Dada la complejidad del problema que implica la detección y segmentación de FCD en MRI, este es un campo de investigación abierto donde se pueden probar otros modelos preentrenados, entrenamiento con mayor cantidad de datos o por medio de técnicas de aumento de datos usadas en visión por computadora, nuevas técnicas de aprendizaje, simulación de volúmenes, entre otros. Con esto en mente, se crearon repositorios públicos en GitHub, donde se encuentran los códigos de los algoritmos de aprendizaje de cada sesión, para que puedan servir como punto de partida de futuros trabajos.

Una limitante en el entrenamiento de modelos de segmentación de imágenes cuando las clases están altamente desbalanceadas, como en el caso de esta investigación, es la elección de una métrica adecuada que pueda usarse como función de costo para evaluar el entrenamiento y para evaluar el aprendizaje en las etapas de validación y prueba. En este trabajo se implementaron métricas de evaluación como exactitud, sensibilidad y coeficiente de similitud Sørensen-Dice, sin embargo, estas métricas mostraron falencias a la hora de evaluar los modelos debido al alto desbalance de la base de datos. Se plantea entonces como trabajo futuro indagar más en el estado del arte en busca de métricas que se adapten mejor al problema, bien sea métricas novedosas o modificar métricas ya existentes para que se adecúen, especialmente en segmentación de FCD, donde el área de la lesión es mucho más pequeña respecto del área de los tejidos sanos en el volumen MRI.

El problema de desbalance en la base de datos puede ser tratado directamente en nuevos trabajos usando técnicas de visión por computadora como rotaciones, recortes, adición de ruido, entre otros. Adicionalmente, para obtener una base de datos balanceada de pacientes con FCD, estas técnicas se pueden combinar con otras como puede ser entrenamiento con parches, extracción de *slices* del volumen MRI y simulación de lesiones de FCD en MRI. Esto bajaría la probabilidad de que los modelos sufran de sobreajuste y, adicionalmente, traería el beneficio de que las métricas de evaluación, especialmente en detección de FCD no se vean afectadas por el desbalance de datos. Sin embargo, persiste la necesidad de encontrar nuevas métricas de evaluación ante la diferencia en área existente entre el área de la malformación respecto del área de tejido sano en todo el volumen MRI. Esto, en el caso de que solo se entrene los modelos usando *slices* completos de la secuencia MRI y pudiéndose evitar en el caso de entrenamiento con parches ya que en este caso se tendría control sobre la proporción del área con lesión que aparece en la imagen.

REFERENCIAS

- [1] Chin-Ann Yang, Mostafa Kaveh, and Bradley J. Erickson. Automated detection of Focal Cortical Dysplasia lesions on T1-weighted MRI using volume-based distributional features. *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 3 2011. doi: 10.1109/isbi.2011.5872541. URL <http://dx.doi.org/10.1109/isbi.2011.5872541>.
- [2] Ravnoor S. Gill, Seok-Jun Hong, Fatemeh Fadaie, Benoit Caldaïrou, Boris C. Bernhardt, Carmen Barba, Armin Brandt, Vanessa C. Coelho, Ludovico d’Incerti, Matteo Lenge, Mira Semmelroch, Fabrice Bartolomei, Fernando Cendes, Francesco Deleo, Renzo Guerrini, Maxime Guye, Graeme Jackson, Andreas Schulze-Bonhage, Tommaso Mansi, Neda Bernasconi, and Andrea Bernasconi. Deep Convolutional Networks for Automated Detection of Epileptogenic Brain Malformations. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 490–497, 2018. doi: 10.1007/978-3-030-00931-1_56. URL http://dx.doi.org/10.1007/978-3-030-00931-1_56.
- [3] Patrick M. House, Mikhail Kopelyan, Natalia Braniewska, Bartosz Silski, Agata Chudzinska, Brigitte Holst, Thomas Sauvigny, Tobias Martens, Stefan Stodieck, and Sirko Pelzl. Automated detection and segmentation of focal cortical dysplasias (FCDs) with artificial intelligence: Presentation of a novel convolutional neural network and its prospective clinical validation. *Epilepsy Research*, 172: 106594, 2 2021. doi: 10.1016/j.eplepsyres.2021.106594.
- [4] Ravnoor S. Gill, Hyo Suk Lee, Benoit Caldaïrou, Hyo Geun Choi, Carmen Barba, Francesco Deleo, Ludovico D’Incerti, Vanessa Félix Nascimento Coelho, Matteo Lenge, Mira Semmelroch, Dewi V. Schrader, Fabrice Bartolomei, Maxime Guye, Andreas Schulze-Bonhage, Horst Urbach, Kyoo Ho Cho, Fernando Cendes, Renzo Guerrini, Graeme D. Jackson, R. Edward Hogan, and Andrea Bernasconi. Multicenter Validation of a Deep Learning Detection Algorithm for Focal Cortical Dysplasia. *Neurology*, 97(16):e1571–e1582, 9 2021. doi: 10.1212/wnl.0000000000012698.
- [5] Chin-Ann Yang, Mostafa Kaveh, and Bradley Erickson. Cluster-based differential features to improve detection accuracy of focal cortical dysplasia. *SPIE Proceedings*, 2 2012. doi: 10.1117/12.905313. URL <http://dx.doi.org/10.1117/12.905313>.
- [6] Cuixia Feng, Hulin Zhao, Yueer Li, and Junhai Wen. Automatic localization and segmentation of focal cortical dysplasia in FLAIR-negative patients using a convolutional neural network. *Journal of Applied Clinical Medical Physics*, 21(9):215–226, 8 2020. doi: 10.1002/acm2.12985. URL <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/acm2.12985>.
- [7] Huiquan Wang, Samsuddin Ahmed, and Mrinal Mandal. Automated detection of focal cortical dysplasia using a deep convolutional neural network. *Computerized Medical Imaging and Graphics*, 79:101662, 1 2020. doi: 10.1016/j.compmedimag.2019.101662. URL <https://doi.org/10.1016/j.compmedimag.2019.101662>.
- [8] Ruslan Aliev, Ekaterina Kondrateva, Maxim Sharaev, Oleg Bronov, Alexey Marinets, Sergey Subbotin, Alexander Bernstein, and Evgeny Burnaev. Convolutional Neural Networks for Automatic Detection of Focal Cortical Dysplasia. *Advances in intelligent systems and computing*, pages 582–588, 10 2020. doi: 10.1007/978-3-030-71637-0_67.
- [9] Yee-Leng Tan, Hosung Kim, Seunghyun Lee, Tarik Tihan, Lawrence Ver Hoef, Susanne G. Mueller, Anthony James Barkovich, Duan Xu, and Robert Knowlton. Quantitative surface analysis

- of combined MRI and PET enhances detection of focal cortical dysplasias. *NeuroImage*, 166: 10–18, 2 2018. doi: 10.1016/j.neuroimage.2017.10.065. URL <http://dx.doi.org/10.1016/j.neuroimage.2017.10.065>.
- [10] Seok-Jun Hong, Boris C. Bernhardt, Dewi S. Schrader, Neda Bernasconi, and Andrea Bernasconi. Whole-brain MRI phenotyping in dysplasia-related frontal lobe epilepsy. *Neurology*, 86(7): 643–650, 1 2016. doi: 10.1212/wnl.0000000000002374. URL <http://dx.doi.org/10.1212/wnl.0000000000002374>.
- [11] Bastian David, Judith Kröll-Seger, Fabiane Schuch, Jan Wagner, Jörg Wellmer, Friedrich Woermann, Bernhard Oehl, Wim Van Paesschen, Tobias Breyer, Albert Becker, Hartmut Vatter, Elke Hattingen, Horst Urbach, Bernd Weber, Rainer Surges, Christian Erich Elger, Hans-Jürgen Huppertz, and Theodor Rüber. External validation of automated focal cortical dysplasia detection using morphometric analysis. *Epilepsia*, 62(4):1005–1021, 2 2021. doi: 10.1111/epi.16853. URL <http://dx.doi.org/10.1111/epi.16853>.
- [12] Cuixia Feng, Hulin Zhao, Jun Zhang, Zhibiao Cheng, and Junhai Wen. Automated localization of Epileptic Focus Using Convolutional Neural Network. *International Conference Big Data Engineering and Technology*, 1 2020. doi: 10.1145/3378904.3378928.
- [13] KM Bijay Dev, Pawan S Jogi, S Niyas, S Vinayagamani, Chandrasekharan Kesavadas, and Jeny Rajan. Automatic detection and localization of focal cortical dysplasia lesions in mri using fully convolutional neural network. *Biomedical Signal Processing and Control*, 52:218–225, 2019. doi: 10.1016/j.bspc.2019.04.024. URL <https://www.sciencedirect.com/science/article/pii/S1746809419301211>.

