



Programa de Doctorado en Electrónica: Sistemas Electrónicos  
Avanzados. Sistemas Inteligentes

---

**Diagnóstico de esclerosis múltiple  
mediante aprendizaje automático  
usando datos de tomografía de  
coherencia óptica**

---

Tesis Doctoral presentada por:  
**Carlo Cavaliere Ballesta**

Dirigida por:  
**Dr. Luciano Boquete Vázquez**  
**Dr. Miguel Ortiz del Castillo**

Enero de 2022, Alcalá de Henares



## Resumen

**Objetivo:** El objetivo de esta tesis doctoral es desarrollar nuevos métodos de diagnóstico de la enfermedad de la esclerosis múltiple (EM), basados en la aplicación de procesos de aprendizaje automático sobre medidas de espesores retinales.

**Base de datos:** Se dispone de un registro compuesto por 47 sujetos de control y 56 pacientes de EM con diagnóstico precoz (menos de 9 meses tras los primeros síntomas de la enfermedad). Mediante tomografía de coherencia óptica, se obtienen datos de espesores de ambos ojos provenientes de cinco estructuras retinales: RNFL, Retina completa, GCL+, GCL++ y Coroides.

**Métodos:** Se realiza un estudio sobre el registro para establecer zonas de un mayor interés diagnóstico mediante el cálculo de AUCs (área bajo la curva) y el análisis del tamaño de efecto, estableciendo distintas modalidades de análisis: imágenes completas, ventanas cuadradas y regiones de Cohen. Esta información se utiliza como entrada de diferentes clasificadores con aprendizaje supervisado: clasificadores KNN y SVM, redes neuronales tradicionales MLP y RBF y redes profundas CNN, ResNet y V-Net.

**Resultados:** Se alcanzan exactitudes mayores de 90% al analizar imágenes completas, y cercanas a 100% para la modalidad de ventanas cuadradas y regiones de Cohen. Las estructuras que permiten obtener la mejor exactitud en el diagnóstico son la RNFL, Retina completa y GCL++.

**Conclusiones:** La alta capacidad predictiva obtenida usando técnicas de aprendizaje automático sugiere que el uso de espesores retinales como biomarcadores sí resulta de interés para el diagnóstico de la esclerosis múltiple.

**Términos clave:** Esclerosis múltiple, tomografía de coherencia óptica, aprendizaje automático, redes neuronales, espesores retinales, capacidad discriminante, AUC, tamaño de efecto.



## Abstract

**Objective:** The main aim of this doctoral thesis is to develop new methods of diagnosis for multiple sclerosis disease (MS), based on the application of machine learning processes on retinal thickness measurements.

**Database:** A clinical data registry composed of 47 control subjects and 56 MS patients with early diagnosis (less than 9 months since the disease's first symptoms) is evaluated with optical coherence tomography. Thickness measurements are obtained from both eyes and for five retinal structures: RNFL, Full retina, GCL+, GCL++ and Choroid.

**Methods:** The registry is evaluated to establish zones with high discriminatory capacity, by calculating the AUC (area under the curve) and by analyzing the effect size, hence establishing three different modes of analysis: full images, square windows and Cohen regions. This information is used as the input of classifiers based on supervised learning: KNN and SVM classifiers, traditional neural networks (MLP, RBF) and deep networks (CNN, ResNet and V-net).

**Results:** The full images modality leads to accuracies higher than 90% for some of the analyzed structures, while both square windows and Cohen regions modalities reach 100% accuracy in several structures. The structures that obtain the best results are the RNFL, Full Retina and GCL++.

**Conclusions:** The high predictive capacity obtained by applying machine learning techniques suggests that the use of retinal thicknesses as biomarkers is indeed useful for the diagnosis of multiple sclerosis.

**Keywords:** Multiple sclerosis, optical coherence tomography, machine learning, neural networks, retinal thickness, discriminatory capacity, AUC, effect size.



## Agradecimientos

En primer lugar, al Ministerio de Ciencia, Innovación y Universidades por proveer la financiación necesaria para la realización de esta tesis doctoral (proyecto de investigación DPI2017-88438-R. AEI/FEDER, EU).

En segundo lugar, a mi tutor y director, el Doctor Luciano Boquete, por la oportunidad de realizar la tesis y por su dirección a lo largo de los tres años de su duración, tanto en lo que respecta al redactado de este manuscrito como a su seguimiento, así como en la elaboración de artículos, y por sus explicaciones sobre conceptos clave que han sido necesarios para la obtención de resultados.

De igual forma, agradecer al Doctor Miguel Ortiz, codirector, por su seguimiento e ideas en lo que respecta a aspectos estructurales y estadísticos de la tesis.

Agradecer también al Doctor Rafael Barea, por su guía en las primeras fases de la organización de la base de datos y de la aplicación de procesos de aprendizaje automático.

Al Doctor Luis de Santiago, por sus recomendaciones y explicaciones de la terminología y señales asociadas al diagnóstico de la esclerosis múltiple durante los principios de mi estancia con el grupo.

Al Grupo de Investigación e Innovación Miguel Servet Oftalmología (GIMSO), el Hospital Universitario Miguel Servet de Zaragoza, y en especial, a la Doctora Elena García Martín (coordinadora de Neurooftalmología), por suministrar las bases de datos, así como disponerse para consultas realizadas sobre los registros.

Al Doctor Carlos Amo por su ayuda durante la elaboración de las diversas presentaciones a congresos y de esta memoria, así como a los otros miembros del Grupo de Ingeniería Biomédica por la buena estancia durante los años de realización de la tesis.

Por último, a mi familia, por su apoyo incondicional en los momentos difíciles de estos tres años.





## Preámbulo. Estructura de la tesis y otras publicaciones

La presente tesis doctoral, llevada a cabo en el marco del Programa de Doctorado en Electrónica: Sistemas Electrónicos Avanzados. Sistemas Inteligentes” (D441), del Departamento de Electrónica de la Universidad de Alcalá, se presenta bajo la modalidad de tesis tradicional, no en compendio de artículos. Para su realización, se han seguido las directrices generales del *Reglamento de elaboración, autorización y defensa de la tesis doctoral*, aprobado por el Consejo de Gobierno de la Universidad de Alcalá el día 28 de septiembre de 2016.

La presente memoria se ha estructurado de la siguiente manera:

- Un capítulo de **introducción**, con conceptos generales sobre los tres temas principales tratados en la tesis: enfermedad, prueba diagnóstica con la que se generan las bases de datos y métodos de procesado de estos.
- **Estado del arte**, ampliando lo visto en el capítulo introductorio y justificando la necesidad de la investigación realizada en la tesis mediante su contraste con los estudios existentes.
- **Definición e hipótesis de trabajo.**
- **Metodología** utilizada, tanto en lo que respecta a bases de datos (así como su tratamiento) como su descomposición en diferentes modalidades de estudio.
- La implementación final de las técnicas de aprendizaje automático empleadas y los **resultados** que se han obtenido mediante estas, así como una **discusión** de estos y los posibles **trabajos futuros** para dar continuidad a la tesis realizada.

A lo largo de los tres años de la tesis, se han realizado las siguientes publicaciones:

- Artículo: **Computer-aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features** (2019) Cavaliere C., Vilades E., Alonso-Rodríguez M. A. C., Rodrigo M. J., Pablo L. E., Miguel J. M., López-Guillén E., Sánchez Morla E. M. A., Boquete L., & García-Martin E. *Sensors (Switzerland)*, 19(23). DOI: 10.3390/s19235323
- Acta en congreso (póster): **Analysis of retinal layers using SS-OCT registries for diagnosing multiple sclerosis** (2019). Cavaliere C., Ortiz M., Barea R., García-Martín E., Rodrigo M.J., Vilades E., Boquete L. *27th Annual Meeting of the European Charcot Foundation*.
- Acta en congreso (presentación): **Diagnóstico de esclerosis múltiple a partir de imágenes de tomografía de coherencia óptica** (2019). Cavaliere C., Ortiz M., Barea R., García-Martín E., Rodrigo M.J., Vilades E., Boquete L. *XXXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*.
- Acta en congreso (presentación en vídeo): **Detección de pacientes con esclerosis múltiple mediante procesamiento de imágenes de tomografía de coherencia óptica utilizando redes neuronales** (2020). Barea R., Ortiz M., García-Martín E., Rodrigo M.J., Vilades E., López Dorado A., Gutiérrez R., Cavaliere C., Boquete L. *27º Seminario anual de automática, electrónica industrial e instrumentación*.
- Acta en congreso (presentación): **Diagnóstico de esclerosis múltiple mediante el análisis de imágenes de tomografía de coherencia óptica con redes neuronales convolucionales** (2020). Barea R., Ortiz M., García-Martín E., Rodrigo M.J., Vilades E., López Dorado A., Gutiérrez R., Cavaliere C., Boquete L. *XXXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*.
- Artículo: **Early diagnosis of multiple sclerosis by OCT analysis using Cohen's d method and a neural network as classifier** (2021). Garcia-Martin E., Ortiz M., Boquete, L., Sánchez-Morla E. M., Barea R., Cavaliere C., Vilades E., Orduna E., & Rodrigo M. J. *Computers in Biology and Medicine*, 129. DOI: 10.1016/j.combiomed.2020.104165

- Artículo: **Early diagnosis of multiple sclerosis by OCT analysis using Swept-Source Optical Coherence Tomography and Convolutional Neural Networks Trained with Data Augmentation** (2022). López-Dorado, A., Ortiz, M., Satue, M., Rodrigo, M. J., Barea, R., Sánchez-Morla, E. M., Cavaliere, C., Rodríguez-Ascariz, J. M., Orduna-Hospital, E., Boquete, L., & Garcia-Martin, E. *Sensors*, 22(1), 167. DOI: 10.3390/s22010167



# Índice de contenido

Resumen .....	1
Abstract.....	3
Agradecimientos.....	5
Preámbulo. Estructura de la tesis y otras publicaciones.....	7
Índice de contenido.....	11
Acrónimos .....	15
1 Introducción. Conceptos generales .....	19
1.1 Esclerosis Múltiple .....	20
1.1.1 Fundamentos y epidemiología.....	20
1.1.2 Tipos de Esclerosis Múltiple .....	22
1.1.3 Efectos sobre el organismo.....	25
1.1.4 Tratamientos.....	28
1.1.5 Diagnóstico. Criterios y pruebas actuales .....	29
1.1.6 Diagnóstico. Apoyo y mejoras planteadas .....	33
1.2 La técnica de OCT .....	34
1.2.1 Introducción a la OCT.....	34
1.2.2 Física y modalidades de OCT .....	36
1.2.3 Limitaciones de la OCT .....	47
1.3 Aprendizaje automático .....	47
1.3.1 Conceptos básicos. Inteligencia artificial.....	47
1.3.2 Contexto y evolución .....	49
1.3.3 Categorías de aprendizaje automático.....	51
1.3.4 Procesos ML tradicionales .....	54

1.3.5	Redes neuronales. Conceptos básicos .....	60
1.3.6	Redes empleadas .....	70
2	Estado del arte .....	93
2.1	Evaluación de la retina.....	94
2.1.1	La vía óptica como extensión del cerebro .....	94
2.1.2	Estructuras retinales de interés .....	95
2.2	OCT para el diagnóstico de EM.....	96
2.2.1	Reducción del espesor retinal observada por OCT. Evolución. ....	96
2.2.2	Equipos de OCT y protocolos de exploración.....	98
2.3	Aprendizaje automático. Estudios e implementación en esclerosis múltiple .....	99
3	Hipótesis y objetivos .....	103
4	Método .....	105
4.1	Base de datos .....	106
4.2	Formato de los registros.....	111
4.3	Softwares usados y otras especificaciones.....	113
4.4	Determinación de la región de mayor capacidad discriminante .....	114
4.4.1	Introducción .....	114
4.4.2	Ventanas cuadradas. Cálculo de AUC .....	114
4.4.3	<i>Effect Size</i> . La <i>d</i> de Cohen.....	119
4.4.4	<i>Effect Size</i> . La Delta de Cliff .....	123
4.4.5	Conclusiones .....	126
5	Implementación y resultados de la aplicación de aprendizaje automático aplicado a la EM	129
5.1	Formato de los resultados. Variables y entrenamiento .....	130
5.1.1	Método <i>leave one out</i> .....	130

5.1.2	Variables de red.....	130
5.1.3	Entrenamiento .....	131
5.2	Aplicación de ML a EM. Algoritmos y redes tradicionales .....	133
5.2.1	Algoritmo <i>k-nearest neighbors</i> .....	133
5.2.2	Perceptrón Multicapa .....	136
5.2.3	Radial Basis Function Networks .....	141
5.2.4	<i>Support Vector Machine</i> .....	146
5.3	Aplicación de ML a EM. Redes neuronales profundas .....	151
5.3.1	Redes Neuronales Convolucionales .....	151
5.3.2	Redes Neuronales Residuales.....	156
5.3.3	V-Nets .....	160
5.4	Discusión de los resultados. Conclusiones .....	164
6	Discusión general. Trabajos futuros.....	169
7	Bibliografía.....	173





## Acrónimos

<i>Tabla 1. Lista de acrónimos.</i>	
MS/EM	Multiple Sclerosis (Esclerosis Múltiple)
SNC	Sistema Nervioso Central.
OMS	Organización Mundial de la Salud.
NMSS	<i>National Multiple Sclerosis Society</i> (Sociedad Nacional de Esclerosis Múltiple)
RIS/SRA	<i>Radiologically Isolated Syndrome</i> / Síndrome Radiológico Aislado
CIS/SCA	<i>Clinically Isolated Syndrome</i> / Síndrome Clínico Aislado
RRMS/EMRR	<i>Relapsing – Remitting MS</i> / Esclerosis Múltiple Remitente-Recurrente
SPMS/EMSP	<i>Secondary Progressive MS</i> / Esclerosis Múltiple Secundaria Progresiva
PPMS/EMPP	<i>Primary Progressive MS</i> / Esclerosis Múltiple Primaria Progresiva
mfVEP	<i>Multi Focal Visual Evoked Potentials</i> (Potenciales Evocados Visuales Multifocales)
TAC	<i>Tomografía Axial Computerizada</i>
CSF/LCR	<i>Cerebrospinal fluid</i> / Líquido cefalorraquídeo
MRI/IMR	<i>Magnetic Resonance Imaging</i> / Imagen por Resonancia Magnética
ON	<i>Optical Neuritis</i> (neuritis óptica)

OCT	<i>Optical Coherence Tomography</i> (Tomografía de Coherencia Óptica)
TD-OCT	<i>Time Domain OCT</i> (OCT en el Dominio del Tiempo)
FD-OCT	<i>Fourier Domain OCT</i> (OCT en el Dominio de Fourier)
SD-OCT	<i>Spectral Domain OCT</i> (OCT en el Dominio Espectral)
SS-OCT	<i>Swept Source OCT</i> (OCT de Barrido de Fuente)
OCT-A	<i>OCT-Angiography</i>
ETDRS	<i>Early Treatment Diabetic Retinopathy Study</i> (Estudio de Retinopatía Diabética de Tratamiento Temprano)
AUC	<i>Area Under the Curve</i> (Área Bajo la Curva)
ES	<i>Effect Size</i> (Tamaño del efecto)
ERG	<i>Electroretinography</i> (Electrorretinografía)
EP	<i>Evoked Potentials</i> (Potenciales Evocados)
RNFL	<i>Retinal Nerve Fiber Layer</i>
mRNFL	<i>macular Retinal Nerve Fiber Layer</i>
<i>pRNFL</i>	<i>Peripapillar RNFL</i>
<i>T-pRNFL</i>	<i>Temporal pRNFL</i>
GCL	<i>Ganglionic Cell Layer</i> (Capa de Células Ganglionares)
GCIPL/GCL+	<i>Ganglionar Cell Inner Plexiform Layer</i>
GCL++	<i>Ganglionar Cell Inner Plexiform Layer</i> <i>with mRNFL</i>

EDSS	<i>Expanded Disability Status Scales</i>
AI/IA	<i>Artificial Intelligence</i> / Inteligencia Artificial
ML	<i>Machine Learning</i> (Aprendizaje Automático)
FF	<i>Feed Forward</i> (Alimentación Hacia Delante)
NN	<i>Neural Network</i> (Red Neuronal)
DN	<i>Deep Network</i> (Aprendizaje profundo)
ANN	<i>Artificial Neural Network</i> (Red Neuronal Artificial)
KNN	<i>K-Nearest Neighbors</i> (K vecinos más próximos)
SVM	<i>Support Vector Machine</i> (Máquina de Vector-Soporte)
MLP	<i>Multi Layer Perceptron</i> (Perceptrón Multi Capa)
RBF	<i>Radial Basis Function</i> (Función de Base Radial)
OLS	<i>Orthogonal Least Squares</i>
CNN	<i>Convolutional Neural Network</i> (Red Neuronal Convolutacional)
ResNet	<i>Residual Neural Network</i> (Red Neuronal Residual)
V-Net	<i>Volumetric Neural Network</i> (Red Neuronal Volumétrica)





# 1 Introducción. Conceptos generales

*Con tal de poder afrontar con propiedad el campo de la enfermedad cuyo diagnóstico se le da apoyo mediante esta tesis, se realiza un primer capítulo introductorio que compila los fundamentos mínimos establecidos que han sido necesarios estudiar para la realización de este trabajo.*

*El capítulo está estructurado en tres secciones: (1) Esclerosis Múltiple, donde se da una breve introducción sobre esta enfermedad, su alcance y sus métodos de diagnóstico; (2) la OCT, prueba con la que se obtienen los datos que se tratan en la tesis y por lo tanto el diagnóstico en la que se centra; y (3) el Aprendizaje Automático, en el que se introducirá el modo de funcionamiento general de los métodos de procesado de datos que se han implementado y con los que se han obtenido los resultados finales.*

## 1.1 Esclerosis Múltiple

### 1.1.1 Fundamentos y epidemiología.

La esclerosis múltiple (*EM*) es una enfermedad desmielinizante del Sistema Nervioso Central (*SNC*) que impide su normal funcionamiento y tiene grandes afectaciones al sistema visual, auditivo y motor ([Sospedra M. & Martin R., 2005](#)). Es de etiología autoinmune y de curso crónico.

El origen etimológico de la esclerosis múltiple se remonta del griego (*skleros*, “duro” y *osis*, “enfermedad”), y hace referencia al endurecimiento de un tejido u órgano ([Pérez-Porto J. & Merino M., 2011](#)). El origen de la enfermedad en sí no está del todo claro. Aunque se hayan descubierto crónicas que se remontan a la Islandia del siglo XII y la Holanda de los siglos XIV-XV - ambas describiendo síntomas semejantes a los asociados con lo que conocemos hoy en día como EM -, no fue hasta 1868 cuando Jean-Martin Charcot, conocido como “el padre de la neurología”, creó el término y definió por primera vez la enfermedad ([Covo P., 2015](#)).

Según la Organización Mundial de la Salud (*OMS*), en 2008 había unos 2,1 millones de personas afectadas con esta enfermedad en el mundo ([Atlas de EM, 2008](#)), con una prevalencia de 30 por cada 100 000 individuos, y 2,3 millones de afectados en 2013 ([Atlas de EM, 2020](#)). En 2018, la Sociedad Nacional de Esclerosis Múltiple, organización fundada en Estados Unidos que promueve la investigación en el campo de esta enfermedad, reportó un aumento a 2,5 millones de pacientes. En 2020, la Federación Internacional de Esclerosis Múltiple da una aproximación de unos 2,8 millones de casos en el mundo y 55000 en España, tal y como puede observarse en la figura 1, siendo este un país con una alta prevalencia de EM (101-200 por cada 100 000) cuando se le compara con las proporciones globales.



Figura 1. Prevalencia de esclerosis múltiple en el mundo y en España en el año 2020 (Atlas de EM, 2020).

Por lo tanto, la esclerosis múltiple es una enfermedad que se encuentra en auge en lo que respecta a números absolutos, aunque según las mismas fuentes, “diversos factores contribuyen a este incremento, incluyendo mejores métodos de recuento nacional e internacional, así como mejoras en el diagnóstico, el aumento de esperanza de vida tanto general como de pacientes de EM y el aumento de la población mundial”.

Frecuentemente se diagnostica en pacientes con edades de 20 a 45 años, con una edad media global de aproximadamente 32 años. Solo el 5% de pacientes a los que se les diagnostica EM tienen edades menores a los 10 años o mayores a los 50.

La enfermedad se presenta con mayor frecuencia en mujeres que en hombres. Aunque la proporción exacta se desconoce, las aproximaciones van de entre 2:1 (F:M) (Bahrun U. & Wijaya C., 2008) a 3:1 (Sospedra M. & Martin R., 2005), aunque dependiendo del país puede llegarse a 4:1 (por ejemplo, Egipto e Irán). Entre algunos factores que pueden

explicar estas proporciones se encuentran las diferencias hormonales y genéticas entre poblaciones, así como la exposición a diferentes estilos de vida y a entornos tanto sociales como laborales.

Con relación a la epidemia del COVID-19 en el año 2020, ha habido preocupaciones sobre el posible riesgo aumentado en los pacientes de EM que contraigan el virus ([Rosenblum K., 2020](#)). Después de todo, y a pesar de que esta enfermedad se diagnostique principalmente en adultos jóvenes, un número sustancial de pacientes de EM pertenece a la tercera edad, demografía en la que se ha observado el mayor riesgo de mortalidad por el virus COVID.

Sin embargo, diversos análisis estadísticos no reportan una especial prevalencia de sujetos con alto riesgo de mortalidad por COVID cuando se comparan con la población general ([Bsteh G. et al., 2020](#)). En Austria, donde se realizó el estudio estadístico y comparativo, el riesgo alto de muerte por COVID se da en menos de un 1% de pacientes con EM. Así, el consenso actual es que no se observa un mayor riesgo de contraer el virus en pacientes de EM, ni tampoco un mayor índice de mortalidad en los mismos.

A pesar de este consenso internacional, hay excepciones. Un estudio en la Rioja ([Castillo F. et al., 2020](#)) indica que la mortalidad de pacientes COVID entre sujetos con EM y la población general es similar, mientras que el porcentaje de infección es aproximadamente el doble en pacientes EM.

### 1.1.2 Tipos de Esclerosis Múltiple

La esclerosis múltiple es una enfermedad cuya presentación varía dependiendo de cada individuo, por lo que no se puede predecir con exactitud el estado en que se encontrará cada paciente al cabo de un período determinado ni los ataques que puede contraer. Sin embargo, a lo largo de los años sí se han definido formas diferenciadas que se ajustan a un modelo general ([NMSS, 2020](#)). Las cuatro formas generales, según la Sociedad Española de Neurología, son: Síndrome Clínico Aislado (*SCA*), también llamada fase de alto riesgo ([Kantarci O.H., 2019](#)), la forma remitente-recurrente (*EMRR*), y la forma progresiva, dividida en secundaria progresiva (*EMSP*) y primaria progresiva (*EMPP*).



- La fase SCA (*CIS* en inglés) constituye el primer episodio de síntomas neurológicos y se caracteriza por una inflamación y desmielinización del SNC. Pese a esto, no puede ser considerada un diagnóstico definitivo de EM, puesto que no todos los pacientes en esta fase acaban desarrollando EM posteriormente. De hecho, entre el 60 y el 70% de los pacientes con SCA desarrollan EM clínicamente definida en 20 años ([Fisniku L.K. et al. 2008](#)). La identificación de los riesgos de conversión de SCA a EM son de gran importancia, ya que la conversión en un corto período de tiempo se asocia con un peor pronóstico y un alto grado de discapacidad ([López-Gómez, J. et al., 2021](#)).
- La fase EMRR (*RRMS* en inglés) es la más común y se caracteriza por ataques o brotes de síndromes neurológicos nuevos o aumentados, llamados *recurrentes*, seguidos de recuperaciones parciales o completas, llamados *remitentes*. Durante las remitentes, los síntomas desaparecen parcial o totalmente. La fase EMRR se puede clasificar, a su vez, en *activa* o *inactiva* dependiendo de si se observa o no nueva actividad neuronal mediante MRI, así como también en *deteriorante* o *no deteriorante* si se observa una creciente discapacidad. En el 85% de personas con EM, el diagnóstico inicial se corresponde a esta fase.
- La fase EMSP (*SPMS* en inglés) sigue un curso progresivo. En algunos de los pacientes diagnosticados con EMSP la EM acaba adoptando un curso de decaimiento de la función neurológica (acumulación de discapacidad) a lo largo del tiempo. Al igual que en el caso de la EMRR, la fase EMSP se puede subdividir en activa o inactiva, dependiendo de si se observa o no nueva actividad mediante MRI y también de si hay brotes; así como también con progresión o sin progresión, dependiendo de si hay evidencia de acumulación de la discapacidad observada inicialmente.
- La fase EMPP (*PPMS* en inglés) se caracteriza por un decaimiento adicional de la función neurológica, sin recurrentes previas. La fase EMPP se observa en un 15% de los pacientes con EM. Su subdivisión es equivalente a la de la fase anterior.

A pesar de que, como ya se ha hecho constar, la enfermedad se puede presentar de forma diferente en cada paciente, en la figura 2 se encuentran ejemplos generales de la evolución de la enfermedad en los diferentes tipos definidos.

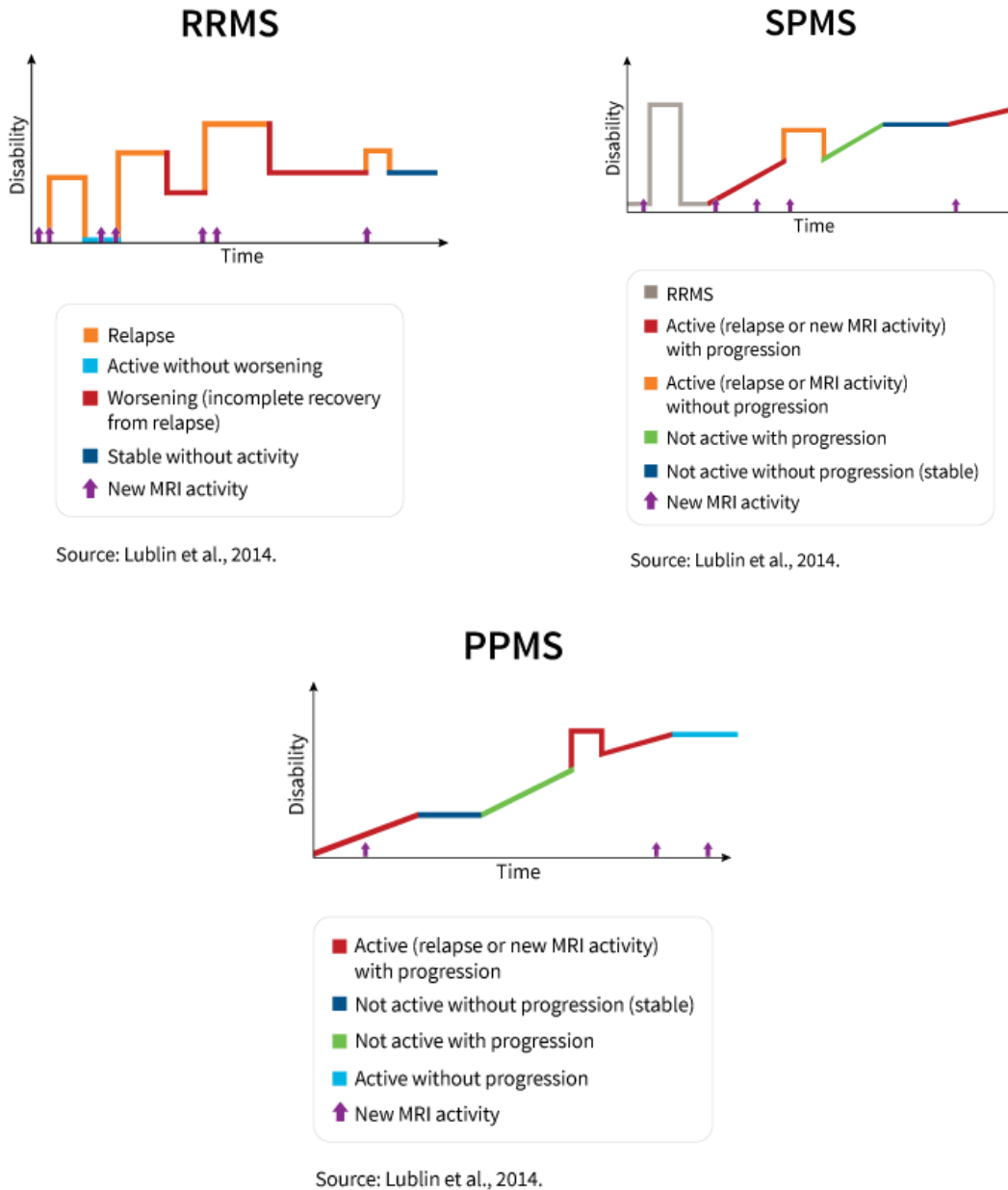


Figura 2. Evolución de la discapacidad en diversos tipos de EM (NMSS, 2020). a) EMRR, b) EMSP, c) EMPP.

A destacar que se considera *ataque recurrente* o *brote* a aquellos cuyos déficits neurológicos duran más de 24 horas, que tengan una localización anatómica asociada (que comúnmente se observaría en MRI) y que a ellos no los acompañe una fiebre o infección que pueda explicarlos.

El estado de Síndrome Radiológico Aislado (*SRA*, *RIS* en inglés), se usa para clasificar aquellos sujetos con lesiones en la sustancia blanca del sistema nervioso central (encéfalo, médula espinal), detectadas mediante una resonancia magnética y que son propias de la esclerosis múltiple (cumplen con los criterios de Barkhof/Tintoré para diseminación en espacio), pero la persona no muestra signos ni síntomas de la enfermedad, y por lo tanto no procede su diagnóstico como EM. Estas anomalías son equivalentes a las observadas en casos de EM y que no pueden explicarse con otros diagnósticos, ni tampoco constan de otros síndromes neurológicos u otras anomalías ([NMSS, 2020](#)).

Más de la mitad de los sujetos en estado SRA acaban desarrollando EM ([NMSS, 2020](#)), pero no existen guías sobre tratamientos concretos a aplicar a pacientes SRA. Debido a que los síntomas son diferentes, es necesario llevar a cabo estudios para determinar si efectivamente la fase RIS puede llegar a considerarse una fase general de la EM.

### 1.1.3 Efectos sobre el organismo

Con tal de poder diagnosticar, monitorizar y tratar la enfermedad es necesario conocer sus afectaciones más allá de sus manifestaciones sobre la vida diaria de los pacientes, las cuales se dan tanto a nivel estructural como a nivel funcional.

#### 1.1.3.1 Efectos sobre la estructura

Los nervios son estructuras conductoras de impulsos nerviosos situadas fuera del sistema nervioso central y formados por un conjunto de axones, cada uno de los cuales procede de una neurona. Se encargan de transmitir información en forma de impulsos sensoriales entre el cerebro y la periferia.

La EM se caracteriza por la desmielinización y pérdida de los axones, tanto en los nervios como en el cerebro, así como también por la infiltración perivascular de células inflamatorias, y por la formación de placa en el cerebro y en la médula espinal ([Bahrun U. & Wijaya C., 2008](#)). La figura 3 muestra un esquema de la desmielinización y consecuente daño en los axones.

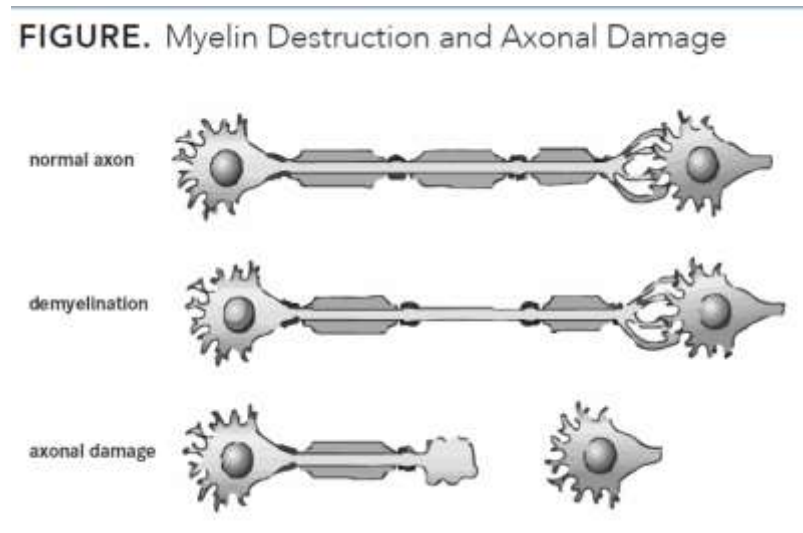


Figura 3. Desmielinización de los axones ([AJMC, 2018](#))

La desmielinización se evaluará midiendo la pérdida de espesor de las zonas afectadas mediante las pruebas de diagnóstico estructurales como la resonancia magnética (*MRI*) y la Tomografía de Coherencia Óptica (*OCT*).

### 1.1.3.2 Efectos sobre la conducción

La desmielinización se origina en las vainas de mielina, que envuelven los axones y actúan como aislante, impidiendo el intercambio de iones a través de la membrana del axón.

Los puntos de contacto entre células contiguas donde no hay vaina de mielina (la cubierta queda interrumpida) se denominan nodos de Ranvier. El impulso nervioso avanza a saltos, de nodo en nodo, por lo que progresa más deprisa. A este fenómeno se le denomina conducción saltatoria del impulso. En realidad, lo que sucede es que en cada nodo de Ranvier se regenera el potencial de acción, debido a sus especiales características

estructurales (gran cantidad de canales de sodio) ([Ferrús A. et al., 2007](#)). Esto hace que el impulso se propague con más velocidad y menor gasto energético, pues requiere la despolarización y repolarización de pequeñas partes de la membrana, llegando a alcanzar velocidades de 100 m/s en las fibras mielínicas más rápidas ([Ferrús A. et al., 2007](#)).

En cada brote de esclerosis múltiple, los nervios se inflaman y desaparece la vaina de mielina que recubre las fibras. Al no haber mielina, el impulso nervioso avanza recorriendo todo el axón (conducción continua). La propagación del impulso nervioso es más lenta, ya que cada segmento del axón debe despolarizarse y repolarizarse, lo que implica mayor movimiento de iones a través de la membrana y, por lo mismo, un mayor gasto de energía. Aunque la conducción pueda restaurarse al cabo de un tiempo, seguirá habiendo un cierto peligro de retraso, pudiendo llegar incluso a bloquearse ([Smith K.J. & McDonald W.I., 1999](#)), lo que lleva a los síntomas negativos durante los brotes (por ejemplo, parálisis y ceguera). La probabilidad de que la conducción se bloquee depende tanto de cuánta mielina se ha perdido como del tiempo desde que se ha producido la desmielinización ([Smith K.J. & McDonald W.I., 1999](#)).

En la figura 4 puede observarse la alteración de la conducción que ocurre durante la EM.

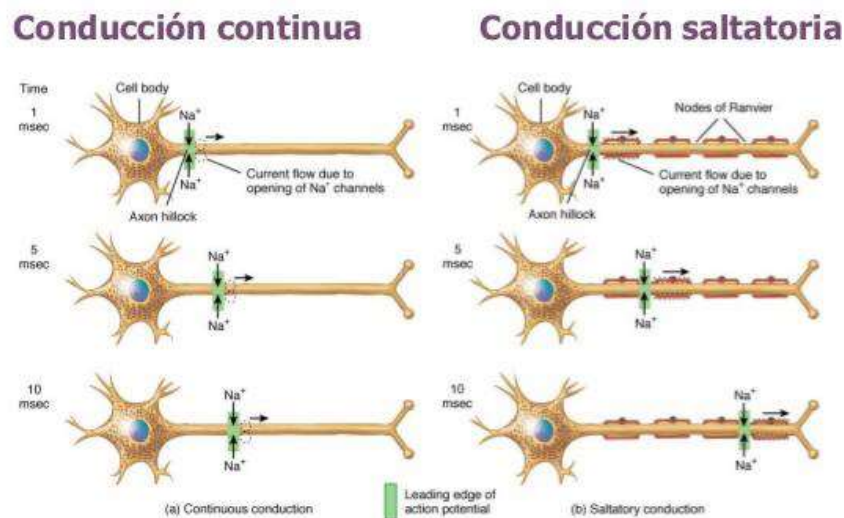


Figura 4. Conducción continua vs conducción saltatoria ([Bioestelle, 2013](#))

Esta alteración con frecuencia se evalúa midiendo el retraso del impulso nervioso mediante técnicas de diagnóstico funcionales como los Potenciales Evocados (auditivos, visuales, somatosensoriales). También resulta de interés estudiar la variación de esta alteración en la conducción durante el proceso terapéutico, para poder medir el efecto de dicho tratamiento ([Farley B.J. et. al, 2019](#)).

#### 1.1.4 Tratamientos

No hay cura conocida para pacientes con EM, pero a lo largo de los años se han desarrollado terapias capaces de mejorar la condición de los que la padecen ([Gohil K., 2015](#)), siempre dependiendo de la fase de la enfermedad en la que se encuentre el paciente. Debido a que no existe una cura definitiva, uno de los mayores factores que influyen en el éxito de estas terapias es establecer un correcto diagnóstico lo antes posible, con tal de comenzar con el tratamiento en el momento adecuado.

La finalidad de los tratamientos existentes es conseguir una rápida recuperación tras los ataques, frenar la progresión de la enfermedad y disminuir el impacto de los brotes. Para estos fines, se busca conseguir la disminución de la respuesta inflamatoria del organismo y el consecuente daño a los axones, con tal de prevenir la discapacidad resultante ([Ömerhoca S. et al., 2018](#)).

Las terapias que existen en la actualidad pueden dividirse en tres categorías, dependiendo de su finalidad principal: tratamientos de exacerbaciones (*el agravamiento de un síntoma*), terapias de modificación de la enfermedad y terapias sintomáticas ([Gohil K., 2015](#)), y a su vez cada categoría incluye tratamientos específicos dependiendo de la fase de la enfermedad en la que se encuentre cada paciente. En casi todos los casos consisten en tratamiento mediante fármacos, aunque como apoyo se suele usar también tratamiento por fisioterapia. Se ha estudiado el efecto de este otro tipo de tratamiento, y efectivamente ha habido estudios recientes que sugieren la posibilidad de que, además de aliviar los síntomas y mejorar la vida cotidiana de los pacientes, también contribuyen a la disminución de la progresión de la enfermedad ([Dalgas U. & Stenager E., 2012](#)).

### 1.1.5 Diagnóstico. Criterios y pruebas actuales

En medicina, un biomarcador es una característica que puede ser medida objetivamente y evaluada como un indicador de procesos biológicos normales, procesos patogénicos o respuestas farmacológicas a intervenciones terapéuticas ([Puntmann V.O., 2009](#)).

No es posible realizar un diagnóstico exacto e infalible de EM mediante un solo biomarcador específico ([Ömerhoca S. et al., 2018](#)), sino que depende del historial médico de cada paciente y de toda una serie de pruebas neurológicas. El diagnóstico erróneo de EM es relativamente frecuente, debido tanto a este hecho como, principalmente, a que muchos síndromes clínicos imitan a la EM. Los síntomas identificados incorrectamente como EM incluyen mielitis transversa completa, náuseas intratables y neuritis óptica (ON) bilateral con escasa recuperación de la visión.

Con tal de identificar correctamente los desórdenes asociados a la EM y discriminar aquellos que no lo están pero que pueden observarse con las mismas pruebas diagnósticas, desde los años 50 ha habido múltiples intentos para establecer una serie de criterios para mejorar la sensibilidad y especificidad del diagnóstico de EM. En algunos casos, la EM puede diagnosticarse fácilmente mediante ensayos clínicos y simples pruebas de laboratorio. Sin embargo, en etapas tempranas de la enfermedad, o en caso de pacientes con historial clínico y radiológico atípico, realizar un diagnóstico exacto es complicado.

Los primeros criterios se propusieron en 1965, estableciendo que para definir un diagnóstico de “probable EM” debía cumplirse que ocurriesen dos ataques neurológicos de duración igual o superior a 24 horas y con una separación temporal de al menos un mes, y por otro lado que se observase más de una región anatómica afectada ([Ömerhoca S. et al., 2018](#)). En 1983 estos criterios se afinaron, dividiendo el diagnóstico entre “definido clínicamente” y “definido con pruebas de laboratorio”, esta última categoría siendo de mayor exactitud, y refiriéndose comúnmente a pruebas como análisis de sangre y otros tipos de biopsias.

En 2001 se definieron por primera vez los criterios cuyas bases siguen usándose en la actualidad: los criterios de McDonald, que incorporaron imagen de resonancia magnética al conjunto de pruebas que hasta entonces se usaban para diagnosticar EM ([Polman C.H. et al., 2005](#)). Con los años estos criterios se fueron refinando, definiendo el intervalo óptimo entre pruebas de imagen y la distribución anatómica de las posibles afectaciones asociadas a la enfermedad, así como añadiendo el análisis de líquido cefalorraquídeo (*LCR*) como importante componente de las pruebas a usar como parte del diagnóstico. También se cambió el criterio de 1965 de solo considerar un diagnóstico probable en caso de varios ataques separados un cierto tiempo. Las categorías de diagnóstico se cambiaron a “MS definido”, “sospecha de EM” y “sin EM” ([Ömerhoca S. et al., 2018](#)). En 2015 se cuantificó ulteriormente parámetros relacionados a la MRI, como el espesor mínimo entre imágenes y otras características a la hora de tomar las imágenes.

Posteriores actualizaciones se realizaron en los años 2005, 2011 y 2018. En 2005 se expresó interés para simplificar y acelerar el diagnóstico y al mismo tiempo mantener la especificidad y sensibilidad adecuadas ([Polman C.H. et al., 2005](#)). Para ello, un cuerpo de expertos se reunió y volvió a analizar las consideraciones que se tomaron al definir las, mayoritariamente a nivel de historial clínico, a pruebas de resonancia magnética y a pruebas de líquido cefalorraquídeo. En base a los mismos criterios, posteriores actualizaciones fueron tomadas en 2011, en las que, entre otras modificaciones, se redujeron los requisitos de resonancia magnética para el diagnóstico final ([Polman C.H. et al., 2011](#)); y en 2018, simplificando lo que debe observarse en MRI para diagnosticar EM y concluyendo que futuras modificaciones deberían orientarse hacia el desgaste del nervio óptico ([Thompson A.J. et al., 2018](#)). **Estas últimas modificaciones reafirman la importancia de pruebas como la OCT para mejorar los métodos de diagnóstico de EM,** pues tal y como se verá en próximos capítulos, constituye la prueba de diagnóstico más usada para la evaluación de los cambios estructurales (degradación del nervio óptico) que se da durante el transcurso de la EM.



Por otro lado, para estudiar las disfunciones de la vía óptica - afecten o no en ese momento a la vida diaria del paciente - se usan pruebas electrofisiológicas, en las que se mide los efectos en la conducción. Las pruebas usadas con más frecuencia son la de potenciales evocados, destacando las pruebas de potenciales evocados visuales (VEP) así como la electroretinografía (ERG) ([Forooghian F. et al., 2006](#)).

Las pruebas (tanto en lo que se refiere a ensayos clínicos como ensayos de laboratorio y técnicas de imagen) que pueden usarse y se han definido a lo largo de los años al refinar los criterios de diagnóstico aparecen listadas en la tabla 2.

- Entre las pruebas primarias están los análisis de sangre, las ya mencionadas resonancias magnética, las pruebas de CSF, tanto craneal como cervical y torácico, y, en caso de pacientes con neuritis óptica, pruebas de VEP y OCT.
- Las pruebas secundarias pueden consistir en ulteriores pruebas VEP y OCT, tanto si el paciente ha tenido neuritis óptica o no, pruebas urodinámicas (análisis de orina, entre otras) y exámenes cognitivos.
- Otros *tests* que suelen realizarse son biopsias más allá del análisis de sangre, así como otros análisis bioquímicos; pruebas de oído, rayos X...

La OCT, la técnica de diagnóstico a partir de la que se obtiene la base de datos tratada en la tesis, se estudiará más en profundidad en las siguientes secciones.

<i>Tabla 2. Pruebas usadas en el diagnóstico de EM (Ömerhoca S. et al., 2018).</i>
<b>Primary Tests</b>
<ol style="list-style-type: none"> <li>1. Blood tests (hemogram, renal and liver function tests, electrolyte levels, sedimentation, CRP, B12, folate and vitamin D, thyroid function tests, lipid panel, viral serology (anti-HIV, anti HCV, HbsAg, anti-Hbs), VDRL-RPR, ANA (1/320 titer and patterns), if ANA positive ENA profile, antiphospholipid antibodies, anti-ds DNA)</li> <li>2. MRI (cranial, cervical, and thoracic)</li> <li>3. CSF analyses (CSF protein, CSF and concurrent blood glucose, CSF albumin and IgG, CSF lactate, serum albumin and IgG, CSF IgG index, CSF OCB analysis with IEF electrophoresis)</li> <li>4. In patients with optic neuritis: VEP and optic coherence tomography</li> </ol>
<b>Secondary Tests</b>
<ol style="list-style-type: none"> <li>1. Evoked potentials (VEP vs SEP)</li> <li>2. Optic coherence tomography</li> <li>3. Urodynamics testing</li> <li>4. Cognitive testing</li> </ol>
<b>Other tests for differential diagnosis</b>
<ol style="list-style-type: none"> <li>1. Further biochemical tests (if there is suspicion of vasculitis, wide autoantibody panel, 24-hour urine analysis, GFR evaluation, for rheumatological disorders anti CCP, serum complement levels, for lymphoma serum anti beta2 microglobulins, for sarcoidosis blood and CSF ACE levels, for adrenoleukodystrophy adrenal hormone levels, long/very long chain fatty acids, for mitochondrial diseases serum pyruvate, lactate levels, for neuromyelitis optica anti-aquaporin4 and anti-MOG tests)</li> <li>2. Specific tests for infectious etiologies (antibodies for Lyme disease and Brucellosis, PPD and quantiferon tests for tuberculosis)</li> <li>3. Angiography (cerebral, fluorescein, MRA)</li> <li>4. Biopsy (skin, lymph node, brain and/or leptomeninges, peripheral nerve, other)</li> <li>5. Eye examination (retina evaluation for metabolic disorders, overevaluation for sarcoidosis and Behçet's disease)</li> <li>6. Hearing tests (for Susac)</li> <li>7. Electrophysiology (nerve conduction studies, EMG)</li> <li>8. Chest X-ray (for chronic latent/sequel infectious lung disorders, and hilar adenopathy)</li> <li>9. Cardiac examination (echocardiography for SLE, and mitochondrialopathies)</li> <li>10. Others (Schirmer test for Sjögren's disease, and salivary gland scintigraphy, for malignancies and metabolic disorders SPECT and PET)</li> </ol>

La resonancia magnética es la primera prueba de diagnóstico por imagen que suele realizarse si se sospecha EM. Debido a que durante la EM se observa una importante desmielinización del sistema nervioso, la importancia de estas estructuras radica en la

pérdida de espesor cortical. Mediante RMI se mide este espesor en ciertas zonas – frecuentemente cráneo, cervix y tórax ([Palombit A. et al., 2017](#)) – y se compara posteriormente con una serie de referencias para determinar si se ha observado una suficiente desmielinización como para considerarse EM.

#### 1.1.6 Diagnóstico. Apoyo y mejoras planteadas

Ulterior investigación sobre los biomarcadores de EM para permitir el diagnóstico temprano y una objetiva evaluación del tratamiento constituye una prioridad para que los pacientes se beneficien de una terapia eficaz lo antes posible.

Tal y como se ha visto en el apartado anterior, el diagnóstico de EM envuelve un elevado número de pruebas para determinar si finalmente el paciente padece la enfermedad o no, algunas de ellas invasivas, como el análisis de líquido cefalorraquídeo (obtenido mediante punción lumbar). Si en cambio se promueve y apoya el uso de pruebas no invasivas que sirvan como biomarcadores diagnósticos se mejoraría su fiabilidad, reduciendo la necesidad de algunas de las pruebas invasivas existentes, lo cual también conduciría a un diagnóstico y un tratamiento más temprano, conllevando una mejora en la calidad de vida de los pacientes y minimizando el impacto de la enfermedad sobre su salud.

La presente tesis se centra en el tratamiento de datos obtenidos a partir una de las pruebas de diagnóstico ya mencionada: la OCT. En el caso de diagnóstico de EM, es usada principalmente en la evaluación de la vía visual. Esta técnica se explorará en profundidad en el siguiente apartado.

## 1.2 La técnica de OCT

### 1.2.1 Introducción a la OCT

#### 1.2.1.1 Definición y concepto general

La Tomografía de Coherencia Óptica (comúnmente conocida por su acrónimo inglés, *OCT*) es una técnica no invasiva de diagnóstico (*por lo que no involucra instrumentos que penetren físicamente el cuerpo para tomar las medidas*), y una nueva modalidad de imagen médica. Introducida a principios de la década de los 90, fue usada por primera vez para obtener imágenes *in vivo* de ciertas regiones retinales y de las arterias coronarias ([Huang D. et al., 1991](#)). Es una prueba de diagnóstico estructural, y como tal, con ella se obtienen datos sobre las dimensiones de aquello que quiere estudiarse.

La OCT se basa en la interferometría óptica: se emite energía en forma de un haz de luz hacia los tejidos y se analiza el haz cuando este regresa. Gran parte de la energía se dispersa o absorbe; la energía restante será reflejada y posteriormente recolectada, y en base a su interacción con el medio que haya atravesado formará la imagen médica ([Hamdan R. et al., 2012](#)).

#### 1.2.1.2 Etimología. Términos importantes

Antes de avanzar, es necesario hablar de los diferentes conceptos que componen el término OCT:

1. **Óptica** hace referencia al medio que usa para generar la imagen: la luz (en contraste a, por ejemplo, los ultrasonidos, que usan el sonido).
2. **Coherencia** es una propiedad de la luz, y aparece esquematizada en la figura 5. Dos ondas son coherentes si sus picos (o “crestas”) se alinean, en cuyo caso se diría que su interferencia sería constructiva. En el otro extremo, la interferencia de dos ondas de la misma longitud de onda pero que estén completamente desfasadas (ondas incoherentes) sería destructiva (se cancelarían). En la OCT se quiere observar el desfase producido al hacer incidir un haz de luz sobre el tejido humano.

3. **Tomografía** se refiere a producir imágenes por secciones. Esto permite analizar tejido en tres dimensiones.

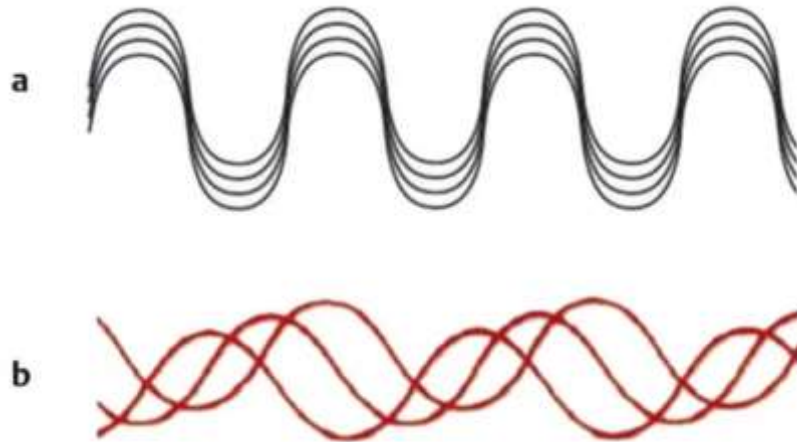


Figura 5. Coherencia en ondas. a) ondas coherentes b) ondas incoherentes ([Khumar D.A. & Agarwal A, 2020](#)).

Por lo tanto, el término completo, Tomografía de Coherencia Óptica, define una técnica de imagen que utiliza la propiedad de coherencia de la luz para tomar imágenes en tres dimensiones (imágenes 2D a lo largo de una longitud).

### 1.2.1.3 Principio de funcionamiento general

El principio de funcionamiento de la OCT se ve esquematizado en la figura 6. Mediante un láser se emite luz hacia un divisor de haces o espejo parcial, que tal y como su nombre indica, divide la luz en dos haces diferentes. Parte de la luz viaja de vuelta a un espejo de referencia, que la reflejará directamente de vuelta hacia el divisor. Otra parte de la energía se dirige al tejido que se pretende estudiar; dependiendo de las propiedades ópticas del mismo, parte de la luz será absorbida, refractada o reflejada. La parte reflejada será la que finalmente viajará de vuelta al divisor, donde se encontrará con el haz de referencia y ambos viajarán hacia un fotodetector.

Al emitirse, los haces son coherentes, y en caso de viajar la misma distancia por el mismo medio, al regresar al divisor seguirían manteniendo esta propiedad. Sin embargo, al interactuar uno de los haces con el tejido humano no se mantendrá la coherencia. Este nivel

de incoherencia es lo que se medirá, mediante la interferencia de estos dos haces, y se procesará para formar finalmente las imágenes de OCT ([Hamdan R. et al., 2012](#)).

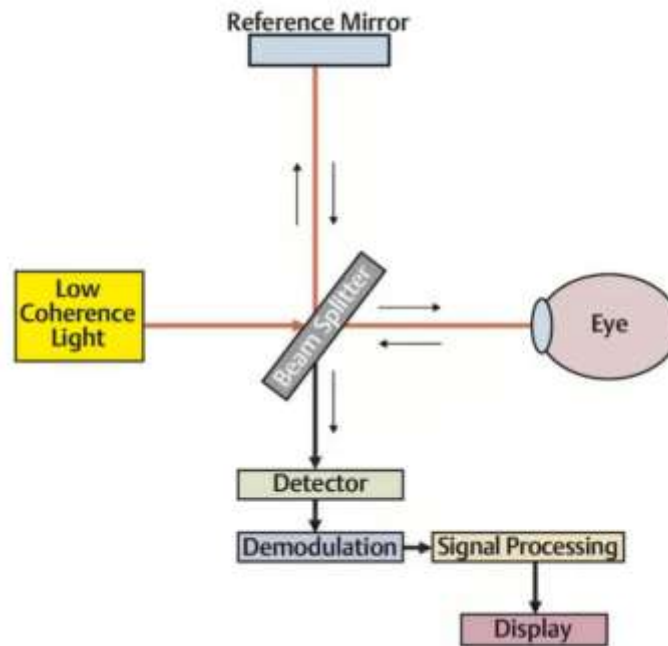


Figura 6. Principio general de funcionamiento de OCT. Óptica. ([Khumar D.A. & Agarwal A, 2020](#)).

Dependiendo del nivel de incoherencia se asignará un valor u otro al punto estudiado en cada momento del proceso. Debe tenerse en cuenta que la coherencia solo puede mantenerse entre haces con la misma longitud de onda, puesto que, en caso contrario, tras un ciclo de una de las dos ondas las crestas dejan de estar alineadas. Aunque pueda haber cierta coherencia entre haces de longitudes de onda muy similares, tradicionalmente en la OCT se han usado haces monocromáticos, es decir, haces con la misma longitud de onda ([Nandler Z., 2015](#)).

## 1.2.2 Física y modalidades de OCT

### 1.2.2.1 Física envuelta

Con tal de elegir o al menos determinar el sistema de OCT adecuado para usar en una situación dada, así como elegir los parámetros de medida, hay que entender las diferencias

entre ellos, pero para ello es conveniente conocer más en profundidad los principios físicos, empezando por la interacción de los haces de luz, pues a partir de estos se forman las imágenes. La interacción dependerá en gran parte de las propiedades del haz que finalmente se recoge al final, que variarán dependiendo a lo ocurrido en su interacción con la materia de la que se quiere obtener imágenes.

Tal y como se observa en la figura 7, de óptica básica se sabe que cuando un haz de luz pasa de un medio a otro (en este caso, tejido humano), una parte de la luz se refleja: rebota con un ángulo de reflexión igual al ángulo de incidencia. En el caso de la OCT, al incidir en ángulo recto (perpendicular al tejido), al reflejarse lo hará en esta misma dirección, y obviamente en sentido contrario. Otra parte de la luz se refracta: consigue pasar de un medio a otro y viajará con un ángulo de refracción diferente al ángulo de incidencia. El resto de energía se absorbe.

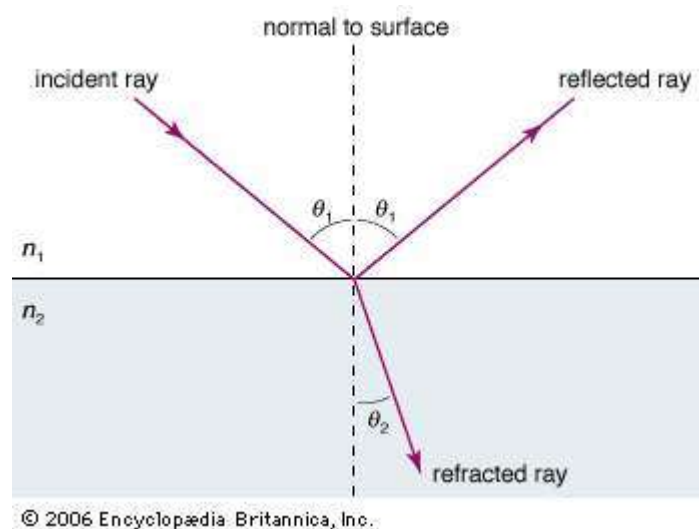


Figura 7. Reflexión y refracción ([Encyclopedia Britannica, Inc, 2006](#))

La cantidad de luz que se refleja o refracta viene dada por factores como el ángulo de incidencia y el medio por el que pasa. Puesto que la fuente de luz más común en la OCT es un láser y una de sus características es el elevado grado de direccionalidad (y su consecuente mínima dispersión), ([O'Brien W., 2013](#)) el ángulo de incidencia es un factor que puede considerarse despreciable.

Con el fin de minimizar la absorción de energía producida por los tejidos (*que no solo no aporta nada si no puede dificultar la interferencia posterior debido a que reduce la energía del haz que regresa*), la energía utilizada está centrada en una longitud de onda de alrededor de 1300 nm [*franja infrarroja*] ([Hamdan R. et al., 2012](#)). Tal y como se observa en la figura 8, la luz infrarroja no se encuentra en la franja de radiación ionizante, la dañina para los tejidos humanos, por lo que su uso constituye otra de las ventajas de la OCT frente a otras técnicas de imagen que, al contrario, sí irradian al paciente y por lo tanto constituyen un cierto riesgo, como es el caso de la Tomografía Axial Computarizada (TAC).

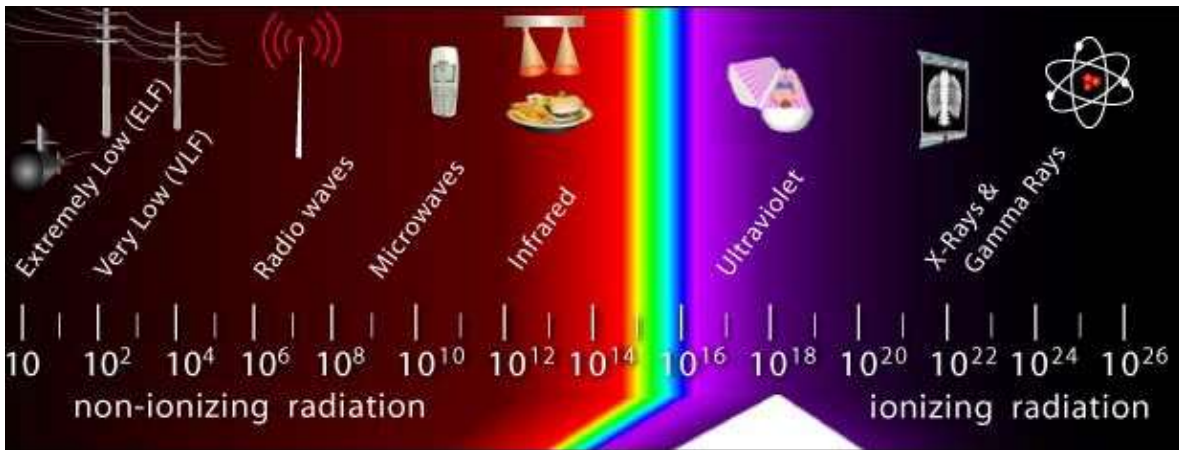


Figura 8. Espejo electromagnético: franjas ionizantes y no ionizantes ([REMM, 2020](#))

La parte de energía que en cambio se refracta continuará su viaje hacia la parte más profunda del ojo, reflejándose continuamente desde tejidos con diferentes propiedades ópticas a medida que atraviesa las distintas estructuras que lo componen. Las distancias y dimensiones de estas diferentes estructuras se ven determinadas mediante la cuantificación del retraso al volver, que dependerá de la distancia recorrida y la velocidad.

El medio concreto por el que viaja la luz en un determinado momento (en este caso, los tejidos de las diferentes estructuras oculares) influirá no solo en la cantidad de luz que se refleja y refracta, sino en la velocidad a la que viajará, puesto que la luz viaja más rápidamente en un medio con un bajo índice de refracción. Este cambio de velocidad es lo que producirá cierto desfase en la posterior interferencia con el haz de referencia, y por lo



tanto es uno de los factores principales que tiene en cuenta la OCT. A partir del retraso se puede reconstruir a qué profundidad fue reflejado ese haz en concreto.

Ahora bien, la primera reflexión ocurre al incidir con la capa más superficial del ojo. Al viajar más allá se estará recorriendo mayor distancia, por lo que los haces que van regresando ya no podrían interferir con el haz de referencia, pues debido a este incremento en distancia recorrida ya no serían coherentes entre ellos.

Es por ello por lo que tanto la distancia entre el ojo y el divisor como la distancia entre el espejo de referencia y el divisor son de gran importancia. Solo en caso de ser iguales puede darse la coherencia, pues en un caso ideal, si los haces viajan la misma distancia, al interferir entre ellos las crestas estarían alineadas perfectamente, puesto que lo hacen también con la misma velocidad.

Para contrarrestar este efecto y mantener solo una variable (la velocidad), la solución tradicional, y la que emplean las OCT en el dominio del tiempo, es aumentar la distancia entre el divisor y el espejo de referencia en cada una de las interferencias, en cada una de las medidas ([Nandler Z., 2015](#)). De esta forma se puede tomar un punto de medida (puede medirse una interferencia) para cada desplazamiento entre el divisor y el espejo.

La solución más común para esto es emplear un pulso de luz en vez de un haz constante, de tal forma que cada pulso que se envíe corresponda a una medida. Es decir, que de todas las reflexiones que, tal y como se ha visto, ocurren al incidir la luz, solo se tenga en cuenta una para cada pulso.

Otra solución, la empleada por los sistemas de OCT en el dominio espectral, es usar un espectrómetro. La tercera alternativa es emplear una técnica de barrido de fuente, empleada en las OCTs de barrido de fuente, conocidas comúnmente por su acrónimo inglés, SS-OCTs (*Swept Source-OCTs*). Estas diferentes modalidades de OCT se estudiarán en apartados posteriores de este mismo capítulo.

Debido al orden de magnitud de la velocidad de la luz, no es fácil medir el retraso de la luz reflectada. Es por ello por lo que los sistemas de OCT utilizan sistemas indirectos mediante la interferometría óptica de baja coherencia; en la siguiente sección se estudiará esto en más profundidad.

#### 1.2.2.2 Resolución y profundidad de las medidas

La ventaja crítica de la OCT frente a otras técnicas de imagen médicas como la ecografía y la resonancia magnética reside en su resolución micrométrica ([Hamdan R. et al., 2012](#)), aunque a diferencia de estas otras, las imágenes que pueden obtenerse de OCT están restringidas a imágenes de tejido superficial, a poca profundidad. Por ello, su uso suele limitarse principalmente a especialidades como la oftalmología y la dermatología ([Nandler Z., 2015](#)).

La **resolución lateral** [*es decir, cuántos píxeles tiene cada imagen*] viene dada por la óptica y es independiente de la resolución axial. Dependerá del tamaño mínimo de punto de la fuente del rayo de luz, y es inversamente proporcional al diámetro de apertura de las lentes focales, de las que procede el pulso de luz.

La **resolución axial** [*el incremento de profundidad cada cuanto se toma una nueva imagen, lo que equivale a cada cuanto se mide una nueva interferencia*], también llamada “resolución de profundidad” o “puerta de coherencia”, viene dada por la fuente de luz usada. Con frecuencia se usan diodos superluminiscentes semiconductores de baja coherencia (*SLD*) ([Popescu D.P. et al., 2011](#)). El motivo de su uso se explica a continuación.

La resolución axial equivale a la longitud de coherencia, que es la distancia hasta la que aún puede considerarse coherente la interferencia entre el haz de luz reflectado y el haz de referencia. “Baja coherencia” suele limitarse a longitudes de coherencia del orden de micras, que es la que se desea con tal de que las diferencias de desfase se observen en distancias muy cortas. El concepto aparece esquematizado y formulado en la figura 9.

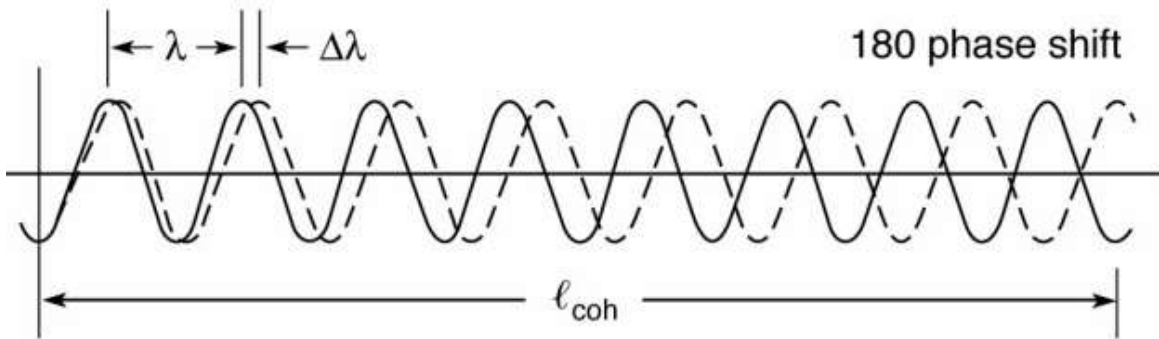


Figura 9. Longitud de coherencia (Attwood D., 2016).

Como se observa, la longitud de coherencia,  $l_c$ , viene dada por:

$$l_c = \frac{\lambda_0^2}{2 \cdot \Delta\lambda} \quad (1)$$

, donde  $\lambda_0$  es la longitud de onda del haz de referencia y  $\Delta\lambda$  el ancho de banda, dado por la diferencia entre la longitud de onda del haz de referencia y la longitud de onda del haz reflectado.

Haces de distintas longitudes de onda mantienen la coherencia a lo largo de esta longitud. En el caso de tener dos haces (prácticamente) monocromáticos, la longitud de coherencia es muy alta, puesto que el ancho de banda es mínimo. En cambio, la longitud de coherencia es más baja si se usan haces de distinta longitud de onda.

El problema de longitudes de coherencia altas es que el desfase no ocurre hasta haber recorrido una larga distancia. Esto quiere decir que, para poder observarse desfase, y, por lo tanto, hasta poder emitir otro pulso de luz, se tendría que llegar a una gran profundidad. Se quiere precisamente lo contrario: ser capaces de observar el desfase en un incremento axial mínimo.

Si se conoce la longitud de coherencia se puede dictar cada cuándo hay que medir una nueva interferencia, y con ello, cada cuánto puede medirse una nueva imagen. Por lo tanto, puede entenderse como el intervalo entre imágenes, que es proporcional a esta longitud de

coherencia. Puesto que se quiere tomar el máximo número de imágenes posible (mayor resolución en profundidad), se **busca una longitud de coherencia baja**. Así, **la resolución axial es inversamente proporcional a la longitud de coherencia**.

En la figura 10 puede observarse la resolución axial para longitudes de onda de 840 y 1310 nm respectivamente, ambas en la franja infrarroja cercana ([Lee S.-W. et al., 2009](#)). Para obtener una alta resolución (del orden de unas pocas micras), debe minimizarse la longitud de coherencia, y por lo tanto disminuir la longitud de onda y usarse un ancho de banda elevado.

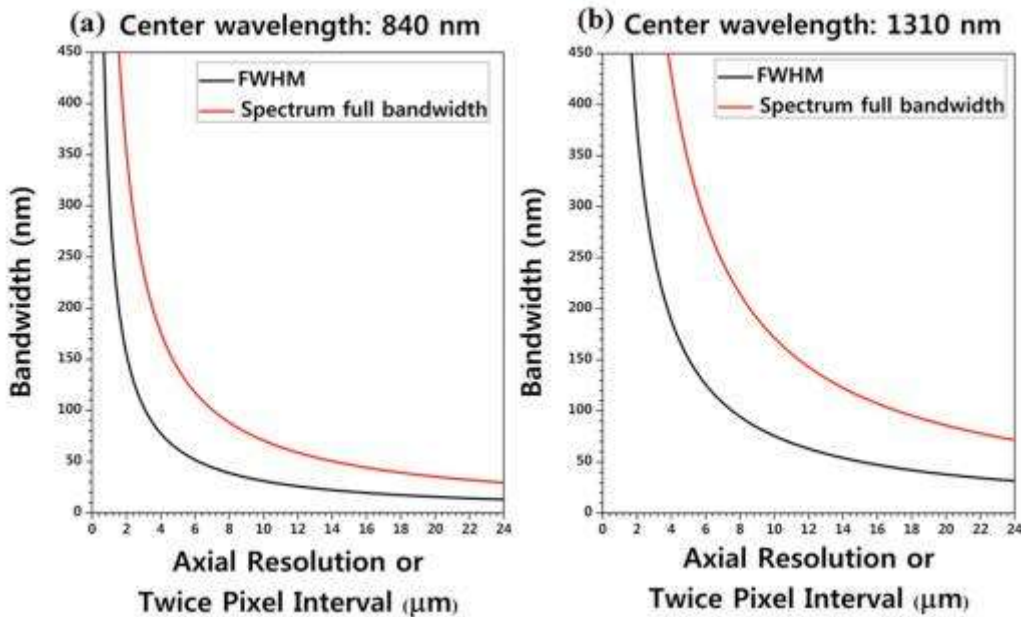


Figura 10. Resolución axial dependiendo de la longitud de onda y el ancho de banda de la fuente de luz usada ([Lee S.-W. et al., 2009](#)).

Una vez definida la resolución axial, así como su formulación, es posible comparar una vez más la OCT con otras técnicas de imagen médica similares como son los ultrasonidos. El motivo por el que las imágenes de OCT tienen mayor resolución a las de ultrasonidos es que la velocidad de cualquier haz de luz que se use será mayor a la del sonido ([Khumar D.A. & Agarwal A, 2020](#)). Esto se demuestra por teoría básica de las ondas:

$$\lambda = \frac{v}{f} \quad (2)$$

, donde  $\lambda$  es la longitud de onda,  $v$  es la velocidad y  $f$  es la frecuencia. De la ecuación, a mayor longitud de onda, mayor longitud de coherencia, lo que lleva a menor intervalo de imágenes y por lo tanto menor resolución. En el caso de ultrasonidos, se tiene que la velocidad del sonido es del orden de los miles de metros por segundo, mientras que su frecuencia es del orden de kHz ([Lucas V.S. et al., 2014](#)), dando longitudes de onda del orden de las centésimas de metros por segundo. En cambio, en el caso de la luz infrarroja, la velocidad es del orden de los billones de metros por segundo, pero la frecuencia está en los terahercios, dando longitudes de onda del orden de las micras. Puesto que esta proporción es varios órdenes de magnitud inferior en el caso de la OCT, se da una resolución axial de varios órdenes de magnitud superior.

Sin embargo, la resolución axial no es lo único a tener en cuenta, sino que también hay que considerar la profundidad máxima de escaneo, que como se ha mencionado anteriormente, es inferior a la de técnicas como los ultrasonidos. Eso es debido a su formulación:

$$z_{max} = \frac{\pi}{2\delta k} \quad (3)$$

, donde  $z_{max}$  es el rango de profundidad máximo, definido como la profundidad más allá de la cual la cantidad de luz reflejada es demasiado pequeña como para ser detectada. El intervalo entre imágenes,  $\delta k$ , es proporcional a la longitud de coherencia definida en (1), y como se ha visto, al aumentar la longitud aumenta también el intervalo de imágenes, lo que conlleva, según (3), una disminución de la profundidad máxima.

Por lo tanto, **la elección de la longitud de onda constituirá un compromiso entre resolución axial y profundidad máxima de escaneo.**

### 1.2.2.3 Tipos de OCT

Hay dos tipos de sistemas de OCT: los de primera generación, también llamados OCT en el dominio del tiempo (*TD-OCT*), y los de nueva (segunda) generación, también llamados

OCT en el dominio de la frecuencia, o de Fourier (*FD-OCT*) ([Hamdan R. et al., 2012](#)), a su vez divididos entre OCT *Swept Source* (*SS-OCT*) y OCT en el dominio espectral (*SD-OCT*) ([Kajlkman J., 2017](#)).

Ambas modalidades comparten el mismo principio de funcionamiento general: usan un interferómetro para medir el desfase de la luz reflejada del tejido comparándola con un haz de referencia. Difieren en la fuente de luz usada y en la manera en que cambian las distancias en cada medida, explicadas en apartados anteriores de este capítulo. En ambas modalidades, además, es muy importante conocer en profundidad e incorporar técnicas de análisis y procesado de señales avanzadas ([Kajlkman J., 2017](#)). Las diferentes modalidades de OCT aparecen representadas en la figura 11.

Por un lado, en el caso de las OCTs de primera generación (dominio en el tiempo), la información de profundidad de la retina se mide moviendo el espejo de referencia en cada medida ([Leng C.K. et al., 2008](#)) [*es decir, tal y como se explica en el primer subapartado de esta sección*]. Se usa un emisor de luz con gran ancho de banda y considerado de baja coherencia; de esta manera, solo se obtiene una señal de interferencia cuando la longitud de camino óptico de la muestra y el brazo de referencia se ajustan a la misma longitud de coherencia, muy estrecha por ser de baja coherencia. La señal de interferencia se demodula [*proceso que extrae la señal original que interesa de una onda modulada que contiene varias señales*] para generar un perfil de profundidad de reflexión, también llamado *A-scan* ([Drexler W. et al., 2014](#)).

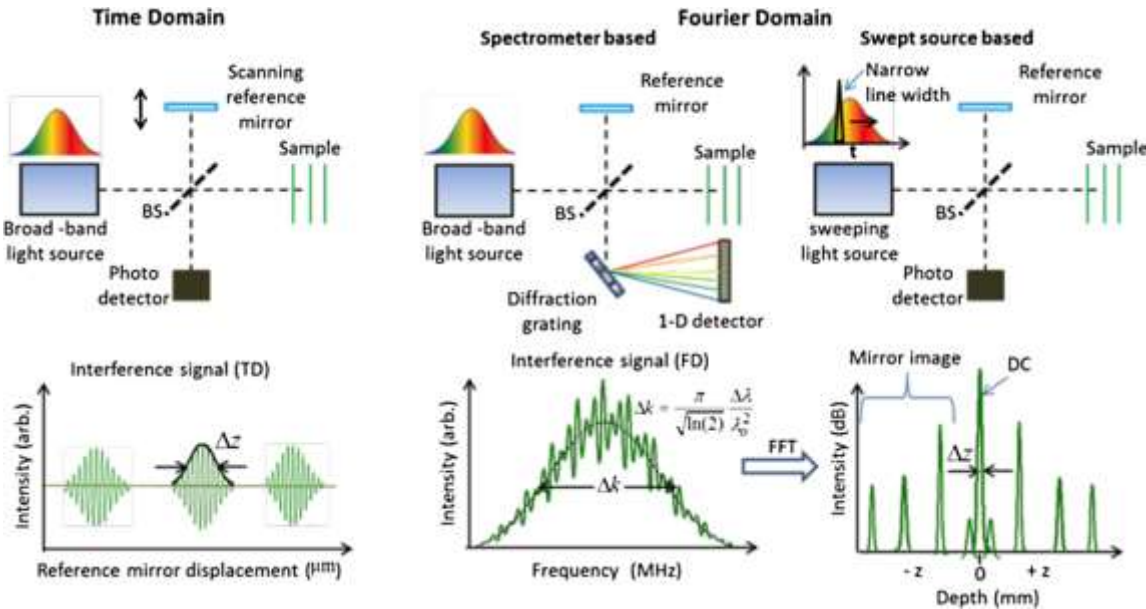


Figura 11. Modalidades de OCT: diferencias en los sistemas de medida empleados y las señales obtenidas (Drexler W. et al., 2014).

Por otro lado, en los sistemas de OCT en el dominio de Fourier el espejo de referencia permanece estático, y el problema de localizar en profundidad cada medida se soluciona estudiando el espectro de la interferencia producida, que es recogida en detectores separados espectralmente, y transformada mediante procesamiento de señales (realizando una transformada de Fourier inversa) en medidas axiales (Kajikman J., 2017). Esta última fuente recoge el análisis y proceso matemático, y no se incluirá en la presente tesis. Dependiendo de cómo se realiza el proceso, los sistemas de OCT en el dominio de la frecuencia pueden dividirse en SD-OCT o SS-OCT:

- En el caso de SD-OCT, el proceso se realiza en el dominio del espacio mediante un espectrómetro. Se usa, al igual que en los sistemas en el dominio del tiempo, una fuente de luz de un cierto ancho de banda. La señal de interferencia se divide en diferentes frecuencias ópticas mediante una red de difracción [que es a su vez otro divisor de haces] y detectada por un dispositivo de carga acoplada [un sensor integrado compuesto por un conjunto de condensadores usado ampliamente en imagen digital, como es este caso] (Drexler W. et al., 2014).

- En el caso de SS-OCT, la señal de interferencia se detecta por un fotodiodo, y su principal característica es que **la fuente de luz varía con la frecuencia**, centrada en muchas ocasiones en una longitud de onda de 1300 nm. El espectro de la fuente de luz se escanea con un pulso instantáneo en el tiempo, y posteriormente se hace la transformada de Fourier para obtener la imagen de profundidad.

Al introducirse, los equipos de nueva generación [*es decir, las OCT en el dominio de la frecuencia*] fueron capaces de superar las limitaciones de los primeros equipos de OCT, mejorándose la calidad de imagen y la rapidez en la que se toman las medidas, puesto que con ellos es posible registrar información espacial de forma simultánea ([Aumann S. et al., 2019](#)). Aún así, el consenso en la actualidad es que tanto un sistema como el otro puede generar imágenes de una resolución adecuada para muchas de las aplicaciones actuales ([Gao W. & Wu X., 2017](#)) ([Wang C. et al., 2015](#)). Dicho esto, estudios preliminares determinaron que se desconoce si las mayores frecuencias de muestreo y la mejor resolución de escaneo mejoran definitivamente su fiabilidad de medida ([Leng C.K. et al., 2008](#)). Ha habido estudios que han demostrado que la OCT en el dominio del tiempo se ajusta mejor a ciertas situaciones y la OCT en el dominio de Fourier a otras ([Sánchez-Dalmau B. et al., 2010](#)). Esta última fuente constituye un ejemplo de que dependiendo de la profundidad en la retina que se estudie (qué capas retinales se analicen) y del cuadrante longitudinal en que uno se encuentre (temporal, nasal, etc.), así como el modelo concreto de cada una de las dos modalidades de OCT, se obtienen mejores resultados con un sistema u otro.

Sin embargo, sí parece existir un consenso según el cual SD-OCT tiene una mejor repetibilidad en cuanto a medidas realizadas en diferentes tipos de escaneos en un mismo equipo y entre equipos de la misma modalidad ([Chan H.H. et al., 2015](#))([Moon S.W. et al., 2009](#)). Los trabajos y estudios realizados previamente y durante la presente tesis se han hecho en base a registros de espesores retinales de controles y pacientes de EM medidos por equipos del tipo SS-OCT ([Cavaliere C. et al., 2019](#)) ([Barea R. et al., 2020](#)), mayoritariamente debido a que son los equipos disponibles en el hospital que ha proporcionado los datos.



### 1.2.3 Limitaciones de la OCT

También hay limitaciones de esta tecnología a tener en cuenta, así como posibles errores, que deben tomarse en consideración.

En primer lugar, diferentes equipos de OCT tienen diferentes protocolos, como ya se ha mencionado, así que para hacer comparaciones precisas, debe usarse siempre el mismo tipo de dispositivo, con los mismos protocolos. En segundo lugar, toda medida de OCT se hacen comparadas a bases de datos normalizadas, con frecuencia formadas casi en su totalidad por sujetos caucásicos de mediana edad, por lo que esto debe tenerse en cuenta si lo que se quiere es evaluar a pacientes fuera de estas demografías. Otro punto a tener en cuenta es que en la interpretación de las medidas automáticas deben considerarse posibles artefactos, por lo que es importante ceñirse a un área concreta e ir haciendo ajustes manuales para ese fin. Por último, la repetibilidad de la segmentación automática varía dependiendo tanto del equipo de OCT usado como del protocolo de exploración y los programas usados por el mismo. Los algoritmos de los equipos de OCT detectan automáticamente diferentes estructuras retinales de interés, pero en casos de ciertas patologías oculares o cuando la medida actual es de escasa calidad, la segmentación puede fallar ([Minakaran N. et al., 2021](#)).

## 1.3 Aprendizaje automático

### 1.3.1 Conceptos básicos. Inteligencia artificial

La capacidad de realizar decisiones complejas en base a la realidad que nos envuelve es una de las principales diferencias entre humanos y máquinas ([Case R, 2019](#)). El desarrollo de los ordenadores hizo posible que, por primera vez, una máquina pudiese tomar decisiones a cierto nivel, pues antes de su aparición, las máquinas estaban limitadas a una repetición de una misma tarea una y otra vez.

Sin embargo, tradicionalmente los ordenadores solo pueden tomar decisiones cuando la respuesta “correcta” se ha programado de antemano, lo cual funciona para problemas

sencillos, pero no aquellos de alta complejidad, aquellos problemas en los que la lógica de la solución nos es parcialmente desconocida (Case R., 2019). Esto puede deberse o bien a la necesidad de un modelo complejo para la solución del problema o bien a la falta de parámetros para que un modelo funcione. El primer problema es aquello hacia lo que el término general *Inteligencia Artificial* (IA, o *AI* en inglés) suele orientarse, pero para el segundo, tal y como puede observarse en la figura 12, se usa un subconjunto de la inteligencia artificial, concretamente *aprendizaje automático*, conocido más comúnmente por su término inglés, *Machine Learning* (ML). El ML, a su vez, contiene un subconjunto de algoritmos llamados de *aprendizaje profundo* (DL en inglés), más modernos y complejos.

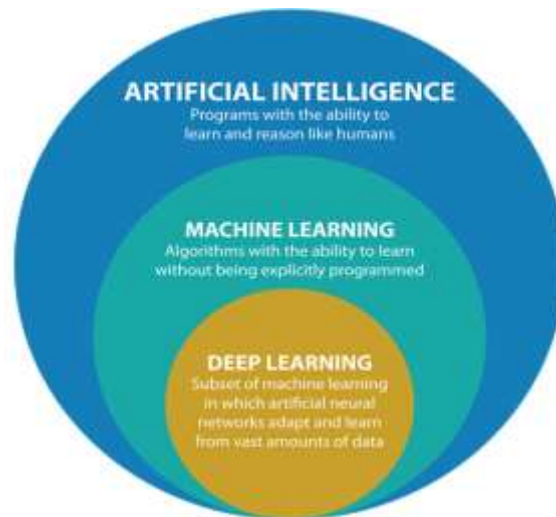


Figura 12. AI, ML y DL: diferencias. De (Opperman A., 2019)

Con tal de solucionar el problema al que se les aplica, los procesos de ML intentan encontrar, en un conjunto de datos determinado, la información necesaria para tomar el conjunto de decisiones que lleva a la resolución de dicho problema, o al menos una aproximación de la solución. Para ello el conjunto de datos seleccionado por el usuario debe contener la información que se considera de interés para llegar a la solución. La manera en la que se llega a la aproximación de la solución depende de la naturaleza del problema y el algoritmo usado, pero todos se basan en la estadística.

En resumen, dada una decisión en la que la lógica nos es desconocida, pero en la que tenemos observaciones sobre su proceso, es posible usar análisis estadísticos de dichas observaciones para crear una aproximación de la lógica de dicha decisión. O en otros términos más sencillos, pueden definirse las técnicas de aprendizaje automático como modelos computacionales que operan con los datos, buscando patrones con el objetivo de transformarlos en decisiones lógicas que permitirán resolver problemas complejos.

### 1.3.2 Contexto y evolución

Las técnicas de ML están basadas, en parte, en la actividad neuronal del cerebro humano ([Foote K.D., 2019](#)), y se remontan concretamente a un modelo definido por Donald Hebb ([Hebb D.,1949](#)). Este autor detalla: “cuando una célula asiste a que otra se active, el axón de la primera célula desarrolla perillas sinápticas (o las agranda si ya existen) en contacto con el soma de la segunda célula”. Esta y posteriores definiciones pueden describirse como una manera de alterar la relación entre diferentes neuronas, conectadas entre ellas, y los cambios en neuronas individuales. Esta relación se hace mayor si las dos neuronas se activan al mismo tiempo (mismo signo) y menor si se activan por separado (signos opuestos). Es aquí donde se introduce por primera vez el término “peso”, usado para describir dichas relaciones. Pesos elevados positivos se corresponden a neuronas que tienden a tener el mismo signo, mientras que pesos elevados negativos describirían la relación de neuronas con signos contrarios (en el caso más sencillo:  $1 \times 1 = 1$ ,  $-1 \times -1 = 1$ ,  $-1 \times 1 = -1$ ). Esta caracterización de los pesos y las neuronas serían algunos de los elementos fundamentales de las redes neuronales.

El término *machine learning* propiamente dicho se usó por primera vez en 1952 ([Foote K.D., 2019](#)), aplicándose a un algoritmo diseñado para ganar juegos de damas, que aprendía de partidas anteriores registrando todas las posiciones presenciadas hasta la fecha.

Sin embargo, lo que se conoce hoy como ML no se concibió hasta 1957, cuando se combinó el modelo neuronal de 1949 con el algoritmo de 1952, creándose el Perceptrón Mark 1, conocido hoy en día simplemente como *perceptrón*. Diseñado inicialmente como

máquina en vez de como programa, fue orientado al reconocimiento de imágenes, aunque debido a sus dificultades para reconocer un elevado número de patrones faciales se abandonó, y por ello este tipo de aprendizaje automático, que posteriormente se conocería como *redes neuronales*, no recuperaría su prominencia hasta los años 90.

A pesar de esto, hubo ciertos avances posteriores a su origen. En los 60' se descubrió que, al proporcionar dos o más capas al perceptrón, la capacidad de procesamiento aumentaba sustancialmente; es en este momento que nace el perceptrón multicapa (*MLP*) y el concepto de red neuronal artificial, conocidas por su acrónimo inglés, *ANN (Artificial Neural Networks)*, o más simple y comúnmente *NN (Neural Networks)*. A su vez, con el incremento del número de capas de la red neuronal fue necesario el desarrollo del algoritmo de entrenamiento de *backpropagation (propagación del error hacia atrás)* en la década de los 70', usado ampliamente en la actualidad para entrenar redes neuronales profundas (NNs con un elevado número de capas). Estos conceptos se estudiarán más en profundidad en el capítulo de aplicación de NNs sobre una base de datos concreta.

Hasta mediados de los 70', los algoritmos de ML simplemente se habían usado como programas de entrenamiento para IA, pero a partir de ahí ML y AI tomaron rumbos diferentes ([Foote K.D., 2019](#)), pues las décadas posteriores fueron caracterizadas por la importancia del análisis de los datos: en 1977 se publica *Exploratory Data Analysis* ([Tukey J.W., 1977](#)), y el mismo año se crea la Asociación Internacional de Computación Estadística. Este cambio se dio, entre otras causas, debido a la creciente cantidad de datos que se generaban incluso antes del auge de internet y la importancia de guardarlos, estructurarlos y procesarlos correctamente, aunque el término Ciencia de Datos no se utilizó hasta 1996 ([Duelo L., 2021](#)).

En gran parte para tratar con estos datos, nacieron las distintas categorías de ML (*que se verán en el apartado siguiente*) y se desarrollan procesos de ML no centrados en NNs, como el algoritmo *k vecinos más cercanos*, creado en 1967 y comúnmente conocido por su acrónimo inglés, *KNN (k-nearest neighbors)*, que constituyó el comienzo del reconocimiento básico de patrones. Este algoritmo consiste en situar un número *k* de

centros y calculando sus distancias a las entradas y en base a esta agrupándolas a un grupo u otra ([Dasgupta A. & Nath A., 2016](#)). Fue usado ampliamente para planificar las rutas de viaje más eficientes, y continúa siendo utilizado en la actualidad. También se crearon los llamados algoritmos *boosting*, diseñados para reducir el sesgo y la varianza durante el aprendizaje supervisado y para intentar transformar un conjunto de clasificadores débiles (aquellos con una correlación pequeña con la clasificación real) en uno más robusto.

En la década de los 90' fueron desarrollados algoritmos de reconocimiento del habla, utilizando redes neuronales profundas, es decir, redes con capas que pueden ser muy diferentes y con frecuencia con un elevado número de cada una de ellas. En los 00', el reconocimiento facial volvió a estar en alza desde sus fracasos de los 50', dando resultados hasta 100 veces más exactos que aquellos que se conseguían tan solo una década atrás. El reconocimiento facial se siguió desarrollando, con distintas grandes entidades (entre ellas Google y Amazon) creando y mejorando sus propios algoritmos y llegando a resultados de exactitud semejante a la de un humano.

El desarrollo de algoritmos de ML sigue en auge en la actualidad, constituyendo una necesidad inmediata debido a la realidad del *Big Data*, un gran número de datos siendo generados, guardados y procesados anualmente. Las aplicaciones de ML van desde reconocimiento de lenguaje hasta recomendaciones personalizadas de productos.

### 1.3.3 Categorías de aprendizaje automático

Tras introducir y contextualizar los modelos de aprendizaje automático, solo queda definir las tres categorías a las que pueden pertenecer: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo ([Dasgupta A. & Nath A., 2016](#)).

- Durante el **aprendizaje supervisado**, a la máquina se le provee tanto las entradas como las salidas que se desean obtener, y tras estudiarlas, encuentra una función general que permite obtener las salidas a partir de las entradas. En problemas de clasificación, se dan los posibles grupos a los que puedan pertenecer los datos. El aprendizaje supervisado se aplica a (1) problemas de

**regresión**, en los que se desea obtener una salida numérica (con frecuencia para hacer previsiones en el tiempo); y (2) a problemas de **clasificación** en los que se desea agrupar los datos en una categoría u otra. En general, **las redes neuronales entran en esta categoría**, al igual que las máquinas de soporte vectorial (SVM) y los algoritmos *KNN* (introducido en el apartado anterior) y *Neighborhood Component Analysis (NCA)*. El proceso de aprendizaje supervisado aparece esquematizado en la figura 13.



Figura 13. Esquema del proceso que ocurre durante el aprendizaje supervisado en problemas de clasificación.

- El objetivo del **aprendizaje no supervisado** es encontrar una representación interna correcta de los datos. El aprendizaje no supervisado se aplica a (1) problemas de **agrupación** (*clustering*), que se traduciría en encontrar los grupos a los que pueden pertenecer los datos; y a (2) problemas de **asociación**, en los que se busca una regla de asociación de los datos (por ejemplo, “la gente que compra X también compra Y”). El algoritmo *k-means clustering* es un ejemplo de técnica de aprendizaje no supervisado (Loukas S., 2020). Esta categoría de ML no supervisado aparece esquematizada en la figura 14.



Figura 14. Esquema del proceso que ocurre durante el aprendizaje no supervisado en problemas de clasificación.

- El **aprendizaje por refuerzo** no trata de generar una salida concreta, sino que consiste en hacer que el algoritmo aprenda cómo actuar dada una entrada, sin saber si se ha alcanzado el objetivo o no. En otras palabras, la salida es una acción o una secuencia de acciones que debe tomar un agente determinado, con un sistema de retroalimentación que da “recompensas” al algoritmo para que este aprenda qué acción es la mejor en una situación u otra, en base a la respuesta del entorno a la acción durante la iteración anterior. El proceso aparece esquematizado en la figura 15. Una de las técnicas más comunes que usan aprendizaje por refuerzo es el *Q-learning*.

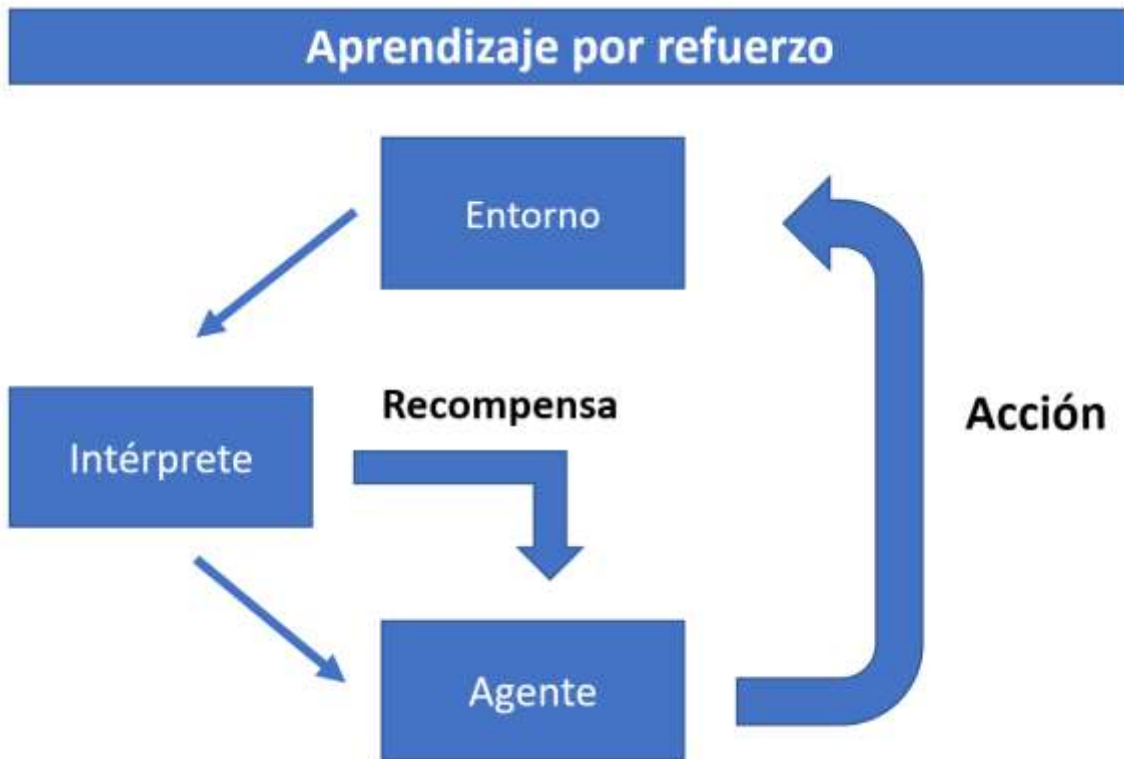


Figura 15. Esquema del proceso que ocurre durante el aprendizaje por refuerzo.

Aunque en la bibliografía puedan aparecer otros grupos, como el aprendizaje semisupervisado, son siempre una combinación de las tres categorías explicadas.

Puesto que en la presente tesis se quiere dar apoyo al diagnóstico, los grupos en los que se dividen los datos son conocidos: pacientes y enfermos. Por lo tanto, en la tesis se implementarán procesos de ML en problemas de **clasificación**, y por ello los procesos ML utilizados son del tipo **aprendizaje supervisado**. Las redes neuronales, siendo los algoritmos de esta categoría con la mayor prevalencia y uso más extendido, serán el foco del siguiente apartado y en general de la parte práctica de esta tesis.

#### 1.3.4 Procesos ML tradicionales

Tal y como ya se ha mencionado, las redes neuronales más prevalentes en la actualidad son las redes neuronales profundas. Sin embargo, también son de mayor complejidad y las que más tiempo tardan en entrenar y clasificar datos. Por ello, en la tesis se estudian primero no



las redes neuronales más actuales sino aquellos algoritmos de ML tradicionales que también puedan considerarse de interés dado el problema que afronta la tesis.

#### 1.3.4.1 Algoritmo k-nearest neighbors

El algoritmo *k vecinos más cercanos*, conocido por su acrónimo inglés, KNN (*K-Nearest Neighbors*) es un simple algoritmo de clasificación creado en los años 50' ([Lopez-Bernal D. et al., 2021](#)), constituyendo así uno de los primeros algoritmos de ML concebidos. Este método ya se ha usado antes para la clasificación de sujetos según tuviesen EM o no, concretamente usando MRI ([Rebbah S. et al., 2019](#)).

El método KNN guarda los datos de entrada junto a los grupos a los que pertenecen, y los usa a la hora de clasificar una nueva observación en un grupo u otro basándose en la similitud con los datos cuyo grupo es conocido, concretamente en las etiquetas de los puntos más próximos al del nuevo dato de entrada. Concretamente se buscan los *k* vecinos más próximos a ese nuevo dato y, considerándose sus etiquetas, se clasificará el nuevo punto según si la mayoría de las etiquetas de vecinos pertenecen a un grupo o a otro. A la hora de clasificar, con tal de evitar casos en los que haya un empate se considera de buenas prácticas elegir un número *k* impar.

Para medir los puntos más próximos, se calcula la distancia entre el nuevo punto y los datos de entrada; la forma más común de hacerlo es mediante la distancia euclídea, que no es más que la longitud de la línea recta que une ambos puntos y que, en dos dimensiones, se obtiene mediante:

$$d(x, y) = \sqrt{(x - y)^2} \quad (4)$$

, donde *d* representa la distancia entre dos puntos, y *x* e *y* representan las coordenadas de los puntos ([Lopez-Bernal D. et al., 2021](#)). Esta distancia se calcula para todos los datos de entrada; las *k* distancias menores son aquellas que señalan los puntos que serán considerados como los vecinos más cercanos.

La clasificación resultante del KNN depende en gran medida del número de vecinos más cercanos considerados, y de hecho probar diferentes valores de  $k$  hasta encontrar el óptimo para el registro que se está tratando es la tarea que lleva más tiempo al aplicar KNN. Sin embargo, cuando este valor óptimo de  $k$  ha sido hallado, el tiempo de entrenamiento pasa a ser muy bajo, lo que constituye una de las principales ventajas del KNN frente a otros métodos de clasificación. La segunda gran ventaja es la simpleza de implementar e interpretar. También constituye un algoritmo útil a la hora de tratar problemas con datos linealmente no separables ([Lopez-Bernal D. et al., 2021](#)), a diferencia de, por ejemplo, el perceptrón lineal, pues al considerar los vecinos más próximos no importa si estos están perfectamente agrupados a un lado u otro de una superficie separable por una línea, solo si están próximos.

El algoritmo tiene también algunas desventajas: su exactitud depende en gran medida de la calidad de los datos y el hecho de que es altamente sensible a ruido, y al aumentar el número de datos de entrada el tiempo de ejecución aumenta en enorme medida, más que otros algoritmos de ML.

#### 1.3.4.2 Support Vector Machine

Se plantea también el uso de un algoritmo ML ajeno a las NN: una máquina de soporte vectorial, conocida como *Support Vector Machine (SVM)*, desarrollado inicialmente para problemas de clasificación binaria, pero que se ha extendido a clasificación múltiple y regresión. Además, se considera competitivo incluso para conjuntos de datos multidimensionales ([Tzotsos, A. & Argialas, D., 2008](#)), que se basan en combinar varios clasificadores binarios. Las SVM ya se han usado antes en el diagnóstico de EM; en 2018 se usó para hacer la clasificación teniendo como datos imágenes de MRI ([Zhurita M. et al., 2018](#)).

Las SVM se basan en la idea de encontrar el hiperplano que mejor divida un conjunto de datos multidimensionales en dos clases, previamente conocidas. Las dimensiones del hiperplano dependen del número de características de entrada. Si hay dos, el hiperplano solo es una línea. En caso de haber tres, consistiría en un plano 2D. Por lo tanto, puede

definirse que, en un espacio euclídeo de dimensión  $n$ , un hiperplano es un subconjunto de dimensiones  $n-1$  de ese espacio, que divide dos partes de este (Pupale R., 2018).

En resumen, en un plano 2D el proceso que realiza un algoritmo SVM consiste en obtener la línea que mejor separe dos nubes de puntos. Para ello, se trazan una serie de líneas que separan estas dos distribuciones, y se encuentra el punto de cada una de las dos distribuciones más cercano a cada línea que se ha trazado (Pupale R., 2018). Una vez hecho esto se comparan las distancias, llamadas *vector soporte*, y se elige aquella línea que maximice su suma. El proceso presenta el mismo problema de limitación de la linealidad que se ha visto anteriormente: solo se puede encontrar el hiperplano ideal en problemas que permitan clasificación lineal; en la mayoría de los problemas prácticos, el algoritmo maximiza el margen flexible permitiendo un pequeño número de clasificaciones erróneas. A destacar que aparte de los dos puntos que definen los vectores soporte, el resto no se utiliza y por tanto puede ser excluido sin afectar al resultado.

Para la formulación del proceso, se usará (Fletcher T., 2008). Supóngase un conjunto de  $L$  puntos de entrenamiento, en el que cada punto  $x_i$  tiene un número  $D$  de atributos o características (por lo que tendrá  $D$  dimensiones), y es o bien de clase  $y_i=-1$  o bien de clase  $y_i=+1$ . Se asume que los datos son linealmente separables, por lo que es posible trazar una línea cuando  $D=2$ , un plano para  $D=3$  y un hiperplano para  $D>3$ . De la ecuación de un plano, este hiperplano puede describirse por:

$$w \cdot x + b = 0 \quad (5)$$

, en el que  $w$  es el vector normal a dicho hiperplano y  $\frac{b}{|w|}$  es la distancia perpendicular al origen de coordenadas. La figura 16 muestra el proceso para el caso descrito.

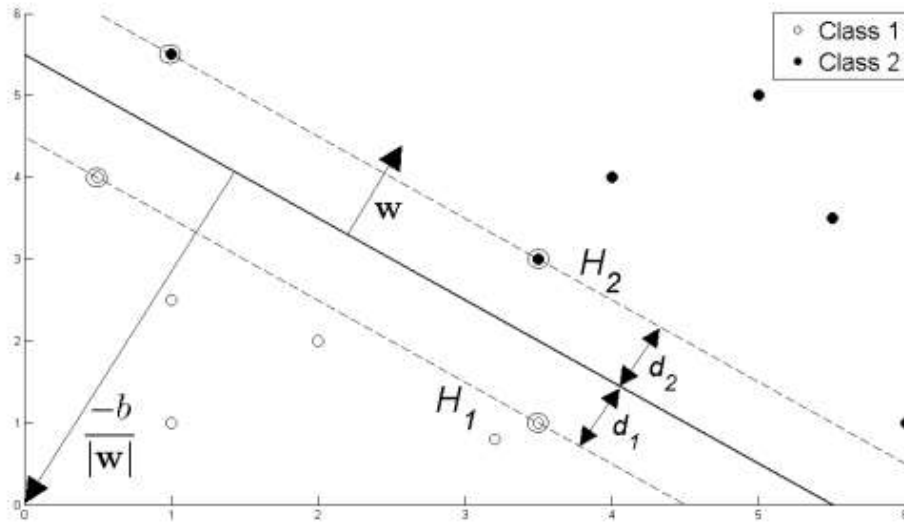


Figura 16. Hiperplano para el caso de dos clases linealmente separables (Fletcher T., 2008).

Implementar el SVM se resume en **seleccionar las variables  $w$  y  $b$**  tales que nuestro conjunto de entrenamiento pueda describirse como:

$$\begin{cases} x_i * w + b \geq 1 \text{ para } y_i = +1 \\ x_i * w + b \leq -1 \text{ para } y_i = -1 \end{cases} \quad (6)$$

, por lo que precisamente se tendrían dos subconjuntos de datos que, dependiendo de su posición, estarían en dos lados diferentes del hiperplano.

Estas ecuaciones pueden combinarse como:

$$y_i(x_i w + b) - 1 \geq 0 \quad \forall_i \quad (7)$$

Tal y como se ha explicado, debe encontrarse el punto de cada una de las dos distribuciones más cercano al hiperplano (marcados en la figura anterior mediante un círculo). Para ello, deben considerarse los planos que contengan estos dos puntos:

$$\begin{cases} x_i * w + b = +1 \text{ para } H_1 \\ x_i * w + b = -1 \text{ para } H_2 \end{cases} \quad (8)$$

Lo siguiente es referirse a las distancias  $d_1$  y  $d_2$  que se muestran en la figura anterior. Puesto que el hiperplano es equidistante a  $H_1$  y  $H_2$  se cumple que:

$$d_1 = d_2 - \frac{1}{\|w\|} \quad (9)$$

, donde la expresión  $\frac{1}{\|w\|}$  constituye el llamado *margen* del SVM. Para obtener el hiperplano que se encuentre a la mayor distancia posible de los vectores soporte, se debe **maximizar este margen**. Este es el objetivo del algoritmo de SVM, y aunque aquí no se seguirá presentando el proceso mediante formulación, maximizar el margen es similar a resolver un problema de optimización cuadrática. Para ello, se utilizan las llamadas funciones *kernel*, que asignan los datos a un espacio diferente, con frecuencia de mayores dimensiones, con la expectativa de que resulte más fácil separar las clases después de esta transformación, simplificando potencialmente los límites de decisión complejos no lineales para hacerlos lineales en el espacio dimensional de características superior asignado, solucionando su ya mencionada incapacidad inicial de resolver problemas más complejos.

En problemas de clasificación no lineal, es inevitable que haya clasificaciones erróneas, por lo que se añade un factor de penalización y las llamadas variables de holgura, que son las distancias desde el margen opuesto a la instancia inclasificable.

$$\min \frac{1}{2} \|W\|^2 + C \sum_{i=1}^L \varepsilon_i \quad (10)$$

, donde  $\varepsilon_i$  es cada una de las variables de holgura y C es la llamada *box constraint*, que aumenta el peso de las clasificaciones erróneas, lo que conlleva una separación más estricta. El resultado es un problema de optimización que siempre tiene solución, aunque los datos no sean linealmente separables.

Sin embargo, antes de poder aplicarse las funciones *kernel* debe hacerse antes la ya mencionada **selección de características**, que debe realizarse tanto sobre el conjunto de datos de entrenamiento como sobre el de validación. La extracción de características no es

inmediata, y otros estudios antes utilizan distintos tipos de imágenes y distintas regiones anatómicas como una característica u otra ([Zhurita M. et al., 2018](#)), entendiéndose como características, en este caso, un tipo u otro de imagen de entrada. En este estudio en concreto se descartan las características irrelevantes en base al denominado criterio de la puntuación de Fischer, y se construye un elevado número de clasificadores, mediante el método *leave one out*, el mismo usado en esta tesis (entrenar el algoritmo con todos los sujetos excepto uno y validarlo con el que se ha descartado; repetir el proceso para todos los sujetos). En el caso de esta tesis se tienen imágenes de diferentes capas, por lo que *a priori* podría pensarse en considerarse cada capa un tipo de imagen, un tipo de característica. Otra solución es extraer características de cada imagen, independientemente de la capa a estudiar, mediante otro algoritmo, y pasar estos resultados a la función *kernel* para que las transforme a parámetros del hiperplano.

### 1.3.5 Redes neuronales. Conceptos básicos

#### 1.3.5.1 Arquitecturas comunes

En los últimos años las redes neuronales, concretamente las redes neuronales profundas, se han convertido en el método de ML más usado para la mayoría de las aplicaciones ([Alpaydin E., 2020](#)). Puesto que, como se ha expuesto en la sección anterior, es frecuente su uso en procesos de ML entrenados con aprendizaje supervisado, las NN serán también el foco principal de la implementación de ML en la tesis.

Su función, exactitud y velocidad de ejecución dependerá de la arquitectura de la red y del conjunto de datos que se use para entrenarla, que deberán ajustarse a la aplicación concreta de cada estudio. Antes de la implementación sobre la base de datos, que se presentará en los capítulos sucesivos, se expondrán conceptos básicos sobre la arquitectura, aprendizaje y tipos de redes neuronales.

A pesar de que sus funciones y parámetros pueda variar en gran medida entre un modelo u otro, las redes neuronales comparten la misma arquitectura básica: se componen por una capa de entrada, que recibirá los datos; por una serie de capas intermedias, comúnmente

referidas como *capas ocultas*, que operarán sobre ellos; y por una capa de salida, que presentará los resultados en el formato conveniente, ya sea en una serie de pesos ponderados o un porcentaje (%). En la figura 17 puede observarse este esquema.

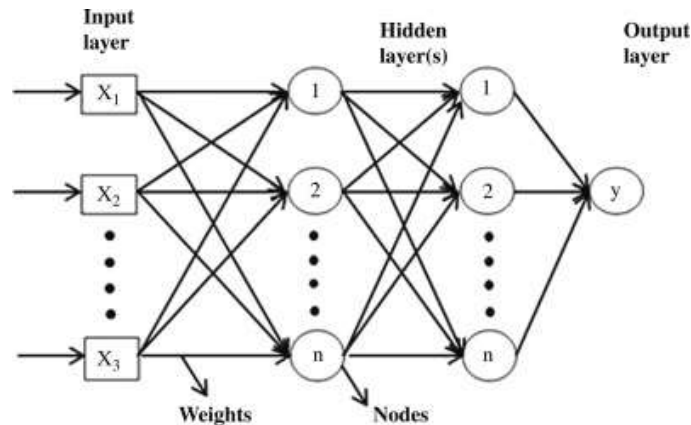


Figura 17. Arquitectura de una red neuronal común.

Las capas ocultas son las encargadas de operar con los datos y buscar patrones sobre ellos. Concretamente son sus nodos, también referidos como *neuronas*, los que realizan este proceso, y lo hacen mediante su **función de activación**, que aceptará las entradas de forma ponderada, realizará una operación determinada sobre ellas y pasará la salida a uno o más nodos de otra capa de la red neuronal.

En la bibliografía, el término Red Neuronal Profunda, más conocido como su nombre inglés, *Deep Network (DN)* suele usarse a toda NN con “múltiples capas ocultas”, por lo que con 2 es suficiente para considerarse tal ([Bevilacqua V. et al., 2019](#)). Otras fuentes no son del todo específicas, llamando arquitecturas profundas a aquellas que “están compuestas por múltiples niveles de operaciones no lineales, como ocurre las redes neuronales con muchas capas ocultas” ([Bengio Y., 2009](#)). Aunque de nuevo, el consenso sobre cuándo usar esta nomenclatura no está del todo claro, en esta tesis **se considerará DN a toda red con al menos dos tipos distintos de capas ocultas**, como las redes neuronales convolucionales, que se verán en el siguiente apartado.

Cada capa de una red neuronal puede tener un número diferente de nodos, y cada nodo consiste en una función de activación, pero en una capa determinada todos los nodos

realizan la misma función. Por ello, a menudo es la capa la que recibe una designación diferente dependiendo de la operación que realizan sus nodos. Ejemplos de esto son las capas convolucionales (*convolutional layers*) y las capas de agrupación (*pooling layers*), que se verán en profundidad en capítulos posteriores.

En la figura 18 pueden verse dos funciones de activación: una lineal y una no lineal, concretamente del tipo sigmoïdal, ampliamente usada en aquellas aplicaciones en las que se quiere obtener una probabilidad concreta, pues su resultado será siempre entre 0 y 1.

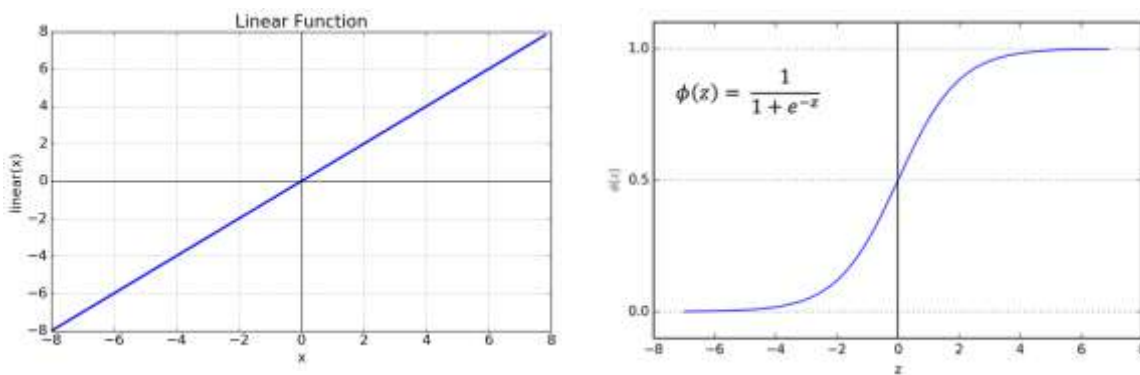


Figura 18. Ejemplos de funciones de activación. (a) Lineal (b) No lineal, sigmoïdal ([Sharma S., 2017](#)).

El uso de una función de activación u otra dependerá de la aplicación que se quiera dar a la red, pero en general, se usan las funciones de activación no lineales, puesto que son las que permiten resolver problemas no triviales ([Sharma S., 2017](#)). Después de todo, las NN con ecuaciones lineales pueden simplificarse en una sola capa y una sola capa no es capaz de resolver problemas con la exactitud deseada. Esto es debido a que independientemente de las capas que se encadenen a una NN cuyas capas contengan exclusivamente neuronas con funciones de activaciones lineales, la ecuación final será siempre lineal.

Para entender mejor estos conceptos se explicará en detalle el caso de la primera red neuronal: el perceptrón. El modelo del perceptrón, tal y como se observa en la figura 19, consiste en una combinación lineal de una serie de entradas: multiplica cada entrada,  $x$ , por su correspondiente peso,  $w$ , y suma los resultados de cada uno de estos productos para

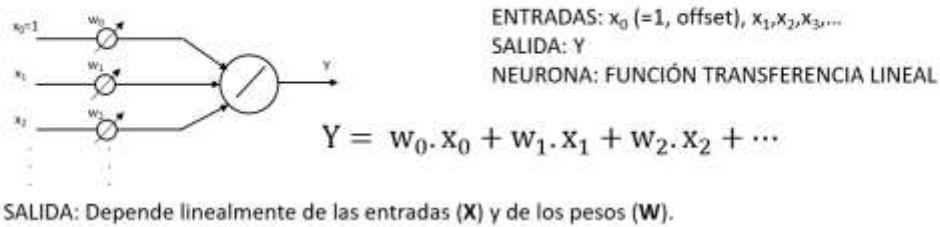


obtener una salida Y. El número de entradas indica también el número de dimensiones; para explicar los conceptos, se supondrán dos entradas,  $X_1$  y  $X_2$ , junto a sus respectivos pesos,  $W_1$  y  $W_2$ . Dado esto, la ecuación del perceptrón es:

$$Y = W_0 \cdot X_0 + W_1 \cdot X_1 + W_2 \cdot X_2 \quad (11)$$

, donde Y es la salida del perceptrón. Además de los pesos correspondientes a cada una de las entradas también hay una constante,  $X_0=1$ , con la que se acaba formando la ecuación de la recta (para  $W_0=0$ , la recta pasa por el origen de coordenadas).

**PERCEPTRÓN**



Con dos entradas ( $x_1, x_2$ ) y suponiendo que la salida Y se compara con 0 ( $Y > 0$  DECISIÓN1,  $Y < 0$  DECISIÓN2):

$$W_0 + W_1 \cdot X_1 + W_2 \cdot X_2 = 0$$

$$X_2 = -\frac{W_1}{W_2} \cdot X_1 - \frac{W_0}{W_2} \cdot 1$$

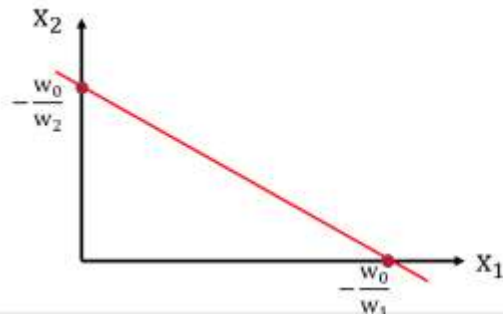


Figura 19. Perceptrón, ideas generales.

Así, para cada nuevo punto, de coordenada ( $X_1, X_2$ ), que se quiera clasificar en dos grupos, si la Y resultante de la combinación lineal es superior a 0, el punto pertenece al grupo 1, y si es inferior a 0, pertenece al grupo 2. Lo que está haciendo el perceptrón mediante esta operación es separar dos áreas mediante una recta; suponiendo la salida 0, la ecuación de la

recta dictada por la fórmula de la figura; por lo tanto, el valor de los pesos condiciona la pendiente de la recta, y por lo tanto la decisión que toma el perceptrón. Este ejemplo se hace en dos dimensiones, pero para dimensiones mayores, la recta se convierte en planos, hiperplanos, etc.

### 1.3.5.2 Aprendizaje de una red neuronal. El perceptrón.

Ahora bien, con tal de construir la recta que permita tomar la decisión correcta como clasificador (generar la salida adecuada a las entradas), es necesario un valor determinado para los pesos, diferente al que se genera aleatoriamente al principio. Para lograr esto, primero debe operar sobre un conjunto de entrenamiento cuya resolución, sea la categoría a la que pertenecen o algún otro parámetro que se quiera calcular, sea conocida, es decir, debe antes aprender de salidas conocidas y correctas a unas entradas determinadas. A este proceso se le llama **aprendizaje de una red neuronal**. Y con el fin de realizar este aprendizaje, es necesario saber en cada iteración cuánto se desvía el resultado de la red neuronal, dada una entrada de salida conocida, de dicho valor del que se dispone. En otras palabras, es necesario conocer por un lado el error y por otro el algoritmo que permita actualizar los pesos a partir del error calculado.

En primer lugar, supóngase entradas y salidas binarias, por ejemplo, las correspondientes a la función AND, tal y como se observa en la figura 20.

Función: AND		
$x_2$	$x_1$	$y_{DESEADA}$
0	0	0
0	1	0
1	0	0
1	1	1

Figura 20. Función lógica AND aplicada al perceptrón.

Teniendo esto en cuenta la función de error, comúnmente conocida por su nombre inglés, *loss function*, se calcula mediante:

$$E(k) = \frac{1}{2}(y_0 - y_{deseada0})^2 + \frac{1}{2}(y_1 - y_{deseada1})^2 + \frac{1}{2}(y_2 - y_{deseada2})^2 + \frac{1}{2}(y_3 - y_{deseada3})^2 \quad (12)$$

, donde  $k$  es la iteración en la que se encuentra el algoritmo de entrenamiento en un momento dado;  $y_0, y_1, y_2$  e  $y_3$  son las salidas de la red neuronal para las entradas  $X_0$  ( $x_1, y_2$ ),  $X_1$  ( $x_1, x_2$ ),  $X_2$  ( $x_1, x_2$ ) y  $X_3$  ( $x_1, x_2$ ) respectivamente, mientras que cada  $y_{deseada}$  corresponde a la salida, conocida, para cada entrada  $X$ . Sin embargo, en vez de usarse (12), con frecuencia se utiliza el error instantáneo tras obtener la salida a cada una de las entradas por separado:

$$E(k) = \frac{1}{2} ( y(k) - y_{deseada}(k) )^2 \quad (13)$$

Puesto que es un error instantáneo, se cumple que:

$$\frac{dE(k)}{dy} = (y(k) - y_{DESEADA}(k)) = e(k) \quad (14)$$

En segundo lugar, el ajuste de pesos, que con frecuencia se hará mediante el llamado **descenso por el gradiente**, calculado mediante la derivada de la función del error respecto a cada peso:

$$\Delta W(k) = -\alpha \frac{dE(k)}{dW} \quad (15)$$

, donde  $\Delta W$  es el cambio a realizar en cada peso ( $w_0, w_1$  y  $w_2$ ) durante cada iteración  $k$  del algoritmo, y  $\alpha$  es el factor de aprendizaje, más comúnmente conocido como *learning rate*, que regula este cambio. Así, cada peso de la iteración actual de entrenamiento quedará definido por:

$$W(k) = W(k - 1) - \alpha \frac{dE(k)}{dW} \quad (16)$$

De (7) y (8), puesto que el error depende de la salida:

$$\begin{aligned} \Delta W(k) &= -\alpha \frac{dE(k)}{dW} = -\alpha \frac{dE(k)}{dY(k)} \frac{dY(k)}{dW} = \\ &= -\alpha (Y(k) - Y_{DESEADA}(k)) x(k) \end{aligned} \quad (17)$$

Así, este método actualiza los valores en la dirección contraria del gradiente del error con respecto al peso que se ajusta, es decir, en la dirección hacia la que el error decrece. Aunque el algoritmo no asegure que el error decrezca lo más rápidamente posible, sí asegura que decrezca en esa dirección.

En la figura 21 aparece la gráfica del error a medida que el peso se va actualizando. Iteración tras iteración el error se va desplazando hacia el mínimo absoluto de la curva.

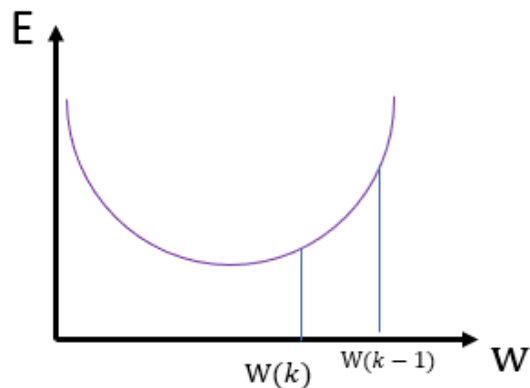


Figura 21. Error en función del peso.

Mediante el parámetro *learning rate*, se puede controlar la actualización de pesos; haciendo que sea pequeño, será más fácil llegar al mínimo del error sin pasarse (en cuyo caso podría ocurrir que se oscila continuamente alrededor del mínimo error sin lograr la detención), pero también se requerirá un elevado número de iteraciones para minimizar el error, lo que aumenta el tiempo de entrenamiento. Es para intentar solucionar este potencial problema que se originó una variante de este método, la de *gradient descent with momentum*. En este

caso, tras cada actualización de pesos, no se cambia la dirección completa sino parcialmente a la del nuevo gradiente, evitando el cambio brusco. De (8):

$$\Delta W(k) = -\alpha \frac{dE(k)}{dW} + \beta \Delta W(k - 1) \quad (18)$$

, donde  $\beta$  es una constante que controla la actualización del *momento* (y por eso también se le suele llamar así). Al igual que  $\alpha$ , el momento se especifica al principio del algoritmo, pero puede actualizarse a medida que el algoritmo se acerca al error.

En el descenso por el gradiente realmente lo que más importa es el sentido (+ o -) del incremento que debe darse a cada peso, puesto que la magnitud supondrá que se tarde más o menos ciclos de entrenamiento. Además, la curva del error en el perceptrón no tiene mínimos locales, por lo que se evita el problema de que la actualización de pesos se encasille en el mínimo encontrado en una determinada posición de la curva, sino que siempre se tenderá hacia el mínimo absoluto (aunque puede ocurrir que de forma natural sea muy elevado).

La diferencia entre ambos procesos aparece esquematizada en la figura 22.

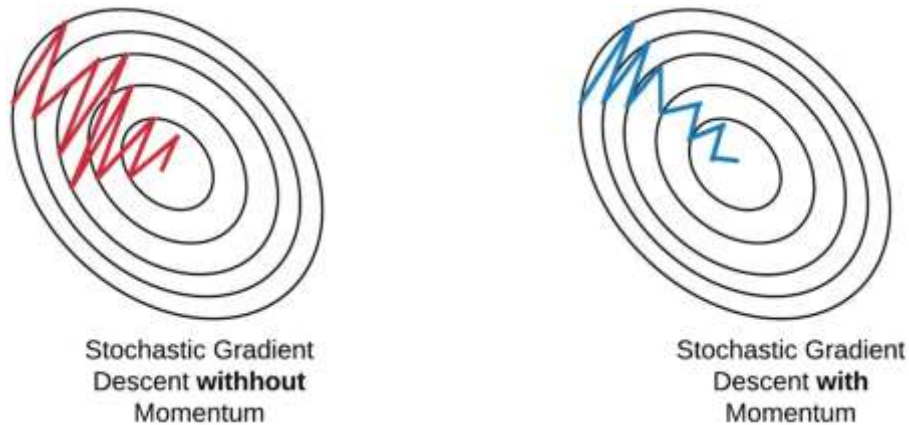
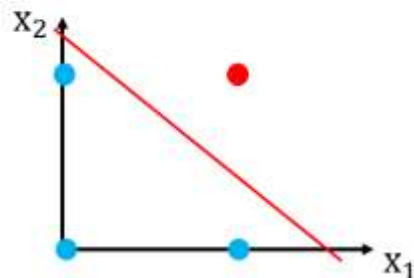


Figura 22. Entrenamiento usando el gradiente de forma total (izquierda) o de forma gradual (derecha).

Este modelo básico de perceptrón soluciona problemas linealmente separables. Representando entradas y salidas bipolares, tal y como se observa en la figura 23, estos serían los correspondientes a las salidas de las funciones lógicas AND y OR, pues es en estos casos que existe una recta que clasifica correctamente las entradas binarias como del grupo 1 o 2. Lo mismo no pasa en casos como los de la función XOR, en los problemas no linealmente separables.

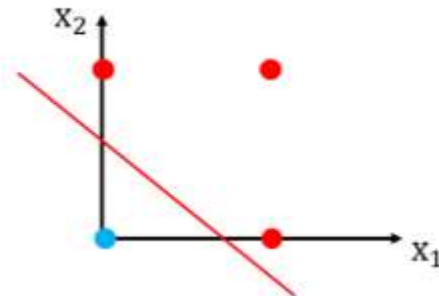
Función: AND		
$x_2$	$x_1$	$y_{DESEADA}$
0	0	0
0	1	0
1	0	0
1	1	1

Función AND: problema LINEALMENTE SEPARABLE .



Función: OR		
$x_2$	$x_1$	$y_{DESEADA}$
0	0	0
0	1	1
1	0	1
1	1	1

Función OR: problema LINEALMENTE SEPARABLE .



Función: XOR		
$x_2$	$x_1$	$y_{DESEADA}$
0	0	0
0	1	1
1	0	1
1	1	0

Función XOR: problema NO LINEALMENTE SEPARABLE .

El perceptrón no tiene capacidad para resolver este tipo de problemas

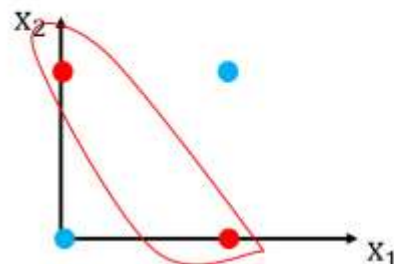


Figura 23. Implementación de las funciones lógicas AND, OR y XOR usando el perceptrón.

Para solventar esto, se introduce el perceptrón no lineal, capaz de solucionar problemas más complejos. Para ello deben introducirse no linealidades en la ecuación de la red, lo que se consigue cambiando la función de transferencia.

De la ecuación del perceptrón:

$$S = w_0 \cdot x_0 + w_1 \cdot x_1 + w_2 \cdot x_2 + \dots \quad (19)$$

De (12), si la función de transferencia no lineal elegida es la sigmoide, se tiene que:

$$Y = f(S) = \frac{1}{1+e^{-S}} = \frac{1}{1+e^{-(w_0 \cdot x_0 + w_1 \cdot x_1 + w_2 \cdot x_2 + \dots)}} \quad (20)$$

Y así, se obtiene una NN cuya salida depende de forma no lineal de las entradas y los pesos. El error, al igual que con el perceptrón básico, viene dado al aplicar la derivada, en este caso de (13):

$$\frac{dy}{dS} = \frac{e^{-S}}{(1+e^{-S})^2} = y(1 - y) \quad (21)$$

El entrenamiento también ocurre de forma equivalente al que se ha visto. Desarrollando (15) de la misma forma que se ha hecho en (17) pero aplicando (21):

$$\begin{aligned} \Delta W(k) &= -\alpha \cdot \frac{dE(k)}{dW} = -\alpha \cdot \frac{dE(k)}{dy(k)} \cdot \frac{dy(k)}{dS(k)} \cdot \frac{dS(k)}{dW} = \\ &= -\alpha(Y(k) - Y_{DESEADA}(k))Y(k)(1 - Y(k))X(k) \end{aligned} \quad (22)$$

Para finalizar el apartado, el tipo de redes que con más frecuencia usan el método de *backpropagation* son las redes neuronales *Feed Forward (FF)*. Estas redes se caracterizan por mandar la información de una capa a otra sin que pasen por una capa anterior; por lo tanto, para generar una salida las entradas pasarán siempre hacia los nodos posteriores. Las redes FF son el tipo de red más antiguo, como se vio en apartados anteriores, y también uno

de los más usados hoy en día. Muchas redes profundas, como las *Redes Neuronales Convolucionales (RNC)*, especializadas en problemas de clasificación, son del tipo FF. Estas redes, así como el modelo más complejo del perceptrón, el *Perceptrón Multicapa*, se verán en el siguiente apartado de esta sección.

### 1.3.6 Redes empleadas

Vistos los algoritmos de ML tradicionales, se plantea también el estudio de redes neuronales profundas. A diferencia de los algoritmos vistos hasta ahora, estas redes aprenden un conjunto de representaciones o características de los datos de entrada de forma jerárquica. Las hay simples, como el perceptrón multicapa, pero también más complejas, como las redes convolucionales, con una complejidad mayor debido al elevado y diverso número de capas que componen sus arquitecturas, aunque también son capaces de alcanzar una exactitud mayor, como se ha mencionado en secciones anteriores. Mediante arquitecturas que contienen un diverso número de capas, aprenden características de bajo, de medio y de alto nivel. En el caso de que los datos sean imágenes, esto es análogo de aprender bordes (bajo nivel), contornos completos (medio nivel) y finalmente objetos enteros (alto nivel).

#### 1.3.6.1 Perceptrón Multicapa.

Tal y como su nombre indica, el perceptrón multicapa (*MLP*) es una variante de mayor tamaño del perceptrón (compuesto por una sola capa oculta y un solo nodo y ya visto en el apartado anterior). En la figura 24 puede verse la arquitectura y principio del MLP: cada entrada viene multiplicada por el peso correspondiente a cada capa en la que entra.



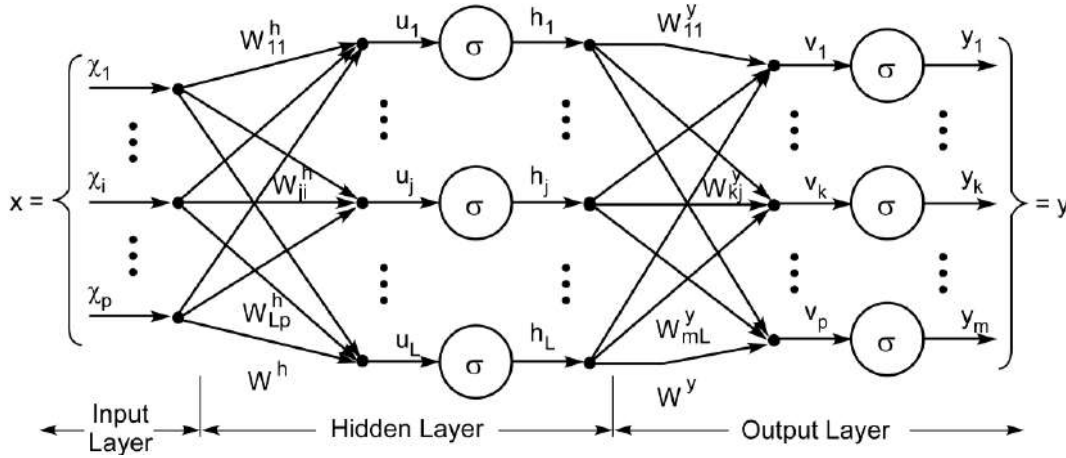


Figura 24. Arquitectura general de un MLP (Faghfour A.E. et al., 2011)

En el MLP cada capa interna puede considerarse lo que en otras redes se llama *Fully Connected Layer (FLC)*, puesto que la salida de cada neurona está conectada a todas las neuronas de la siguiente capa oculta.

De la figura anterior, la entrada  $u$  a cada neurona  $j$ , puede definirse como:

$$u_j = w_j \sum_{i=1}^{i=p} x_i \quad (23)$$

, donde  $w_j$  es el peso de esa neurona en concreto y  $x_i$  cada una del total de entradas,  $p$ , a la red neuronal. A partir de aquí, la salida de cada neurona,  $h$ , vendrá dada por la función de activación de las neuronas de esa capa,  $\sigma$ :

$$h_j = \sigma(u_j) \quad (24)$$

Como ya se ha establecido, cada función de activación será más o menos adecuada a la red neuronal dependiendo del problema con que se esté tratando. Sin embargo, dadas estas ecuaciones es fácil discernir el problema de las funciones de activación lineales, introducido en la sección anterior. De la ecuación anterior, y teniendo en cuenta la ecuación de la línea sin la constante, se tendría:

$$h_j = a * u_j \quad (25)$$

, siendo “a” la pendiente. Al sustituir de la primera, se pasaría a tener:

$$h_j = a * w_j \sum_{i=1}^{i=p} x_i \quad (26)$$

, y por lo tanto no se tiene más que una combinación lineal, independientemente de las capas que se le añadan, puesto que al sumatorio de las entradas simplemente se le multiplicaría a un coeficiente por cada una de las capas por las que iría pasando la entrada original. Como ya se ha mencionado en el apartado anterior, no es capaz de solucionar problemas lineales por más que se le añadan capas y neuronas.

Supóngase ahora un ejemplo particular de MLP con función de activación sigmoide (y por lo tanto no lineal), y una arquitectura de una sola capa con tres neuronas y dos salidas, tal y como la que se observa en la figura 25.

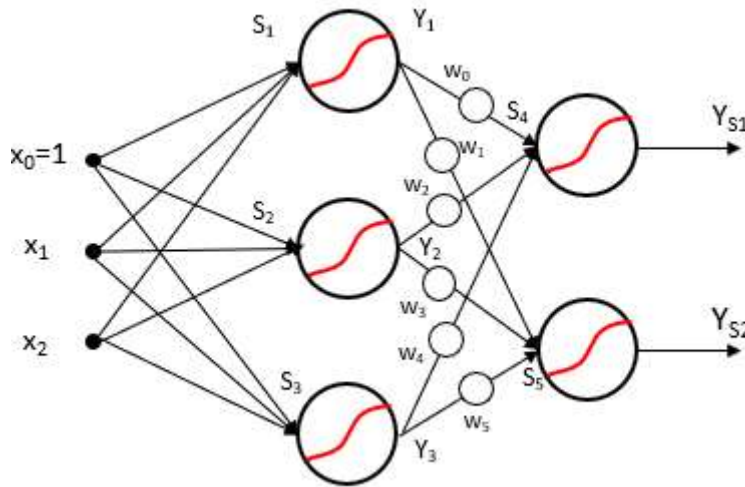


Figura 25. Ejemplo concreto de arquitectura de MLP. Función de activación sigmoideal (no lineal).

Para obtener las salidas se tendría:

$$S_1 = S_2 = S_3 = x_0 + x_1 + x_2$$

$$Y_1 = f(S_1) = \frac{1}{1+e^{-S_1}}; Y_2 = f(S_2) = \frac{1}{1+e^{-S_2}}; Y_3 = f(S_3) = (S_3) = \frac{1}{1+e^{-S_3}};$$

$$S_4 = w_0 Y_1 + w_2 Y_2 + w_4 Y_3$$

$$S_5 = w_1 Y_1 + w_3 Y_2 + w_5 Y_3$$

$$Y_{S1} = f(S_4) = \frac{1}{1+e^{-S_4}} ; Y_{S2} = f(S_5) = \frac{1}{1+e^{-S_5}} \quad (27)$$

Al haber introducido no linealidades en la función de activación, ahora la salida depende de forma no lineal de las entradas. Ahora es posible solucionar problemas más complejos.

El entrenamiento se realiza de forma semejante a lo visto al introducir el perceptrón básico, con *backpropagation*. Adaptando (13) a este ejemplo:

$$E(k) = \frac{1}{2} ( Y_{S1}(k) - Y_{DESEADA_1}(k) )^2 + \frac{1}{2} ( Y_{S2}(k) - Y_{DESEADA_2}(k) )^2 \quad (28)$$

Para cada iteración k, el ajuste de pesos se realiza mediante:

$$\begin{aligned} \Delta W &= -\alpha \frac{dE}{dW} = -\alpha \frac{dE}{d(Y_{S1} + Y_{S2})} \frac{d(Y_{S1} + Y_{S2})}{dW} = \\ &= -\alpha \frac{dE}{d(Y_{S1})} \frac{d(Y_{S1})}{dW} - \alpha \frac{dE}{d(Y_{S2})} \frac{d(Y_{S2})}{dW} = \\ &= -\alpha (Y_{S1}(k) - Y_{DESEADA_1}) \frac{dY_{S1}}{d(S_4)} \frac{d(S_4)}{dW} - (-\alpha) (Y_{S2}(k) - Y_{DESEADA_2}) \frac{dY_{S2}}{d(S_5)} \frac{d(S_5)}{dW} \quad (29) \end{aligned}$$

Por lo que por ejemplo para el caso de  $w_0$  la actualización de pesos quedaría como:

$$\Delta w_0(k) = -\alpha \left( Y_{S1}(k) - Y_{DESEADA_1}(k) \right) Y_{S1}(k) (1 - Y_{S1}) Y_1 \quad (30)$$

El ajuste de cada peso en el instante (k) puede realizarse teniendo en cuenta el ajuste que ha tenido en (k-1), y puede añadirse, además del *learning rate*, el *momento*.

A destacar que en una red multicapa sí puede haber mínimos locales. En un mínimo local la derivada del error con respecto a los pesos es 0, y la red queda atrapada en esa solución subóptima. La solución es o bien entrenar la red con un *learning rate* diferente o entrenar la

red varias veces; al inicializarse los pesos de forma aleatoria, en cada entrenamiento se generarán valores diferentes, por lo que simplemente se elige la opción con el mínimo error ([Atakulreka A. & Sutivong D., 2007](#)).

### 1.3.6.2 Radial Basis Function Networks

Las *Radial Basis Function Networks* (*RBFN* o simplemente *RBF*) son redes neuronales de dos capas que, al igual que los MLP, pueden usarse como solución a problemas de clasificación. Ya se ha habido estudios que hacen uso de estas RNs como apoyo al diagnóstico médico, por ejemplo, al de enfermedades cardíacas ([Alsalamah M. et al., 2015](#)).

Para entender fácilmente el método que emplean las RBF para este tipo de problemas, supóngase un plano en el que se dispone una serie de puntos a clasificar. A grandes rasgos, el algoritmo traza una serie de círculos en el dominio en el que se encuentran los datos, y divide estos datos en un grupo u otro según se encuentren dentro o fuera de estos círculos. Este proceso puede verse esquematizado en la figura 26, en la que también puede observarse la arquitectura común a estas redes.

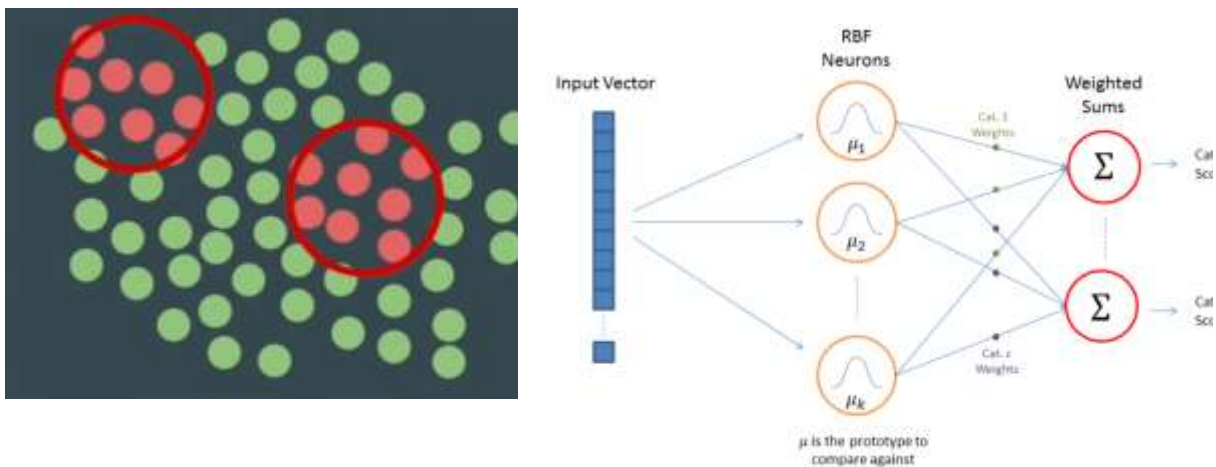


Figura 26. Fundamentos de las RBF. (a) Modo de operación en 2D (b) Arquitectura común

Aunque al igual que el MLP todas las RBF sean redes del tipo FF, tienen en común que su arquitectura es de dos capas:

- Una capa oculta, cuyas neuronas comparten una misma función de activación con base radial (de ahí el nombre de la red), aunque ubicadas en puntos diferentes del espacio de entrada. Dada la naturaleza de dicha función, esta capa hará una transformación no lineal de los datos, otorgando a las RBF la capacidad de resolver problemas no lineales. Es una FCL, por lo que todos los datos de entrada llegan a todas las neuronas.
- Una capa de salida, que, en problemas de clasificación, mediante una función lineal obtiene la clase más probable a la que pertenecen las entradas.

En problemas de clasificación, el objetivo es determinar si un nuevo dato de entrada se halla en una categoría u otra. Suponiendo un espacio 2D y dos categorías en los que clasificar los datos, como ya se ha mencionado la RBF traza una serie de círculos que contienen los puntos de una de las categorías, dejando fuera los puntos de la otra categoría; esta es la fase de entrenamiento. Una vez realizado este proceso, para cada nuevo punto cuya clasificación es desconocida se calcula la distancia entre este nuevo punto y el centro del círculo; si la distancia es demasiado elevada, el algoritmo decidirá que no se halla dentro de la circunferencia y por lo tanto no formará parte de la clase contenida en dicho círculo.

Sin embargo, a menudo no se realiza el proceso con umbrales estrictos a partir de los cuales se dicte con toda seguridad que un nuevo punto se encuentre fuera de ellos, sino que se usan círculos suavizados, con determinadas transiciones que informen sobre la *probabilidad* de que el punto se halle fuera o dentro del círculo. Con esta idea en mente, en las RBF no se usa la distancia por sí sola, sino que al resultado se le multiplica un valor  $b$ , llamado frecuentemente *factor de escala*, con tal de poder modularlo:

$$d_i = \left| |p - c_i| \right| \cdot b_i \quad (31)$$

, donde  $c$  es el centro correspondiente a cada neurona  $i$  y  $p$  es cada una de las entradas, pues al igual que en el MLP, toda neurona de la que se compone la primera capa de la RBF recibirá todas las entradas. El factor de escala  $b$  puede ser el mismo para todas las neuronas.

Las distancias de cada punto a los centros se calculan en todas las dimensiones de entrada; en el caso de que lo que se quiera clasificar sean imágenes, el número de dimensiones será de tantos píxeles como tenga dicha imagen.

El resultado de (31) será pasado a la función de activación, que en RBFs suele consistir en una función gaussiana, la llamada función de base radial, definida por:

$$g_i = e^{-d_i^2} \quad (32)$$

Una propiedad importante de esta función es que es local: a mayor distancia ( $d$ ) menor será la salida de la función de activación ( $g$ ). Esta es una de las mayores diferencias con las funciones sigmoideas, usadas en los MLP, en los que la salida tiende a 1 cuando la entrada se acerca a infinito.

El resultado de la función de base radial pasa a la segunda capa, la capa de salida de la RBF, correspondiente a un clasificador lineal, en la que se realiza la suma ponderada de cada salida de las neuronas; es aquí donde participan los pesos de la función, teniendo que finalmente la salida es:

$$y_j = w_{ij} \cdot g_i \quad (33)$$

, donde  $y$  es una salida  $j$  de la red neuronal y  $w$  es el peso asociado a la neurona  $i$  de la capa oculta de la red y a la salida  $j$ .

Los pesos con frecuencia son inicializados de forma aleatoria. Para actualizarlos, puede usarse el algoritmo de descenso por el gradiente como se ha visto en el caso del MLP; la diferencia es que en vez de trabajar con las entradas directas se trabaja con las salidas de las funciones de base radial. Aplicando de nuevo (13), se tiene que hay que resolver, para cada instante  $k$ :

$$E(k) = \frac{1}{2} ( y(k) - y_{deseada}(k) )^2$$

$$y_j(k) = w_{1j} \cdot g_1 + w_{2j} \cdot g_2 + \dots + w_{ij} \cdot g_i$$

$$\Delta W(k) = -\alpha \cdot \frac{dE(k)}{dW} = -\alpha \cdot \frac{dE(k)}{dy_j(k)} \cdot \frac{dy_j(k)}{dW} = -\alpha \cdot e_j(k) \cdot g_i(k) \quad (34)$$

Otro de los algoritmos de entrenamientos más común en estas redes, aún usado a pesar de su antigüedad, es el llamado *Orthogonal least squares learning algorithm (OLS)* ([Chen S. et al., 1991](#)). No se incluirá la formulación, pero con este método durante el entrenamiento se actualizan tanto los pesos de esta capa de salida como el número y la posición de los centros en la capa oculta, los cuales son añadidos uno a uno hasta dar con el error mínimo.

De esta capa de salida se obtiene la probabilidad de que el nuevo punto de entrada se halle fuera o dentro del círculo, y por lo tanto la clase más probable a la que corresponde la entrada, de forma equivalente al MLP.

### 1.3.6.3 Redes Neuronales Convolucionales

Las Redes Neuronales Convolucionales (*RNC*), conocidas comúnmente por su término inglés, *Convolutional Neural Networks (CNN)* son redes neuronales profundas usadas extensamente hoy en día en problemas de clasificación. Aunque fueron inventadas en los años 80', debido a sus requisitos en cuanto a velocidad de procesamiento, su auge no ocurrió hasta la década de los 2000 ([Chellapilla K. et al., 2006](#)).

En la figura 27 aparece la arquitectura de una CNN común. Las capas características de las CNN son las capas de convolución, las *ReLU*s, las *Pooling* y las FCL.

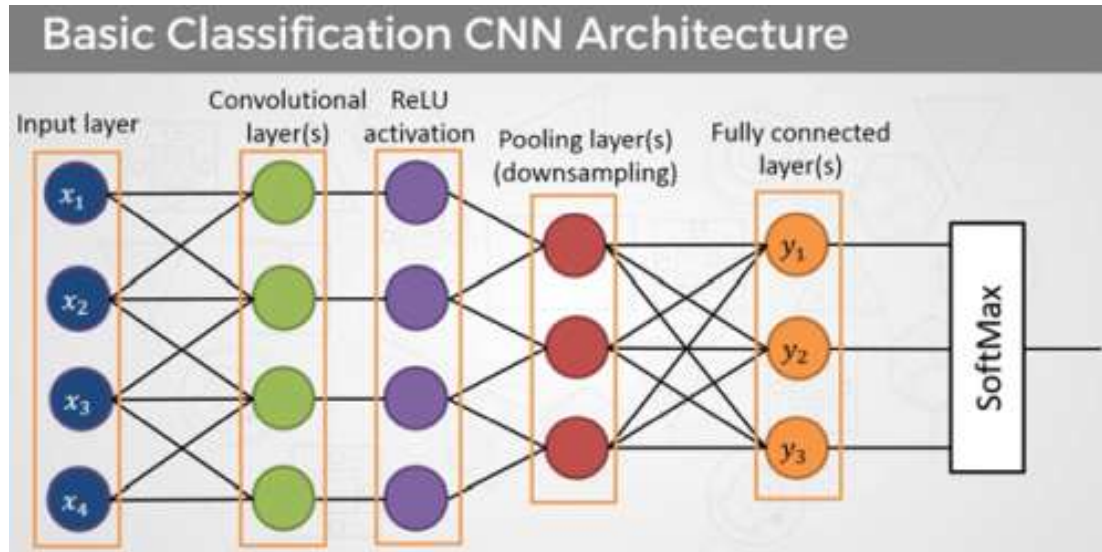


Figura 27. Arquitectura de una red neuronal convolucional de una sola capa ([Intelligency, 2021](#)).

Al igual que las dos redes vistas en la sección anterior, las CNN son redes del tipo FF, aunque las redes vistas hasta ahora no hacían más que aplanar las imágenes de entrada, es decir, reconvertirlas a un vector de una dimensión. Haciendo esto se pierde información espacial, puesto que no se consideran vecinos. Las CNN, en cambio, sí consideran esta información.

En las CNN no todas las capas son FCLs, a diferencia del caso de los MLP y las RBF, sino que tras la capa de entrada se pasa a la **capa convolucional**, que realiza sobre cada una de las subdivisiones de datos una convolución, la operación que le da nombre a este tipo de red. A grandes rasgos, la convolución es una herramienta usada en procesamiento de señales para obtener la respuesta de un sistema a una señal conocida ([Kim H., 2019](#)), que equivale a grandes rasgos a cómo se superponen. En el uso de redes neuronales esto se traduce en una operación para convertir una imagen a un mapa de sus características ([Kim H., 2019](#)). En la figura 28 puede verse el proceso de convolución para imágenes 2D, en cuyo caso equivale a realizar el producto escalar de la matriz de datos tomada y un filtro.



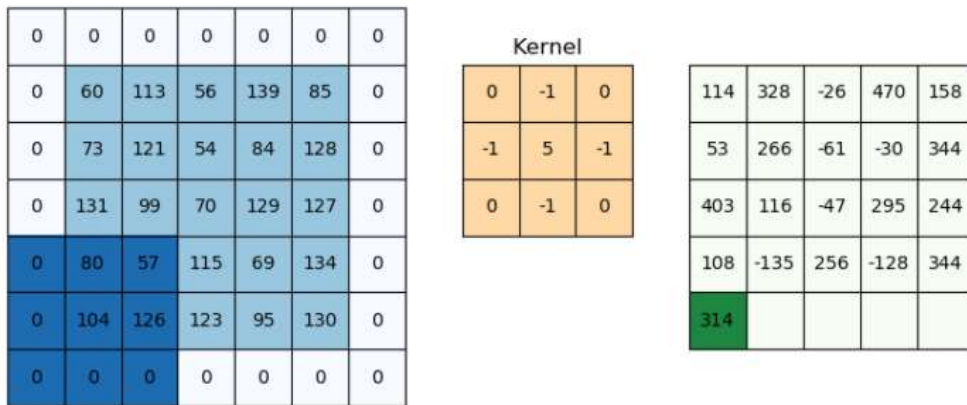
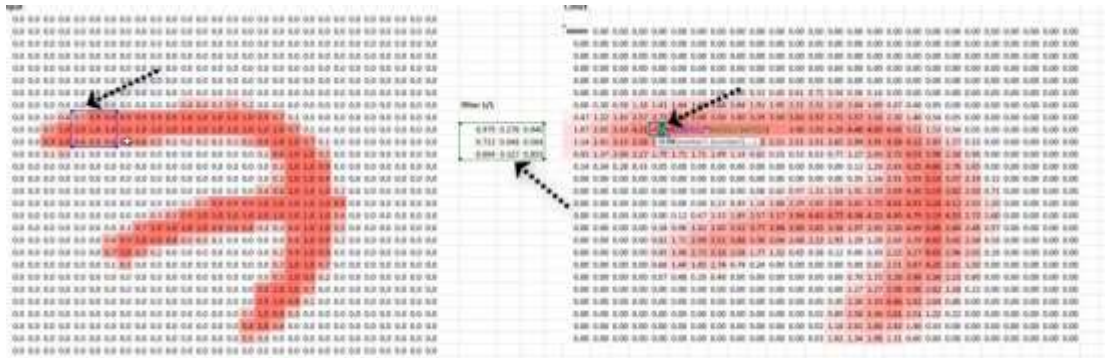


Figura 28. Convolución. (a) Producto escalar (convolución en 2D) (Howard J., 2018). (b) Padding en los bordes.

Así, la convolución se lleva a cabo en la capa convolucional, en la que cada nodo consta de un filtro sobre el que se realiza dicha operación. Los filtros tienen unas dimensiones determinadas que pueden definirse como parte de la arquitectura CNN. Una vez realizada la convolución sobre un conjunto de datos, se mueve el filtro una distancia determinada, pasando a tomar una región distinta de los datos, y vuelve a realizarse la convolución. Debido a que, de realizarse de esta forma, no se conseguiría un resultado para los bordes, al convolucionar también es necesario realizar *padding*, es decir, añadir ceros a los bordes. De esta forma así también se consigue mantener las dimensiones de la imagen. Realizar más o menos *padding* dependerá de las dimensiones del filtro. De forma opcional puede realizarse una operación de *striding* si no quiere analizarse los píxeles uno a uno, en cuyo caso el filtro no se moverá de uno en uno, sino que se saltará algunas zonas.

Tras la convolución, se usa una capa cuyas neuronas tienen como función de activación una **Rectified Linear Unit (ReLU)**, que fija la salida a 0 para entradas con signo negativo y consiste en una recta para entradas de signo positivo, y que puede observarse en la figura 29. Con frecuencia suele considerarse no una capa por separado, sino parte de la capa convolucional.

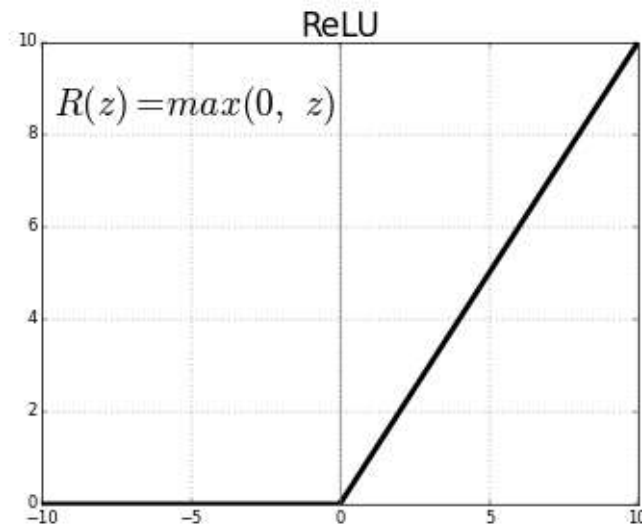


Figura 29. Función de activación ReLU.

Estas capas, a las que solo entra la salida de la anterior, se utilizan para eliminar el cambio gradual de la imagen, por lo que puede entenderse que son un tipo de filtro paso alto, con el objetivo de quedarse solo con los cambios bruscos dentro de la imagen. Esto es debido a que el objetivo del proceso de convolución y por extensión de las capas convolucionales es el de extraer características de las imágenes; detectando los cambios más bruscos de la imagen es como suelen obtenerse estas características, con frecuencia contornos de diferentes objetos que hay en la imagen. A las características extraídas se las llama, en ocasiones, *activaciones*.

En procesos de segmentación es común detectar como características bordes de objetos y otras formas simples en las primeras capas convolucionales, y formas más complejas en capas más profundas. Asignando diferentes pesos a diferentes regiones de las imágenes se puede, además, delimitar aquellas regiones que se consideren de interés. Es mediante esta

combinación de convolución y ReLU que pueden detectarse, por ejemplo, diferentes objetos en una imagen.

Sin embargo, como ya se ha mencionado las activaciones extraídas tienen las dimensiones de la imagen original. Puesto que de cada imagen se extraerán un número de activaciones que puede ser del orden de los centenares, conservar el tamaño de la imagen original conllevaría un aumento insostenible del tiempo de entrenamiento. Este problema se evita reduciendo las dimensiones de cada activación; es posible hacer esto sin pérdida de información puesto que, al menos en las capas convolucionales más superficiales, las características extraídas consisten en bordes, perfiles simples que pueden simplificarse a un bajo número de píxeles. Es por ello que tras la operación de *ReLU* con frecuencia se introduce una **pooling layer**, con el objetivo de reducir las dimensiones de cada activación y con ello la cantidad de parámetros en una red, con una pérdida de exactitud negligible ([Santos C.F.G. et al., 2020](#)). Una manera común de hacer este proceso es con un filtro 2x2, que realiza un proceso de *striding* como el ya descrito anteriormente: a lo largo de la imagen se desliza este filtro y se toma el máximo valor en cada matriz definida por él. En la figura 30 aparece este proceso esquematizado.

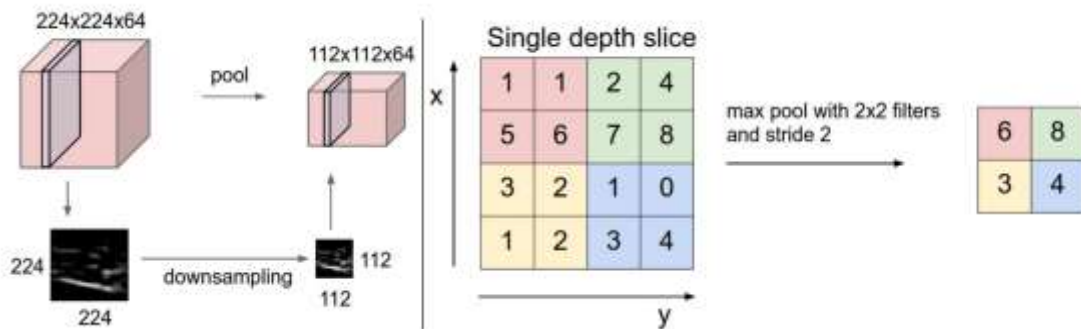


Figura 30. Submuestreo llevado a cabo en las pooling layers.

Por lo tanto, en una red neuronal, pueden reducirse las dimensiones de la entrada aplicando *striding* o bien al hacer la convolución o bien en las *pooling layers*, cuyo único objetivo es el de reducir el tamaño de los datos.

Tras el proceso, todas las activaciones pasan a cada neurona de la **FCL** que, al igual que en las redes neuronales anteriores, asigna pesos a cada activación, realizando una suma ponderada de cada una de ellas, cada una de las neuronas con pesos diferentes. A su vez, cada neurona genera como salida los pesos de una de las categorías en la que puede clasificarse la imagen original. Finalmente, una **capa softmax** acaba convirtiendo estos pesos en probabilidades. Con frecuencia esta capa suele considerarse parte de la FCL.

En la figura 31 se ve un ejemplo de aplicación de una CNN, en el que se usan dos capas convolucionales con filtros de dimensiones 5x5 para extraer activaciones, cada una seguida de una *pooling layer*, y finalmente por una FCL que acaba obteniendo las probabilidades de que la imagen de entrada constituya un dígito u otro.

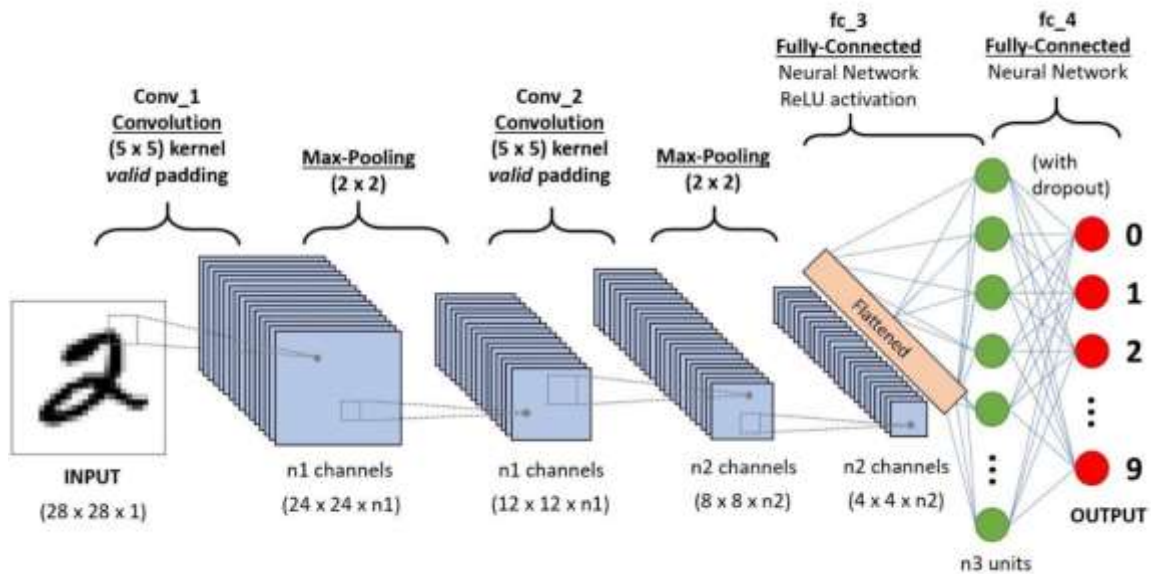


Figura 31. Ejemplo de aplicación de una CNN: reconocimiento de dígitos ([Phani, 2021](#))

Este es el tipo de CNNs que se implementará en la tesis, aunque antes de pasar a la implementación hay algunos conceptos adicionales a tener en cuenta si se quiere alcanzar la máxima exactitud posible. A destacar que las CNN también se entrenan con proceso de *backpropagation*, ya explicado en apartados anteriores.

El tiempo de entrenamiento de una red profunda no solo es superior al de ML tradicionales debido a sus arquitecturas más complejas, sino también debido al llamado Cambio en la Covarianza Interna, comúnmente conocido como su término inglés, *Internal Covariance Shift (ICS)*, problema que se hace mayor cuanto más profunda sea una RN, e introducido por primera vez en ([Ioffe S.&Szegedy C., 2015](#)). Esta fuente define el ICS como el fenómeno por el que la distribución de las entradas de cada capa cambia durante el entrenamiento, lo que ralentiza el entrenamiento de la red al necesitar menores *learning rates*. En otras palabras: durante cada iteración de entrenamiento, tras cada actualización de pesos, pasan a procesarse datos cuyo rango puede diferir sustancialmente al de los datos anteriores, y estos cambios se propagan entre neuronas, viéndose amplificados en gran medida si se trata de una DN. Las neuronas tienen que adaptarse constantemente a estos cambios, lo que provoca esta ralentización del entrenamiento de la RN ([Vinod R., 2020](#)).

La solución a este problema es usar lo que se llama *Normalización en lote*, comúnmente conocida por su nombre inglés, *Batch Normalization*, e implementada dentro de una DN en forma de capas.

Tal y como su nombre indica, durante la *batch normalization* se normalizan los datos no en su totalidad sino en múltiples grupos, correspondiendo a una normalización por neurona de la capa en la que se implementa este proceso ([Riva M., 2021](#)). La normalización se define como:

$$z^N = \left( \frac{z - m_z}{s_z} \right) \quad (35)$$

, donde  $z$  es el dato a normalizar,  $m$  su media y  $s$  su desviación estándar. En una red neuronal, la normalización se aplica en cada neurona, tal y como se observa en la figura 32.

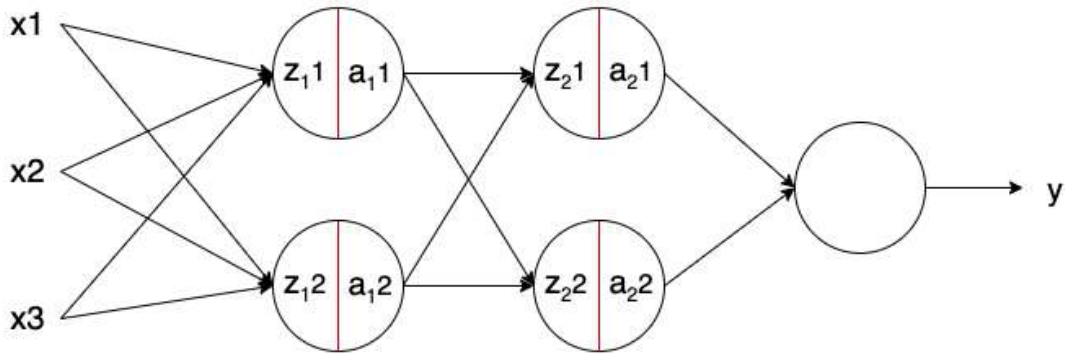


Figura 32. Implementación de Batch Normalization en una NN. Las verticales en color rojo de cada neurona indican el proceso de normalización. De ([Riva M, 2021](#))

Las entradas,  $x$ , entran a cada neurona mediante una transformación lineal, ya vista en anteriores secciones.

$$z = g(w, x) + b \quad (36)$$

, donde  $w$  son los pesos de esa neurona y  $b$  su sesgo. Sin normalización,  $z$  pasaría directamente a la función de activación,  $f$ :

$$a = f(z) \quad (37)$$

, pero si se añade *Batch Normalization* se tiene, en cambio:

$$z^N = \left( \frac{z - m_z}{s_z} \right) * \gamma + \beta \quad (38)$$

, donde  $z^N$  es la salida del proceso de la normalización,  $m_z$  y  $s_z$  la media y la desviación estándar respectivamente de las salidas de las neuronas ([Riva M, 2021](#)). Nótese que, durante la normalización, el parámetro original  $b$ , se elimina debido a la sustracción de la media; en su lugar están  $\gamma$  y  $\beta$ , los parámetros de entrenamiento de esta normalización que se actualizarán durante cada iteración de la red con tal de hacer disminuir el error de predicción.

Hay cierto debate sobre cuándo implementar *Batch Normalization* en las NN, si hacerlo antes de la función de activación o después ([Kilcher Y., 2020](#)), pero en la tesis se aplicará

siempre antes, en base al artículo de origen de una de las redes profundas que se han implementado (He K. et al., 2015).

Usando *batch normalization* también se soluciona un segundo problema. Para introducirlo, obsérvese la figura 33, en la que aparece el proceso de entrenamiento durante una de las iteraciones de entrenamiento de una CNN ya implementada y en la que se observa el progreso de la minimización de la función de error.



Figura 33. Detalles sobre el entrenamiento de una CNN.

De la figura puede observarse que, llegados a cierto punto (~40 iteraciones), el error ya no decrece. Es importante visualizar varios de los procesos de entrenamiento de esta manera para no causar un error común que se comete al aplicar redes neuronales, y concretamente redes Neuronales profundas, es el de llegar al sobreajuste (*overfitting*), y provocado por sobreentrenar una red (Ying X., 2019). Este fenómeno puede observarse en la figura 34.

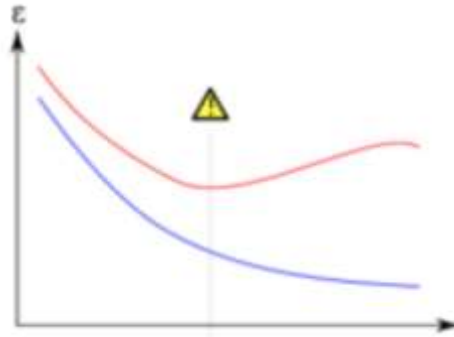


Figura 34. *Overfitting*. En azul, el error de entrenamiento en función de los datos usados. En rojo, el error de validación. Identificado por la vertical está el punto óptimo de detención de entrenamiento ([Ying X., 2019](#)).

El sobreajuste impulsa al algoritmo de ML a memorizar todos los datos, incluyendo ruido que en ocasiones no es posible eliminar, en vez de aprender el patrón, la disciplina, oculta dentro de los datos. Además de usar *batch normalization*, se recomienda identificar un punto de detención más allá del cual no haya una disminución del error a la hora de aplicar la red al conjunto de datos de validación. En Matlab, esto puede especificarse a la hora de establecerse el entrenamiento: que lo detenga cuando, pasadas unas pocas iteraciones (2 o 3) ya no aumente la exactitud.

#### 1.3.6.4 ResNet.

Las redes neuronales profundas, particularmente las CNN, han protagonizado un gran número de estudios desde su auge debido a su exactitud en los problemas de clasificación. Particularmente, en los últimos años se ido construyendo CNNs con arquitecturas cada vez más complejas, apoyando la teoría de que, al aumentar el número de capas en una red neuronal, se acaba obteniendo un mayor número de características para unos datos de entrada, y características mejores. En 2015, sin embargo, se publicó un artículo mostrando lo contrario: llega un punto en que, si se añaden más y más capas, el entrenamiento empeora ([He K. et al., 2015](#)). Fue en esta misma publicación donde se introdujeron las Redes Neuronales Residuales, comúnmente conocidas como *ResNets* (*Residual Neural Networks*).



Tal y como se observa en la figura 35, llega un punto en que el error no solo no disminuye, sino que es superior al correspondiente a arquitecturas más sencillas. Puesto que dicho error se observa tanto durante el entrenamiento como en la validación, este fenómeno no se trata del ya introducido *overfitting*, sino que es más probable que se trate de un fallo de optimización, y es que cuanto más profunda es una red, más difícil es no solo llegar a configurar manualmente los parámetros de la arquitectura sino también es más difícil que el algoritmo actualice los pesos de cada nodo correctamente.

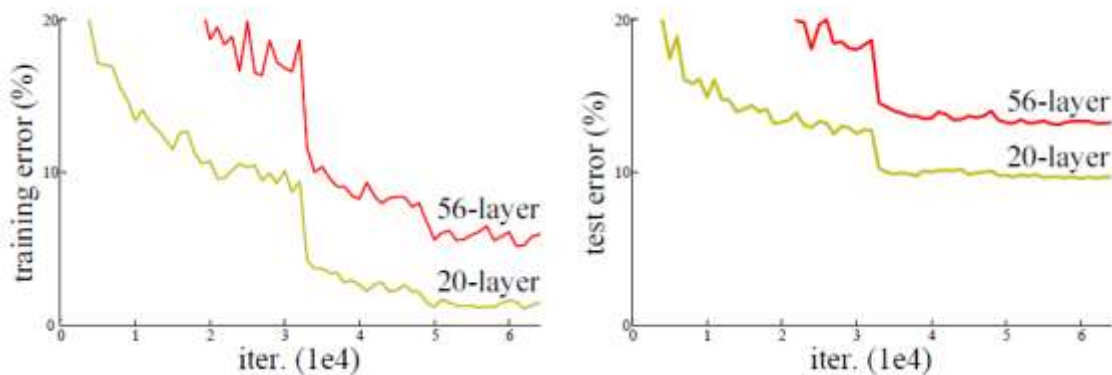


Figura 35. Evolución del error en función de las iteraciones de entrenamiento para una arquitectura de dimensiones medias (en amarillo) y una de grandes dimensiones (en rojo). De [\(He K. et al., 2015\)](#).

El artículo propone como solución a dicho problema considerar dos redes: una red de arquitectura menor, en la que no se haya observado pérdidas de exactitud debido a sus dimensiones, y una más profunda, cuya estructura consista en una copia de la menor por un lado y con capas adicionales por otro. Si a las capas adicionales son entrenadas de tal forma que aprendan la **función identidad** con tal de que devuelvan como salida la misma entrada que reciben (a esto también se le llama *identity mapping*). Para ello, los pesos tienen que establecerse a 1; en capas convolucionales, esto consiste en tener filtros cuyo centro sea 1 y todos los otros valores, 0.

Sin embargo, normalmente los pesos se inicializan a valores aleatorios próximos a 0, no a 1 [\(Kilcher Y., 2020\)](#). Por lo tanto, de forma predefinida la red aprende la función cero, no la

función identidad, y hacer que la aprenda sin modificar la entrada no es obvio ni inmediato. La solución propuesta es usar la conexión residual, que es en lo que se basa una Red Neuronal Residual, tal y como su nombre indica. Esto consiste no en intentar igualar la salida a la entrada mediante un conjunto de capas, sino en entrenar esas capas por separado para que produzcan una desviación cercana a 0. La arquitectura resultante es el bloque que aparece en la figura 36.

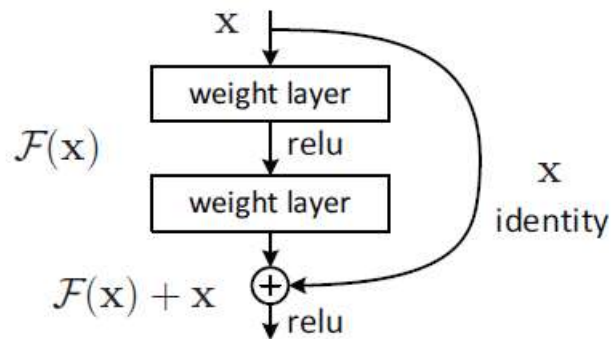


Figura 36. Bloque de aprendizaje residual. De ([He K. et al., 2015](#)).

Tal y como se observa, aparte de la estructura FF lineal hay una conexión aparte que hace que la entrada no pase por capas y por lo tanto no se vea alterada mediante los pesos de los nodos de estas; es decir, esta conexión realiza la función de identidad. Por ello, la salida no es una  $F(x)$  como producto de pasar por otras capas, sino  $F(x)+x$ , en el artículo llamada  $H(x)$ . Puesto que es fácil configurar los pesos para que aprendan la función cero, es fácil hacer que  $F(x)=0$ , y por lo tanto se consigue  $H(x)=x$ , precisamente aquello que se buscaba para solucionar el problema, puesto que la función predefinida de esta red ya no es la función cero sino, aproximadamente (puesto que después viene una función ReLU) la función identidad. Con esto se consigue una mejor propagación del gradiente desde la capa de salida de una iteración hacia las capas más externas de la red.

Combinando estos bloques se obtiene una ResNet, a la que puede pasarse a partir de una CNN normal simplemente añadiendo las conexiones residuales ya vistas, tal y como se ve en la figura 37.

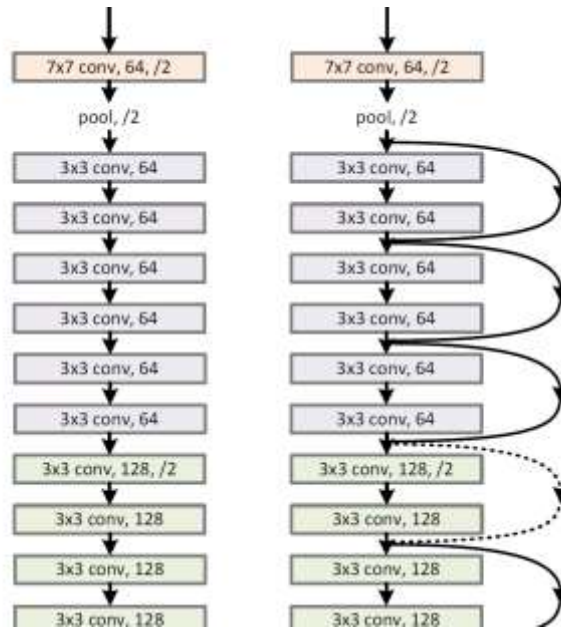


Figura 37. Ejemplo de ResNet. De (He K. et al., 2015).

A destacar también que, tal y como puede observarse por la única línea discontinua de la figura, no es posible colocar conexiones residuales en aquellas capas que vayan seguidas de *maxPoolingLayers*. Esto es porque como ya se ha visto, las *pooling layers* cambian las dimensiones, pero obviamente la función identidad obviamente requiere las mismas dimensiones que la entrada. Los resultados del estudio se recogen en la figura 38.

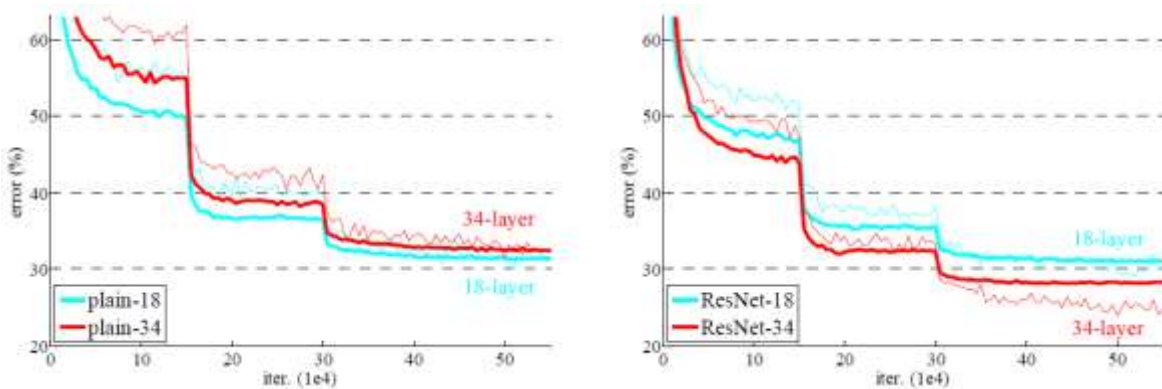


Figura 38. Cambios en el error de entrenamiento (líneas delgadas) y validación (líneas en negrita) al aplicar conexiones residuales.

Tal y como se observa, no solo el error es inferior al aplicar ResNet, sino que el efecto es proporcional a las dimensiones de la red; a más profunda una red, más disminuirá el error, teóricamente. Este es el efecto que se busca obtener también en esta tesis.

#### 1.3.6.5 V-Net

Las Redes Neuronales Volumétricas, comúnmente conocidas por su acrónimo inglés, V-Net, son una segunda variante de las CNNs que, al igual que las ResNets, fueron conceptualizadas con el objetivo de resolver ciertas limitaciones o defectos intrínsecos de las redes convolucionales; en este caso, el hecho de que las CNNs no son capaces de considerar información volumétrica, sino solo imágenes 2D por separado. Este es el mismo problema que tienen los algoritmos de ML tradicionales y redes neuronales no profundas con dos dimensiones, tal y como se vio en la sección anterior: deben descomponer una imagen en un vector, eliminando la información que puede obtenerse al considerar el área circundante del valor actual. Del mismo modo, las CNNs tradicionales consideran toda imagen, pertenezca o no a un mismo volumen, totalmente independiente, cuando es probable que no sea el caso.

En cambio, tal y como se explica en el artículo en que fueron introducidas ([Milletari et al., 2016](#)), en el diagnóstico médico por imagen, que es precisamente el campo que estudia la tesis, se tiene muchas veces no imágenes independientes sino una multitud de cortes 2D a lo largo de una longitud de escaneo, como es el caso de las imágenes de OCT, tal y como se ha visto en el capítulo anterior. Las V-Net se crean para tener en cuenta no solo la información retenida en cada imagen sino también la proximidad espacial entre ellas.

El estudio aplica estas redes a la segmentación de imágenes de MR de próstata, y para ello hace uso de lo que llaman convoluciones volumétricas. Tal y como se observa en la figura 39, se hace pasar un filtro tridimensional, en este caso un cubo de longitud de lado 2 a lo largo de todo el volumen, con un *stride* de 2.

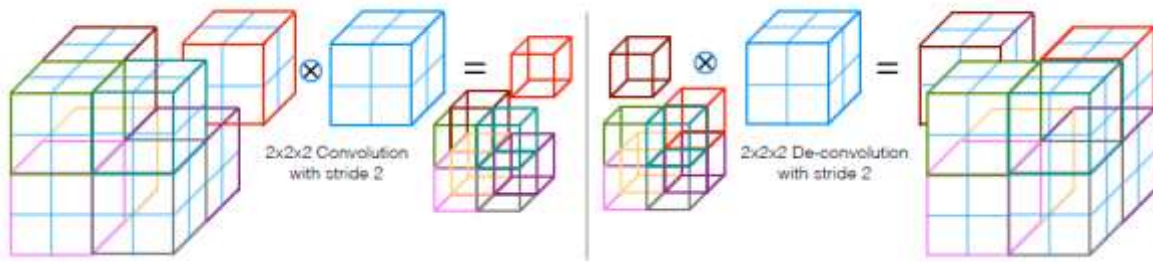


Figura 39. Convoluciones volumétricas en V-Nets. Izquierda: convolución. Derecha: deconvolución. De ([Milletari et al., 2016](#)).

El estudio detalla posteriormente la red implementada, usándose una arquitectura de dos módulos principales, uno que reduce el tamaño de la imagen y el otro que lo aumenta de nuevo, empleándose convoluciones y deconvoluciones respectivamente. Sin embargo, tanto las deconvoluciones como las otras particularidades de esta arquitectura no son de gran interés en esta tesis debido a la diferente naturaleza del problema que se afrontan los autores citados (segmentación frente a la clasificación). Después de todo, en problemas de clasificación es necesario que, llegada la entrada a la capa final, esta sea de tamaño unitario, que solo indique la clase a la que pertenece la entrada, y por lo tanto no hay que volver a aumentar las dimensiones del volumen dentro de la V-Net.





## 2 Estado del arte

*Además de asimilar los conceptos necesarios para introducirse al diagnóstico de la EM, también ha sido necesario ponerse al corriente del estado actual de la investigación para poder así orientar esta tesis hacia fines novedosos que permitan nuevos avances en este campo.*

*El estado del arte se ha desglosado en tres secciones: la evaluación de la retina, la evolución que ha tenido la OCT como técnica de diagnóstico de EM y el uso de aprendizaje automático sobre los datos más comúnmente usado para estos fines.*

## 2.1 Evaluación de la retina

### 2.1.1 La vía óptica como extensión del cerebro

Debido a la complejidad y la dificultad que hay a la hora de estudiar el sistema nervioso y por consiguiente realizar el diagnóstico sobre él, las pruebas no se limitan a la evaluación del sistema nervioso central sino también extensiones de este, como son el sentido del oído o, principalmente, el sistema visual humano.

La retina es la parte más profunda del ojo. Constituye el elemento clave de la vía visual, siendo responsable de recibir la luz incidente, procesarla y transmitirla al sistema nervioso central. La retina y el cerebro derivan del neuroectodermo tras alrededor de 23 días de gestación, y por lo tanto tienen orígenes embrionarios comunes ([Kashani A.H. et al., 2021](#)) ([Jacobson M. & Hirose G., 1978](#)) y ambos se consideran parte del SNC, encontrándose que las neuronas de la retina tienen muchas similitudes estructurales y funcionales con las del cerebro ([Kashani A.H. et al., 2021](#)).

Y efectivamente, hay ciertas lesiones y patologías ópticas que están altamente correlacionada con la EM, siendo la más común la Neuritis Óptica (ON), pues aproximadamente un 20% de los pacientes a los que se les acaba diagnosticando EM tienen ON como primera patología ([Kale N., 2016](#)). Sus síntomas incluyen pérdida visual unilateral y dolorosa, frecuentemente no acompañada de otros síntomas neurológicos. La ON, al igual que la EM, es una enfermedad desmielinizante del sistema nervioso, afectando concretamente al nervio óptico, que también se da principalmente en mujeres jóvenes.

Un hecho que refuerza ulteriormente el interés que hay en estudiar la vía visual es que análisis *post-mortem* indican que el 94-99% de los pacientes diagnosticados con EM tienen lesiones ópticas, incluso en aquellos casos en los que no se les diagnostica previamente ON ([Britze J. et al., 2017](#)). La desmielinización de los axones del nervio óptico puede conllevar la disminución de espesor en los axones y/o la degeneración del nervio óptico, con la consiguiente pérdida de células ganglionares. Es debido a estos estudios y observaciones,



por lo tanto, que se considera que evaluar la vía visual es de particular interés de cara al diagnóstico de EM.

### 2.1.2 Estructuras retinales de interés

Tras las justificaciones del apartado anterior, se concluye que la vía óptica en particular es de especial interés en lo que respecta a la EM, pues su evaluación es de gran importancia para el diagnóstico de esta enfermedad. Así, pruebas funcionales como los potenciales evocados visuales o pruebas estructurales (por ejemplo, análisis del adelgazamiento de diferentes capas de la retina) pueden aportar información clínica que apoye decisiones basadas en el estudio por imagen, u otros marcadores bioquímicos. Sin embargo, también es necesario definir, dentro de lo que constituye el estudio de la vía visual, aquellas zonas anatómicas que sean de mayor importancia para este fin. Después de todo, la retina no es un tejido uniforme, sino que se divide en una serie de capas de distinta composición, lo que les da distintas funciones.

Puesto que la principal afectación se da al nervio óptico, suele considerarse de interés analizar aquellas estructuras ópticas con la mayor concentración de fibras nerviosas, y por ello las capas comúnmente analizadas son la *Retinal Fiber Layer (RNFL)* y la *Ganglionar Cell Layer (GCL)*, de la que comúnmente se estudia su parte macular, llegándose a llamar en la bibliografía, en ocasiones, *mGCL*. A pesar de esto, por contraste y para estudiar ulteriormente su relación con la EM, también suelen estudiarse otras capas, como la coroides, que forma parte de la túnica uveal ([The Retinal Tunic, 2007](#)). Estas capas retinales pueden observarse en la figura 40.

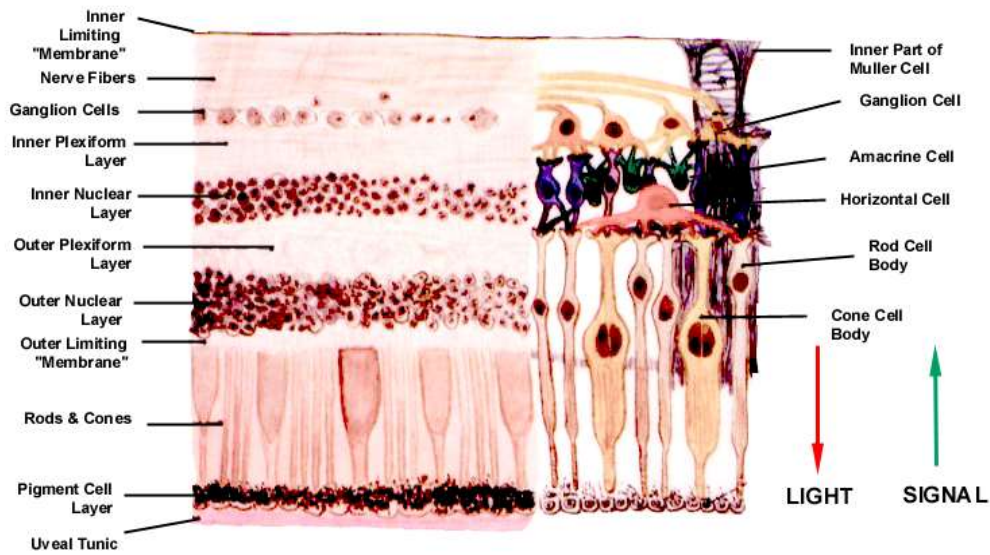


Figura 40. Anatomía de la retina. De ([The Retinal Tunic, 2007](#))

En ocasiones la GCL no se estudia por separado, sino que se forman dos estructuras, las comúnmente llamadas *GCL+* (equivalente a la capa de las células ganglionares en conjunto con la capa interior plexiforme, también llamada en la bibliografía *Ganglionar Cell Inner Plexiform Layer* o *GCIPL*), y *GCL++* (equivalente al conjunto de RNFL macular, o *mRNFL*, y *GCL+*) ([Mihaylova B. & Dimitrova G., 2019](#)). Debido a estas agrupaciones de varias capas, durante el resto de la tesis se usará el término **estructuras retinales** tanto para capas singulares como para agrupaciones de las mismas. Las estructuras retinales de los registros que finalmente tratará la tesis se presentarán más adelante, en el apartado en que se introduce la base de datos.

## 2.2 OCT para el diagnóstico de EM

### 2.2.1 Reducción del espesor retinal observada por OCT. Evolución.

La evaluación de la vía visual debida a la desmielinización producida durante el transcurso de la EM se puede realizar mediante Tomografía de Coherencia Óptica (*OCT*), y al igual que en el caso de Resonancia Magnética, se medirán espesores de tejido, en este caso espesores retinales. A lo largo de los años se han realizado numerosos estudios siguiendo este método, y se continúan haciendo en la actualidad ([Shurz N. et al., 2021](#)), indicando que

se trata de una prueba de diagnóstico y control fiable. En 1999, se verificó mediante la técnica de OCT que existe una reducción significativa de espesores de RNFL en pacientes de EM comparada con sujetos de control ([Parisi V. et al., 1999](#)). Pacientes con EM en fases tempranas o intermedias tienen aproximadamente 10 micras menos de espesor medio de RNFL que sujetos sanos de la misma edad y sexo ([García-Martín E. et al., 2017](#)).

Estudios recientes de OCT ([Wicki C.A. et al., 2018](#)) reportan una reducción del espesor retinal a medida que la EM avanza, mientras que el análisis de las diferentes capas de la retina refleja diferentes aspectos de la actividad de la enfermedad durante EM preclínica y clínica. Medidas de espesores de RNFL peripapilar (*pRNFL*) parecen ser biomarcadores ideales en lo que respecta a integridad de los axones, mientras que espesores de GCL+ permiten una detección temprana de la degeneración axonal; por otro lado, espesores de la capa nuclear interna macular parecen constituir biomarcadores de la inflamación debido a la enfermedad y de la respuesta al tratamiento ([Wicki C.A. et al., 2018](#)). Con frecuencia, tres meses después de un episodio agudo de ON en pacientes de EM, se desarrolla atrofia *pRNFL*, de nuevo observable por OCT, observándose reducción de sus espesores sobre todo en la región temporal ([Minakaran N. et al., 2021](#)). Sin embargo, debido a lo elevadas que resultan las medidas de *pRNFL* en contraste a la reducción de varios meses después, puede ser difícil usar solo la *pRNFL* para investigar la integridad axonal en concreto. Además, está el llamado “efecto suelo” de OCT, por el que la media de espesores de *pRNFL* no se reduce a menos de 30  $\mu\text{m}$  independientemente del desgaste del nervio óptico, siendo difícil detectar cambios en atrofas ópticas severas u ON recurrentes ([Minakaran N. et al., 2021](#)).

En ([Petzold A. et al., 2017](#)), tras revisar estudios con 1667 pacientes EM con ON y 4109 pacientes EM sin ON, los autores concluyen que las diferencias más grandes y robustas entre ojos de sujetos con ON y controles se encuentran en la *pRNFL* y la GCL+ macular. En ([Britze J. et al 2017](#)), donde se analizan 2118 ojos, los investigadores encuentran una reducción significativa (6,73  $\mu\text{m}$ ) en el espesor de la GCL en pacientes EM sin ON, y que esta reducción del espesor está asociada con la función visual y mayores índices de

discapacidad, según la escala *Expanded Disability Status Scales* (EDSS). Asimismo, en casos de ON aguda se detectó una reducción significativa del espesor de GCL antes que de RNFL. En ([Alonso R. et al., 2018](#)), los autores citan estudios que demuestran que afectaciones de la RNFL y la GCL están relacionadas con discapacidades físicas y cognitivas, así como con actividad detectada con MRI. La revisión de ([Yap T.E. et al., 2019](#)) resume los biomarcadores asociados con EM, remarcando la importancia de medidas retinales debido a la repetibilidad de las mismas (sobre todo en ciertas estructuras retinales de interés como la RNFL), así como otros biomarcadores obtenidos, como son medidas del líquido cefalorraquídeo.

Por consiguiente, medidas retinales usando OCT pueden considerarse un potencial biomarcador de lo que respecta a la pérdida axonal y a la neurodegeneración que ocurre durante la EM.

### 2.2.2 Equipos de OCT y protocolos de exploración

En los últimos años varios equipos de OCT disponibles en el mercado se han habilitado con protocolos de adquisición para hacer posibles analizar medidas en grandes áreas de la retina. El dispositivo DRI Triton SS-OCT (Topcon, Tokio, Japón), de la modalidad SS-OCT, tiene un protocolo de escaneo centrado en una rejilla de 45x60 puntos, correspondientes a 9 mm de altura y 12 mm de anchura, abarcando tanto las zonas peripapilares (es decir, la cabeza del nervio óptico) como la mácula.

En general, el análisis de OCT se lleva a cabo en regiones retinales predefinidas. Por ejemplo, en el área macular las regiones analizadas suelen estar definidas por una rejilla ETDRS (*Early Treatment Diabetic Retinopathy Study*) ([Chew E.Y., 1996](#)), mientras que en las medidas alrededor de la cabeza del nervio óptico es común usar un escaneo circular peripapilar de 3,4 mm de radio. Usando esta área, se demostró ([Birkeldh U. et al, 2017](#)) que (1) el espesor de la pNRFL temporal (*T-pRNFL*) es altamente sensible, (2) a él se asocian tanto discapacidades física como cognitivas y (3) se considera una de las estimaciones de OCT más importantes en el diagnóstico de EM.

Sin embargo, hay casos en los que no se han usado estas regiones. Estudios recientes han buscado regiones arbitrarias de retina usando varios equipos de OCT. En ([Hu H. et al., 2019](#)) los autores identifican una reducción de mRNFL y mGCIPL en una forma descrita como una herradura de caballo (la “zona U”) y en ella encuentran una mayor reducción de espesores en pacientes con ON con una larga duración de enfermedad ( $9,1 \pm 6,8$  años). La zona U muestra una alta capacidad discriminante (una *Area Under the Curve*, o AUC, que vendrá explicada en capítulos posteriores, de 0,97) y una correlación con disfunciones visuales.

En otro estudio ([Pérez Del Palomar A. et al., 2019](#)), los autores analizan el protocolo usado por equipos DRI Triton SS-OCT y, usando técnicas de aprendizaje automático, identifican la mRNFL como la estructura de mayor capacidad discriminante (AUC=0,99) en lo que respecta a diagnosticar la enfermedad en pacientes sin ON con larga duración de enfermedad ( $7,12 \pm 2,66$  años). En ([Shi C. et al., 2019](#)), los investigadores usan un equipo OCT personalizado usando un protocolo cúbico de 512x128 espesores, encontrando de nueva una “zona U” en la que hay una importante disminución de espesor de mGCIPL, particularmente en el área nasal, e identificando un área focal circular de 1 mm de diámetro en el sector inferior nasal (llamándola “zona M”), con AUC de 0,75.

En todos estos estudios, la EMRR no fue diagnosticada en una fase temprana, y la enfermedad duró varios años. Sin embargo, un diagnóstico temprano de EM sigue siendo de importancia vital, pues el tratamiento aplicado durante el inicio de la enfermedad es capaz de retrasar el curso clínico y de reducir el número de brotes ([Kavaliunas A. et al., 2017](#)).

### **2.3 Aprendizaje automático. Estudios e implementación en esclerosis múltiple**

En apartados anteriores se ha establecido ya que el tipo de problema que afronta la tesis (clasificación) requiere ML con aprendizaje supervisado, siendo este tipo de aprendizaje de uso muy frecuente en la aplicación de redes neuronales y aprendizaje automático en general.

Sin embargo, con tal de hacer una ulterior selección de qué procesos de ML es conveniente aplicar, es útil ser consciente de la literatura existente en cuanto a su aplicación al diagnóstico de la EM.

En el pasado las redes neuronales se han usado con éxito en el estudio de EM para clasificar sujetos en pacientes o controles ([Zhang Y.D. et al., 2018](#)), o incluso para detectar en qué fase de la enfermedad se encuentran ([Marzullo A. et al., 2019](#)) y la probabilidad de un ataque ([Karaka Y. & Hayta Ş., 2015](#)). Con frecuencia estos estudios se hacen a partir de imágenes del SNC tomadas por RM, no de espesores retinales tomados por OCT.

Los tipos de redes usadas varían, pero es común el uso de MLPs ([Karaka Y. & Haytaş., 2015](#)) así como de redes más complejas como las CNN ([Zhang Y.D. et al., 2018](#)), aunque los hay también que usan Redes Neuronales de Gráficos (GNN) ([Marzullo A. et al., 2019](#)). A lo largo de la realización de esta tesis, el grupo también ha realizado pruebas con otros algoritmos, como las SVM ([Barea R. et al., 2020](#)).

Las técnicas de ML aplicadas a EM se han centrado en el diagnóstico temprano, incluyendo análisis de conversión potencial de posibles estados preliminares de la enfermedad, como SRA, a EM definitiva ([Miguel J.M. et al., 2021](#)), o predicciones de progresión de la enfermedad y de sus consecuencias ([Pinto M.F. et al., 2020](#)). Según la revisión sistemática de la aplicación de ML a EM de ([Vázquez-Marrufo M. et al., 2021](#)): “a pesar de la corta trayectoria de aplicación de los algoritmos de ML a la EM, es posible observar un claro cambio de tendencia en su breve historia sobre los objetivos perseguidos. En los estudios iniciales, el principal reto era la clasificación de los pacientes dentro de la enfermedad (subtipos) [*SCA, EMRR...*].” En la figura 41 puede observarse la evolución en número de artículos de los estudios de estas tres naturalezas.

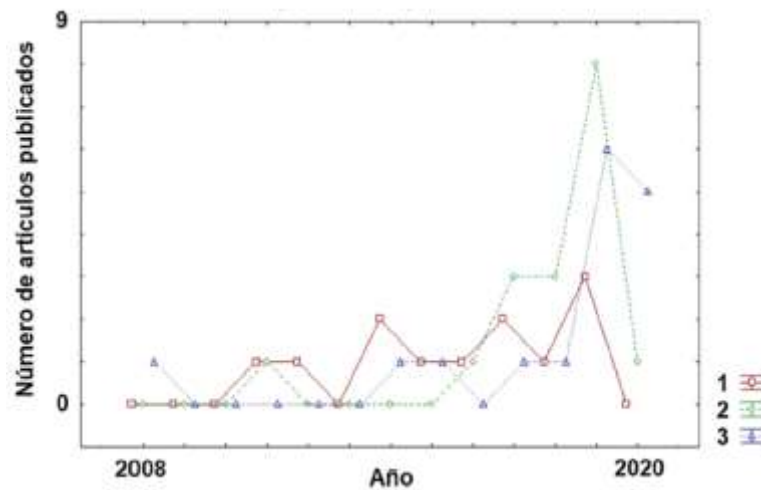


Figura 41. Evolución anual de estudios publicados en Pubmed empleando las palabras clave “Machine learning” y “Multiple Sclerosis” y para cada una de las categorías analizadas (1: clasificación; 2: diagnóstico; 3: predicción “. De ([Vázquez-Marrufo M. et al., 2021](#)).

Tal y como observan en dicha revisión: “los estudios con este objetivo se presentan de manera más homogénea a lo largo de la pasada década respecto a otras aplicaciones (diagnóstico o predicción) que han aumentado considerablemente su número en los últimos años (2017-2020).”

Dependiendo de las variables empleadas para la clasificación y los subtipos de la enfermedad que se pretenden identificar se obtienen rangos de precisión de entre el 70,6 hasta el 96% ([Vázquez-Marrufo M. et al., 2021](#)). Así, la naturaleza de los datos analizados varía entre estudios, aunque también las herramientas de análisis y clasificación. Por ejemplo, en ([Miguel J.M. et al., 2021](#)) se usa el algoritmo de muestreo RUSBoost. En ([Pinto M.F. et al., 2020](#)) se analiza información clínica (edad, edad de diagnóstico, manifestaciones de EM iniciales y diagnóstico clínico que a su vez llevó al diagnóstico de EM, por ejemplo, MRI o potenciales evocados) usando cuatro clasificadores: SVM, KNN, árboles de decisión y regresión lineal. En ([De Brouwer E. et al., 2021](#)), se usan redes neuronales recurrentes para predecir la progresión de la enfermedad en pacientes de EM a lo largo de una ventana de dos años.

Así, los datos procedentes de RMI constituyen la principal variable que se ha empleado para mejorar la capacidad de diagnóstico de EM ([Vázquez-Marrufo M. et al., 2021](#)), habiendo muy pocos estudios que hayan investigado el potencial para el diagnóstico de EM que tiene usar aprendizaje automático concretamente sobre espesores retinales. Uno reciente, ([Montolío A. et al., 2021](#)) investiga el uso de aprendizaje automático para diagnosticar y predecir el curso de la enfermedad en pacientes EM en fase mediana o avanzada ( $10,15 \pm 8,30$  años). Solo se analizan datos de RNFL (espesores medios en los cuatro cuadrantes sobre los que se divide la zona peripapilar), y se toma en consideración datos del paciente como la edad y el sexo. Para el diagnóstico, los autores usaron 9 características y los mejores resultados se obtuvieron con clasificadores que combinan varios tipos de ML, obteniéndose una exactitud de 87,7%.

En conclusión, se puede indicar que los trabajos que utilizan técnicas de inteligencia artificial para el análisis de los espesores de la retina son escasos, más aún para el diagnóstico de EM en fase precoz, por lo que se requieren ulteriores aplicaciones del aprendizaje automático sobre este tipo de datos.





## **3 Hipótesis y objetivos**

Dados los fundamentos y visto el estado del arte, se presenta la siguiente hipótesis:

- Es posible avanzar en la obtención de biomarcadores objetivos de EM en fases precoces de la enfermedad, mediante el análisis con técnicas de inteligencia artificial de las medidas de los espesores de diferentes capas de la retina, obtenidas mediante OCT.

A su vez, se trazan los siguientes objetivos:

- Estudiar las técnicas actuales de diagnóstico de EM basadas en el análisis de los espesores de capas de la retina obtenidos mediante OCT.
- Decodificar los protocolos de datos de los equipos de OCT.
- Disponer de una base de datos suficientemente representativa de controles y pacientes de diagnóstico reciente.
- Implementar los principales tipos de redes neuronales convencionales y convolucionales.
- Investigar técnicas de procesamiento de los datos entregados por los equipos de OCT.
- Evaluar la capacidad diagnóstica de los diferentes tipos de modelos de inteligencia artificial implementados, cuyos resultados sean presentados, justificados y contrastados debidamente.
- Proponer futuras líneas de investigación en base a lo visto en la tesis.



## 4 Método

*En este capítulo se declaran las características clínicas de la base de datos utilizada y su composición. También se desarrolla su segmentación en distintos subconjuntos que puedan resultar de más interés de cara a su análisis mediante aprendizaje automático. Se obtendrán tres modalidades distintas: imágenes completas (45x60), ventanas cuadradas de cierta longitud, y una selección basada en el tamaño del efecto, cada una de las cuales estará desarrollada en su sección correspondiente.*

## 4.1 Base de datos

El estudio ha sido aprobado por el Comité Ético del Hospital Universitario Miguel Servet (HUMS), se han seguido los protocolos recogidos en la declaración de Helsinki y los participantes ratificaron su consentimiento por escrito.

Los registros de espesores retinianos se han adquirido con un equipo de tomografía de coherencia óptica de barrido modelo Topcon DRI OCT Triton Swept Source OCT: SS-OCT (Topcon, Japón). El valor medio de la longitud de onda utilizado es de 1050 nm, con velocidad de escaneo de 100 000 cortes/s. El protocolo de exploración consiste en una rejilla con 45x60 puntos de medida para cada una de las estructuras exploradas que abarca el área macular y la cabeza del nervio (zona peripapilar), cubriendo una superficie de 9 mm de alto x 12 mm de ancho. El sistema de adquisición OCT implementa sus propios algoritmos de segmentación de las capas analizadas (Topcon Advanced Boundary Software®) y permite exportar los datos de espesor para su posterior visualización y/o análisis.

En la tabla 3 vienen recogidos los descriptivos de la base de datos. Se ha realizado la prueba de normalidad de Shapiro-Wilk ([Shapiro SS. & Wilk M.B., 1965](#)) sobre los valores de la edad de los pacientes, así como su distribución de sexos. De ella, se obtiene una *p* valor inferior a 0,05 en la distribución de edades de los pacientes con EM, indicando que siguen una distribución normal y por lo tanto se presentan media y desviación estándar; en la base de datos de controles ocurre lo contrario (la edad de los sujetos no sigue una distribución normal).

Al encontrarse que una de las dos distribuciones no es normal, se pasa a realizar la prueba de Mann-Whitney para compararlas; puesto que la *p* valor obtenida (0,37) es superior a 0,05 no hay diferencias significativas entre las medianas de edad.

*Tabla 3. Base de datos: descriptivos y distribución de los sexos.*

<b>Edades (años): Sujetos de control</b>		<b>Distribución de sexos: Sujetos de control</b>		<b>Edades (años): Pacientes con EM</b>		<b>Distribución de sexos: Pacientes con EM</b>	
Mediana	52,18	Hombres	14	Media	46,23	Hombres	11
Rango intercuartil	26,40	Mujeres	33	Desviación estándar	12,80	Mujeres	45

Tal y como se observa en la tabla, se dispone de una base de datos de ambos ojos de pacientes consistente en **56** registros OCT con diagnóstico precoz (menos de 9 meses desde los primeros síntomas en la fecha del registro) de EM de tipo EMRR sin neuritis óptica previa (edad:  $46,23 \pm 12,80$  años), entre los cuales hay 11 hombres y 45 mujeres, y **47** registros de sujetos de control (edad: 52,18 [26,40] años), 14 de los cuales son hombres y 33 son mujeres.

Se ha realizado también la prueba Chi cuadrado sobre ambas distribuciones de sexo, obteniendo una p valor de 0,23. Al ser superior a 0,05, se concluye que no existen diferencias significativas entre la distribución de sexos entre controles y pacientes.

Para cada ojo se dispone de imágenes correspondientes a los espesores de las siguientes estructuras de la retina:

**Espesor de la RNFL (Retinal Nerve Fiber Layer):** desde la ILM (ILM) hasta el límite interno de la capa de células ganglionares.

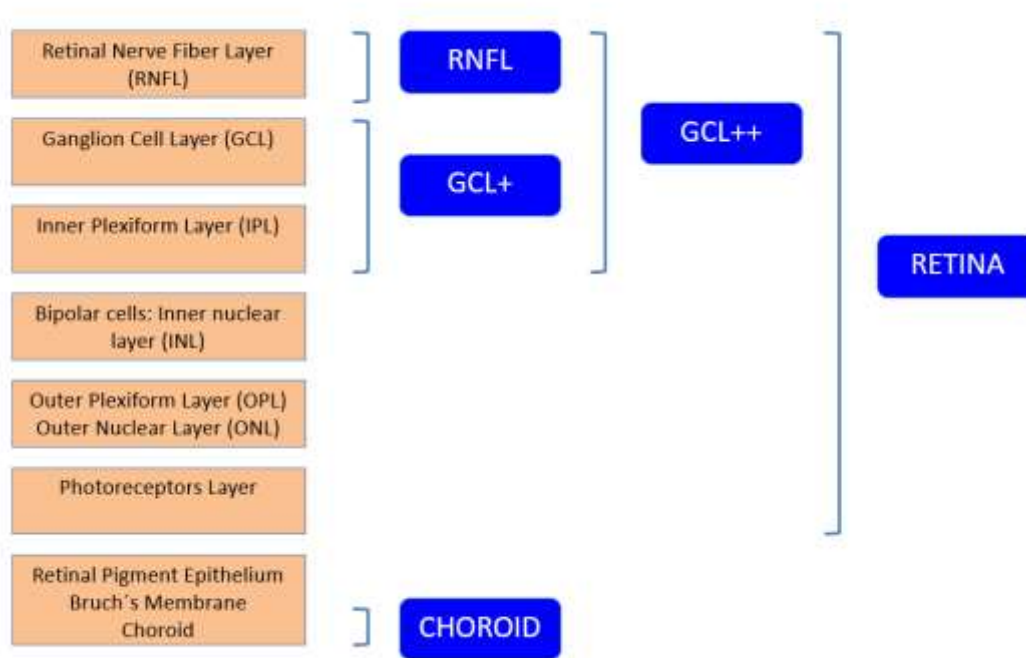
**Espesor de la retina completa:** desde la membrana limitante interna hasta el límite entre el epitelio pigmentario y la capa de fotorreceptores.

**Espesor GCL+:** desde el límite interno de la capa de células ganglionares hasta el límite externo de la IPL (capa plexiforme interna).

**Espesor GCL++:** desde la ILM hasta la línea IPL/INL, englobando así tanto la RNFL como la capa de células ganglionares.

**Espesor coroides:** desde la membrana de Bruch hasta la superficie escleral de la coroides.

En la figura 42 aparecen esquematizadas estas estructuras retinales respecto a qué capa/s retinales individuales corresponde cada una de ellas. Como puede observarse, se tienen medidas que corresponden a una sola capa retinal y medidas que engloban un conjunto de capas, como por ejemplo la GCL+ y la GCL++, tal y como se ha visto en la sección 2.1.



*Figura 42. Correspondencia de capas del registro con las estructuras retinales*

A su vez también se muestra, en la figura 43, la localización de estas estructuras en una imagen anatómica, obtenida por el mismo equipo de OCT.

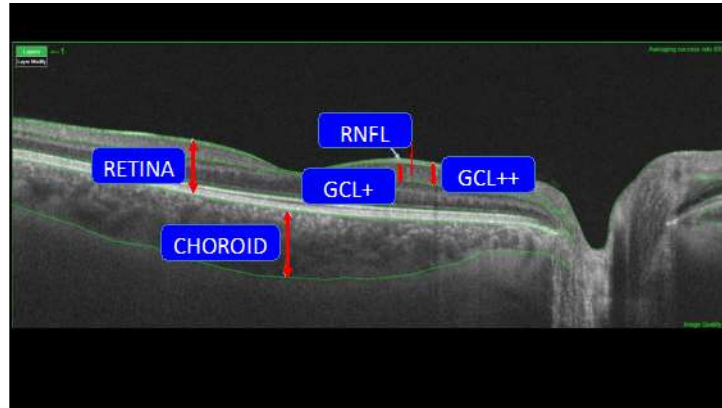
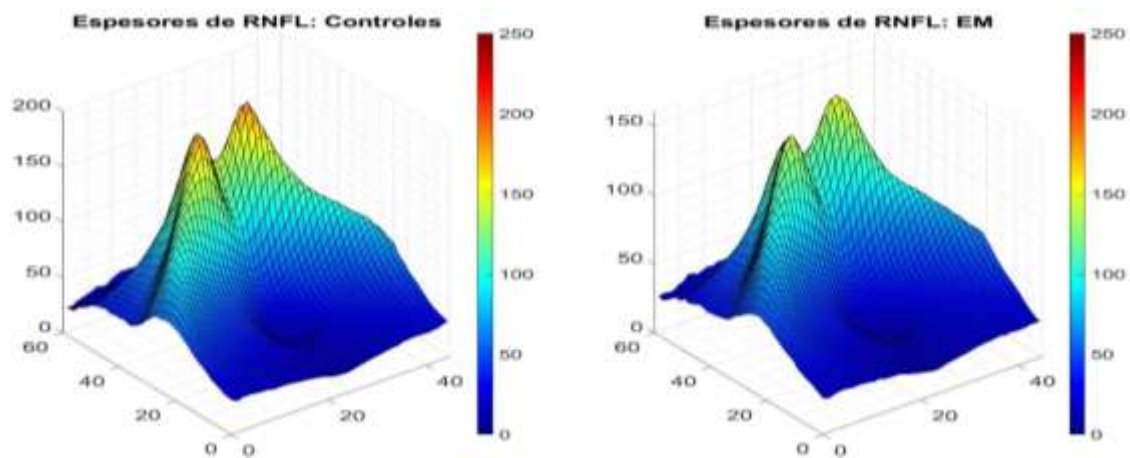
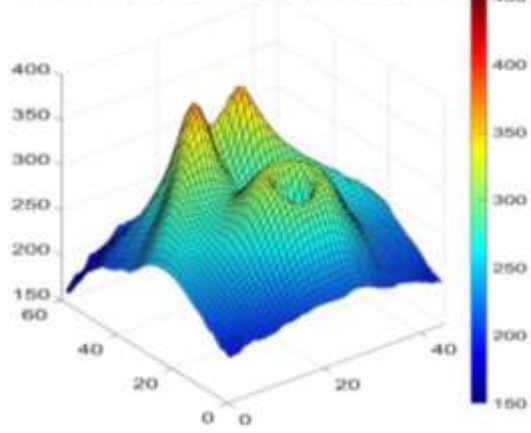


Figura 43. Estructuras retinales usadas: correspondencia anatómica. Imagen obtenida por SS-OCT.

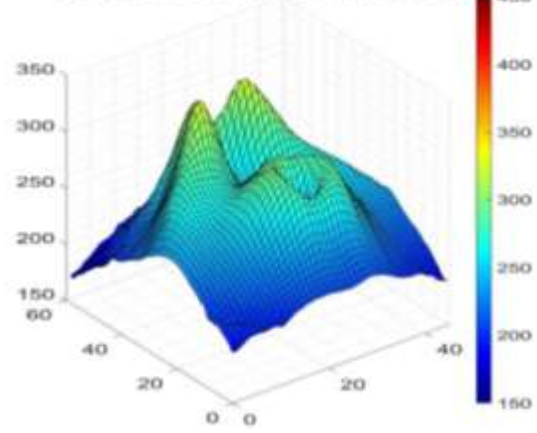
Como ejemplo para visualizar los datos de los que se dispone, en la figura 44 se realiza la representación volumétrica de las medidas tomando la media de espesores de todos los pacientes y controles, para cada una de las cinco estructuras. Tonalidades correspondientes a colores cálidos indican espesores retinales más altos, correspondiente a las zonas de nervio óptico; tonalidades correspondientes a colores fríos indican espesores retinales más bajos, indicando una menor cantidad o la ausencia de nervio óptico.



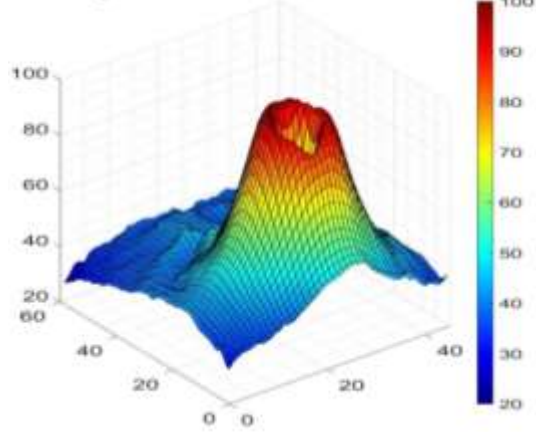
**Espesores de Retina completa: Controles**



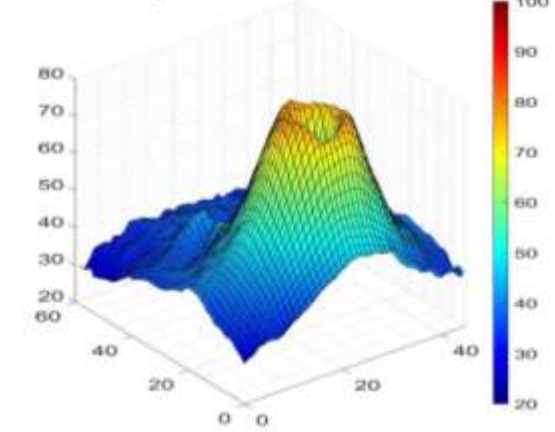
**Espesores de Retina completa: EM**



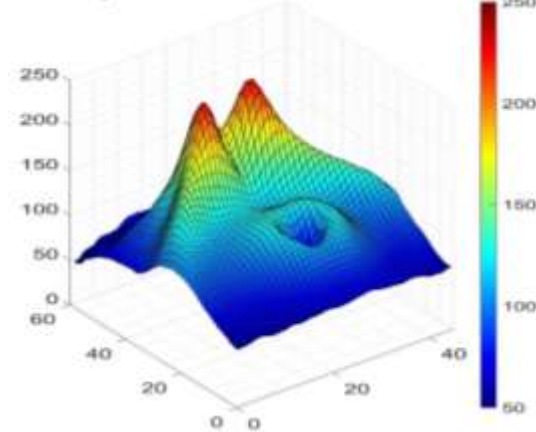
**Espesores de GCL+: Controles**



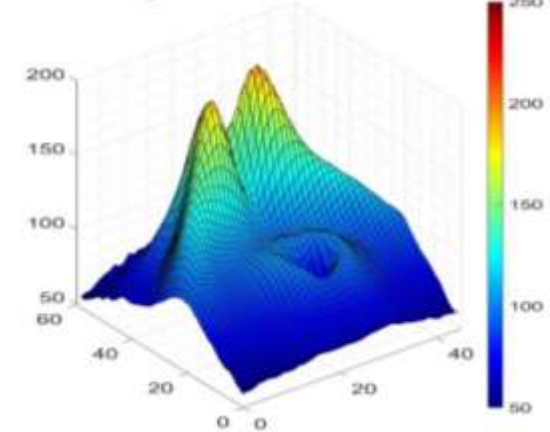
**Espesores de GCL+: EM**



**Espesores de GCL++: Controles**



**Espesores de GCL++: EM**





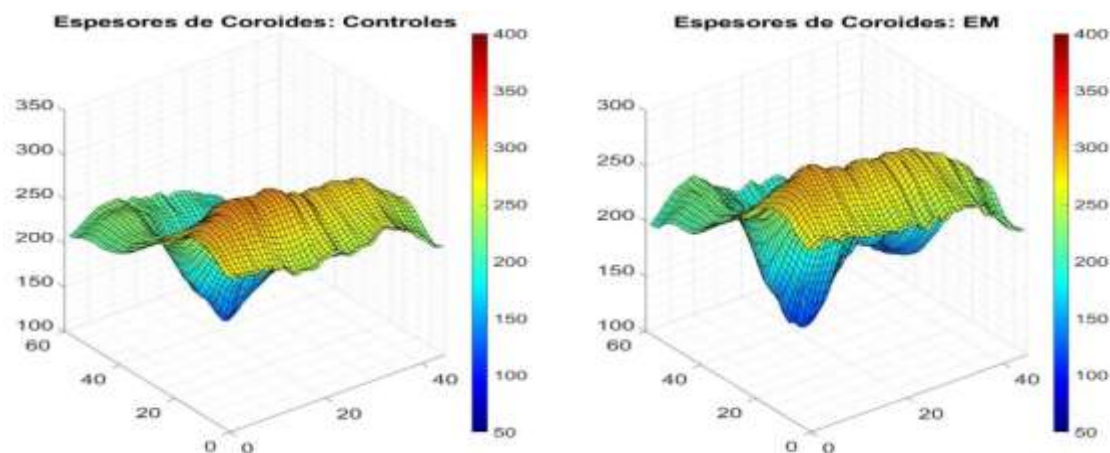


Figura 44. *Media de espesores retinales ( $\mu\text{m}$ ) de cada capa: representación volumétrica. A la izquierda, controles; a la derecha, pacientes EM. De arriba abajo: RNFL, Retina completa, GCL+, GCL++, Coroides.*

Se dispone de una sola imagen para cada ojo y capa, por lo que el total de imágenes será la suma de los 56 pacientes de EM y los 47 sujetos de control, para un total de 103 imágenes. Como registro, **se estudiará la media de ambos ojos.**

Este es el registro principal que se usará a lo largo de la tesis, aunque puesto que también se han realizado artículos con otras bases de datos, se citarán también estudios con otras bases de datos semejantes.

## 4.2 Formato de los registros

Un equipo de OCT obtiene medidas puntuales de las diferentes capas de la retina. Según el protocolo establecido, estas medidas pueden abarcar una superficie rectangular (por ejemplo, 45x60 puntos mediante el protocolo 3DH Wide, como es el caso del registro de la tesis) o circular. Al operador clínico, generalmente los resultados se presentan como valores medios en determinadas regiones predefinidas, como las ya mencionadas ETDRS y TSNT; sin embargo, desde el punto de vista de análisis de la información obtenida, es más útil analizar los datos originales, pues de esa manera se posibilita realizar agrupamientos más flexibles, que incluso pueden diferir entre las estructuras analizadas.

Tal y como se ha indicado previamente, un gran número de trabajos que estudian el espesor de las capas retinales se basan en unas plantillas predefinidas de espesores medios, sobre todo de plantillas denominadas ETDRS (acrónimo de *Early Treatment Diabetic Retinopathy Study*) ([Lee, K. et al., 2016](#)) ([Tántrai E. et al., 2012](#)), que tal y como puede observarse en la figura 45, dividen cada imagen en una serie de circunferencias concéntricas, a su vez divididas en cuadrantes, obteniendo 9 regiones distintas y por lo tanto estos registros recogerán un total de 9 valores de espesores medios.

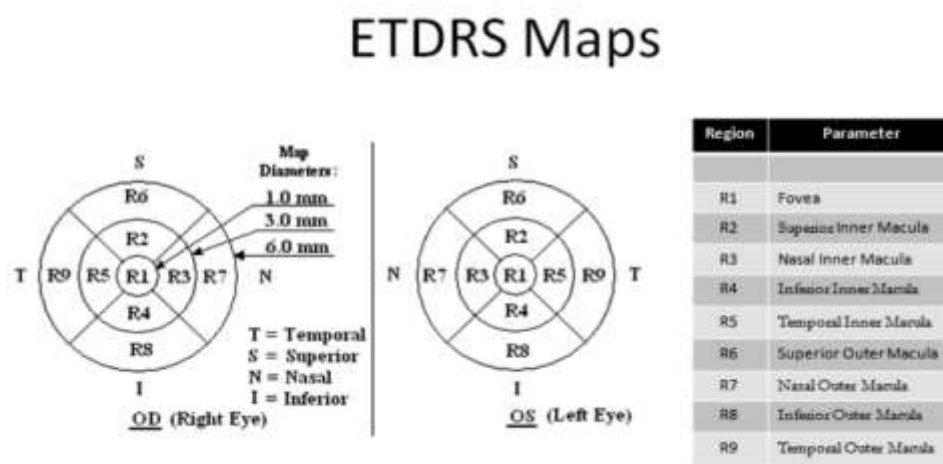


Figura 45. Esquema de las plantillas ETDRS.

Otro tipo de plantillas de espesores medios son las denominadas TSNIT (Temporal, Superior, Nasal, Inferior, Temporal), que también subdividen cada imagen en una serie de regiones diferenciadas. Durante la realización de esta tesis se ha trabajado también con espesores medios en los formatos ETDRS y TSNIT, pero se han recogido únicamente los resultados correspondientes al estudio de espesores puntuales.

Tal y como se ha visto en el capítulo 4, se tienen registros de espesores retinales, obtenidos mediante OCT, de cinco estructuras retinales. El equipo Tripton, en su protocolo Wide, permite exportar los datos a formato *csv*, hojas de *Excel* que pueden leerse y ordenarse posteriormente en Matlab. En estos registros, la información principal reside en los espesores retinales de cada una de las capturas axiales. Para identificar su posición exacta en el corte también contiene las coordenadas x-y de cada espesor, listándose las medidas de

columna a columna y de fila a fila. Dependiendo del registro puede haber uno o más imágenes por cada capa, en cuyo caso se hace la media de estas.

Tanto para registros de espesores puntuales como registros de espesores medios se tiene, además, que para cada medida los registros contienen un identificador de paciente codificado, fecha de nacimiento, fecha del registro, un conjunto de parámetros que indiquen el estado del hardware y el software de captura y capa a la que corresponde la medida. No se utiliza toda la información sino solo aquella que se considera de interés como son las fechas de nacimiento y fecha de realización de la medida, a partir de las cuales puede obtenerse la edad de cada paciente.

### **4.3 Softwares usados y otras especificaciones**

Las implementaciones de redes neuronales, cálculos de AUC, diseños de interfaces gráficas para la visualización de imágenes de OCT (así como de otros datos) y otros programas planteados para la resolución de esta tesis, hayan sido incluidos o no en esta memoria, se realizan en el software Matlab, actualizado a lo largo de los tres años de duración de este trabajo, albergando todas las versiones entre 2019a y la 2021a. Como software de apoyo, se utiliza también Microsoft Excel para abrir los registros OCT y el software estadístico SPSS para la comprobación del correcto funcionamiento del código implementado para el cálculo de AUCs sobre los registros OCT.

El software se ejecuta en un ordenador con una memoria RAM de 16 GB y procesador Intel Core i5-3570K a 3,4 GHz y 3.8 GHz, así como en un ordenador personal, con memoria RAM de 16 GB y un procesador Intel Core i7-4790K a 4 GHz. Ambos ordenadores son de 64 bits, y cuentan con GPU, con la que es posible reducir los tiempos de ejecución de programas, recurso importante al considerar el tiempo de entrenamiento de los procesos de ML implementados.

## 4.4 Determinación de la región de mayor capacidad discriminante

### 4.4.1 Introducción

Tal y como se ha visto en la sección anterior, se disponen de bases de datos de un área de dimensiones determinadas correspondiente a cinco estructuras retinales diferentes. Estas imágenes se usarán como entrada en programas que se implementarán en el siguiente capítulo de la tesis. Sin embargo, hay que tener en cuenta que, por un lado, es posible que la información de interés diagnóstico no se encuentre en toda la imagen, que haya zonas que puedan considerarse ruido. Se considera, por lo tanto, la posibilidad de subdividir la base de datos en zonas más pequeñas. Estas subdivisiones constituirán lo que a partir de ahora se considerarán y denominarán distintas **modalidades** de estudio.

Por otro lado, puesto que para diagnosticar EM se mide la desmielinización, y debido a que algunas estructuras retinales tienen mayor presencia de fibras nerviosas que otras, no todas tienen la misma importancia para este diagnóstico. En estudios de esta naturaleza, se consideran de mayor interés las que tengan una mayor concentración de este tipo de fibras en contraste a, por ejemplo, una mayor concentración de vasos sanguíneos, como es el caso de la coroides.

En otras palabras, distintas estructuras retinales y distintas zonas dentro de las imágenes tienen mayor **capacidad discriminante** (o **diagnóstica**) que otras. Así, esta sección estará dedicada a estudiar la capacidad diagnóstica de las diferentes subdivisiones de la base de datos disponible.

### 4.4.2 Ventanas cuadradas. Cálculo de AUC

#### 4.4.2.1 AUC. Fundamentos

Para evaluar la capacidad discriminante de un conjunto de datos se utiliza la AUC (*Area Under the Curve*), el parámetro que mide el grado de solapamiento entre dos distribuciones determinadas. Se calcula sobreponiendo la distribución de un parámetro determinado (en este caso, espesores retinales) para los dos grupos que se espera que dicho parámetro pueda

separar (sujetos de control y pacientes de EM). En la figura 46 aparece este concepto esquematizado.

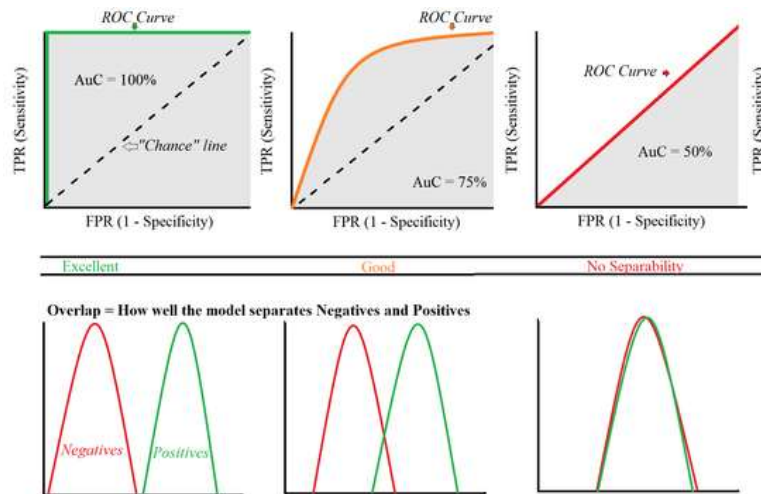


Figura 46. AUC dependiendo de la calidad del clasificador estudiado.

Conjuntos de datos con baja capacidad discriminante tienen distribuciones con un elevado grado de solapamiento, obteniéndose una AUC resultante de alrededor de 0,5. Es el caso de parámetros que no son de interés diagnóstico, espesores retinales en regiones que no contengan el número necesario de fibras nerviosas y distribuciones con un elevado número de medidas erróneas. Buenos clasificadores, por otro lado, son aquellos que permiten separar estas dos distribuciones lo máximo posible.

En el caso de esta tesis, el objetivo es encontrar dos distribuciones de espesores que estén lo suficientemente separadas como para que se pueda discernir entre ambos grupos con el mínimo error posible, o, en otras palabras, identificar la distribución de espesores que determine, dado un nuevo sujeto, si este tiene o no EM, con las mínimas probabilidades de error posibles.

#### 4.4.2.2 AUC. Implementación

El cálculo de AUC se realiza sobre el registro ya presentado en la base de datos. Se analizan los pacientes mediante 4 modalidades: ojos izquierdos, ojos derechos, ojos aleatorios (en cuyo caso para cada paciente se elige aleatoriamente uno de los dos ojos), y

la media de ambos ojos, pero finalmente solo se presentan los resultados correspondientes a **la media de ambos ojos**, puesto que es la modalidad con la que se obtienen los mayores valores de AUC.

Se considera ahora el registro de espesores puntuales. Puesto que se tienen espesores puntuales, es necesario tener en cuenta que la AUC no puede calcularse valor a valor, puesto que dicho número corresponde a cada medida de espesor puntual, que por sí sola no tiene significado diagnóstico. Por ello, debe considerarse un conjunto de espesores de las imágenes de los sujetos de control, que corresponderán a los espesores de una zona determinada de la imagen, y calcular la AUC con sus correspondientes espesores para los pacientes de EM.

El método para establecer cada una de estas distribuciones que se decide usar en esta tesis se lleva a cabo mediante el algoritmo que aparece en la figura 47, y se le llama **análisis de regiones cuadradas**.

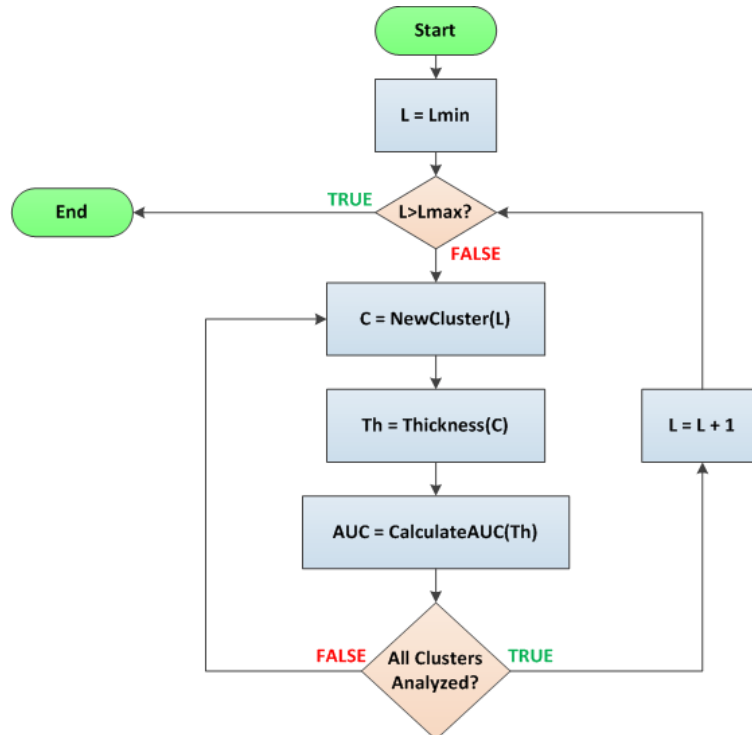


Figura 47. Algoritmo para el cálculo de AUCs mediante regiones cuadradas.

Este proceso se basa en el uso de ventanas cuadradas ( $L \times L$ ), que se irán deslizando a lo largo de la imagen. Para cada una de estas posiciones, se consideran los espesores contenidos dentro de la ventana, tanto para el caso de sujetos de control como de pacientes de EM, y se calcula la AUC entre los dos grupos. Una vez hecho el cálculo, se desplaza la ventana en un píxel, y se vuelve a hacer el cálculo hasta haber recorrido toda la imagen. El proceso se repite para todas las longitudes de ventanas, desde  $L_{min}$  hasta  $L_{max}$ , y para todas las imágenes de las que se dispone. A recalcar que se empieza dejando un margen a cada lado de la imagen; esto es porque se ha considerado los bordes como ruido, puesto que la mayoría de las fibras nerviosas y por tanto información de interés se halla en la parte más central de la imagen.

Además de la AUC máxima para cada imagen, también se guarda la posición en X e Y de la ventana, así como la longitud L mediante la que se obtiene la mayor AUC. Los resultados se recogen en la tabla 4, y la evolución de la AUC al aumentar el tamaño de la ventana cuadrada aparece en la figura a continuación.

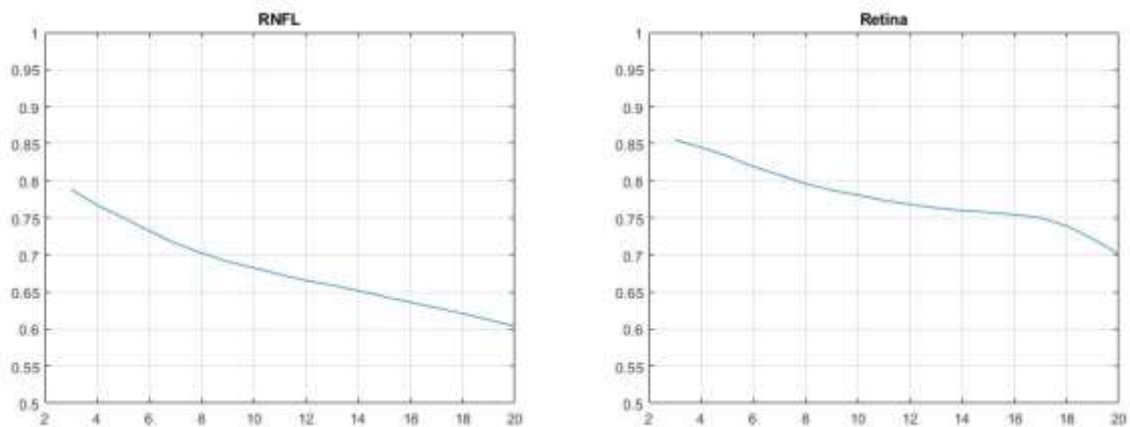
Se ha realizado la prueba de Shapiro-Wilk a los datos de los pacientes para comprobar si siguen una distribución normal. Al obtenerse una *p value* inferior a 0,05, se considera que no siguen dicha distribución. Por ello, al presentarse descriptivos sobre la distribución de espesores, se hará mediante la mediana y el rango intercuartílico (IQR).

<i>Tabla 4. Detección de la región con máxima capacidad discriminante, ventanas cuadradas: media de ambos ojos. Longitud óptima de ventana, capa y coordenadas.</i>					
<b>Capa</b>	<b>Mediana de espesores (<math>\mu\text{m}</math>)</b> <b>[IQR]: Controles</b>	<b>Mediana de espesores (<math>\mu\text{m}</math>)</b> <b>[IQR]: EM</b>	<b>L óptima</b> <b>(píxeles)</b>	<b>Coordenada inicio (C, F)</b>	<b>Valor de AUC máxima</b>
RNFL	42,38 [39,92]	38,47 [32,23]	3	(37, 28)	<b>0,7888</b>
Retina	247,33 [66,23]	239,20 [53,01]	3	(24, 26)	<b>0,8550</b>
GCL+	41,77 [15,27]	39,34 [12,92]	3	(24, 26)	<b>0,8695</b>

GCL++	97,51 [46,03]	88,48 [33,28]	3	(26, 27)	<b>0,8785</b>
Coroides	229,02 [85,12]	222,75 [82,86]	3	(7, 43)	<b>0,5894</b>

Se ha recopilado, además de las AUCs máximas y los tamaños de ventana, la mediana de espesores correspondientes a los controles y los de los pacientes. Puede observarse que, como era de esperar, los espesores retinales correspondientes a los sujetos de control son superiores, puesto que en los pacientes ha tomado lugar la desmielinización, es decir, el desgaste de las fibras nerviosas, y por tanto la reducción de los espesores.

Respecto al tamaño óptimo de ventana, tal y como se observa en la tabla, el máximo de AUC para el registro tratado en la tesis se encuentra en la longitud de ventana mínima (3) para todas las estructuras, con resultados cada vez peores a medida que se aumenta la L. Esto se observa en la figura 48.





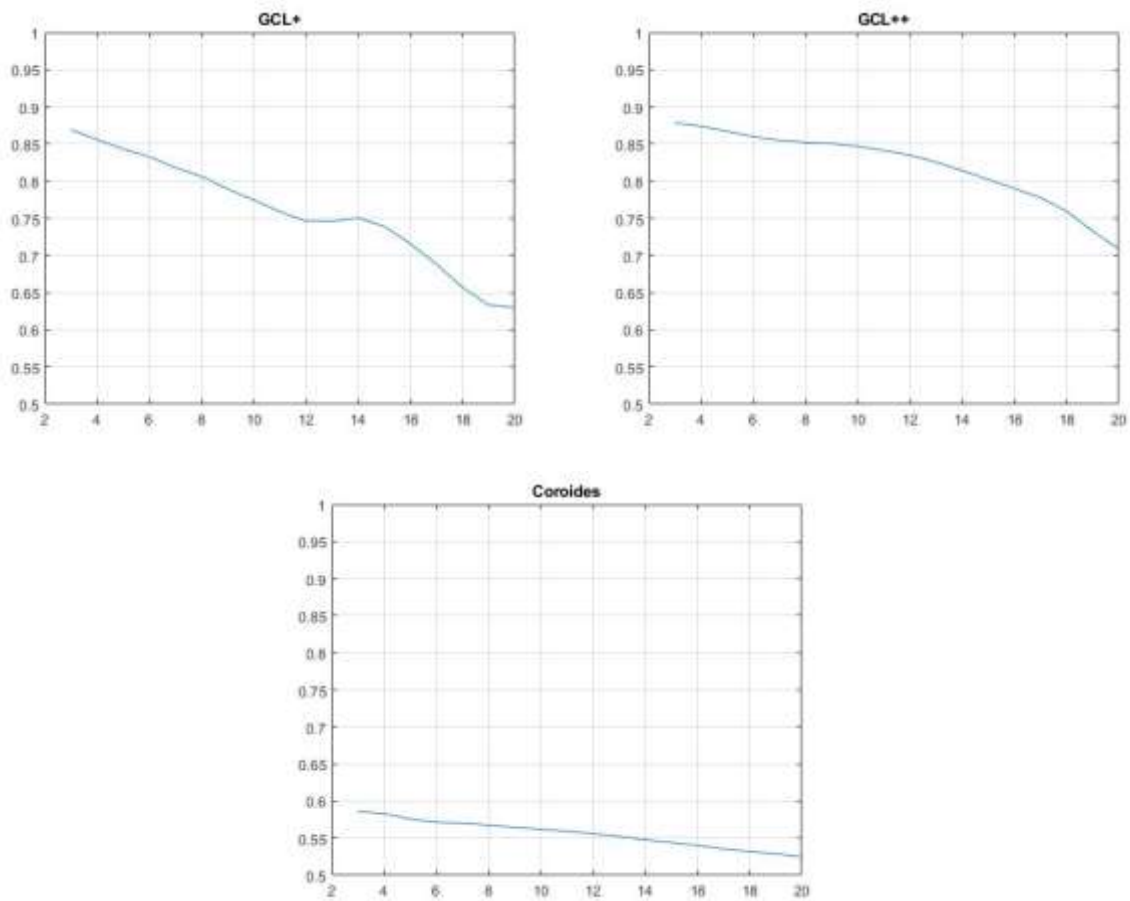


Figura 48. AUC en función del tamaño de ventana cuadrada. Media de ambos ojos.

#### 4.4.3 Effect Size. La $d$ de Cohen.

##### 4.4.3.1 Fundamentos

Además del análisis de regiones arbitrarias mediante ventanas cuadradas y AUC, descrito en la sección anterior, se plantea una segunda manera de encontrar una región de notable interés diagnóstico: estudiando el Tamaño del Efecto, conocido comúnmente por su término inglés, *Effect Size* (*ES*).

El ES mide el efecto de un tratamiento. A diferencia de otros *tests* de significancia, el ES es independiente del tamaño de la muestra (Becker L.A., 2000). Es una medida que ya se ha implementado antes en el estudio de EM (García-Martin, E. et al., 2020), y en esta tesis se espera replicar y mejorar los resultados obtenidos con tal de mejorar el diagnóstico de EM.

La forma más común de medir el ES es mediante el coeficiente de Cohen, que se basa en calcular la diferencia estandarizada entre dos medias ([Cohen J., 1998](#)), mediante la siguiente ecuación:

$$d = \frac{M1-M2}{\sigma} \quad (39)$$

, donde d es el coeficiente de Cohen, M1 y M2 son las medias de los grupos 1 y 2 y  $\sigma$  es la desviación estándar (suponiendo desviaciones estándar iguales). Hay que tener en cuenta que en el caso de la tesis se tiene una distribución para cada valor de espesor, para cada coordenada dentro de cada imagen. Teniendo esto en cuenta, el coeficiente de Cohen se calcularía mediante:

$$d(x, y) = \frac{M1(x,y)-M2(x,y)}{\sigma} \quad (40)$$

La desviación estándar,  $\sigma$ , se obtiene mediante:

$$\sigma = \sqrt{\frac{\sum(X-M)^2}{N}} \quad (41)$$

, donde X es cada valor de una de las distribuciones, M su media y N el número de muestras.

Hay distintas variaciones de esta fórmula. En la práctica es común usar no la desviación estándar de una de las dos distribuciones sino la media cuadrática de las desviaciones estándares de los dos grupos ([Becker L.A., 2000](#)), tal que:

$$\sigma_{pooled} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}} \quad (42)$$

, y será también la manera según la que se calcularán los coeficientes de Cohen en la tesis. A tener en cuenta que, cuando las dos distribuciones estándar son similares, el resultado de la media cuadrática no será muy diferente al de la media de ambas distribuciones.

#### 4.4.3.2 Implementación

El cálculo de ES se realiza en el mismo programa que calcula las medias y las AUCs; al final, además de extraer imágenes completas y las subdivisiones en ventanas cuadradas, también se exportará la subdivisión resultante de esta implementación. Tras calcular píxel a píxel los coeficientes de Cohen, vistos en el apartado anterior, es necesario considerar un umbral a partir del cual se pasará a considerar un espesor dado como parte de la región de Cohen. En la tesis, se tomará como umbral el máximo ES de la estructura menos discriminante de entre las cuatro que tienen un interés diagnóstico (que según las AUCs son RNFL, Retina, GCL+ y GCL++). Para cada una de estas cuatro estructuras, se calculará el máximo ES, y se tomará, de entre esos 4 máximos, el más pequeño. Todo espesor cuyo ES supere este umbral se considerará parte de la nube de puntos que constituirá la tercera modalidad del registro estudiado. Puesto que las redes neuronales convolucionales aceptan imágenes como entradas, y no nubes de puntos, para poder aplicar este tipo de redes a las regiones de Cohen se transformarán las nubes de punto a una imagen 45x60 cuyos espesores que no correspondan a los puntos pertenecientes a las regiones delimitadas por este método valgan 0. En la figura 49 puede observarse el resultado.

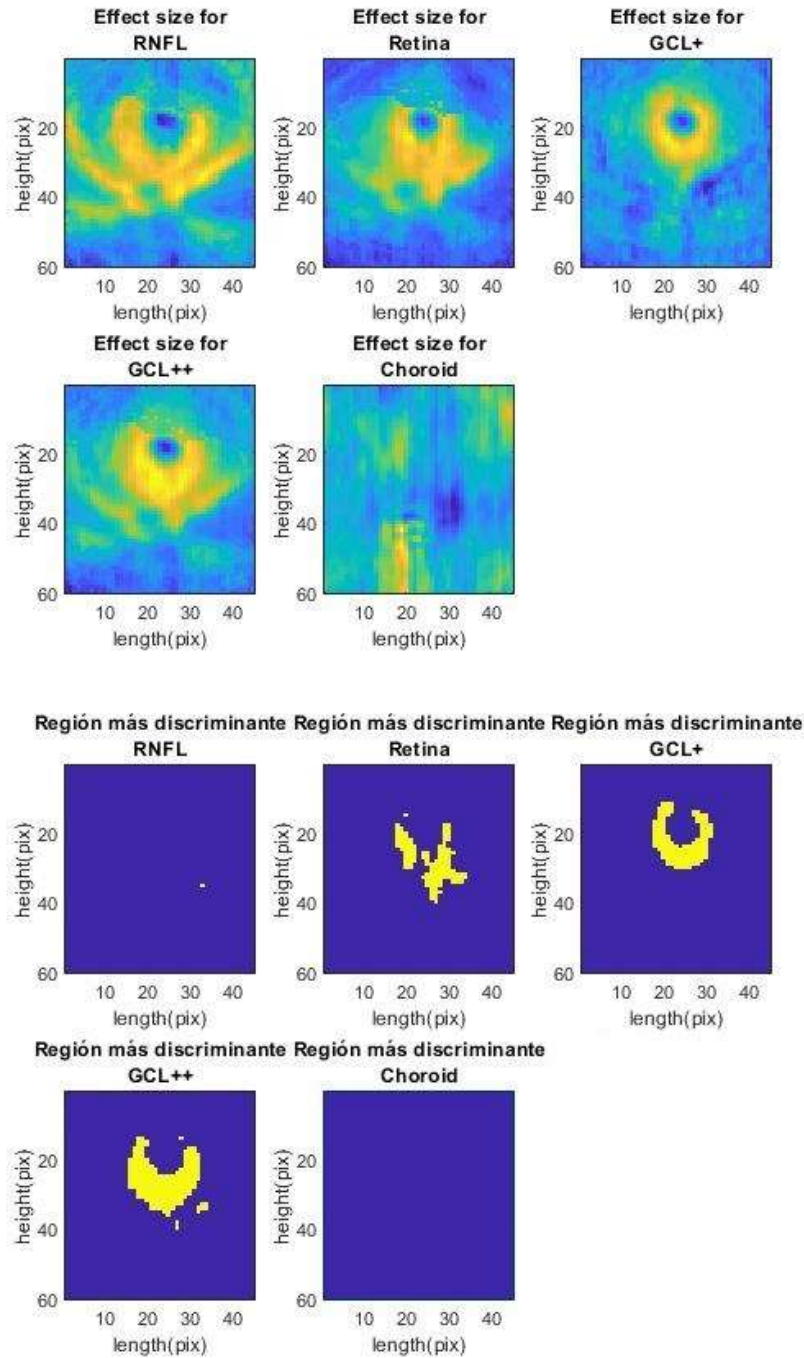


Figura 49. Regiones de Cohen: media de ambos ojos. Arriba: ES calculado para cada píxel. Abajo: nube de puntos que supera el umbral.

Tal y como ocurre con la AUC, la coroides es siempre la estructura con el menor interés diagnóstico. En la media de ambos ojos, que tal y como se ha visto en la sección anterior, es con la que se obtienen mejores resultados de AUC, la Retina sustituye a la RNFL como una de las tres estructuras de mayor interés diagnóstico. Las nubes de punto se concentran en la parte central de cada imagen, que es donde reside la mayor cantidad de fibras nerviosas y por ello la información de interés.

#### 4.4.4 *Effect Size*. La Delta de Cliff

##### 4.4.4.1 Fundamentos

Tal y como se ha visto en el apartado anterior, la *d* de Cohen se utiliza para cuantificar el tamaño del efecto y así obtener una medida adicional de la capacidad discriminante de cada punto (x,y), que a su vez puede usarse para seleccionar regiones de interés en una imagen. Sin embargo, la *d* de Cohen presupone distribuciones normales de datos.

El estadístico Delta de Cliff, en cambio, permite cuantificar la magnitud de la diferencia entre dos grupos de observaciones que resultan incompatibles con el presupuesto de normalidad. Por lo tanto, constituye un parámetro con capacidad para calcular el valor del tamaño del efecto no paramétrico para dos grupos de observaciones ([Macbeth G. et al.,2011](#)).

La Delta de Cliff se define por:

$$Delta = \frac{\#(x_1 > x_2) - \#(x_1 \leq x_2)}{n_1 n_2} \quad (43)$$

, donde  $n_1$  y  $n_2$  son el tamaño de las distribuciones 1 y 2 respectivamente,  $x_1$  y  $x_2$  son valores correspondientes a los grupos 1 y 2 y el símbolo # indica conteo. Esta expresión estima la probabilidad de que un valor de uno de los dos grupos sea mayor a un valor del otro grupo, menos la probabilidad inversa; por lo tanto, puede entenderse como otra medida adicional del solapamiento existente entre dos distribuciones ([Macbeth G. et al.,2011](#)). Al igual que ocurre con la AUC, valores de Delta muy cercanos a 1 o -1 indican solapamiento mínimo y por lo tanto que se está estudiando una variable con gran capacidad discriminante.

Una forma alternativa de calcular la Delta de Cliff entre dos distribuciones es:

$$Delta = \frac{2*U}{n_1 n_2} - 1 \quad (44)$$

, donde  $n_1$  y  $n_2$  son el tamaño de las distribuciones 1 y 2 respectivamente y “U” es el estadístico obtenido a partir de la prueba no paramétrica de Mann-Whitney ([Goedhart J., 2016](#)).

#### 4.4.4.2 Implementación

El cálculo del ES en este caso se realiza de forma equivalente al de la d de Cohen: píxel a píxel, se calcula primero el estadístico de Mann-Whitney sobre las distribuciones dadas por el valor de espesor de esa coordenada para cada paciente y sujeto de control, con el que se obtiene posteriormente la delta de Cliff aplicando.

Tras calcular píxel a píxel las Deltas de Cliff, de igual forma que se ha hecho en el apartado anterior, se selecciona como umbral el máximo ES de la estructura menos discriminante entre las cuatro que tienen interés diagnóstico. Para cada una de estas cuatro capas, se calculará el máximo ES, y se tomará, de entre esos 4 máximos, el más pequeño. Todo espesor cuyo ES supere este umbral se considerará parte de la nube de puntos que constituirá la tercera modalidad del registro estudiado. En la figura 50 aparece la región definida por estas deltas, que se llamará a partir de ahora la **región de Cliff**.

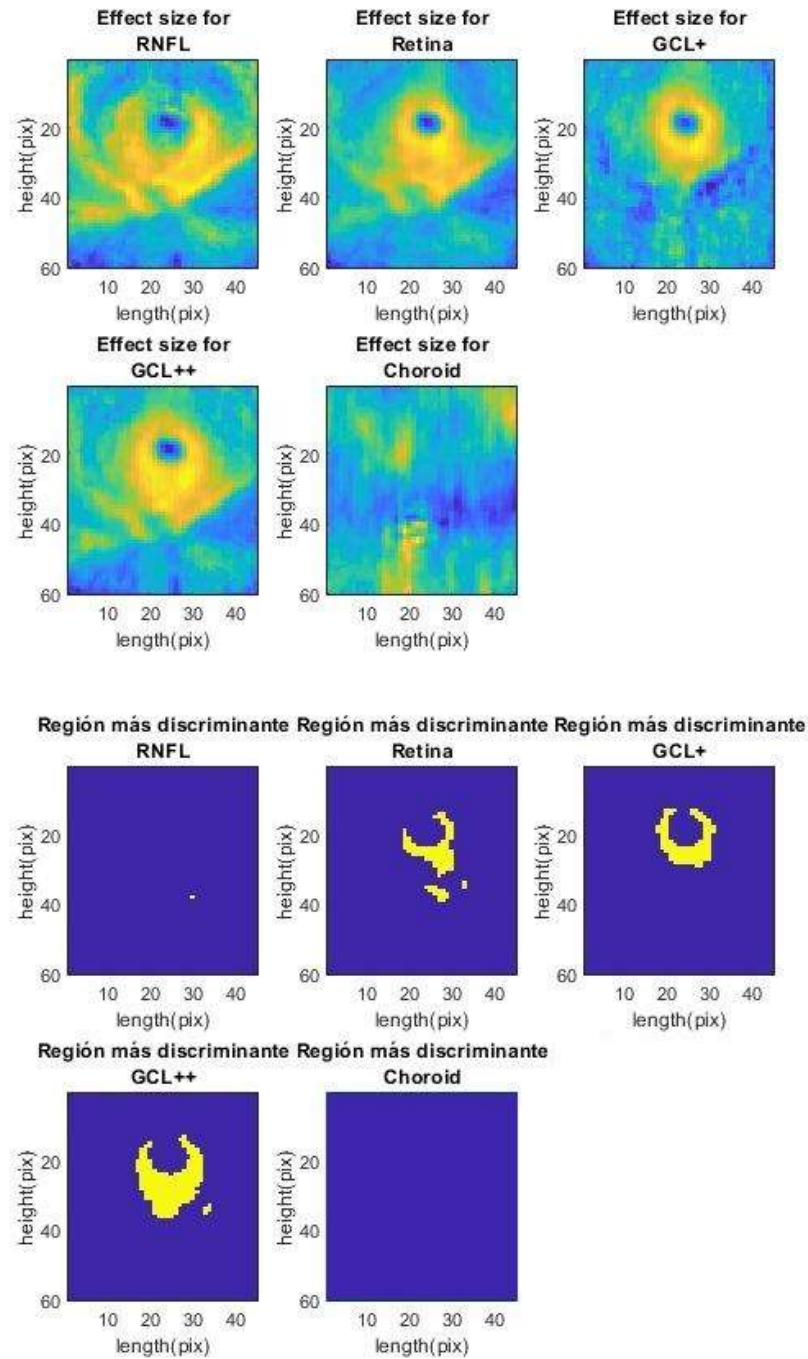


Figura 50. Regiones de Cliff: media de ambos ojos. Arriba: ES calculado para cada píxel. Abajo: nube de puntos que supera el umbral.

Al hacer el cálculo del tamaño del efecto mediante el método de la Delta de Cliff, se han formado regiones de forma equivalente a las generadas mediante la  $d$  de Cohen. Al igual que en el método anterior, la RNFL es la estructura con tamaños de efectos menores. Las formas de las regiones de las estructuras GCL+ y GCL++ son semejantes a las obtenidas por Cohen, pero en el caso de la Retina, esta vez se observa una forma de herradura, similar a la de las otras estructuras.

Las nubes de punto vuelven a concentrarse en la parte central de cada imagen, que es donde reside la mayor cantidad de fibras nerviosas y por ello la información de interés.

#### 4.4.5 Conclusiones

En este capítulo, se ha presentado dos métodos para identificar, en cada una de las estructuras de la retina disponibles, regiones concretas con alta capacidad discriminante. El primero, buscando regiones cuadradas mediante el cálculo de AUC a lo largo de una ventana deslizante; el segundo, buscando regiones arbitrarias mediante el tamaño del efecto (métodos de Cohen y de Cliff). De los resultados puede establecerse que:

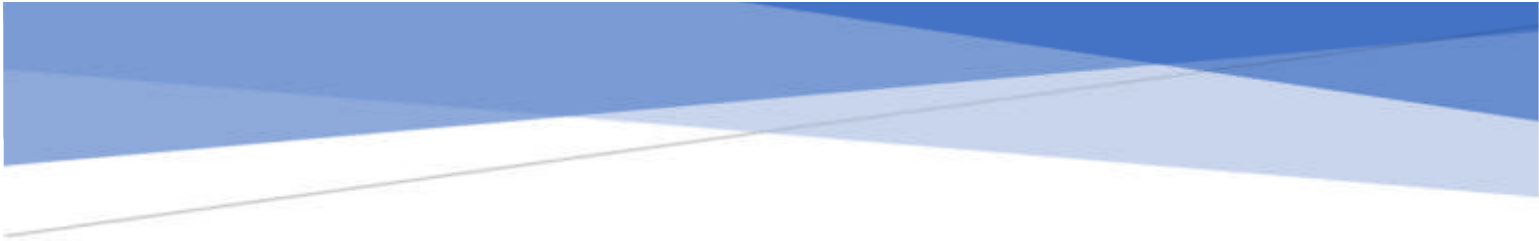
- Las estructuras retinales que presentan mayor AUC y por lo tanto mejor capacidad discriminante son la Retina completa (AUC = 0,8550) y capas ganglionares (AUC = 0,8695 para GCL+ y AUC = 0,8785 para la GCL++), seguidas de la RNFL (0,7888) y, por último, como era de esperarse, la Coroides (0,5894). Este resultado es congruente con el obtenido en las regiones de Cohen, pues solo se han obtenido nubes de puntos de interés para la Retina Completa y capas ganglionares.
- Los altos valores de AUC obtenidos sugieren que **el análisis de regiones cuadradas, no solo el análisis sobre ETDRS, puede contribuir significativamente al diagnóstico de EM.**
- Se ha estudiado el tamaño de efecto mediante dos métodos:  $d$  de Cohen y Delta de Cliff, obteniéndose nubes de puntos para las estructuras retinales Retina, GCL+ y GCL++. Las regiones resultantes tienen forma de herradura, coincidiendo esto con resultados de estudios como ([Hu H. et al., 2019](#)), en el que se observa que las zonas



en las que se ha producido una mayor desmielinización siguen esta misma forma. Debido a su similitud, **como entrada a los clasificadores basados en ML se usarán solo las regiones de Cohen.**

Con tal de examinar con más detalle la capacidad discriminante de las estructuras retinales, el siguiente capítulo mirará más allá de la AUC. Se entrará con cierta profundidad en las redes neuronales, algoritmos que tomarán como entrada los registros de OCT y que se esperan entrenar para poder desarrollar un método de apoyo al diagnóstico. Habiendo determinado la región más discriminante de los registros de espesores puntuales en cada una de las estructuras, puede probarse no solo a introducir la totalidad de la imagen, sino esta región concreta que se ha determinado de mayor interés, con la que se espera obtener resultados todavía mejores.





## **5 Implementación y resultados de la aplicación de aprendizaje automático aplicado a la EM**

*Durante este capítulo se irán introduciendo los diferentes modelos de ML que se han decidido usar en esta tesis, y los resultados obtenidos mediante cada uno de ellos. Puesto que su funcionamiento general ha sido detallado ya, aquí solo se explicará cómo se han implementado, tanto en lo que respecta a arquitecturas concretas como a parámetros de entrenamiento.*

## 5.1 Formato de los resultados. Variables y entrenamiento

### 5.1.1 Método *leave one out*

Los resultados de la aplicación de ML sobre las tres modalidades de registro (imágenes completas, ventanas cuadradas y regiones de Cohen) obtenidas en el capítulo anterior se obtienen mediante validación cruzada de los sujetos dejando uno fuera: se subdividen las imágenes totales en un conjunto de entrenamiento correspondiente al número total de imágenes menos una, y se utiliza la imagen restante para comprobar si el clasificador obtiene o no un resultado correcto en la clasificación. El resultado de este proceso constituye una entrada para la matriz de confusión que permitirá evaluar la capacidad diagnóstica. El proceso se repite tantas veces como imágenes haya, y por lo tanto se tiene un total de 103 iteraciones.

Al final del programa de cada clasificador, se guardan los resultados de exactitud de dicha iteración, parámetros de arquitectura y matriz de confusión correspondiente al número de sujetos y pacientes clasificados correcta e incorrectamente. Las matrices de confusión finales presentan el número de imágenes clasificadas correctamente en la primera diagonal y aquellas que no han sido clasificadas correctamente en la segunda diagonal, separando así ambos grupos. La exactitud se obtiene dividiendo el número de casos clasificados correctamente por el número de sujetos totales. A este método, ya mencionado en capítulos anteriores, se le llama *leave one out*.

### 5.1.2 Variables de red

Cada una de las tablas presentadas refleja los resultados de aplicar los programas de ML desarrollados durante la realización de la tesis sobre la base de datos. Los programas de ML se aplican sobre las tres modalidades de imagen ya especificadas: imágenes completas, ventanas cuadradas y regiones de Cohen, que en ocasiones en los comentarios se les referirá como modalidad 1, modalidad 2 y modalidad 3 respectivamente. Respecto a las regiones de Cohen, debido al criterio para delimitar dichas regiones, ya explicado en su

apartado, se obtienen nubes de puntos para la Retina completa, GCL+ y GCL++, pero no para la RNFL ni la Coroides. Para reiterar, las redes neuronales convolucionales aceptan imágenes como entradas, y no nubes de puntos, por lo que las nubes de puntos de Cohen se transforman a imágenes en las que todo espesor fuera de la región de Cohen vale 0.

Además de las iteraciones correspondientes al número total de pacientes, se ha añadido un bucle adicional en el programa con el objetivo de que se pruebe un elevado número de arquitecturas diferentes, y se presentarán solo los mejores resultados de cada proceso de ML; a continuación, se realiza una serie de observaciones en base a los mismos. Ya que cada tipo de ML es de naturaleza distinta y por ello tiene un distinto número de parámetros personalizables, cada tabla tiene sus propias variables en cuanto a la arquitectura óptima. En el caso del MLP, se cambia el número de capas y el número de nodos por capas; en el del RBF, se cambia el parámetro de la distribución normal *spread*. Respecto a las redes profundas: en las CNNs se variará el número de capas convolucionales, en las ResNets el número de conexiones residuales (con la que se añade un total de tres capas convolucionales nuevas) y en las V-Nets el número de capas convolucionales, así como qué estructuras retinales incluye el volumen considerado en cada caso.

### 5.1.3 Entrenamiento

El entrenamiento de las redes depende de cada proceso de ML, aunque tanto para MLP como para CNN, ResNet y V-Net se lleva a cabo usando la variante *stochastic gradient descent with momentum*, ya introducida en su apartado correspondiente. Para estas redes se utiliza un *learning rate (alpha)* de 0,001 y un momento de 0,9 (los parámetros que actualizan los pesos durante el entrenamiento); un *batch size* de 102 (el número total de imágenes de entrenamiento) y un máximo de *epochs* de 1000. En la figura 51 puede verse un ejemplo de entrenamiento, concretamente de una CNN.



Figura 51. Ejemplo del entrenamiento de una CNN.

En la columna derecha de la imagen aparecen los parámetros que intervienen durante el entrenamiento. *Batch size* se refiere al número de muestras que se propaga a lo largo de la red antes de actualizar los parámetros del modelo, es decir, en una iteración de entrenamiento. Grandes *batch sizes* disminuyen el tiempo total de entrenamiento, pero requieren mejores especificaciones de hardware (RAM y GPU).

En este ejemplo, se tienen 102 imágenes para el entrenamiento, un *batch size* de 102, el número máximo de veces que se haya usado el conjunto de datos, o *epochs*, es de 1000. Según estos parámetros, se tomarán las 102 imágenes durante cada iteración de entrenamiento, por lo que una sola iteración es suficiente para que todos los datos de entrenamiento recorran la red neuronal, y por lo tanto equivaldrá también a 1 *epoch*. El proceso se repite un máximo de 1000 veces (1000 *epochs*), pero se le especifica al programa que se detenga si la exactitud alcanzada se repite tres veces seguidas, puesto que se considera la exactitud máxima que pueda alcanzar. La decisión de detener o no el entrenamiento se hace cada 10 iteraciones.

## 5.2 Aplicación de ML a EM. Algoritmos y redes tradicionales

### 5.2.1 Algoritmo *k-nearest neighbors*

#### 5.2.1.1 Implementación

El algoritmo KNN no es una red neuronal, pero se incluye el mismo en el estudio debido a su interés en aplicaciones de pocos datos y de sencilla solución, para así contrastarlo con las redes neuronales y el SVM, procesos de ML de mayor complejidad. Su implementación en Matlab se realiza mediante la función *fitcknn* tras guardar el grupo al que pertenece cada uno de los 103 sujetos (paciente o control) de cada imagen y convertir los espesores a formato *double*, el formato que pide la función.

Como entrada, además de esta clasificación ya conocida también recibe un vector de longitud correspondiente al número de espesores de cada modalidad. Para la modalidad de imágenes completas, 2700 (60x45) espesores; para las ventanas cuadradas, 9 (3x3); para Cohen, 131 espesores en el caso de Retina completa, 139 en el caso de GCL+ y 219 en el caso de GCL++.

Como salida, devuelve el modelo KNN entrenado, que a continuación se usa para predecir de qué clase es la imagen sobrante. El resultado de este proceso es una puntuación correspondiente a la probabilidad de que dicha imagen sea de un grupo u otro (control o paciente); la mayor corresponde a la clase elegida por el clasificador.

Para evitar los casos en que haya tantos vecinos de un grupo como de otro, se prueban valores de *k* impares entre 3 y 9, y se tabula el caso que lleva a la exactitud más alta. Los resultados se tabulan a continuación.

#### 5.2.1.2 Exactitudes obtenidas

En la tabla 5 aparecen las exactitudes resultantes de aplicar el clasificador KNN implementado sobre la modalidad de imágenes completas.

*Tabla 5. Aplicación de KNN: resultados óptimos. Media de ambos ojos, imagen completa (Modalidad 1).*

<b>Estructura retinal analizada</b>	<b>k óptima k=3,5,7,9</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL	7	EM (real)	56	0	97,0874
		Sanos (real)	3	44	
Retina completa	3	EM (real)	55	1	97,0874
		Sanos (real)	2	45	
GCL+	5	EM (real)	45	11	75,7282
		Sanos (real)	14	33	
GCL++	3	EM (real)	56	0	97,0874
		Sanos (real)	3	44	
Coroides	9	EM (real)	43	13	70,8738
		Sanos (real)	17	30	

En este experimento, los mejores resultados de exactitud se han obtenido en la RNFL, Retina completa y GCL++, obteniéndose una exactitud del 97% al analizarse toda la imagen. No existe un número  $k$  de vecinos común entre los mejores resultados de las estructuras analizadas, variando en casi todos los casos.

A continuación, la modalidad de ventanas cuadradas, cuyos resultados aparecen en la tabla 6.

*Tabla 6. Aplicación de KNN: resultados óptimos. Media de ambos ojos, ventanas cuadradas (Modalidad 2).*

<b>Estructura retinal analizada</b>	<b>k óptima k=3,5,7,9</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL	3	EM (real)	52	4	92,2330



		Sanos (real)	4	43	
Retina completa	3	EM (real)	56	0	100,0000
		Sanos (real)	0	47	
GCL+	5	EM (real)	44	12	79,6117
		Sanos (real)	9	38	
GCL++	9	EM (real)	56	0	99,0291
		Sanos (real)	1	46	
Coroides	3	EM (real)	29	27	54,3689
		Sanos (real)	20	27	

Utilizando una región cuadrada de la retina completa, con  $k=3$  se obtiene un resultado perfecto en la clasificación, prácticamente igual que en la estructura GCL++, en la que el clasificador solo falla en uno de los 103 casos. A destacar que los resultados no se corresponden con los valores de AUC máximos, obtenidos en el capítulo anterior, pues fueron para la Retina completa, GCL+ y GCL++.

En tercer lugar, la tabla 7, con los resultados correspondientes a la tercera modalidad de estudio.

<i>Tabla 7. Aplicación de KNN: resultados óptimos. Media de ambos ojos, regiones de Cohen (Modalidad 3).</i>					
<b>Estructura retinal analizada</b>	<b>k óptima k=3,5,7,9</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
Retina completa	3	EM (real)	49	7	89,3204
		Sanos (real)	4	43	
GCL+	3	EM (real)	51	5	88,3495
		Sanos (real)	7	40	
GCL++	3	EM (real)	45	11	83,4951
		Sanos (real)	6	41	

Como se ha indicado, en la modalidad 3 solamente se analizan la Retina completa, GCL+ y GCL++, obteniendo el mejor resultado (exactitud=89,32%) en la Retina completa. Esta modalidad es aquella con la que se obtienen los peores resultados en el caso de usar KNN.

### 5.2.1.3 Resumen de resultados: KNN

El método KNN no es una red neuronal, pero sí un sistema de clasificación supervisado ampliamente utilizado. El proceso de entrenamiento es sencillo (almacenar el conjunto de datos de entrenamiento) y el proceso de decisión consiste en calcular dos distancias euclídeas, de ahí a que el tiempo de ejecución sea pequeño (el menor de los programas de ML implementados). En cuanto a los resultados del clasificador, el mejor resultado (exactitud=100%) se ha obtenido analizando la Retina completa mediante la modalidad 2.

## 5.2.2 Perceptrón Multicapa

### 5.2.2.1 Implementación

Para la implementación de un MLP se usa la función de Matlab *feedforwardnet*. Las variables principales a tener en cuenta son las correspondientes a la arquitectura: número de capas, número de nodos en cada capa y función de activación de las neuronas, esta última debiendo especificarse aparte. Además, también pueden modificarse las opciones de entrenamiento, entre las cuales está el modo concreto de entrenamiento (tipo de *backpropagation* usado), así como los tiempos máximos en cada iteración y cuánto se modifica la dirección del gradiente (a mayor cambio de dirección del vector gradiente, mayor velocidad de ejecución, pero menor exactitud). En la figura 52 aparecen representadas las tres funciones de activación usadas con la función *feedforwardnet*, así como la salida para un rango de entradas  $n$  que va desde  $-\infty$  a  $+\infty$ .

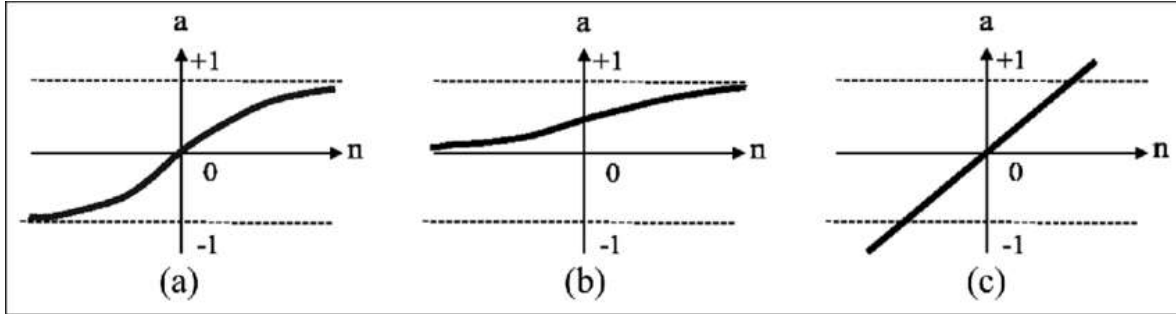


Figura 52. Funciones de activación disponibles para la función usada. (a) log-sigmoid (b) tan-sigmoid (c) purelin. De (Shim J.-Y. et al., 2018).

Así, se tienen las funciones: logarítmica sigmoideal, tangente sigmoideal y lineal pura. La primera (logarítmica sigmoideal) se define como:

$$y = \frac{2}{1+e^{-2x}} - 1 \quad (45)$$

El rango de la salida de esta función va de -1 a 1.

La segunda (tangente sigmoidea) y la tercera (lineal) se han definido ya en apartados anteriores y son:

$$y = \frac{1}{1+e^{-x}} \quad (46)$$

$$y = kx \quad (47)$$

Dada la naturaleza de un problema de clasificación (determinar la probabilidad de que una imagen pertenezca a un grupo u otro), se usará una función tangente sigmoideal como la función de activación de todos los nodos que componen las capas del MLP. Esto es debido a que, como se ha mencionado ya en la sección anterior, las funciones sigmoideales de este tipo son ampliamente usadas en aquellas aplicaciones en las que se quiere obtener una probabilidad concreta, pues su resultado será siempre entre 0 y 1, como se ha observado en la figura anterior.

En la figura 52 aparece representada la arquitectura de esta RN. El MLP tiene 2700 entradas, correspondientes al número total de píxeles de cada imagen (45x60), dos capas

ocultas con neuronas de función tangente sigmoidal y dos salidas: las probabilidades de cada una de las dos clases (pacientes y enfermos). Se establece de forma manual que la mayor de las dos salidas obtenidas corresponda a la clase a la que pertenece la imagen actual. Si el primer componente del vector es el mayor, dicha imagen corresponde a un paciente enfermo de EM; si en cambio el mayor valor es el segundo componente, a un sujeto sano.

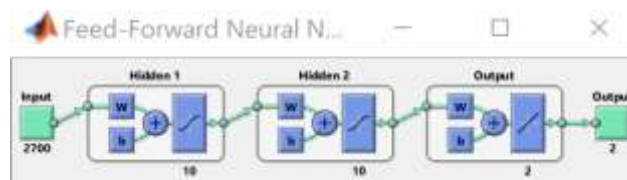


Figura 53. Ejemplo de MLP implementado. Arquitectura

El tiempo de ejecución está correlacionado con el tamaño de la arquitectura utilizada: a más capas y mayor número de neuronas, más tarda en ejecutarse. Debido a que se trata de una red sencilla, el tiempo de ejecución es bajo, por lo que es posible probar un número mayor de arquitecturas que en redes más complejas. Así, se generan dos sencillos bucles para probar un distinto número de capas (de 1 a 4) y nodos por capa (de 1 a 64).

A su vez, ([Zhang Y.D. et al., 2018](#)) introduce la posibilidad de arquitecturas con modelos semejantes a embudos: que la primera capa tenga un número determinado de nodos (neuronas) y las siguientes un número superior. Se establece, así, una segunda ejecución del programa con este modo, aumentando al doble el número de neuronas para capas sucesivas y redondeando hacia arriba para asegurar un número entero. A modo de prueba también se hará la inversa: considerar un número alto de neuronas al principio, que irá disminuyendo a medida que se alcanzan las capas más profundas.

#### 5.2.2.2 Exactitudes obtenidas

La tabla 8 presenta los resultados de la aplicación del MLP sobre la modalidad de imágenes completas.

*Tabla 8. Aplicación de MLP: resultados óptimos. Media de ambos ojos, imagen completa (Modalidad 1).*

Estructura retinal analizada	Arquitectura (N.º capas, N.º nodos por capa)		Matriz de confusión	EM (pred.)	Sanos (pred.)	Exactitud (%)
	RNFL	Capas		1	EM (real)	
	Nodos	64	Sanos (real)	1	46	
Retina completa	Capas	4	EM (real)	55	1	98,0583
	Nodos	7, 14, 28, 56	Sanos (real)	1	46	
GCL+	Capas	3	EM (real)	43	13	76,6990
	Nodos	15, 30, 59	Sanos (real)	11	36	
GCL++	Capas	2	EM (real)	56	0	100,0000
	Nodos	32, 64	Sanos (real)	0	47	
Coroides	Capas	1	EM (real)	44	12	72,8155
	Nodos	12	Sanos (real)	16	31	

Al aplicar el MLP sobre las imágenes completas, se obtienen exactitudes muy elevadas en la RNFL, Retina completa y GCL++, usando 1, 4 y 2 capas respectivamente.

Para ver si se obtienen mejores resultados en áreas más pequeñas, se estudia la modalidad de ventanas cuadradas, con resultados tabulados en la tabla 9.

*Tabla 9. Aplicación de MLP: resultados óptimos. Media de ambos ojos, ventanas cuadradas (Modalidad 2).*

Estructura retinal analizada	Arquitectura (N.º capas, N.º nodos por capa)		Matriz de confusión	EM (pred.)	Sanos (pred.)	Exactitud (%)
	RNFL	Capas		2	EM (real)	
	Nodos	25, 50	Sanos (real)	2	45	
Retina	Capas	2	EM (real)	56	0	100,0000

completa	Nodos	32, 63	Sanos (real)	0	47	
GCL+	Capas	1	EM (real)	49	7	83,4951
	Nodos	57	Sanos (real)	10	37	
GCL++	Capas	1	EM (real)	56	0	100,0000
	Nodos	16	Sanos (real)	0	47	
Coroides	Capas	2	EM (real)	43	13	68,9320
	Nodos	19, 38	Sanos (real)	19	28	

Al hacer la misma aplicación sobre la modalidad 2, se alcanza el 100% en exactitud en dos de las estructuras: Retina completa y GCL++. Aunque se obtengan mejores resultados en GCL+, esta estructura continúa conllevando los peores resultados de exactitud después de la Coroides.

Por último, la tabla 10 con la modalidad de regiones de Cohen. El número de entradas al MLP es de 131 espesores en el caso de Retina completa, 139 en el caso de GCL+ y 219 en el caso de GCL++.

<i>Tabla 10. Aplicación de MLP: resultados óptimos. Media de ambos ojos, región de Cohen (Modalidad 3).</i>						
<b>Estructura retinal analizada</b>	<b>Arquitectura (N.º capas, N.º nodos por capa)</b>		<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
Retina completa	Capas	3	EM (real)	56	0	100,0000
	Nodos	1, 1, 1, 1	Sanos (real)	0	47	
GCL+	Capas	3	EM (real)	44	12	85,4369
	Nodos	1, 1, 1, 1	Sanos (real)	3	44	
GCL++	Capas	2	EM (real)	56	0	100,0000
	Nodos	1, 1	Sanos (real)	0	47	

Se obtienen dos exactitudes de 100%, substancialmente superiores a los resultados obtenidos con KNN (80-85%).

### 5.2.2.3 Resumen de resultados: MLP

El MLP constituye la primera red neuronal aplicada; se ha implementado de tal forma que durante su ejecución se pruebe un elevado número de arquitecturas, variando número de capas y número de neuronas por cada capa. Este segundo clasificador obtiene cinco exactitudes del 100%, un resultado ligeramente mejor al KNN (que obtenía el 100% en un caso) con un tiempo de ejecución notablemente mayor. Se obtienen dos exactitudes de 100% tanto con las regiones de Cohen como con ventanas cuadradas.

## 5.2.3 Radial Basis Function Networks

### 5.2.3.1 Implementación

Tal y como se ha explicado en su apartado, las RBF son redes neuronales que constan de una capa interna, cuyas neuronas comparten la misma función de activación (función de base radial) y una capa de salida que consiste en un clasificador lineal a partir de las salidas de las neuronas.

La implementación en Matlab de esta red se puede realizar mediante dos funciones internas: *newrbe* y *newrb*. A diferencia de muchos de los otros tipos de funciones que crean redes neuronales, ambas funciones se definen al mismo tiempo que se entrena la red. Por lo tanto, Matlab buscará automáticamente la arquitectura (esto es, el número de neuronas) óptima que minimice el error de entrenamiento. En ambos casos, el programa fija los centros de las funciones gaussianas de forma uniforme a lo largo del espacio de los datos.

La primera función, *newrbe*, forma una RBF con tantas neuronas como número de imágenes se dispone, la única variable con la que se puede experimentar es el parámetro, *spread*. Esta variable es inversamente proporcional a la *b* vista en el apartado de teoría, y por lo tanto es equivalente a pensar en ella como la desviación estándar de la distribución

gaussiana (y por lo tanto su anchura): a mayor *spread*, más tardará en tender a 0 la salida de la gaussiana, tal y como puede verse en la figura 54.

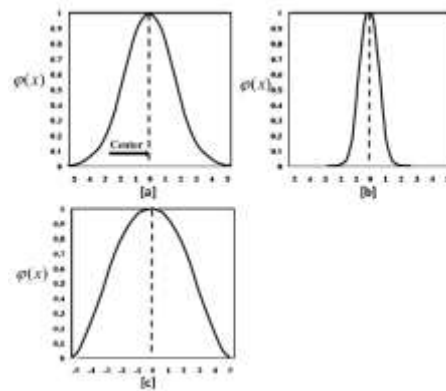


Figura 54. Salida de la función gaussiana para distintos valores de *spread*. (a) *spread* medio (b) *spread* bajo (c) *spread* alto. (El-Shafie et al., 2012)

En otras palabras, esta variable controlará la sensibilidad de la red a los puntos cerca de un centro; a mayor *spread*, menor sensibilidad a las entradas (El-Shafie et al., 2012). En problemas de clasificación esto equivale a que al disminuir esta variable, mayor será la probabilidad de que una entrada determinada forme parte del grupo asignado al centro cuya distancia a dicha entrada se ha calculado.

La segunda función, *newrb*, además de realizar el proceso de optimización de los pesos descrito para *newrbe*, prueba arquitecturas de menor a mayor tamaño, añadiendo una neurona en cada iteración. A pesar de que, en una primera instancia, pueda pensarse que esto permite una mayor personalización, a la práctica el error mínimo se encuentre precisamente en el número de neuronas asignado por la primera función, *newrb*, tal y como se observa en la figura 55.



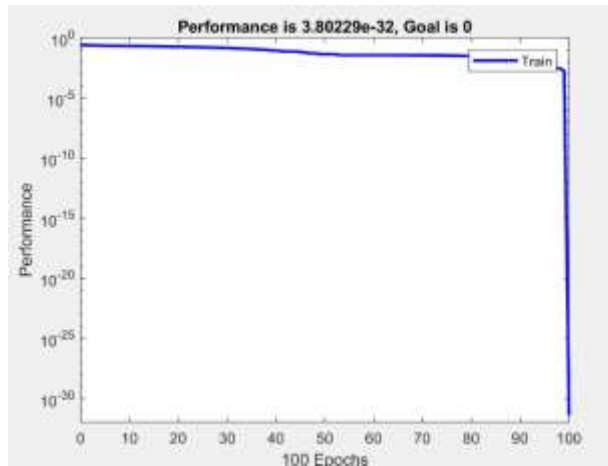


Figura 55. Evolución del error de entrenamiento en función del número de neuronas.

Considerando el aumento del tiempo de ejecución del programa al probar arquitecturas neurona a neurona, y habiéndose comprobado que el error mínimo se alcanza al asignar tantas neuronas como imágenes del subconjunto que se usa para entrenar, tal y como hace la primera función, *newrbe*, la **implementación final se realiza con esta primera función.**

Como algoritmo de entrenamiento, aunque también pueda usarse *backpropagation* como hace el MLP, en Matlab se usa el OLS, ya mencionado en el subapartado de teoría de las RBFs. Este algoritmo añade los centros uno a uno, y durante el entrenamiento se actualizan tanto los pesos de esta capa de salida como el número y la posición de los centros en la capa oculta, aunque como ya se ha establecido, el número de centros óptimo es el mismo que el de imágenes de entrenamiento.

Durante la ejecución del programa, se prueban diferentes valores de *spread* hasta encontrar el valor correspondiente al resultado óptimo, encontrado mediante validación cruzada, al igual que se ha realizado con el MLP.

### 5.2.3.2 Exactitudes obtenidas

En la tabla 11 aparecen las exactitudes obtenidas al aplicar RBF sobre la primera modalidad.

<p>Tabla 11. Aplicación de RBF: resultados óptimos. Media de ambos ojos, imagen completa</p>
--

<i>(Modalidad 1).</i>					
<b>Estructura retinal analizada</b>	<b>Arquitectura (spread)</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL	638	EM (real)	54	2	97,0874
		Sanos (real)	1	46	
Retina completa	718	EM (real)	55	1	91,2621
		Sanos (real)	8	39	
GCL+	574	EM (real)	51	5	78,6408
		Sanos (real)	17	30	
GCL++	720	EM (real)	56	0	97,0874
		Sanos (real)	3	44	
Coroides	574	EM (real)	45	11	71,8447
		Sanos (real)	18	29	

Para la primera modalidad, se alcanzan de nuevo exactitudes elevadas (aproximadamente 97% para RNFL y GCL++). Lo que destaca de estos primeros resultados es que se obtienen mediante anchuras de gaussianas muy elevadas: para la retina completa, cuyos espesores tienen una media de 250 micras, el resultado óptimo se alcanza mediante una anchura de 718; para la GCL++, cuya media de espesores es de aproximadamente 97, se alcanza con un spread de 720.

A continuación, las exactitudes al aplicar el mismo método sobre la modalidad de ventanas cuadradas.

*Tabla 12. Aplicación de RBF: resultados óptimos. Media de ambos ojos, ventanas cuadradas (Modalidad 2).*

<b>Estructura retinal analizada</b>	<b>Arquitectura (spread)</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL	242	EM (real)	46	10	78,6408
		Sanos (real)	12	35	
Retina completa	252	EM (real)	54	2	98,0583
		Sanos (real)	0	47	
GCL+	504	EM (real)	33	23	68,9320
		Sanos (real)	9	38	
GCL++	204	EM (real)	52	4	92,2330
		Sanos (real)	4	43	
Coroides	202	EM (real)	30	26	53,3981
		Sanos (real)	22	25	

En este caso, las anchuras de las gaussianas son notablemente inferiores, aunque también se obtienen peores resultados en varias de las capas, pasando de 97% a 78% para la RNFL y de 97% a 92% para la GCL++.

Por último, se aplica el RBF sobre las regiones de Cohen. Los resultados aparecen en la tabla 13.

*Tabla 13. Aplicación de RBF: resultados óptimos. Media de ambos ojos, región de Cohen (Modalidad 3).*

<b>Estructura retinal analizada</b>	<b>Arquitectura (spread)</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
Retina completa	180	EM (real)	56	0	100,0000
		Sanos (real)	0	47	
GCL+	186	EM (real)	47	9	72,8155
		Sanos (real)	19	28	
GCL++	180	EM (real)	55	1	100,0000
		Sanos (real)	1	46	

Al igual que con las MLP, se obtiene una exactitud del 100% tanto para la retina completa como para la GCL++.

### 5.2.3.3 Resumen de resultados: RBF

Las RBF son redes neuronales sencillas, al igual que los MLPs. De una sola capa oculta, permiten probar un elevado número de arquitecturas, en este caso modificando la anchura de la gaussiana hasta encontrar la óptima en cada caso. A pesar de ello, puesto que se parte siempre de tantas neuronas como número de imágenes, el tiempo es considerablemente mayor al de algoritmos simples como el KNN. La exactitud máxima se encuentra al analizar la modalidad de Cohen (100% en dos de los tres casos), seguidas por ventanas cuadradas de imágenes de retina completa (98%).

## 5.2.4 Support Vector Machine

### 5.2.4.1 Implementación

El SVM se ha implementado mediante la función *ftcsvm* para ajustar este algoritmo a aplicaciones a problemas de clasificación binaria, como la clasificación tratada en la tesis. Al igual que en el caso de KNN, el SVM recibe como entrada un vector correspondiente a los 2700 espesores de cada una de las imágenes de entrenamiento (102 en total), así como

las etiquetas en formato binario. Como salida, devuelve el modelo SVM entrenado, que a continuación se usa para predecir de qué clase es la imagen sobrante. El resultado de este proceso es una puntuación correspondiente a la probabilidad de que dicha imagen sea de un grupo u otro (control o paciente); la mayor corresponde a la clase elegida por el clasificador.

Sin embargo, en el caso de SVM, tal y como se ha visto en la teoría, para obtener los hiperparámetros adecuados se aplica una función *kernel* u otra. Matlab admite varios *kernel*, algunos de ellos correspondientes a las funciones de activación vistas hasta ahora: RBF, lineal y polinómica. Otras funciones, como la sigmoide, deben ser programadas aparte. La función polinómica, que hasta ahora no se ha aplicado, se define como:

$$K(x_1, x_2) = (x_1^T x_2 + 1)^\rho \quad (48)$$

, donde  $\rho$  es el orden del polinomio, a especificarse como parámetro. Durante la implementación, se usarán polinomios de orden 3.

La última variable a tener en cuenta es el llamado *box constraint*. Tal y como se ha visto en la teoría, es la variable dentro de la función objetivo que aumenta el peso de las clasificaciones erróneas:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^L \varepsilon_i \quad (49)$$

, donde C es el susodicho *box constraint*; aumentarla conlleva una separación más estricta de los datos. El resultado es un problema de optimización que siempre tiene solución, aunque los datos no sean linealmente separables.

Así, se usarán las cuatro funciones mencionadas y se variará el *box constraint* en cada iteración, guardándose las exactitudes obtenidas mediante cada una de ellas y resaltando finalmente el mejor resultado, junto a la función con la que se ha obtenido.

## 5.2.4.2 Exactitudes obtenidas

Las tablas 14, 15 y 16 recogen los resultados de exactitud obtenidos por el clasificador SVM al aplicarlo sobre cada una de las tres modalidades de análisis.

*Tabla 14. Aplicación de SVM: resultados óptimos. Media de ambos ojos, imagen completa (Modalidad 1).*

<b>Estructura retinal analizada</b>	<b>Función <i>kernel</i></b>	<b>Box constraint</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL	Lineal	2	EM (real)	56	0	99,0291
			Sanos (real)	1	46	
Retina completa	RBF	1	EM (real)	53	3	97,0874
			Sanos (real)	0	47	
GCL+	Polinómica (p=3)		EM (real)	47	9	82,5243
			Sanos (real)	9	38	
GCL++	RBF	1	EM (real)	56	0	100,0000
			Sanos (real)	0	47	
Coroides	Lineal	1	EM (real)	46	10	74,7573
			Sanos (real)	16	31	

Al aplicar SVM sobre la modalidad 1, se obtiene perfecta exactitud al analizar GCL++, con exactitudes también muy altas con la RNFL y la Retina completa (99% y 97% respectivamente). A destacar que no hay una función *kernel* concreta con la que se obtenga un resultado óptimo en todos los casos, aunque la RBF es con la que se alcanzan dos de las tres exactitudes más altas. Por otro lado, aumentar la rigidez de separación de los datos no conlleva una mejora de exactitud, excepto en el caso de la RNFL.

Los resultados correspondientes a la segunda modalidad de análisis aparecen en la tabla 15.

*Tabla 15. Aplicación de SVM: resultados óptimos. Media de ambos ojos, ventanas cuadradas (Modalidad 2).*

<b>Estructura retinal analizada</b>	<b>Función <i>kernel</i></b>	<b>Box constraint</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL	RBF	2	EM (real)	51	5	90,2913
			Sanos (real)	5	42	
Retina completa	RBF	6	EM (real)	56	0	100,0000
			Sanos (real)	0	47	
GCL+	Lineal	1	EM (real)	46	10	79,6117
			Sanos (real)	11	36	
GCL++	RBF	7	EM (real)	56	0	99,0291
			Sanos (real)	1	46	
Coroides	Lineal	1	EM (real)	43	13	63,1068
			Sanos (real)	25	22	

En el caso de ventanas cuadradas de longitud 3, de nuevo se alcanzan exactitudes elevadas. Con la función RBF se alcanza una exactitud perfecta y los tres mejores resultados (90%, 100% y 99%).

En tercer lugar, la tabla 16 reúne los resultados de la aplicación del SVM sobre las regiones de Cohen.

*Tabla 16. Aplicación de SVM: resultados óptimos. Media de ambos ojos, regiones de Cohen (Modalidad 3).*

<b>Estructura retinal analizada</b>	<b>Función <i>kernel</i></b>	<b>Box constraint</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
Retina completa	RBF	1	EM (real)	55	1	99,0291
			Sanos (real)	0	47	
GCL+	Lineal	2	EM (real)	55	1	98,0583
			Sanos (real)	1	46	
GCL++	Sigmoidal	1	EM (real)	55	1	98,0583
			Sanos (real)	1	46	

Aunque ninguno de los resultados al aplicar SVM sobre las regiones de Cohen llegue al 100% de exactitud, esta vez sí se obtienen muy buenos resultados en las tres estructuras estudiadas, no solamente en dos de ellas.

#### 5.2.4.3 Resumen de resultados: SVM

Los SVM, concebidos históricamente después del algoritmo KNN y de las redes MLP y RBF, obtienen resultados de exactitud muy elevados en las tres modalidades. En la primera, aplicándolo sobre la GCL++; en la segunda sobre la retina completa y en la tercera sobre las dos capas ganglionares.



## 5.3 Aplicación de ML a EM. Redes neuronales profundas

### 5.3.1 Redes Neuronales Convolucionales

#### 5.3.1.1 Implementación

Hay distintas consideraciones a realizar a la hora de decidir la arquitectura de una CNN, pero su elemento fundamental son las capas convolucionales. Teniendo esto en cuenta, en la figura 56 puede observarse una de las arquitecturas que se han usado, vista mediante la función *viewNetwork* de Matlab.

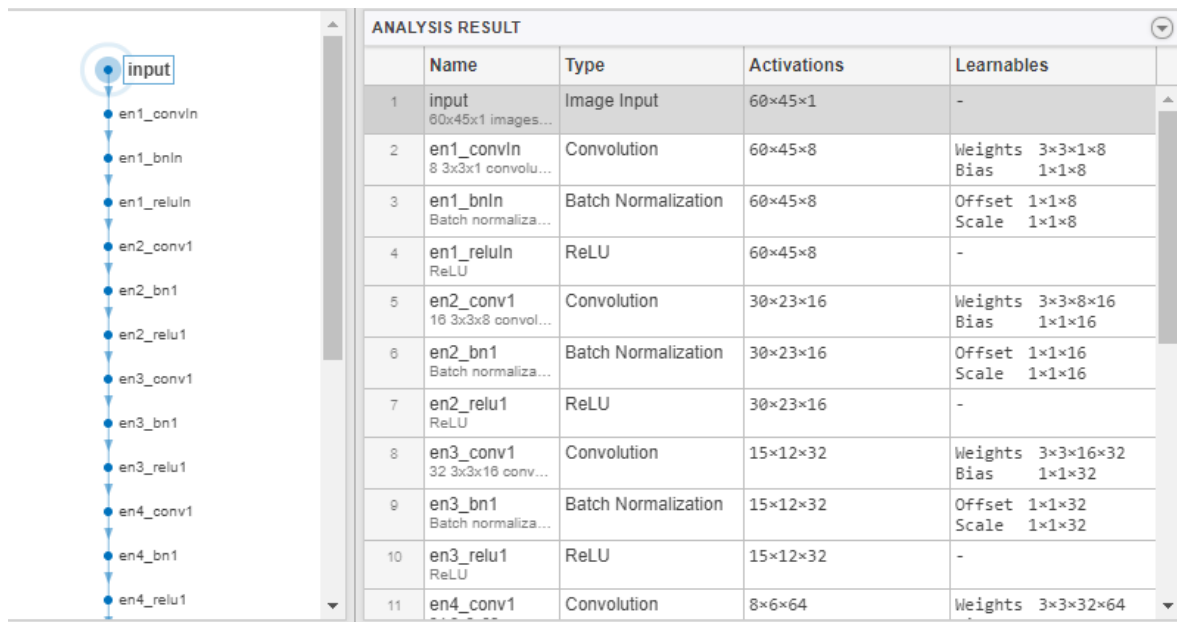


Figura 56. Una de las CNNs implementadas. Arquitectura de 4 capas convolucionales.

Así, debe establecerse cada capa durante la definición de la CNN:

- **Capas convolucionales:** lo primero a considerar es el **número de capas** de este tipo, que son las principales en las CNN. Aunque se han visto estudios con redes que llegan a acumular docenas de estas capas, debido al elevado tiempo de computación se establece el máximo en 10 capas convolucionales. Este umbral coincide con el de otros estudios que se han hecho para analizar EM ([Zhang et](#)

[al., 2018](#)). Las capas convolucionales irán siempre seguidas de *ReLU* y una capa de reducción por muestreo (*poolinglayer*) para compensar los cambios a las entradas que realiza la capa convolucional. El estudio citado también muestra un máximo de exactitud en 7 capas, disminuyendo posteriormente. Es importante tener en cuenta que, al incrementar el número de capas, hay que ajustar los parámetros de las capas *ReLU* y *pooling*. Este proceso de optimización se hace cada vez más complicado a medida que se van incrementando las capas. Lo segundo es el **número de filtros** por capa, equivalentes al número de nodos. Usando de nuevo una **arquitectura de embudo**, se eligen 64 para las capas más profundas, parámetro con el que se obtiene la exactitud óptima tanto en este caso como en otros estudios similares ([Barea R. et al., 2020](#)), y estableciendo 64 como máximo, las otras capas tendrán el número de filtros de la siguiente dividido por dos: 32, 16, 8, 4 y 2. También se prueban arquitecturas mayores a 6 capas, en cuyo caso algunas de las capas tendrán el mismo número de filtros en vez de la mitad. El **tamaño de cada filtro** se fija en 3, tomando como ejemplo ([Zhang et al., 2018](#))

Tal y como se ha visto en la teoría, tras cada capa convolucional se colocan capas que ajustan las dimensiones de los resultados de la convolución para que al final puedan transformarse en probabilidades. Cada capa convolucional junto a estas capas sucesivas (capa de *batch normalization*, capa *ReLU* y *Max Pooling Layer*) forman un bloque convolucional. Al final de la red, se establece una *fully connected layer* para obtener las probabilidades de que la imagen de entrada corresponda a un grupo u otro (paciente o control).

- **Capa Batch Normalization:** sin parámetros. Se coloca siempre antes de la capa *ReLU*.
- **Capa ReLu:** sin parámetros. Esta capa siempre se coloca tras la capa convolucional y la de *Batch Normalization*.

- **Max Pooling layer:** opcional, pero incluido en la tesis. El único parámetro a introducir es el *stride*, que indica la reducción de dimensiones a realizar en esta capa, tal y como se ha explicado en la teoría.
- **Fully Connected Layer:** la que finalmente da los pesos para cada categoría a la que se clasificarán los datos. Solo hay que establecer el **número de clases** a clasificar los datos, que en este caso son 2. Estas se convertirán en probabilidades mediante la capa *softmax*. Finalmente, *classoutput* le asignará a la imagen tratada la categoría correspondiente a la probabilidad máxima. En este caso solo hay dos categorías: “sano” y “enfermo”.

El tiempo de ejecución depende sobre todo del tamaño de la red, y es de entre una hora para una red con dos capas convolucionales a aproximadamente ocho horas para cuatro capas, por lo que aumenta exponencialmente.

Las CNNs se aplican sobre imágenes completas y ventanas cuadradas, así como sobre las regiones de Cohen convertidas a imágenes. Tal y como se ha especificado, se usa un tamaño de filtro de 3 y un número de filtros de 64 para la capa más profunda y un número equivalente a la mitad por cada capa que se añade anteriormente (por lo tanto, 64 en la más profunda, 32 en la siguiente, 16... hasta llegar a 2). El resto de su arquitectura se ha especificado ya en su respectivo apartado.

Para el entrenamiento, se usan los mismos parámetros que en el MLP: entrenamiento con *stochastic gradient descent with momentum*, ya vista en su apartado correspondiente. Se usa un momento de 0,9, un *learning rate (alpha)* de 0,001, un *batch size* de 102 (el número total de imágenes de entrenamiento) y un máximo de *epochs* de 1000.

### 5.3.1.2 Exactitudes obtenidas

A continuación, los resultados de implementar las CNNs sobre las modalidades de estudio. En la tabla 17 aparecen los resultados con las imágenes completas.

*Tabla 17. Aplicación de CNN: resultados óptimos. Media de ambos ojos, imagen completa (Modalidad 1).*

<b>Estructura retinal analizada</b>	<b>Arquitectura: N.º de capas convolucionales</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL	6	EM (real)	54	2	98,0583
		Sanos (real)	0	47	
Retina completa	8	EM (real)	55	1	99,0291
		Sanos (real)	0	47	
GCL+	8	EM (real)	45	11	80,5825
		Sanos (real)	9	38	
GCL++	5	EM (real)	56	0	100,0000
		Sanos (real)	0	47	
Coroides	4	EM (real)	40	16	70,8738
		Sanos (real)	14	33	

En la modalidad de imágenes completas, las CNNs alcanzan algunas de las exactitudes más altas vistas hasta ahora, siendo la RNFL, Retina y GCL++ las estructuras con los mejores resultados, obtenidos con 6, 8 y 5 bloques convolucionales respectivamente.

La tabla 18 reúne los resultados de la aplicación de esta red sobre las ventanas cuadradas de dimensiones 3x3:

*Tabla 18. Aplicación de CNN: resultados óptimos. Media de ambos ojos, ventanas cuadradas (Modalidad 2).*

<b>Estructura retinal analizada</b>	<b>Arquitectura: N.º de capas convolucionales</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL	2	EM (real)	51	5	91,2621
		Sanos (real)	4	43	
Retina completa	4	EM (real)	56	0	100,0000
		Sanos (real)	0	47	
GCL+	2	EM (real)	43	13	81,5534
		Sanos (real)	6	31	
GCL++	6	EM (real)	55	1	99,0291
		Sanos (real)	0	47	
Coroides	2	EM (real)	38	18	65,0485
		Sanos (real)	18	29	

Para esta modalidad se alcanzan exactitudes cercanas o iguales a 100 en dos, no tres de las estructuras. De nuevo se observa el patrón de RNFL, Retina y GCL++ como las mejores estructuras.

En tercer lugar, se aplica el clasificador CNN sobre las regiones de Cohen. Reiterar que el número de entradas es de 2700, puesto que se han fijado a 0 los valores de espesor ajenos a la región de Cohen definida en el capítulo anterior.

*Tabla 19. Aplicación de CNN: resultados óptimos. Media de ambos ojos, regiones de Cohen (Modalidad 3).*

<b>Estructura retinal analizada</b>	<b>Arquitectura: N.º de capas convolucionales</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
Retina	2	EM (real)	55	1	100,0000

completa		Sanos (real)	1	46	
GCL+	2	EM (real)	44	12	81,5534
		Sanos (real)	7	40	
GCL++	1	EM (real)	55	1	98,0583
		Sanos (real)	1	46	

Al aplicar la primera red profunda sobre las regiones de Cohen, se obtiene una exactitud del 100%, no dos como se han obtenido con MLP y RBF.

### 5.3.1.3 Resumen de resultados: CNN

Las redes neuronales convolucionales, el primer tipo de DNs que se aplican, obtienen exactitudes muy elevadas. Analizando ventanas cuadradas no solo se requiere un menor número de bloques convolucionales, sino que también se obtienen resultados ligeramente mayores, llegando al 100% de exactitud en dos de las estructuras analizadas.

## 5.3.2 Redes Neuronales Residuales

### 5.3.2.1 Implementación

Debido a las similitudes con la CNN, para la implementación de las ResNets se ha tomado como base la función ya creada para construir CNNs, y se crea una variante que añade las conexiones residuales vistas en la teoría. Así, por un lado, se crea la rama principal, con la estructura secuencial de capa convolucional, *batchNormalization*, *ReLU* y *maxpool*, formada por una cadena de bloques convolucionales y al final las capas de clasificación. Por otro lado, una rama secundaria con las conexiones residuales. Hay que tener en cuenta (1) de ajustar las conexiones de tal forma que, cuando haya un cambio de dimensión en la rama principal, también tiene que haberlo en la rama secundaria; y (2) que después de cada conexión residual, en Matlab hay que añadir una *additionLayer*. En la figura 57 puede verse el modelo de la red implementada.



Figura 57. Ejemplo de ResNet implementada. Izquierda: totalidad de una ResNet básica (3 bloques, 1 conexión residual). Derecha: parte de una ResNet de mayor tamaño (5 bloques, 3 conexiones residuales).

Tal y como puede verse, se colocan las conexiones residuales entre dos convoluciones, comenzando desde justo antes de la capa convolucional a la que se desea añadir dicho enlace y acabando después de otra convolución (o más concretamente, después de su normalización en lote, si hay). Para ello la *addition layer*, tal y como dice la teoría, realiza la función de suma entre la identidad en el bloque residual y el producto de la convolución en el bloque principal. La red consta del bloque de entrada, los bloques intermedios (cuyo número equivale al de las conexiones residuales de las que consta la red) y el bloque de clasificación, al final.

A destacar que, cuando cambian las dimensiones, es necesaria hacer una convolución también en la rama residual para igualar las dimensiones entre cada uno de los extremos de dichas conexiones residuales. Esto hay que hacerlo tanto cuando varían las dimensiones de la imagen como cuando el número de filtros entre ambos lados de la conexión es diferente.

Las ResNets se aplican sobre las tres modalidades. Se usan los mismos parámetros de entrenamiento que los usados en las CNNs.

## 5.3.2.2 Exactitudes obtenidas

Al aplicar las ResNets sobre la modalidad de imágenes completas, se obtienen los resultados recogidos en la tabla 20.

*Tabla 20. Aplicación de ResNet: resultados óptimos. Media de ambos ojos, imagen completa (Modalidad 1).*

<b>Estructura retinal analizada</b>	<b>Arquitectura: N.º de conexiones residuales</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL	1	EM (real)	55	1	97,0874
		Sanos (real)	2	45	
Retina completa	2	EM (real)	56	0	98,0583
		Sanos (real)	2	45	
GCL+	3	EM (real)	47	9	81,5534
		Sanos (real)	10	37	
GCL++	2	EM (real)	55	1	98,0583
		Sanos (real)	1	46	
Coroides	5	EM (real)	44	12	67,9612
		Sanos (real)	21	26	

Mediante las ResNets aplicadas a la modalidad 1 se obtienen resultados semejantes a los correspondientes a la CNN. La mayor exactitud, alcanzada con la retina completa y en la GCL++ (98%), es inferior a la obtenida en la CNN (99%), pero se obtienen valores semejantes a lo largo de tres de las cinco estructuras. Los mejores resultados se obtienen para arquitecturas más pequeñas (1 o 2 bloques residuales frente a 5 del peor resultado).

En la tabla 21 se observan los resultados correspondientes a la modalidad de ventanas cuadradas:



*Tabla 21. Aplicación de ResNet: resultados óptimos. Media de ambos ojos, ventanas cuadradas (Modalidad 2).*

<b>Estructura retinal analizada</b>	<b>Arquitectura: N.º de conexiones residuales</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL	1	EM (real)	50	6	89,3204
		Sanos (real)	5	42	
Retina completa	1	EM (real)	56	0	99,0291
		Sanos (real)	1	46	
GCL+	1	EM (real)	48	8	80,5825
		Sanos (real)	12	35	
GCL++	1	EM (real)	55	1	99,0291
		Sanos (real)	0	47	
Coroides	1	EM (real)	40	16	66,0194
		Sanos (real)	19	28	

Los resultados mejoran al aplicar la ResNet sobre las ventanas cuadradas, alcanzándose un 99% de exactitud en la Retina completa y la GCL++ y 89% en la RNFL. De nuevo se observa que las exactitudes óptimas se alcanzan con pocos bloques estructurales, en este caso uno solo para todos los casos.

Por último, los resultados al aplicar la red sobre las regiones de Cohen. De nuevo, se tienen 2700 entradas a la red.

*Tabla 22. Aplicación de ResNet: resultados óptimos. Media de ambos ojos, regiones de Cohen (Modalidad 3).*

<b>Estructura retinal analizada</b>	<b>Arquitectura: N.º de conexiones residuales</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
Retina completa	1	EM (real)	56	0	100,0000
		Sanos (real)	0	47	
GCL+	1	EM (real)	46	10	78,6408
		Sanos (real)	12	35	
GCL++	1	EM (real)	55	1	99,0291
		Sanos (real)	0	47	

Al estudiar las regiones de Cohen, de nuevo se obtienen dos exactitudes igual o muy cercanas a 100%

### 5.3.2.3 Resumen de resultados: ResNet

Las ResNets se diferencian de las CNN en que añaden una conexión residual entre pares de capas convolucionales. Con este método también se obtienen exactitudes muy elevadas, aunque tal y como se ha mencionado ya en su apartado y como dice la bibliografía, al tener un número reducido de imágenes y sobre todo de bloques convolucionales, no se observan mejoras significativas respecto a los otros tipos de redes aplicadas.

## 5.3.3 V-Nets

### 5.3.3.1 Implementación

El registro que trata la tesis contiene información volumétrica a lo largo de 5 estructuras retinales. Sin embargo, tanto debido a los resultados de AUC del capítulo anterior, como de varios artículos ya citados ([Cavaliere C. et al, 2019](#)) ([Barea R. et al, 2020](#)), debido a su escasa capacidad discriminante la coroides se considera de poco interés diagnóstico. Por lo tanto, y teniendo en cuenta la anchura y longitud de cada imagen 2D, la primera consideración será considerar volúmenes de dimensiones 45x60x4. Una segunda

posibilidad sería considerar una subselección de las tres estructuras, tomando solo dos o tres de las cuatro estructuras restantes. Sin embargo, los resultados de otras redes, recogidos en el capítulo siguiente, no señalan a una de ellas como estructura de poco interés diagnóstico. Por lo tanto, la implementación se hará con la RNFL, Retina, GCL+ y GCL++.

Tomando como base la misma función implementada en CNNs y ResNets para crear redes neuronales en bucle con tal de probar arquitecturas de varios tamaños, y adaptándola a volúmenes, se implementa la V-Net para la base de datos de la tesis, cuya estructura puede observarse en la figura 58.

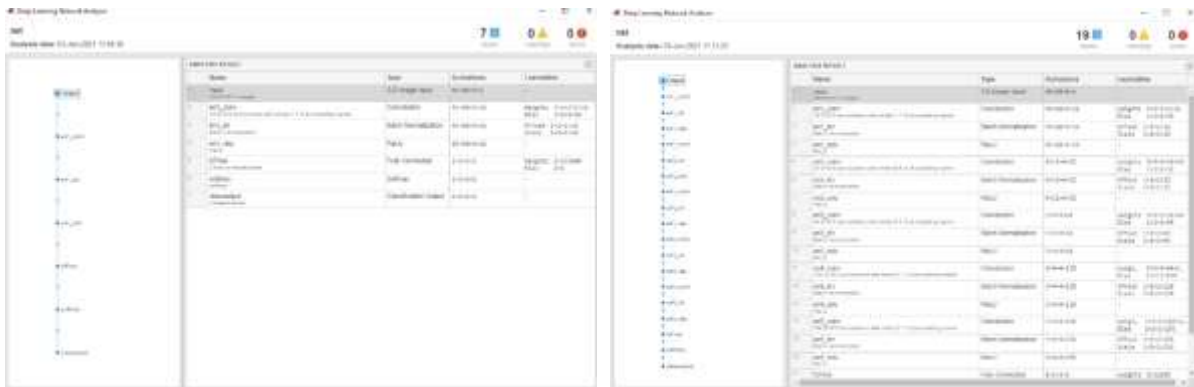


Figura 58. V-Nets implementadas, dependiendo del número de capas convolucionales.

Tal y como se ha hecho en las dos redes profundas anteriores, la primera capa convolucional no reduce las dimensiones de la entrada, pero las dos siguientes sí, aplicando *strides* tridimensionales de 2 para cada nueva capa convolucional, reduciendo así las dimensiones de 45x60x4 a 23x30x2 (redondeo hacia arriba), 12x15x1, 6x8x1 y 3x4x1. Por último, el módulo final clasifica los datos según su clase más probable, tal y como se ha visto en las dos DNs anteriores.

Para entrenar la V-Net, se usan los mismos parámetros que los usados en la CNN y la ResNet. Se aplica, como todas las anteriores, sobre las tres modalidades, y se prueba a aplicarla tanto considerando cuatro estructuras (RNFL, Retina completa, GCL+ y GCL++) como tomando una subselección de estas cuatro, probándose Retina completa, GCL+ y GCL++ (correspondiente a las estructuras de mayor AUC y con las que se obtienen

regiones de Cohen) así como RNFL, Retina completa y GCL++ (correspondientes con las estructuras con las que se obtienen mejores exactitudes al aplicar ML).

### 5.3.3.2 Exactitudes obtenidas

La implementación de la V-Net constituye la última aplicación de un clasificador automático sobre la base de datos de la tesis. Los resultados aparecen en las tablas 23, 24 y 25. A continuación, los correspondientes a la modalidad de imágenes completas:

*Tabla 23. Aplicación de V-Net: resultados óptimos. Media de ambos ojos, imagen completa (Modalidad 1).*

<b>Estructuras retinales analizadas</b>	<b>Arquitectura: N° de capas convolucionales</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL, Retina completa, GCL+, GCL++	4	EM (real)	54	2	98,0583
		Sanos (real)	0	47	
Retina completa, GCL+, GCL++	5	EM (real)	55	1	99,0291
		Sanos (real)	0	47	
RNFL, Retina completa, GCL++	4	EM (real)	55	1	99,0291
		Sanos (real)	0	47	

En la modalidad 1, las exactitudes se acercan a 100%. Al considerar una subselección de las 4 estructuras seleccionadas inicialmente, el programa acierta un diagnóstico más, pasando de 98 a 99% de exactitud de predicción. En el caso de considerar la RNFL (la estructura con los terceros resultados más altos de exactitud) en vez de la GCL+ el resultado es el mismo, aunque se alcanza con una red ligeramente más superficial, con la resultante disminución de tiempo de ejecución del programa.

La tabla 24 reúne los resultados correspondientes a la aplicación de la V-Net sobre las ventanas cuadradas:

*Tabla 24. Aplicación de V-Net: resultados óptimos. Media de ambos ojos, ventanas cuadradas (Modalidad 2).*

<b>Estructuras retinales analizadas</b>	<b>Arquitectura: N.º de capas convolucionales</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL, Retina completa, GCL+, GCL++	1	EM (real)	54	2	98,0583
		Sanos (real)	0	47	
Retina completa, GCL+, GCL++	1	EM (real)	55	1	99,0291
		Sanos (real)	0	47	
RNFL, Retina completa, GCL++	1	EM (real)	55	1	99,0291
		Sanos (real)	0	47	

Se obtienen los mismos resultados al usar ventanas cuadradas que al usar imágenes completas. Al hacer una subselección en la que se remueve la RNFL, la exactitud aumenta ligeramente; lo mismo ocurre al remover la GCL+.

Por último, se aplica V-Net sobre las regiones de Cohen, con resultados recogidos en la tabla 25.

*Tabla 25. Aplicación de V-Net resultados óptimos. Media de ambos ojos, regiones de Cohen (Modalidad 3).*

<b>Estructuras retinales analizadas</b>	<b>Arquitectura: N.º de capas convolucionales</b>	<b>Matriz de confusión</b>	EM (pred.)	Sanos (pred.)	<b>Exactitud (%)</b>
RNFL, Retina completa, GCL+, GCL++	3	EM (real)	54	2	97,0874
		Sanos (real)	1	46	
Retina completa, GCL+, GCL++	5	EM (real)	54	2	98,0583
		Sanos (real)	0	47	
RNFL, Retina completa, GCL++	4	EM (real)	56	0	100,0000
		Sanos (real)	47	0	

La exactitud del 100% se obtiene descartando la GCL+, correspondiente a aquella con la que se han obtenido los peores resultados con Cohen hasta ahora.

### 5.3.3.3 Resumen de resultados: V-Net

Al seleccionar un volumen en vez de una sola estructura retinal, con la V-Net es posible considerar información en tres dimensiones a la vez en lugar de considerar estructuras como entradas totalmente independientes las unas de las otras. Como pueden comprobarse en los resultados, esto conlleva una clara mejora pues se llega, en las tres modalidades, a exactitudes cercanas a 100.

## 5.4 Discusión de los resultados. Conclusiones

En este capítulo se han presentado los datos que muestran la capacidad diagnóstica de diferentes clasificadores automáticos (KNN, MLP, RBF, SVM, CNN, ResNet y V-Net) de acuerdo con las tres modalidades de procesamiento de los datos de espesores, que se presentan en cinco estructuras retinales diferentes (RNFL, Retina completa, GCL+, GCL++ y Coroides). Con la batería de pruebas realizadas, en esta última sección se presentan unas breves conclusiones sobre los mismos.

Para una mayor claridad a la hora de interpretar los resultados de exactitudes, a destacar que existen dos variables principales que diferencian una tabla de otra: **modalidad de imagen estudiada** (imágenes completas, ventanas cuadradas o región de Cohen) y **tipo de clasificador aplicado** (KNN, MLP, RBF, SVM, CNN, ResNets o V-Nets). A su vez, dentro de cada tabla, excepto en el caso de las V-Nets, se presenta un resultado para cada **estructura retinal analizada**. Dicho esto, pueden hacerse un número de observaciones iniciales:

**Observaciones generales:** Hay una mayor diferencia de exactitud entre estructura analizada que entre la modalidad de imagen y los diferentes clasificadores basados en ML aplicados, por lo que constituyen las dos variables de las que depende la exactitud de predicción. Puesto que estas son las dos variables correspondientes a la base de datos y no al aplicado en la tesis, se consideran de interés para establecer qué estructura o conjunto de estructura tienen mayor capacidad discriminante, pero no como consideraciones de cara a mejorar los programas de ML implementados.

**Estructura retinal de interés.** Tal y como se ha remarcado en el capítulo de determinación de la región más discriminante, hay una estructura de poco interés diagnóstico, hecho que coincide en todas las tablas: la coroides. En dicho capítulo también se ha establecido que de entre las otras cuatro estructuras, las capas ganglionares y la Retina completa son las tres de mayor interés. Los resultados de las redes reflejan lo mismo de forma parcial: aunque con RNFL, Retina y GCL++ sí se obtengan buenos resultados a lo largo de las implementaciones, no es el caso de la GCL+. Por ejemplo, aplicar las CNNs sobre registros de imágenes completas resulta en exactitudes de alrededor de 98, 99 y 100% para la RNFL, Retina y GCL++ respectivamente, teniendo un peor resultado en la GCL+ (aproximadamente 80%). Sin embargo, sí se observa una tendencia de tres estructuras con una exactitud considerablemente alta, observándose también en el caso de aplicar SVM sobre la región de Cohen (99% en Retina completa y 98% para GCL+ y GCL+).

**Ojos analizados:** aunque no se hayan mostrado los resultados para los otros dos ojos, la modalidad con la que se obtiene mejores resultados es la media de ambos ojos -

coincidiendo con lo observado al calcular la AUC en capítulos anteriores-; a continuación, el ojo derecho y, por último, el ojo izquierdo.

**Modalidad de imagen estudiada:** no existen grandes diferencias entre analizar imágenes completas o ventanas cuadradas. Hay casos, como al aplicar ResNet sobre la media de ambos ojos, en los que se obtiene mejores resultados en las ventanas cuadradas (89, 99 y 99% en RNFL, Retina y GCL++) frente a los obtenidos en imágenes completas (97, 98 y 98%). Al estudiar las regiones de Cohen ocurre lo mismo excepto en el caso de SVM, en el que se obtienen resultados muy semejantes para las tres estructuras (99, 98 y 98% respectivamente). A destacar que la zona correspondiente a la longitud óptima de ventana (modalidad 2) se trata de un área pequeña (3x3 espesores), que a primera vista podría sugerir que se trata de ruido. En ([Cavaliere C. et al., 2019](#)) el área mínima considerada es de 25 (5x5 espesores). Sin embargo, en el artículo se obtienen valores similares de AUC a los obtenidos en la tesis: 0,8781 para la GCL+ frente a 0,8785 para la GCL++; además, en la tesis se han obtenido mejores valores de exactitud, llegando al 100% en algunas estructuras retinales aplicando MLP, SVM, CNN, ResNet y V-Net frente a los resultados del estudio citado (94% de exactitud).

**Clasificador usado: ML tradicionales.** De los tres procesos de ML tradicionales implementados, los MLP llegan al 100% de exactitud en más ocasiones: cinco veces frente a una del KNN y dos para RBF y SVM. El mejor clasificador para estudiar las regiones de Cohen es el SVM (99% para Retina, 98% para GCL+ y 98% para GCL++) y el peor el KNN (89% para Retina, 88% para GCL+ y 83% para GCL++), aunque debido a la simpleza y al escaso tiempo que requiere el KNN, para las otras dos modalidades se recomienda su uso frente al de MLP, RBF o SVM.

**Clasificador usado: redes profundas.** No existen grandes diferencias entre las tres redes aplicadas, alcanzándose exactitudes altas en los tres casos. Las CNN llegan al 100% en dos ocasiones mientras que en ResNets y V-Nets en solo una, y a destacar que los resultados más altos se alcanzan con redes de uno o dos bloques estructurales, siendo esto congruente con la bibliografía (bases de datos de pocas imágenes no requieren redes de un elevado



número de capas). En las V-Nets se observa una vez más que los resultados de GCL+ son inferiores a las otras tres estructuras de interés, pues la exactitud aumenta en los casos en los que se excluye esta estructura, pasando de 97% a 98 y 100% en el caso de las regiones de Cohen y de 98 a 99% en el caso de imágenes completas y ventanas cuadradas.

A partir de todas estas observaciones, puede concluirse que:

- A pesar de que en un primer momento las dimensiones tan reducidas podrían sugerir facilidad de que ruido se introduzca junto a información de interés, **la mejor modalidad de estudio es la de ventanas cuadradas de dimensiones 3x3.**
- Respecto al proceso de aprendizaje automático usado, aunque hay varios con los que se llega a la máxima exactitud en una de las cinco estructuras, se concluye que **el mejor clasificador es el Perceptrón Multicapa**, pues es el que alcanza el 100% de exactitud más frecuentemente (cinco veces). A pesar de que, tal y como se ha explicado en la teoría, las redes profundas sean más modernas, de uso más extendido y que su uso como clasificadores esté considerado superior al de procesos tradicionales como el MLP, hay que destacar que suelen emplearse en problemas con bases de datos muy grandes, como el reconocimiento de objetos. En cambio, el registro de la tesis no contiene miles de imágenes sino alrededor de 100. El peor clasificador puede considerarse la otra red tradicional usada, el RBF, aunque puesto que su mejor resultado es aproximadamente 98%, no es apropiado descartarlo en futuras aplicaciones.





## **6 Discusión general. Trabajos futuros.**

El principal objetivo de esta tesis doctoral ha sido investigar el potencial de los espesores retinales, obtenidos por OCT, como biomarcador de interés para poder hacer un diagnóstico temprano de esclerosis múltiple. Para ello, se ha hecho un análisis de los registros, tanto a nivel de imagen completa como realizando dos subselecciones de hipotético interés diagnóstico debido a su alta capacidad discriminante, y se ha hecho uso de clasificadores basados en aprendizaje automático para poder predecir si es posible acertar el diagnóstico de un paciente usando estos espesores retinales como entrenamiento de estas redes, con la idea de que disponer de un sistema automático de ayuda al diagnóstico facilita la toma de decisiones por parte del personal médico.

Puesto que en el momento del registro los pacientes sufrían un diagnóstico precoz (menos de 9 meses desde los primeros síntomas) y por lo tanto con desmielinización menos marcada que si se hubiera tratado de un diagnóstico más tardío, cabía la posibilidad de que no fuese suficiente para que los espesores retinales fuesen de gran utilidad en cuanto a exactitud de predicción frente a nuevos casos.

Después de todo, en muchos casos el diagnóstico de esta enfermedad es difícil debido al solapamiento con los síntomas de otras enfermedades ([Tăuțan, A.M. et al., 2021](#)). El número de diagnósticos erróneos es relativamente elevado; en ([Kaisev M. et al., 2019](#)) se reporta un 18% de casos mal diagnosticados; en relación con el retraso en el diagnóstico, medido a partir de los primeros síntomas, en caso extremos puede llegar hasta los 30 meses ([Patti F. et al., 2021](#)), con el consiguiente impacto negativo en la trayectoria de la enfermedad, al retrasar el inicio del tratamiento del paciente.

Sin embargo, mediante el método aplicado en esta tesis, los resultados de exactitud tras aplicar varios procesos de aprendizaje automático han sido cercanos al 100%, por lo que no solo ha sido adecuado utilizar datos tomados menos de 9 meses después de los primeros síntomas, sino que podrían usarse como datos registros incluso más tempranos, si se dispusiese de ellos.

Respecto al tipo de clasificador basado en aprendizaje automático utilizado, debido a la limitación en lo que respecta al número de imágenes del registro, al aplicar redes profundas no se obtienen resultados particularmente altos comparados con las exactitudes obtenidas mediante clasificadores tradicionales, por lo que, dada su simpleza, se recomienda el uso de estos clasificadores, particularmente el perceptrón multicapa.

A su vez, aunque se esperaban obtener resultados de exactitud muy altos con ambas estructuras de capas ganglionares (GCL+ y GCL++), la GCL+ no ha resultado de particular interés diagnóstico, siendo esta la estructura con la que se obtienen los peores resultados después de la coroides (cuyos resultados sí han sido los esperados). Por lo tanto, en estudios futuros se propone centrarse en las otras tres estructuras: RNFL, Retina y GCL++.

En cuanto a la modalidad de estudio, se concluye que es conveniente hacer un análisis de ventanas cuadradas de bajas longitudes (3-5 píxeles) tanto debido al bajo tiempo requerido para entrenar los clasificadores como a las altas exactitudes obtenidas, aunque no existen grandes diferencias en lo que respecta a la modalidad de análisis.

Así pues, tras el trabajo realizado en la tesis, se reafirma la hipótesis de que los espesores retinales constituyen buenos biomarcadores para el diagnóstico de EM, aunque debido a la escasez de estudios en este campo y a la vez a sus posibilidades, también se establecen algunas líneas futuras:

- Aunque la base de datos haya sido mayor a las probadas en otros trabajos ([Cavaliere C., 2019](#)), ([Barea R., 2020](#)), se propone probar los mismos métodos sobre una base de datos con un mayor número de sujetos, tanto controles como pacientes. De esta forma se comprobaría la validez de los resultados y también es probable que las exactitudes obtenidas mediante redes profundas mejorasen, pues en teoría funcionan mejor con bases de datos grandes. También se propone usar bases de datos fusionadas de varios centros (asegurando que el modelo del equipo, las condiciones de medida y el volumen estudiado sean equivalentes), y probar a usar el resto de información clínica disponible: edad, sexo, presencia de síntomas relacionados, etc,

- Habiéndose alcanzado los resultados de exactitud óptima, probar métodos como los propuestos en ([García R. et al., 2019](#)), estudio en el que se usan los llamados *Activation Occurrence Maps* para reducir las dimensiones de las redes y que así el tiempo de ejecución de los programas de ML sea inferior. Después de todo, y pese a no ser una cuestión crítica comparada con un hipotético diagnóstico erróneo, mediante el método de la tesis, dependiendo del equipo los tiempos de entrenamiento pueden llegar a un día por modalidad de estudio aplicada.
- Recientemente se ha introducido otra modalidad de OCT, la tecnología Angio-OCT (Angiografía-OCT, comúnmente conocida como *OCT-A*), y se han realizado distintos estudios en los que se ha encontrado una reducción en la densidad de los vasos retinales y una correlación entre densidad de vasos y EDSS, sugiriendo que la OCT-A podría ser un buen marcador del estado de la EM y de las discapacidades asociadas ([Lo C. et al., 2021](#)). Sin embargo, actualmente no se dispone de suficientes pruebas científicas, en parte debido a que estos equipos se encuentran presentes en un número muy limitado de hospitales e instituciones. Así, otra posible línea futura sería fusionar los resultados obtenidos con medidas de espesores provenientes de la nueva modalidad de OCT, la OCT-A, así como con medidas provenientes de otras técnicas de diagnóstico comúnmente empleadas en el campo de la EM, como el electroretinograma multifocal, los potenciales evocados o niveles de líquido cefalorraquídeo.
- Aplicar los métodos de aprendizaje automático desarrollados y probados en esta tesis sobre datos de espesores de pacientes de otras enfermedades neurodegenerativas. Por ejemplo, al caso de la esquizofrenia, en la que también se han observado anormalidades en datos de OCT, así como en ERG, tal y como se observa en estudios como ([Silverstein S.M. et al., 2019](#)).



## **7 Bibliografía**

- AJMC (2018). Current Strategies in the Treatment of Multiple Sclerosis. Consultado por última vez en octubre de 2021. <https://www.ajmc.com/view/current-strategies-in-the-treatment-of-multiple-sclerosis>
- Alpaydin E. (2020). “Introduction to Machine Learning (Fourth ed.)”. *MIT*. pp. xix, 1–3, 13–18. ISBN 978-0262043793.
- Alonso R., Gonzalez-Moron D., and Garcea O., 2018. Optical coherence tomography as a biomarker of neurodegeneration in multiple sclerosis: A review. *Mult. Scler. Relat. Disord.*, vol. 22, pp. 77–82, doi: 10.1016/j.msard.2018.03.007.
- Alsalamah, M., Amin, S., & Halloran, J. (2015). Diagnosis of heart disease by using a radial basis function network classification technique on patients’ medical records. *Conference Proceedings - 2014 IEEE MTT-S International Microwave Workshop Series on: RF and Wireless Technologies for Biomedical and Healthcare Applications, IMWS-Bio 2014*. <https://doi.org/10.1109/IMWS-BIO.2014.7032401>
- Atakulreka A., Sutivong D. (2007) Avoiding Local Minima in Feedforward Neural Networks by Simultaneous Learning. In: Orgun M.A., Thornton J. (eds) *AI 2007: Advances in Artificial Intelligence. AI 2007. Lecture Notes in Computer Science*, vol 4830. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-76928-6\\_12](https://doi.org/10.1007/978-3-540-76928-6_12)
- Attwood D. (2016). Coherence, Brightness and Time Structure: from Synchrotrons to FELs. *University of California, Berkeley*. <http://indico.ictp.it/event/7594/session/21/contribution/105/material/slides/0.pdf>
- Aumann S., Donner S., Fischer J., Müller F. (2019), “Optical Coherence Tomography (OCT): Principle and Technical Realization”. *Springer*. [https://link.springer.com/chapter/10.1007/978-3-030-16638-0\\_3](https://link.springer.com/chapter/10.1007/978-3-030-16638-0_3)



- Bahrún U., Wijaya C. (2018), IMMUNOLOGY OF MULTIPLE SCLEROSIS. *Indonesian Journal of clinical pathology and medical laboratory*, 24 (2), 191-195.
- Becker L.A. (2000), *Effect Size (ES)*. University of Colorado. <https://www.uv.es/~friasnav/EffectSizeBecker.pdf>
- Bevilacqua, V., Brunetti, A., Guerriero, A., Trotta, G. F., Telegrafo, M., & Moschetta, M. (2019). A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. *Cognitive Systems Research*, 53. <https://doi.org/10.1016/j.cogsys.2018.04.011>
- Bengio Y., (2009). Learning Deep Architectures for AI. *Foundations and Trends® in Machine Learning: Vol. 2: No. 1, pp 1*. <http://dx.doi.org/10.1561/22000000006>
- Bhattacharyya J., 2020. Activation functions in Neural Networks: An Overview. <https://analyticsindiamag.com/activation-functions-in-neural-network/>
- Bioestelles (2013). Anatomía y fisiología del Sistema nervioso. Consultado por última vez en junio de 2021. <http://es.slideshare.net/Bioestelles/tema-10-nervioso>
- Birkeldh, U., Manouchehrinia, A., Hietala, M.A., Hillert, J., Olsson, T., Piehl, F., Kockum, I.S., Brundin, L., Zahavi, O., Wahlberg-Ramsay, M., et al. (2017) The Temporal Retinal Nerve Fiber Layer Thickness Is the Most Important Optical Coherence Tomography Estimate in Multiple Sclerosis. *Front. Neurol.* 8, 675.
- Britze, J., Pihl-Jensen, G., & Frederiksen, J. L. (2017). Retinal ganglion cell analysis in multiple sclerosis and optic neuritis: a systematic review and meta-analysis. *Journal of Neurology*, 264(9), 1837–1853. <https://doi.org/10.1007/s00415-017-8531-y>
- Bsteh, G., Bitschnau, C., Hegen, H., Auer, M., Di Pauli, F., Rommer, P., ... Berger, T. (2020). Multiple sclerosis and COVID-19: how many are at risk? *European Journal of Neurology*. <https://doi.org/10.1111/ene.14555>

- Case R., 2019. “Why machine learning”? *Towards Data Science Blog*. Consultado por última vez en septiembre de 2021. <https://towardsdatascience.com/why-machine-learning-303e6bdaa29d>
- Castillo Álvarez, F., López Pérez, M. Á., & Marzo Sola, M. E. (2020). Risk of SARS-CoV-2 infection and clinical outcomes in multiple sclerosis patients in La Rioja (Spain). *Medicina Clínica (English Edition)*, 155(8), 362–363. <https://doi.org/10.1016/j.medcle.2020.06.017>
- Chan, H. H., Zhao, Y., Tun, T. A., & Tong, L. (2015). Repeatability of tear meniscus evaluation using spectral-domain Cirrus® HD-OCT and time-domain Visante® OCT. *Contact Lens and Anterior Eye*, 38(5), 368–372. <https://doi.org/10.1016/j.clae.2015.04.002>
- Chellapilla K., Puri S., Simard P. (2006). High Performance Convolutional Neural Networks for Document Processing. *Tenth International Workshop on Frontiers in Handwriting Recognition*, Université de Rennes 1, La Baule (France). Université de Rennes 1, Oct 2006, La Baule (France). inria-00112631
- Chen, S., Cowan, C. F. N., & Grant, P. M. (1991). Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks. *IEEE Transactions on Neural Networks*, 2(2), 302–309. <https://doi.org/10.1109/72.80341>
- Chew, E.Y. (1996). Association of Elevated Serum Lipid Levels With Retinal Hard Exudate in Diabetic Retinopathy. *Arch. Ophthalmol.* 114, 1079. doi:10.1001/archophth.1996.01100140281004
- Cohen J. (1988). Statistical power analysis for the behavioural sciences (2<sup>nd</sup> ed.). *Hillsdale, NJ: Lawrence Earlbaum Associates*.

- Covo P., (2015). Introducción a la historia de la esclerosis múltiple. *Acta Neurológica Colombiana*. <http://www.scielo.org.co/pdf/anco/v31n1/v31n1a17.pdf>
- Dalgas, U., & Stenager, E. (2012). Exercise and disease progression in multiple sclerosis: Can exercise slow down the progression of multiple sclerosis? *Therapeutic Advances in Neurological Disorders*, 5(2), 81–95. <https://doi.org/10.1177/1756285611430719>
- Dasgupta A., Nath A. (2016). Classification of Machine Learning Algorithms. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, vol 3. ISSN: 2349-2763
- De Brouwer E., Becker T., Moreau Y., Havrdova E.K., Trojano M., Eichau S., Ozakbas S., Onofrj M., Grammond P., Kuhle J., et al (2021). Longitudinal machine learning modeling of MS patient trajectories improves predictions of disability progression. *Comput. Methods Programs Biomed.* 2021, 208, 106180.
- Dhivya Ashok Kumar and Amar Agarwal (2020). *History, Principles and Instrumentation of Optical Coherence Tomography*. Ento Key. <https://entokey.com/1-history-principles-and-instrumentation-of-optical-coherence-tomography/>
- Dodge S., Karam L. (2016), “Understanding How Image Quality Affects Deep Networks”. Arizona State University.
- Drexler, W., Liu, M., Kumar, A., Kamali, T., Unterhuber, A., & Leitgeb, R. A. (2014). Optical coherence tomography today: speed, contrast, and multimodality. *Journal of Biomedical Optics*, 19(7), 071412. <https://doi.org/10.1117/1.jbo.19.7.071412>
- Duelo L., 2021. *Taller práctico sobre analítica de datos (semana de la digitalización)*. Escuela de Doctorado, UAH, 2021.
- El-Shafie, A., Noureldin, A., Taha, M., Hussain, A., & Mukhlisin, M. (2012). Dynamic versus static neural network model for rainfall forecasting at Klang River Basin,

- Malaysia. *Hydrology and Earth System Sciences*, 16(4), 1151–1169.  
<https://doi.org/10.5194/hess-16-1151-2012>
- Encyclopedia Britannica, Inc* (2006). Consultado por última vez en diciembre de 2021.  
<https://www.britannica.com/science/light/Reflection-and-refraction>
- Evgeniou, T., & Pontil, M. (2001). Support vector machines: Theory and applications. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2049 LNAI, 249–257.  
[https://doi.org/10.1007/3-540-44673-7\\_12](https://doi.org/10.1007/3-540-44673-7_12)
- Faghfouri, A. E., & Frish, M. B. (2011). Robust discrimination of human footsteps using seismic signals. *Unattended Ground, Sea, and Air Sensor Technologies and Applications XIII*, 8046, 80460D. <https://doi.org/10.1117/12.882726>
- Farley, B. J., Morozova, E., Dion, J., Wang, B., Harvey, B. D., Gianni, D., Wipke, B., Cadavid, D., Wittmann, M., & Hajos, M. (2019). Evoked potentials as a translatable biomarker to track functional remyelination. *Molecular and Cellular Neuroscience*, 99. <https://doi.org/10.1016/j.mcn.2019.103393>
- Federación Internacional de Esclerosis Múltiple (2020). Atlas de Esclerosis Múltiple. <https://www.atlasofms.org/map/spain/epidemiology/number-of-people-with-ms>
- Ferrús, A., Amador-Arjona, A., Barco, Á., Obradors, C., Arqué, G., Morgado Berna, I., ... Gallego, X. (2007). Viaje al universo neuronal. *Fundación Española para la ciencia y la tecnología, F.E.*
- Fisniku, L. K., Brex, P. A., Altmann, D. R., Miszkiel, K. A., Benton, C. E., Lanyon, R., Thompson, A. J., & Miller, D. H. (2008). Disability and T2 MRI lesions: A 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain*, 131(3), 808–817.  
<https://doi.org/10.1093/brain/awm329>

- Fletcher T. (2008). Support Vector Machines Explained. *University College of London*.
- Foote K.D., 2019: “A brief history of Machine Learning”. *Data Diversity*. Consultado por última vez en noviembre de 2021. <https://www.dataversity.net/a-brief-history-of-machine-learning/>
- Forooghian, F., Sproule, M., Westall, C., Gordon, L., Jirawuthiworavong, G., Shimazaki, K., & O’connor, P. (2006). Electroretinographic abnormalities in multiple sclerosis: Possible role for retinal autoantibodies. *Documenta Ophthalmologica*, 113(2), 123–132. <https://doi.org/10.1007/s10633-006-9022-0>
- Gao, W., & Wu, X. (2017). Differences between time domain and Fourier domain optical coherence tomography in imaging tissues. *Journal of Microscopy*, 268(2), 119–128. <https://doi.org/10.1111/jmi.12592>
- Garcia, R., Falcao, A. X., Telea, A. C., Da Silva, B. C., Torresen, J., & DÍhl Comba, J. L. (2019). A Methodology for Neural Network Architectural Tuning Using Activation Occurrence Maps. *Proceedings of the International Joint Conference on Neural Networks, 2019-July*. <https://doi.org/10.1109/IJCNN.2019.8852223>
- Gohil, K. (2015). Multiple sclerosis: Progress, but no cure. *P and T*, 40(9), 604–605.
- Goedhart, J. (2016). Calculation of a distribution free estimate of effect size and confidence intervals using VBA/Excel. *BioRxiv*, 073999. <https://doi.org/10.1101/073999>
- Hagan M., “Radial Basis Networks”, pp. 2-3. Consultado por última vez en enero de 2022. <http://hagan.okstate.edu/RadialBasis.pdf>
- Hamdan, R., Gonzalez, R. G., Ghostine, S., &Caussin, C. (2012). Optical coherence tomography: From physical principles to clinical applications. *Archives of Cardiovascular Diseases*, 105(10), 529–534. <https://doi.org/10.1016/j.acvd.2012.02.012>

- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-December, 770–778.  
<https://doi.org/10.1109/CVPR.2016.90>
- Hebb D., 1949. “The organization of behaviour. A neuropsychological theory”. McGill University. *John Wiley&Sons, Inc.*
- Howard J., 2018. “Lesson 4 - Collaborative Filtering, Embeddings, and more”. *Fast AI Course*. Consultado por última vez en noviembre de 2021.  
<https://course17.fast.ai/lessons/lesson4.html>
- Hu, H., Jiang, H., Gameiro, G.R., Hernandez, J., Delgado, S., Wang, J. (2019). Focal Thickness Reduction of the Ganglion Cell-Inner Plexiform Layer Best Discriminates Prior Optic Neuritis in Patients with Multiple Sclerosis. *Investig. Ophthalmology Vis. Sci.* 60, 4257.
- Huang, D., Swanson, E. A., Lin, C. P., Schuman, J. S., Stinson, W. G., Chang, W., ... Fujimoto, J. G. (1991). Optical coherence tomography. *Science*, 254(5035), 1178–1181. <https://doi.org/10.1126/science.1957169>
- Intelligency Blog*. “AI with Convolutional Neural Networks”. Consultado por última vez en julio de 2021. [http://intelligency.org/ai\\_cnn.php](http://intelligency.org/ai_cnn.php)
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd International Conference on Machine Learning, ICML 2015, 1, 448–456.
- Jacobson, M., & Hirose, G. (1978). Origin of the retina from both sides of the embryonic brain: A contribution to the problem of crossing at the optic chiasma. *Science*, 202(4368), 637–639. <https://doi.org/10.1126/science.705349>

- Kaisey, M., Solomon, A. J., Luu, M., Giesser, B. S., & Sicotte, N. L. (2019). Incidence of multiple sclerosis misdiagnosis in referrals to two academic centers. *Multiple Sclerosis and Related Disorders*, 30, 51–56. <https://doi.org/10.1016/j.msard.2019.01.048>
- Kale, N. (2016). Optic neuritis as an early sign of multiple sclerosis. *Eye and Brain*, 8, 195–202. <https://doi.org/10.2147/EB.S54131>
- Kalkman, J. (2017). Fourier-Domain Optical Coherence Tomography Signal Analysis and Numerical Modeling. *International Journal of Optics*, 2017. <https://doi.org/10.1155/2017/9586067>
- Kantarci, O. H. (2019). Phases and phenotypes of multiple sclerosis. *CONTINUUM Lifelong Learning in Neurology*, 25(3), 636–654. <https://doi.org/10.1212/CON.0000000000000737>
- Karaca, Y., & Hayta, Ş. (2015). The Significance of Artificial Neural Networks Algorithms Classification in the Multiple Sclerosis and Its Subgroups. *Iarjset*, 2(12), 1–7. <https://doi.org/10.17148/iarjset.2015.21201>
- Kashani, A. H., Asanad, S., Chen, J., Singer, M., Zhang, J., Sharifi, M., ... & Ringman, J. M. (2021). Past, present and future role of retinal imaging in neurodegenerative disease. *Progress in Retinal and Eye Research*, 100938.
- Kavaliunas, A., Manouchehrinia, A., Stawiarz, L., Ramanujam, R., Agholme, J., Hedström, A.K., Beiki, O., Glaser, A., Hillert, J. (2017). Importance of early treatment initiation in the clinical course of multiple sclerosis. *Mult. Scler.* 23, 1233–1240.
- Kilcher Y., “[Classic]Deep Residual Learning for Image Recognition (Paper Explained)” [video]. *Youtube*. Consultado por última vez en octubre de 2021. <https://www.youtube.com/watch?v=GWt6Fu05voI>

- Kim, H. (2019). The Definition of Convolution in Deep Learning by using Matrix. *Journal of Engineering and Applied Sciences*, 14(7), 2272–2275. <https://doi.org/10.36478/jeasci.2019.2272.2275>
- Lee, K. M., Lee, E. J., Kim, T. W., & Kim, H. (2016). Comparison of the abilities of SD-OCT and SS-OCT in evaluating the thickness of the macular inner retinal layer for glaucoma diagnosis. *PLoS ONE*, 11(1). <https://doi.org/10.1371/journal.pone.0147964>
- Lee, S. W., Jeong, H. W., Kim, B. M., Ahn, Y. C., Jung, W., & Chen, Z. (2009). Optimization for axial resolution, depth range, and sensitivity of spectral domain optical coherence tomography at 1.3  $\mu\text{m}$ . *Journal of the Korean Physical Society*, 55(6), 2354–2360. <https://doi.org/10.3938/jkps.55.2354>
- Leung, C. K. S., Cheung, C. Y. L., Weinreb, R. N., Lee, G., Lin, D., Chi, P. P., & Lam, D. S. C. (2008). Comparison of macular thickness measurements between time domain and spectral domain optical coherence tomography. *Investigative Ophthalmology and Visual Science*, 49(11), 4893–4897. <https://doi.org/10.1167/iovs.07-1326>
- Lo, C., Vuong, L., & Micieli, J. (2021). Recent advances and future directions on the use of optical coherence tomography in neuro-ophthalmology. *Taiwan Journal of Ophthalmology*, 11(1), 3–15. [https://doi.org/10.4103/tjo.tjo\\_76\\_20](https://doi.org/10.4103/tjo.tjo_76_20)
- Lopez-Bernal, D., Balderas, D., Ponce, P., & Molina, A. (2021). Education 4.0: Teaching the basics of knn, lda and simple perceptron algorithms for binary classification problems. *Future Internet*, 13(8). <https://doi.org/10.3390/fi13080193>
- López-Gómez, J., Sacristán-Enciso, B., & Caro-Miró, M. A. (2021). Clinically isolated syndrome: diagnosis and risk of developing clinically definite multiple sclerosis. *Neurologia*. <https://doi.org/10.1016/j.nrl.2021.01.011>



- Loukas S., 2020. “What is Machine Learning: Supervised, Unsupervised, Semi-Supervised and Reinforcement learning methods”. *Towards data science*.  
<https://towardsdatascience.com/what-is-machine-learning-a-short-note-on-supervised-unsupervised-semi-supervised-and-aed1573ae9bb>
- Lucas, V. S., Burk, R. S., Creehan, S., & Grap, M. J. (2014). Utility of high-frequency ultrasound: Moving beyond the surface to detect changes in skin integrity. *Plastic Surgical Nursing*, 34(1), 34–38. <https://doi.org/10.1097/PSN.0000000000000031>
- Marzullo, A., Kocevar, G., Stamile, C., Durand-Dubief, F., Terracina, G., Calimeri, F., & Sappey-Marinier, D. (2019). Classification of multiple sclerosis clinical profiles via graph convolutional neural networks. *Frontiers in Neuroscience*, 13(JUN).  
<https://doi.org/10.3389/fnins.2019.00594>
- Mathworks*. <https://es.mathworks.com>. Consultado a lo largo de toda la tesis.
- Macbeth, G., Razumiejczyk, E. & Ledesma, R. D. (2011). Cliff's Delta Calculator: A non-parametric effect size program for two groups of observations. *Universitas Psychologica*, 10 (2), 545-555.
- Michalewska Z., Nawrocki J. (2017), “Atlas of Swept Source Optical Coherence Tomography”, *Springer*.
- Miguel, J.M., Roldán, M., Pérez-Rico, C., Ortiz, M., Boquete, L., Blanco, R (2021). Using advanced analysis of multifocal visual-evoked potentials to evaluate the risk of clinical progression in patients with radiologically isolated syndrome. *Sci. Rep. 2021*, 11, 2036.
- Mihaylova, B., & Dimitrova, G. (2019). Evaluation of Retinal Nerve Fiber Layer and Inner Macular Layers in Primary Open-Angle Glaucoma with Spectral-Domain Optical Coherence Tomography. *Optic Nerve*. <https://doi.org/10.5772/intechopen.79102>

- Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, 565–571.  
<https://doi.org/10.1109/3DV.2016.79>
- Minakaran, N., de Carvalho, E. R., Petzold, A., & Wong, S. H. (2021). Optical coherence tomography (OCT) in neuro-ophthalmology. *Eye (Basingstoke)*, 35(1), 17–32.  
<https://doi.org/10.1038/s41433-020-01288-x>
- Montolío, A., Martín-Gallego, A., Cegoñino, J.; Orduna, E., Vilades, E., Garcia-Martin, E., Palomar A.P. (2021). Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography. *Comput. Biol. Med.* 2021, 133, 104416.
- Moon, S. W., Kim, E. S., Kim, Y. G., Yoo, S. Y., & Kwak, H. W. (2009). The Comparison of Macular Thickness Measurements and Repeatabilities Between Time Domain and Spectral Domain OCT. *Journal of the Korean Ophthalmological Society*, 50(7), 1050.  
<https://doi.org/10.3341/jkos.2009.50.7.1050>
- National Multiple Sclerosis Society (NMSS) (2020). *Types of MS*.  
<https://www.nationalmssociety.org/What-is-MS/Types-of-MS>
- Nandler Z. (2015, Mar 11). *Optical Coherence Tomography Basic Explanation* [video]. *Youtube*. <https://www.youtube.com/watch?v=HJnNJIUPm4s>
- Ömerhoca, S., YaziciAkkas, S., & Kale Icen, N. (2018). Multiple sclerosis: Diagnosis and Differential Diagnosis. *Archives of Neuropsychiatry*.  
<https://doi.org/10.29399/npa.23418>
- Opperman A. (2019). Artificial Intelligence vs. Machine Learning vs. Deep Learning. *Towards Data Science Blog*. Consultado por última vez en diciembre de 2021.

<https://towardsdatascience.com/artificial-intelligence-vs-machine-learning-vs-deep-learning-2210ba8cc4ac>

Organización Mundial de la Salud. *Atlas de Esclerosis Múltiple*, pp. 14-16, 2008.

Organización Mundial de la Salud. *Atlas de Esclerosis Múltiple*, 2020.

Otsuka K. (2021). 3-D Deep Learning: Lung Tumor

Segmentation (<https://www.mathworks.com/matlabcentral/fileexchange/71521-3-d-deep-learning-lung-tumor-segmentation>), MATLAB Central File Exchange.

Retrieved May 10, 2021.

Palombit, A., Castellaro, M., Calabrese, M., Romualdi, C., Pizzini, F. B., Montemezzi, S., ... Bertoldo, A. (2017). Cortical Thickness variability in Multiple Sclerosis: The role of lesion segmentation and filling. *Proceedings - International Symposium on Biomedical Imaging*, 792–795. <https://doi.org/10.1109/ISBI.2017.7950637>

Pandit, L., & Murthy, J. M. K. (2011). Treatment of multiple sclerosis. *Annals of Indian Academy of Neurology*, 14 (SUPPL. 1). <https://doi.org/10.4103/0972-2327.83094>

Parisi V. et al., 1999. Correlation between morphological and functional retinal impairment in multiple sclerosis patients. *Invest. Ophthalmol. Vis. Sci.*, vol. 40, no. 11, pp. 2520–7, [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10509645>.

Patti, F., Chisari, C. G., Arena, S., Toscano, S., Finocchiaro, C., Fermo, S. Lo, Judica, M. L., & Maimone, D. (2021). Factors driving delayed time to multiple sclerosis diagnosis: Results from a population-based study. *Multiple Sclerosis and Related Disorders*. <https://doi.org/10.1016/j.msard.2021.103361>

Pérez Del Palomar, A.; Cegoñino, J.; Montolío, A.; Orduna, E.; Vilades, E.; Sebastián, B.; Pablo, L.E.; Garcia-Martin, E. (2019) Swept source optical coherence tomography to

early detect multiple sclerosis disease. The use of machine learning techniques. *PLoS One*. 14, e0216410.

Pérez-Porto J., Merino M. (2011, 2014). Definición de esclerosis múltiple. *Definición de*. <https://definicion.de/esclerosis-multiple/>

Petzold A. et al., 2017. Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis. *Lancet Neurol.*, vol. 16, no. 10, pp. 797–812. doi: 10.1016/S1474-4422(17)30278-8.

Pinto, M.F., Oliveira, H.; Batista, S.; Cruz, L.; Pinto, M.; Correia, I.; Martins, P.; Teixeira, C (2020). Prediction of disease progression and outcomes in multiple sclerosis with machine learning. *Sci. Rep.* 2020, 10, 21038.

Phani, 2021. “What is the Convolutional Neural Network Architecture?” *AnalyticsVidhya Blog*. Consultado por última vez en noviembre de 2021. <https://www.analyticsvidhya.com/blog/2020/10/what-is-the-convolutional-neural-network-architecture/>

Polman, C.H., Reingold, S.C., Edan, G., Filippi, M., Hartung, H.P., Kappos, L., Lublin, F.D., Metz, L.M., McFarland, H.F., O'Connor, P.W., Sandberg Wollheim, M., Thompson, A.J., Weinshenker, B.G. and Wolinsky, J.S. (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the “McDonald Criteria”. *Ann Neurol.*, 58: 840-846. <https://doi.org/10.1002/ana.20703>

Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., ... Wolinsky, J. S. (2011). Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Annals of Neurology*, 69(2), 292–302. <https://doi.org/10.1002/ana.22366>

- Popescu, D. P., Choo-Smith, L. P. in., Flueraru, C., Mao, Y., Chang, S., Disano, J., ... Sowa, M. G. (2011). Optical coherence tomography: Fundamental principles, instrumental designs and biomedical applications. *Biophysical Reviews*, 3(3), 155–169. <https://doi.org/10.1007/s12551-011-0054-7>
- Puntmann, V.O., 2009. How-to guide on biomarkers: biomarker definitions, validation and applications with examples from cardiovascular disease. *Postgrad. Med. J.*, vol. 85, no. 1008, pp. 538–545, Oct. 2009, doi: 10.1136/pgmj.2008.073759.
- Pupale R., 2018. “Support Vector Machines (SVM) – An Overview”. *TowardsDataScience Blog*. Consultado por última vez en diciembre de 2021. <https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>
- Radiation Emergency Medical Management (REMM) (2020). *Electromagnetic Radiation: Ionizing and Non-ionizing*. <https://www.remm.nlm.gov/radiationspectrum.htm>
- Rebbah S., Delahaye D., Puechmorel Stéphane, Maréchal P., Nicol F., Berry I. (2019). Classification of Multiple Sclerosis patients using a histogram-based K-Nearest Neighbors algorithm. *OHBM 2019, 25th annual meeting of Organization for Human Brain Mapping, Jun 2019, Rome, Italy*. f4hal02156448f
- Riva M, 2021. Batch Normalization in Convolutional Neural Networks. *Baeldung CS*. Consultado por última vez en agosto de 2021. <https://www.baeldung.com/cs/batch-normalization-cnn>
- Rosenblum K (2020). Multiple Sclerosis and COVID-19: what we know for now. *CedarsSinai Blog*. Consultado por última vez en diciembre de 2021 <https://www.cedars-sinai.org/blog/ms-and-covid-19.html>

- Sánchez-Dalmau, B., Llorenç, V., Ortiz, S., & Martínez, L. (2010). Comparison of three optical coherence tomography devices, one time-domain and two Fourier-domain, for the estimation of the retinal nerve fiber layer thickness. *Archivos de La Sociedad Española de Oftalmología*, 85(5), 165–173. [https://doi.org/10.1016/s0365-6691\(10\)55002-0](https://doi.org/10.1016/s0365-6691(10)55002-0)
- Santiago L. (2016). Análisis avanzado de señales de potenciales evocados multifocales aplicados al diagnóstico de neuropatías ópticas. Universidad de Alcalá, Alcalá de Henares.
- Santos, C. F. G. dos, Moreira, T. P., Colombo, D., & Papa, J. P. (2020). Does Removing Pooling Layers from Convolutional Neural Networks Improve Results? *SN Computer Science*, 1(5). <https://doi.org/10.1007/s42979-020-00295-9>
- Sarra, S.A., 2017. The Matlab Radial Basis Function Toolbox. *Journal of Open Research Software*, 5(1), p.8. <http://doi.org/10.5334/jors.131>
- Schurz N., Sariaslani L., Altmann P., Leutmezer F., Mitsch C., Pemp B., Rommer P., Zrzavy T., Berger T., Bsteh G. Evaluation of Retinal Layer Thickness Parameters as Biomarkers in a Real-World Multiple Sclerosis Cohort. *Eye Brain*. 2021; 13:59-69 <https://doi.org/10.2147/EB.S295610>
- Sharma S., “Activation Functions in Neural Networks”. Towards Data Science Blog, 2017. <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Shi, C., Jiang, H., Gameiro, G.R., Hu, H., Hernandez, J., Delgado, S., Wang, J. (2019). Visual Function and Disability Are Associated With Focal Thickness Reduction of the

- Ganglion Cell-Inner Plexiform Layer in Patients With Multiple Sclerosis. *Invest. Ophthalmol. Vis. Sci.* 60, 1213–1223. DOI: 10.1167/iovs.18-25809
- Shim, Ji-Yeon & Zhang, Jan-Wei & Yoon, Han-Yong & Kang, Bong & Kim, I.s. (2018). Prediction model for bead reinforcement area in automatic gas metal arc welding. *Advances in Mechanical Engineering*. <https://doi.org/10.1177/1687814018781492>
- Silverstein, S. M., Fradkin, S. I., & Demmin, D. L. (2020). Schizophrenia and the retina: Towards a 2020 perspective. *Schizophrenia Research*, 219, 84–94. <https://doi.org/10.1016/j.schres.2019.09.016>
- Smith K.J. & McDonald W.I. (1999). The pathophysiology of multiple sclerosis : the machanisms underlying the production or symptoms and the natural history of the disease. *Phil. Trans. R. Soc. Lond. B (1999)* 354, 1649-1673. *The Royal Society*.
- Solomon A.J., Pettigrew R., Naismith R.T., Chahin S., Krieger S., Weinshenker B. (2021). Challenges in multiple sclerosis diagnosis: Misunderstanding and misapplication of the McDonald criteria. *Mult. Scler. J.*, vol. 27, no. 2, pp. 250–258. doi: 10.1177/1352458520910496.
- Sospedra, M., & Martin, R. (2005). Immunology of multiple sclerosis. *Annual Review of Immunology*, 23, 683–747. <https://doi.org/10.1146/annurev.immunol.23.021704.115707>
- Tátrai, E., Simó, M., Iljicsov, A., Németh, J., DeBuc, D. C., & Somfai, G. M. (2012). In vivo evaluation of retinal neurodegeneration in patients with multiple sclerosis. *PLoS ONE*, 7(1). <https://doi.org/10.1371/journal.pone.0030922>
- Tăuțan, A. M., Ionescu, B., & Santarnecchi, E. (2021). Artificial intelligence in neurodegenerative diseases: A review of available tools with a focus on machine learning techniques. *Artificial Intelligence in Medicine*, 117. <https://doi.org/10.1016/j.artmed.2021.102081>

- The Retinal Tunic. 18 May 2007 at the [Wayback Machine](#) *Virginia–Maryland Regional College of Veterinary Medicine*.  
<https://web.archive.org/web/20070518033845/http://education.vetmed.vt.edu/Curriculum/VM8054/EYE/RETINA.HTM>
- Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., ... Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, 17(2), 162–173.  
[https://doi.org/10.1016/S1474-4422\(17\)30470-2](https://doi.org/10.1016/S1474-4422(17)30470-2)
- Topcon Healthcare. DRI OCT Triton™ Series Swept Source Optical Coherence Tomography. *Equipment specifications*. [https://topconhealthcare.jp/wp-content/uploads/2020/08/Topcon Triton Brochure Rev5 202105 E325-.pdf](https://topconhealthcare.jp/wp-content/uploads/2020/08/Topcon_Triton_Brochure_Rev5_202105_E325-.pdf)
- Tukey J.W. (1977). Exploratory data análisis. *Reading, Mass. Addison-Wesley Pub. Co., @1977*.
- Tzotsos, A., &Argialas, D. (2008). Support vector machine classification for object-based image analysis. *Lecture Notes in Geoinformation and Cartography*, 0(9783540770572), 663–677. [https://doi.org/10.1007/978-3-540-77058-9\\_36](https://doi.org/10.1007/978-3-540-77058-9_36)
- Vapnik, V.N (1999). An overview of statistical learning theory. *IEEE Trans. neural networks* 1999, 10, 988–99. doi:10.1109/72.788640.
- Vázquez-Marrufo, M., Sarrias-Arrabal, E., García-Torres, M., Martín-Clemente, R., & Izquierdo, G. (2021). A systematic review of the application of machine-learning algorithms in multiple sclerosis. *Neurologia*. <https://doi.org/10.1016/j.nrl.2020.10.017>



- Vinod R. (2020). Batch Normalization Explained. *Towards Data Science Blog*.  
<https://towardsdatascience.com/batch-normalisation-explained-5f4bd9de5feb>
- Wang, C., Xia, X., Tian, B., & Zhou, S. (2015). Comparison of Fourier-Domain and Time-Domain Optical Coherence Tomography in the Measurement of Thinnest Corneal Thickness in Keratoconus. *Journal of Ophthalmology*, 2015.  
<https://doi.org/10.1155/2015/402925>
- Wicki, C.A., Hanson, J.V.M., Schippling, S. (2018). Optical coherence tomography as a means to characterize visual pathway involvement in multiple sclerosis. *Curr. Opin. Neurol.* 31, 662–668.
- Yap, T.E., Balendra, S.I., Almonte, M.T., Cordeiro, M.F. (2019). Retinal correlates of neurological disorders. *Ther. Adv. Chronic Dis.* Doi: 10, 204062231988220.
- Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, 1168(2). <https://doi.org/10.1088/1742-6596/1168/2/022022>
- Zhang, Y. D., Pan, C., Sun, J., & Tang, C. (2018). Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU. *Journal of Computational Science*, 28, 1–10. <https://doi.org/10.1016/j.jocs.2018.07.003>
- Zurita, M., Montalba, C., Labbé, T., Cruz, J. P., Dalboni da Rocha, J., Tejos, C., Ciampi, E., Cárcamo, C., Sitaram, R., & Uribe, S. (2018). Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data. *Neuro Image: Clinical*, 20, 724–730.  
<https://doi.org/10.1016/j.nicl.2018.09.002>

**Artículos del Grupo de Investigación en Ingeniería Biomédica, UAH:**

Barea R., Ortiz M., García-Martín E., Rodrigo M.J., Vilades E., López Dorado A., Gutiérrez R., Cavaliere C., Boquete L. (2020). Detección de pacientes con esclerosis múltiple mediante procesamiento de imágenes de tomografía de coherencia óptica utilizando redes neuronales. *27º Seminario anual de automática, electrónica industrial e instrumentación*.

Barea R., Ortiz M., García-Martín E., Rodrigo M.J., Vilades E., López Dorado A., Gutiérrez R., Cavaliere C., Boquete L. (2020) Diagnóstico de esclerosis múltiple mediante el análisis de imágenes de tomografía de coherencia óptica con redes neuronales convolucionales. *XXXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*.

Cavaliere C., Ortiz M., Barea R., García-Martín E., Rodrigo M.J., Vilades E., Boquete L. (2019) Analysis of retinal layers using SS-OCT registries for diagnosing multiple sclerosis. *27th Annual Meeting of the European Charcot Foundation*.

Cavaliere C., Ortiz M., Barea R., García-Martín E., Rodrigo M.J., Vilades E., Boquete L. (2019) Diagnóstico de esclerosis múltiple a partir de imágenes de tomografía de coherencia óptica. *XXXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*.

Cavaliere C., Ortiz M., Barea R., García-Martín E., Rodrigo M.J., Vilades E., Boquete L. (2020). Diagnóstico de esclerosis múltiple mediante procesamiento de imágenes de tomografía de coherencia óptica utilizando redes neuronales. *XXXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*.

García-Martín E. et al. (2017). Retinal and Optic Nerve Degeneration in Patients with Multiple Sclerosis Followed up for 5 Years. *Ophthalmology*, vol. 124, no. 5, pp. 688–696, doi: 10.1016/j.ophtha.2017.01.005.

- García-Martin, E., Ortiz, M., Boquete, L., Sánchez-Morla, E. M., Barea, R., Cavaliere, C., Vilades, E., Orduna, E., & Rodrigo, M. J. (2021). Early diagnosis of multiple sclerosis by OCT analysis using Cohen's d method and a neural network as classifier. *Computers in Biology and Medicine*, 129. <https://doi.org/10.1016/j.combiomed.2020.104165>
- López-Dorado, A., Ortiz, M., Satue, M., Rodrigo, M. J., Barea, R., Sánchez-Morla, E. M., Cavaliere, C., Rodríguez-Ascariz, J. M., Orduna-Hospital, E., Boquete, L., & García-Martin, E. (2022). Early diagnosis of multiple sclerosis by OCT analysis using Swept-Source Optical Coherence Tomography and Convolutional Neural Networks Trained with Data Augmentation *Sensors*, 22(1), 167. DOI: 10.3390/s22010167