



**Universidad**  
Zaragoza

## **Trabajo Fin de Máster**

El uso de la retina como biomarcador en enfermedades neurodegenerativas. Simulación numérica.

Use of retina as a biomarker in neurodegenerative diseases. Numerical simulation.

Autor

**Alejandro Martín Gallego**

Directora

**Amaya Pérez del Palomar Aldea**

Máster Universitario en Ingeniería Mecánica  
Mecánica de Medios Continuos y Teoría de Estructuras  
Escuela de Ingeniería y Arquitectura de la Universidad de Zaragoza  
Curso 2017-2018



## DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD

(Este documento debe acompañar al Trabajo Fin de Grado (TFG)/Trabajo Fin de Máster (TFM) cuando sea depositado para su evaluación).

D./D<sup>a</sup>. \_\_\_\_\_,

con nº de DNI \_\_\_\_\_ en aplicación de lo dispuesto en el art.

14 (Derechos de autor) del Acuerdo de 11 de septiembre de 2014, del Consejo de Gobierno, por el que se aprueba el Reglamento de los TFG y TFM de la Universidad de Zaragoza,

Declaro que el presente Trabajo de Fin de (Grado/Máster)  
\_\_\_\_\_, (Título del Trabajo)

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

es de mi autoría y es original, no habiéndose utilizado fuente sin ser citada debidamente.

Zaragoza, \_\_\_\_\_

Fdo: \_\_\_\_\_

# RESUMEN

---

En la actualidad el diagnóstico de la esclerosis múltiple (EM) se basa en la existencia de criterios clínicos de diseminación espacial (dos o más lesiones independientes en el sistema nervioso central) y temporal (dos o más episodios de disfunción neurológica), con la ayuda de los métodos de investigación paraclínicas como el líquido cefalorraquídeo (LCR, que presenta un aspecto sospechoso en pacientes con EM), la resonancia magnética (RM, que permite descartar otras enfermedades sospechosas) y los potenciales evocados (PE, utilizados para la valoración de la función en algunas vías nerviosas). Este método diagnóstico resulta costoso y en ocasiones muy largo, lo que puede retrasar de forma crítica la aplicación de tratamientos que puedan frenar el avance de los efectos discapacitantes de la enfermedad en los pacientes. En este contexto se pretende conseguir un sistema de diagnóstico rápido, eficaz y fiable. Para ello, se realiza un análisis de la retina como biomarcador en la EM. Ésta forma parte del sistema nervioso, afectado por la enfermedad, y además en diversos estudios bibliográficos se encuentran evidencias de la alteración del grosor de la retina en pacientes que sufren la patología.

Para llevar a cabo el análisis, se estudian los datos obtenidos por el Servicio de Neuro-Oftalmología del Hospital Universitario Miguel Servet de Zaragoza a través de una prueba no invasiva llamada OCT (*Optic Coherence Tomography*), realizada tanto para pacientes con EM como para controles sanos. La OCT nos proporciona datos del grosor de las capas de la retina discriminando entre diferentes capas y zonas de la misma. En primer lugar, se realiza un tratamiento previo a las bases de datos proporcionadas por el Hospital para determinar los datos más relevantes, normalizar las bases de datos y poder realizar un estudio estadístico con las mismas. Además, se realiza un estudio de sensibilidad de los parámetros característicos para observar la importancia de los mismos y su relación con el diagnóstico de la enfermedad.

En segundo lugar, se realizan varios modelos de predicción diagnóstica mediante el entrenamiento y validación de dos algoritmos de clasificación binaria y aprendizaje supervisado: el perceptrón multicapa (MLP, del inglés *MultiLayer Perceptron*) y las máquinas de soporte vectorial (SVM, del inglés *Support Vector Machine*). Se utilizan las librerías de aprendizaje automático implementadas en Matlab y para cada uno de los algoritmos se realiza una optimización de los diferentes parámetros que influyen en su capacidad de predicción con el objetivo de minimizar el error cometido en el diagnóstico.

Por último, y en base a los resultados obtenidos, se compara la capacidad de los dos algoritmos en la clasificación de los pacientes en sanos y enfermos, además se muestra el valor diagnóstico de los diferentes modelos de predicción implementados mediante la representación de curvas ROC (*Receiver Operating Characteristic*) que permiten seleccionar uno de los modelos desarrollados que será el que tiene mayor valor diagnóstico para la patología.

A modo de conclusión, se valora la retina como biomarcador en la EM y se sugiere la inclusión de ésta en el diagnóstico a través de la prueba OCT y un modelo desarrollado mediante algoritmos de predicción. Esto ayudará a los neurólogos a decidir y clasificar un paciente con esclerosis múltiple y llegando a conseguir así un diagnóstico más rápido que permita comenzar un tratamiento precoz ayudando a frenar la progresión de la enfermedad y a mejorar la calidad de vida de los pacientes.

# Índice

1	INTRODUCCIÓN.....	3
1.1	Motivación.....	3
1.2	Estado del arte.....	3
1.3	Objetivos.....	4
2	ENFERMEDAD.....	5
2.1	Esclerosis Múltiple.....	5
2.1.1	Efectos de la EM en el sistema central nervioso.....	5
2.2	Hipótesis de trabajo.....	6
3	ANATOMÍA DEL OJO HUMANO.....	7
3.1	Sistema nervioso visual.....	7
3.2	Ojo humano.....	8
3.3	Retina.....	9
4	REDES NEURONALES ARTIFICIALES.....	11
4.1	Inteligencia artificial.....	11
4.2	Redes neuronales artificiales.....	11
4.2.1	Modos de operación.....	12
4.3	El perceptrón multicapa (MLP).....	12
4.3.1	Propagación de los patrones de entrada.....	12
4.3.2	Algoritmo de retropropagación (Backpropagation).....	13
4.4	Máquinas de soporte vectorial (SVM).....	15
4.4.1	SVM para clasificación binaria de ejemplos separables linealmente.....	15
4.4.2	SVM para clasificación binaria de ejemplos no linealmente separables.....	16
4.4.3	Truco del Kernel para el caso no linealmente separable.....	17
5	MATERIAL Y MÉTODO.....	19
5.1	Sujetos y protocolo de medida.....	19
5.1.1	Tomografía de coherencia óptica.....	20
5.2	Tratamiento de datos.....	20
5.2.1	Selección de variables.....	21
5.3	Normalización de los datos.....	28
5.4	Análisis estadístico mediante el método de las redes neuronales.....	28
5.4.1	Validación cruzada.....	28
5.4.2	Curvas ROC.....	29
6	EXPERIMENTOS REALIZADOS Y RESULTADOS.....	31
6.1	Predicción mediante el Perceptrón Multicapa.....	31

6.2	Predicción mediante SVM. ....	33
6.3	Precisiones alcanzadas. ....	35
6.4	Área bajo las curvas ROC. ....	35
7	CONCLUSIONES. ....	38
8	LÍNEAS DE FUTURO. ....	40
9	BIBLIOGRAFÍA. ....	41
10	ÍNDICE DE FIGURAS. ....	44
11	ÍNDICE DE TABLAS. ....	46

# 1 INTRODUCCIÓN.

## 1.1 Motivación.

La Esclerosis Múltiple (EM) es una enfermedad neurológica, crónica y discapacitante, que afecta principalmente a adultos jóvenes. Se caracteriza por la inflamación y desmielinización de las fibras nerviosas, así como por la degeneración axonal [1,2], la cual se encuentra ya presente en fases tempranas de la enfermedad. La degeneración axonal contribuye al desarrollo de atrofia cerebral y se relaciona con la discapacidad de la enfermedad [3].

Todo esto hace que un diagnóstico temprano de la enfermedad sea de vital importancia en la aplicación de tratamientos que intenten frenar la velocidad de avance de la enfermedad en los pacientes para ayudar al retardo de los efectos neurodegenerativos y discapacitantes de la misma [4].

Con la motivación de contribuir en la mejora de la predicción y un temprano diagnóstico de la EM se realiza el Trabajo de Fin de Máster (TFM) enfocado al estudio de nuevos métodos para predecir y diagnosticar la enfermedad.

## 1.2 Estado del arte.

En la actualidad la esclerosis múltiple solo puede ser diagnosticada con fiabilidad mediante una autopsia post-mortem o una biopsia, pero existen criterios no invasivos para poder diagnosticarla con cierta certeza [5]. Es posible detectar lesiones provocadas por la EM sin que hayan existido nunca fenómenos clínicos, para ello se utilizan métodos de investigación paraclínicos como el análisis del líquido cefalorraquídeo (LCR, que presenta un aspecto sospechoso en pacientes con EM), la resonancia magnética (RM, que permite descartar otras enfermedades sospechosas) y los potenciales evocados (PE, utilizados para la valoración de la función en algunas vías nerviosas) [6]. El diagnóstico de la EM se basa en la concurrencia de diseminación en el espacio (al menos dos lesiones de distinta localización en el SNC) y en el tiempo (al menos dos episodios de disfunción neurológica separados en el tiempo), por ello, el proceso de diagnóstico puede ser largo y con frecuencia puede suponer un reto incluso para el neurólogo más experimentado.

La investigación del origen, tratamiento y evolución de esta enfermedad es un tema de especial transcendencia ya que no existe cura. La progresión de la enfermedad puede ser rápida y los pacientes jóvenes, lo que hace que el diagnóstico precoz sea muy útil para poder instaurar terapias que frenen la progresión de la enfermedad y por tanto disminuya el impacto en la capacidad funcional de los pacientes [4]. Así pues, la búsqueda de un biomarcador no invasivo es una prioridad en la actualidad [7–9].

Los síntomas y signos que aparecen pueden ser debidos a múltiples enfermedades diferentes [5], por lo que se solicita una serie de pruebas para descartar estas enfermedades antes de hacer un diagnóstico definitivo. Las pruebas que se solicitan con más frecuencia son: la analítica, la resonancia magnética cerebral, la puntuación lumbar y los potenciales evocados.

No se dispone de correlaciones clínicas y anatómicas fiables para medir la discapacidad en la esclerosis múltiple. Los biomarcadores estructurales como la imagen de resonancia magnética (IRM) se ven obstaculizados porque no se pueden separar con precisión la desmielinización de los axones de la lesión del tejido dentro del sistema nervioso central (SNC) [10].

Se ha intentado utilizar redes neuronales artificiales (RNA) en diversos estudios para predecir la evolución a corto plazo [11] y del curso de la enfermedad [12] en pacientes con esclerosis múltiple con resultados positivos. También se ha estudiado la habilidad de funciones lineales de discriminación a partir del dominio espectral OCT generando imágenes de la retina [13,14]. Y por último, se han realizado estudios de predicción de diagnóstico para otras enfermedades como la neuritis óptica [15]. Todos estos precedentes hacen pensar que los algoritmos de aprendizaje automático, junto a nuevas pruebas médicas, pueden ser de gran ayuda para los especialistas en el diagnóstico de esta enfermedad neurodegenerativa.

### 1.3 Objetivos.

En el presente trabajo se pretende evaluar la retina como biomarcador en enfermedades neurodegenerativas, concretamente en la esclerosis múltiple (EM). Los estudios realizados han demostrado que existe un adelgazamiento de la capa de fibras nerviosas de la retina (RNFL, del inglés Retinal Nerve Fiber Layer) mayor en los pacientes afectados de EM al compararlos con controles sanos [16,17].

Se pretende conseguir un sistema de diagnóstico rápido, eficaz y fiable para evaluar la posibilidad de que un paciente tenga EM a través del análisis de los datos obtenidos por una prueba no invasiva llamada OCT (Optic Coherence Tomography) que consiste en el escáner de las capas de la retina.

Por último, evaluar la capacidad tanto de las redes neuronales artificiales como las máquinas de soporte vectorial en la clasificación binaria en problemas complejos.

Por un lado, se sabe que la capa RNFL está relacionada con el desarrollo de la enfermedad [2] y se asume que un descenso en el espesor de la RNFL conlleva un descenso general del volumen del sistema nervioso central, por tanto, a través de la observación de la RNFL en la retina se pretende saber si un paciente tiene una gran probabilidad de tener EM.

Para este estudio se dispone de una gran cantidad de datos de OCT realizados en el **Servicio de Neuro-Oftalmología del Hospital Miguel Servet de Zaragoza** a través de los cuales se pretende entrenar y validar una red neuronal artificial (RNA) y una máquina de soporte vectorial (SVM).

Tanto la RNA como la SVM deberán poder calcular la probabilidad de un paciente de estar enfermo de EM con los datos de entrada que proporcionará una simple prueba médica no invasiva y rápida como es la OCT.

## 2 ENFERMEDAD.

### 2.1 Esclerosis Múltiple.

La Esclerosis Múltiple es una enfermedad neurológica autoinmune, crónica y discapacitante, que afecta principalmente a adultos jóvenes. Se caracteriza por la inflamación y desmielinización de las fibras nerviosas, así como por la degeneración axonal, la cual se encuentra ya presente en fases tempranas de la enfermedad. La degeneración axonal contribuye al desarrollo de atrofia cerebral y se relaciona con la discapacidad de la enfermedad [3].

Existen varios tipos de EM según la evolución de la enfermedad en cada paciente [18]:

- EM remitente recurrente (EMRR): Presencia de brotes de la enfermedad claramente definidos con recuperación parcial con secuelas o recuperación total y completa. No existe progresión de la enfermedad entre los brotes.
- EM primaria progresiva (EMPP): Los pacientes con esta forma clínica presentan progresión de la enfermedad desde el inicio, con estabilizaciones ocasionales y mejorías temporalmente menores. Estos pacientes no presentan brotes.
- EM secundaria progresiva (EMSP): El curso inicial es de una forma EMRR, seguido de un incremento de la discapacidad neurológica de forma progresiva con o sin brotes sobreañadidos.
- EM progresiva recurrente (EMPR): Estos enfermos presentan progresión desde el inicio con brotes evidentes, con o sin recuperación completa y con progresión continua en el periodo entre brotes.

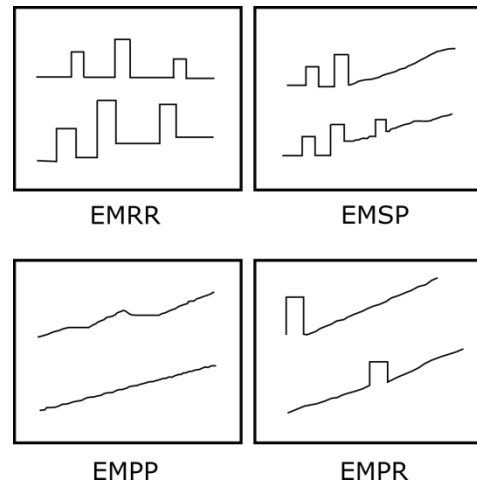


Figura 2.1 Esquema de la evolución temporal de los tipos de EM

#### 2.1.1 Efectos de la EM en el sistema nervioso central.

La enfermedad causa varios efectos en el ser humano:

- Desmielinización: en estado normal existe una barrera entre el sistema nervioso central y la sangre llamada hematoencefálica que ‘tapiza’ los vasos sanguíneos. En pacientes con EM esta barrera no funciona bien, lo que hace que las células T ataquen la mielina del sistema nervioso, sustancia que recubre los axones y que tiene gran importancia en la transmisión nerviosa.
- Daño en los axones: se ha demostrado que el daño a los axones es una de las principales causas del desarrollo de discapacidades permanentes. Se ha demostrado que esta lesión permanente no solo está presente en estadios tardíos de la enfermedad, sino que interviene desde el inicio [2,19].
- Remielinización: el cerebro es capaz de reclutar células madre que migran de otras zonas del cerebro y recrean la cubierta de mielina. Esta nueva cubierta no es tan efectiva como la original, y ataques repetidos tendrán como consecuencia la disminución de la capacidad de remielinización.

La remielinización es una de las causas por las que en fases tempranas de la enfermedad los síntomas tienden a disminuir o desaparecer después de un ataque.

Sin embargo, el daño axonal y la pérdida irreversible de las neuronas aparece muy pronto en el transcurso de la enfermedad. La EM causa lesiones múltiples en el cerebro y la espina dorsal, y los síntomas pueden variar mucho entre individuos, dependiendo de dónde ocurran las lesiones.

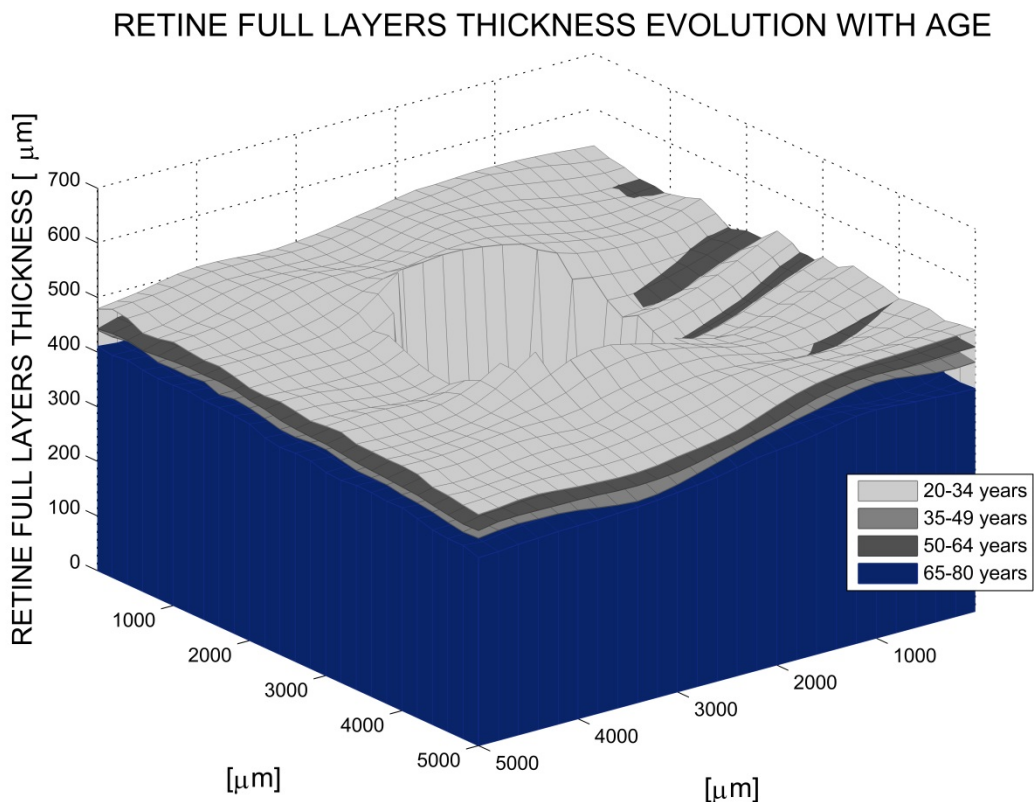


## 2.2 Hipótesis de trabajo.

Para poder realizar un modelo de predicción de la enfermedad hay que asumir ciertas hipótesis de trabajo apoyadas en los efectos que la EM causa en el ser humano para tomarlas como referencia en la predicción.

La pérdida axonal en el sistema nervioso central (SNC) [2,19] va a traducirse en una reducción de la capa de fibras nerviosas de la retina (RNFL), ya que ésta forma parte del sistema nervioso visual, y a su vez, éste forma parte del SNC [20].

Esta reducción del espesor de la RNFL la tomaremos como un biomarcador para la predicción de esclerosis múltiple, sin olvidar que todos los seres humanos sufren una pérdida axonal a lo largo de su vida, debido a que las células van muriendo con el paso de los años, lo que se traduce en una reducción del espesor de la retina con la edad (Ver Figura 2.2).



**Figura 2.2 Evolución del espesor [ $\mu\text{m}$ ] de la retina (todas las capas) con la edad de un grupo de pacientes sanos del Hospital Universitario Miguel Servet de Zaragoza.**

Debido a esta reducción general del espesor de la retina con la edad y el amplio rango de espesores para los pacientes, no resulta inmediato encontrar una relación entre dicho espesor y el diagnóstico de la Esclerosis Múltiple.

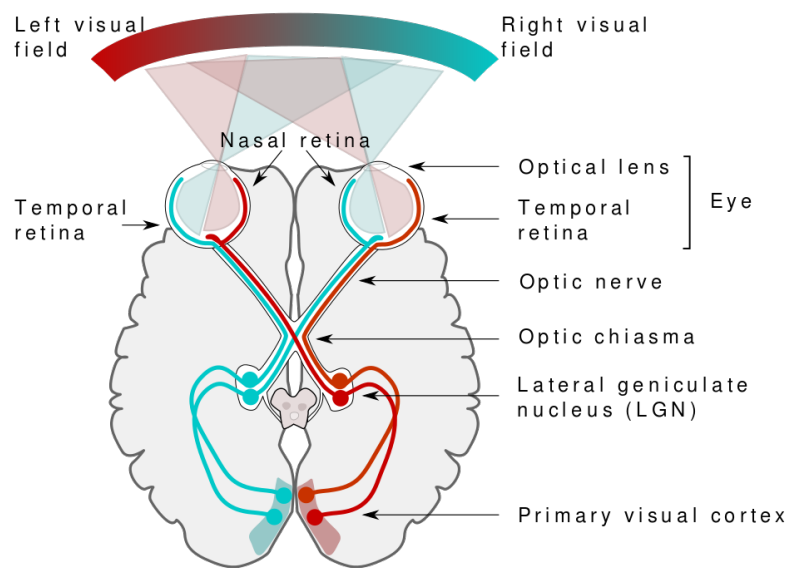
### 3 ANATOMÍA DEL OJO HUMANO.

La relación entre la EM y la disminución del espesor de la capa de fibras nerviosas de la retina (RNFL) nos lleva al estudio de la anatomía humana, concretamente al sentido de la vista y al sistema nervioso visual.

#### 3.1 Sistema nervioso visual.

El sistema visual humano (SVH) es el encargado de convertir las ondas electromagnéticas que pertenecen al espectro visible y que llegan hasta los ojos, en señales nerviosas que son interpretadas por el cerebro[21].

En la Figura 3.1 podemos observar el funcionamiento del sentido de la vista en el ser humano, y como el SVH transmite la información visual desde la retina hasta el córtex visual primario.



**Figura 3.1 Esquema de las principales vías visuales del cerebro**

Las principales partes del sistema visual [22] son las siguientes:

- Nervio óptico: es el nervio sensitivo encargado de transmitir la información visual desde la retina hasta el cerebro. Se origina en la capa de células ganglionares de la retina.
- Quiasma óptico: es la parte del cerebro donde se entrecruzan parcialmente las fibras axónicas de los nervios ópticos. En este entrecruzamiento la mitad de las fibras pasan del nervio óptico derecho a la cintilla óptica izquierda, y viceversa, así las imágenes formadas en cada una de las retinas se cruzan al lado opuesto del cerebro.
- Cintillas ópticas o tracto óptico: es el conjunto de fibras nerviosas que son la continuación del nervio óptico desde el quiasma óptico hasta el núcleo geniculado lateral.
- Núcleo geniculado lateral (NGL): es el centro de procesamiento primario de la información visual recibida por la retina del ojo. El NGL se localiza dentro de la región cerebral conocida como tálamo, y recibe la información directamente a través de las células ganglionares que provienen de la retina a través del tracto óptico. Las neuronas del NGL envían sus axones a través de la radiación óptica hasta el córtex visual.
- Radiación óptica, también conocida como el tracto geniculocalcarino, son los axones que van desde las neuronas del NGL hasta la corteza visual primaria.
- Corteza o córtex visual primario: recibe la información visual del NGL y está localizado en el polo posterior de la corteza occipital. Es el responsable del procesamiento de los estímulos visuales.

### 3.2 Ojo humano.

El ojo, junto al cerebro, es el elemento fundamental del sentido de la visión. Su funcionamiento básicamente consiste en transformar la energía lumínica en señales eléctricas que son enviadas al cerebro a través del nervio óptico. [23]

Las partes del ojo humano son las siguientes [24]:

- El iris: parte coloreada que tiene como función regular la entrada de luz aumentando o disminuyendo su tamaño según la intensidad de la misma.
- La pupila: orificio central del iris. Se dilata o contrae en función de la cantidad de luz existente.
- El cristalino: parte del ojo que enfoca el haz de luz en la retina. Tiene forma de lente biconvexa y es la segunda lente más importante.
- La córnea: es una de las partes externas del ojo y protege al cristalino y al iris permitiendo el paso de la luz.
- La retina: es la parte del ojo sensible a la luz. Está compuesta de conos y bastones. El ojo tiene alrededor de 6 millones de conos y su función es dar información sobre la nitidez y el color. Los bastones son 120 millones y son muy sensibles, con ellos percibimos el brillo y el blanco y negro. Se estimulan en función de la luz que reciben y envían información al nervio óptico. Detalle de la posición de la retina en el ojo humano en la Figura 3.3.
- Nervio óptico: conduce los impulsos nerviosos de los bastones y los conos al cerebro. El mensaje visual es transmitido en forma de señales eléctricas, las cuales el cerebro transformará en sensación visual.

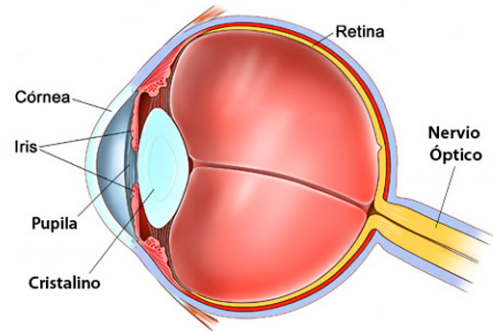


Figura 3.2 Partes del ojo humano. [4]

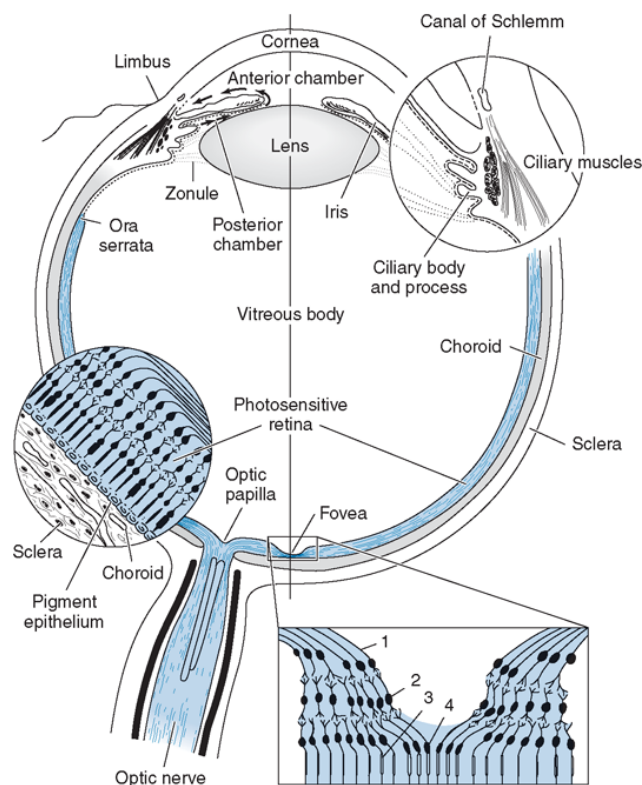


Figura 3.3 Partes del ojo humano y detalle de la retina [25]

### 3.3 Retina.

La reducción de espesor de la retina se va a considerar como un biomarcador en la predicción de la esclerosis múltiple [15], por lo que se procede a explicar las diferentes capas que forman esta importante parte del sistema nervioso visual [24]:

1. Membrana limitante interna: separa la retina del humor vítreo.
2. Capa de fibras nerviosas de la retina (RNFL): formada por axones de las células ganglionares.
3. Células ganglionares: formada por los núcleos de las mismas.
4. Capa plexiforme interna: región de la conexión sináptica entre células bipolares, amacrinas y ganglionares.
5. Capa nuclear interna: formada por los núcleos de las células bipolares, horizontales y amacrinas.
6. Capa plexiforme externa: región de conexión sináptica entre células fotorreceptoras y las células bipolares.
7. Capa nuclear externa: formada por los núcleos de las células fotorreceptoras.
8. Capa limitante externa: uniones intercelulares entre células fotorreceptoras y células Müller.
9. Fotorreceptores: formado por los segmentos más externos de los conos y los bastones.
10. Epitelio pigmentario retina: es la capa más externa de la retina y está formada por las células cúbicas que poseen gránulos de melanina que le dan una pigmentación característica.

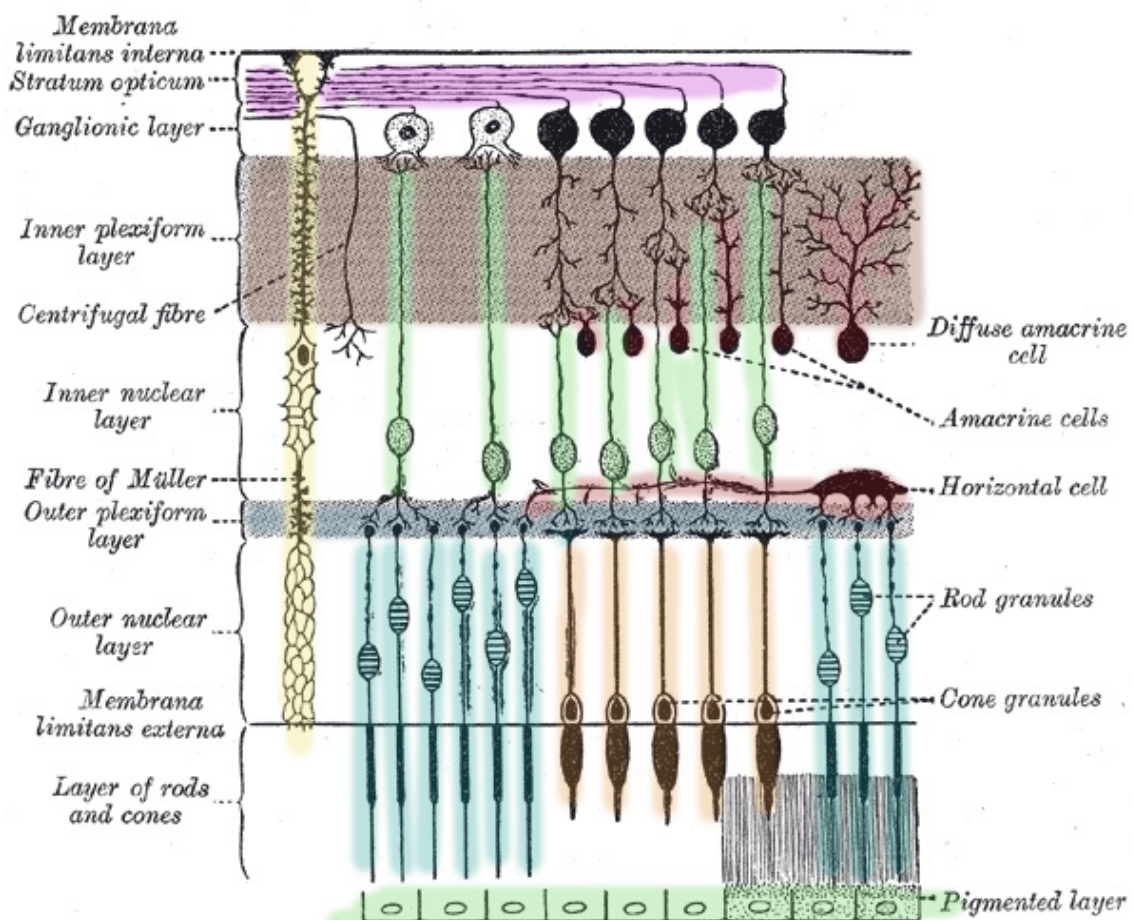
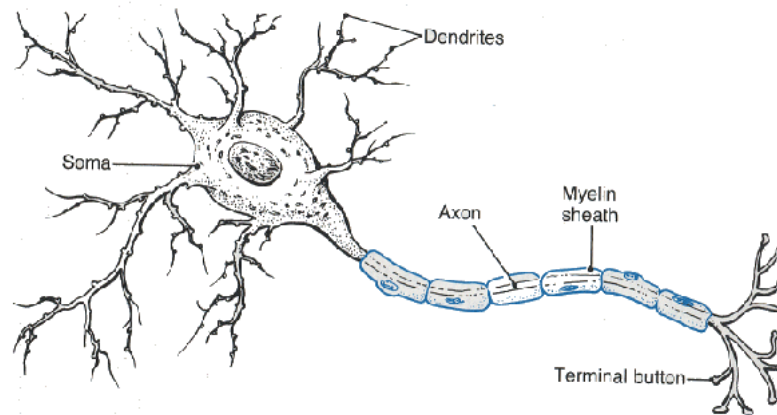


Figura 3.4 Ilustración de las capas de la retina de Ramón y Cajal

La parte más importante de la retina en el estudio de la predicción de la enfermedad son la capa de células ganglionares y la capa de fibras nerviosas de la retina RNFL (Retinal Nerve Fiber Layer).

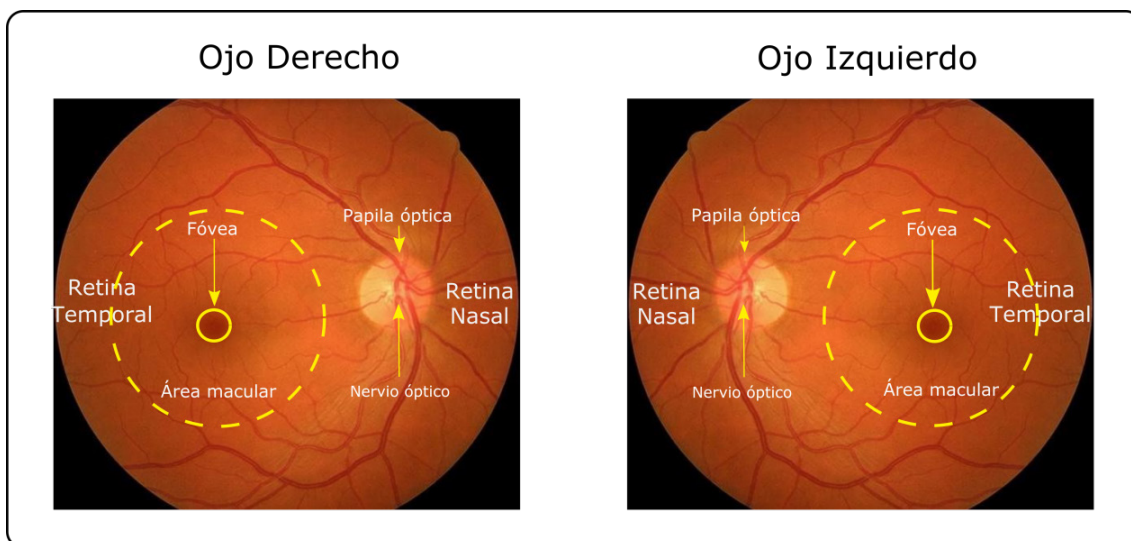
Las células ganglionares representan la salida de la retina, es decir, son las encargadas de transmitir la información visual que percibimos en forma de impulsos eléctricos y como en todas las células, las partes fundamentales de éstas son; el soma (cuerpo de la célula), las dendritas (reciben entradas de otras células) y los axones (que envían salidas a otras células):



**Figura 3.5 Representación de las partes de una célula**

La desmielinización y la pérdida axonal son algunos de los efectos de la esclerosis múltiple [2]. En la retina no hay mielina recubriendo los axones, pero sí se produce pérdida axonal, por lo que un método a través del cual se pueda analizar el espesor de la RNFL, que contiene los axones de las células ganglionares, puede ser clave en la predicción de la EM.

En la siguiente imagen se muestran las zonas más representativas de la retina humana, tanto para el ojo derecho como el izquierdo, en una imagen obtenida mediante una OCT.



**Figura 3.6 Imagen tomada mediante la técnica de OCT (Optical Coherence Tomography) de la retina de un ser humano. A la izquierda, imagen de la retina del ojo derecho y a la derecha, imagen de la retina del ojo izquierdo.**

En la Figura 3.6 podemos distinguir la zona temporal y la zona nasal de la retina. Ésta última corresponde a la zona de la retina que está más cerca de la nariz, de forma que las imágenes de la retina son simétricas entre el ojo derecho e izquierdo. También se marca la posición del nervio óptico, la papila, el área macular y la fovea (centro del área macular), zonas importantes en el estudio de la retina.

## 4 REDES NEURONALES ARTIFICIALES.

### 4.1 Inteligencia artificial.

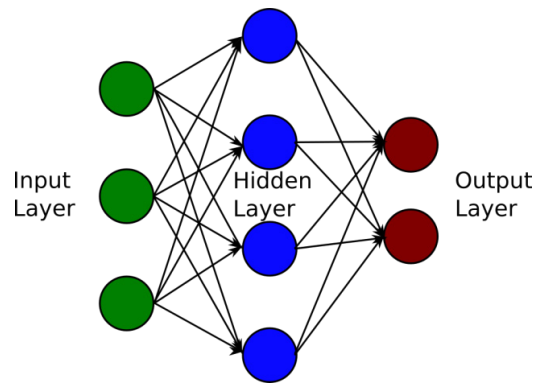
La inteligencia artificial es la simulación de procesos de inteligencia humana por parte de máquinas, esencialmente sistemas informáticos. Estos procesos incluyen el aprendizaje (adquisición de información y reglas para el uso de la misma), el razonamiento (usando las reglas para llegar a conclusiones aproximadas o definitivas) y la autocorrección.

Las aplicaciones de la inteligencia artificial incluyen asistencia sanitaria para hacer diagnósticos mejores y más rápidos, a través de aplicar el aprendizaje de la máquina con datos de pacientes y otras fuentes de datos disponibles.

### 4.2 Redes neuronales artificiales.

Las redes neuronales artificiales (RNA) son modelos computacionales basados en un conjunto de unidades neuronales simples (neuronas artificiales). Cada unidad está conectada con muchas otras y los enlaces entre ellas pueden incrementar o inhibir el estado de actuación de las neuronas adyacentes. El objetivo de las RNA es resolver los problemas de la misma manera que un cerebro humano y lo que da mayor interés a las redes neuronales es la posibilidad de aprendizaje de una determinada tarea a resolver.

El esquema general de una red neuronal artificial muy simple se puede observar en la Figura 4.1. Las neuronas se suelen agrupar en unidades estructurales denominadas capas, y, finalmente, el conjunto de una o más capas constituye la red neuronal. En la RNA de la figura se puede distinguir tres tipos de capas donde se agrupan las neuronas:



**Figura 4.1** Esquema general de una RNA unidireccional de tres capas, de entrada, oculta y de salida.

- Capa de entrada (Input Layer): capa en la que se introducen las entradas del problema a tratar.
- Capas ocultas (Hidden Layer): neuronas ocultas cuya función es “acomodarse” para encontrar la relación entre las entradas y las salidas.
- Capa de salida (Output Layer): estas neuronas darán la respuesta del modelo en cada simulación.

La capa de entrada está compuesta por neuronas que reciben datos o señales procedentes del entorno, mientras que la capa de salida está formada por aquellas neuronas que proporcionan la respuesta de la red neuronal. Una capa oculta es aquella que no tiene conexión directa con el entorno. Este tipo de capa proporciona a la red grados de libertad adicionales, gracias a los cuales puede encontrar representaciones internas correspondientes a determinados rasgos del entorno, proporcionando una mayor riqueza computacional [26].

Las conexiones entre las neuronas pueden ser excitatorias o inhibitorias, según el signo del peso sináptico asociado a la conexión, y por medio del aprendizaje se obtiene un valor para el peso que incluye signo y magnitud.

Para que cierto problema sea susceptible a ser resuelto mediante una red neuronal debe cumplir las siguientes características:

- No se dispone de formalismo ni reglas sistemáticas.
- Se dispone de muchos casos históricos.
- Datos imprecisos o con ruido.
- Problemas no lineales.
- Problemas con elevada dimensionalidad.

### 4.2.1 Modos de operación.

Se parte de una arquitectura de red, neuronas con cierto patrón de conexiones. Al inicio, la red no tiene información útil, es decir, los parámetros que utiliza la red (pesos sinápticos) son aleatorios o nulos, y esta red debe ser entrenada mediante ejemplos de salida-entrada de la tarea que debe realizar [25]. El entrenamiento consiste en el ajuste iterativo de los pesos de la red presentando una serie de ejemplos. Por tanto, se puede decir que hay dos modos básicos de funcionamiento de la red, el modo de entrenamiento o aprendizaje, y el modo de recuerdo o modo de ejecución normal.

Puede definirse como aprendizaje, el proceso por el que se produce el ajuste de los parámetros libres de la red a partir de un proceso de estimulación por el entorno que rodea la red. El proceso de aprendizaje suele ser iterativo, actualizándose los pesos una y otra vez hasta que la red alcanza el rendimiento deseado. Los diferentes tipos de aprendizaje en las RNA son los siguientes:

- Supervisado: los ejemplos consisten en datos de entrada y datos de salida deseados.
- No supervisado: solo se proporcionan los datos de entrada.
- Híbrido: incluye una capa no supervisada y una capa supervisada.
- Reforzado: en el que se impone un ‘castigo’ o un ‘premio’ a los parámetros que se ajustan en función de cómo de bien o de mal han realizado el entrenamiento.
- Arquitecturas evolutivas: se añaden o eliminan neuronas completas.

Generalmente una vez el sistema ha sido entrenado, el aprendizaje se “desconecta”, por lo que los pesos y la estructura quedan fijos para procesar datos. Este modo de operación se denomina modo recuerdo o de ejecución.

### 4.3 El perceptrón multicapa (MLP).

El perceptrón multicapa o red multicapa con conexiones hacia adelante es un tipo de red neuronal que puede ser una solución adecuada para tratar ciertos problemas no lineales. Es una de las arquitecturas de red neuronal más utilizada en la resolución de problemas debido a su capacidad como aproximador universal [26], así como a su fácil uso y aplicabilidad [27].

El perceptrón posee una serie de limitaciones, como el largo proceso de aprendizaje para problemas complejos dependientes de un gran número de variables, la dificultad para realizar un análisis teórico de la red debido a la presencia de componentes no lineales y a la alta conectividad.

#### 4.3.1 Propagación de los patrones de entrada.

El perceptrón define una relación entre los patrones de entrada y las variables de salida de la red. Esta relación se obtiene propagando hacia adelante los valores de la entrada, procesando en cada neurona la información recibida y produciendo una respuesta o activación que se propaga [28].

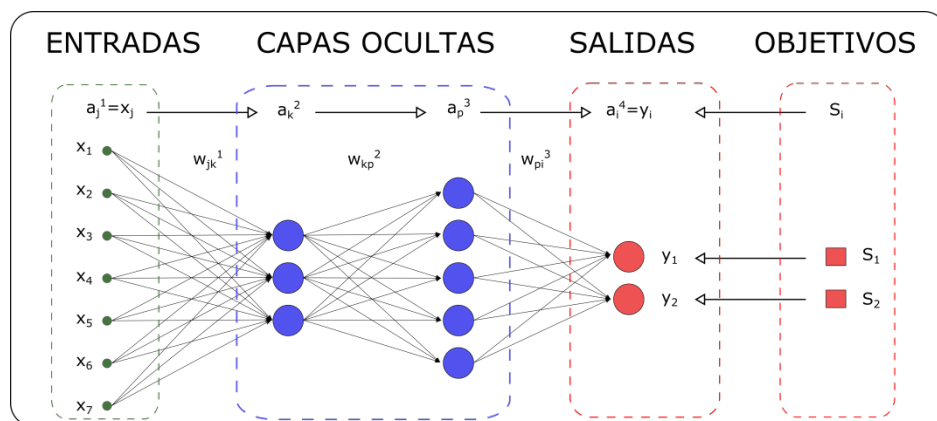


Figura 4.2 Ejemplo de un perceptrón multicapa con 7 entradas, 2 capas ocultas de neuronas y 2 salidas.

Sea un perceptrón multicapa con  $K$  capas,  $K-2$  capas ocultas, y  $n_k$  número de neuronas en la capa  $k$ , para  $k = 1, 2, \dots, C$ . Sea  $W^k = (\omega_{ji}^k)$  la matriz de pesos donde  $\omega_{ji}^k$  representa el peso de la conexión de la neurona  $i$  de la capa  $k$  para  $k = 2, \dots, K$ . Denotaremos  $a_i^k$  y  $u_i^k$  a la función y al umbral de activación respectivamente de la neurona  $i$  de la capa  $k$ .

Las expresiones para el cálculo de las activaciones de cada neurona son las siguientes:

- Activación de las neuronas de la capa de entrada:

$$a_i^1 = x_i \tag{1}$$

- Activación de las neuronas de las capas ocultas:

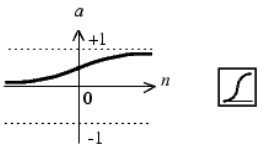
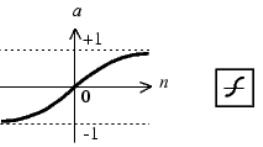
$$a_i^k = f \left( u_i^k + \sum_{j=1}^{n_{k-1}} a_j^{k-1} \cdot \omega_{ji}^{k-1} \right) \tag{2}$$

- Activación de las neuronas de la capa de salida:

$$y_i = a_i^K = f \left( u_i^K + \sum_{j=1}^{n_{K-1}} a_j^{K-1} \cdot \omega_{ji}^{K-1} \right) \tag{3}$$

Donde la función  $f$  es la función de activación, que para el perceptrón multicapa puede utilizarse la función sigmoide y la función tangente:

**Tabla 1 Funciones de activación para el perceptrón multicapa.**

Función de activación	Expresión analítica	Rango de las ordenadas	Representación gráfica
Función sigmoide	$f(x) = \frac{1}{1 + e^{-x}}$	[0, 1]	
Función tangente	$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$	[-1, 1]	

Generalmente la elección de una función de activación u otra no influye en la capacidad de la red para resolver el problema, se suele hacer la elección basándose en el recorrido deseado. En el caso de la red neuronal utilizada se ha escogido la función sigmoide para su diseño.

### 4.3.2 Algoritmo de retropropagación (Backpropagation).

El perceptrón multicapa es un algoritmo de aprendizaje supervisado, es decir, la modificación de los parámetros se realiza para que la salida de la red sea lo más próxima posible a la salida deseada proporcionada por el supervisor.

El algoritmo de retropropagación tiene como objetivo la minimización del error que se produce entre la salida obtenida por la red y la salida deseada, modificando el valor de los pesos sinápticos que unen las neuronas de una capa con la siguiente.



El método de optimización más empleado es el denominado descenso por el gradiente. Para ello se define una función coste que proporcione el error actual  $E$  que comete la red neuronal en función de los pesos sinápticos:  $e(w)$  [26].

El error se calcula como la suma del error cuadrático de las salidas de la red con respecto a las salidas deseadas:

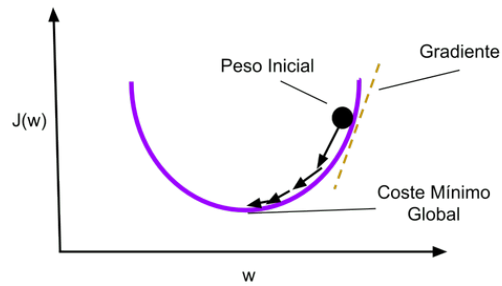


Figura 4.3 Espacio de los pesos  $w$  y el descenso por el gradiente hacia un mínimo.

$$E(n) = \frac{1}{n_K} \cdot \sum_{i=1}^{n_K} (s_i(n) - y_i(n))^2 \tag{4}$$

Siendo  $Y(n) = (y_1(n), y_2(n), \dots, y_{n_K}(n))$  y  $S(n) = (s_1(n), s_2(n), \dots, s_{n_K}(n))$  los vectores de salidas de la red y salidas deseadas para el patrón  $n$ , respectivamente.

Utilizando el método del descenso del gradiente estocástico, cada parámetro de la red  $\omega$  y  $u$  se modifica para cada patrón  $n$  de acuerdo con la ley de aprendizaje:

$$\omega_{ji}^k = \omega_{ji}^k - \alpha \cdot \frac{\partial e}{\partial \omega_{ji}^k} \tag{5}$$

$$u_i^k = u_i^k - \alpha \cdot \frac{\partial e}{\partial u_i^k} \tag{6}$$

Siendo  $\alpha$  la razón o tasa de aprendizaje, parámetro que influye en la magnitud del desplazamiento en la superficie del error.

La ventaja principal del algoritmo *BackPropagation* es que se puede aplicar a multitud de problemas diferentes, proporcionando buenas soluciones sin requerir un conocimiento matemático de la función que relaciona los patrones de entrada y salida.

La principal desventaja es la lentitud de convergencia además de la posibilidad de converger en un mínimo local de la función de error (ver Figura 4.4), por lo que no se obtendrían los valores de pesos sinápticos óptimos.

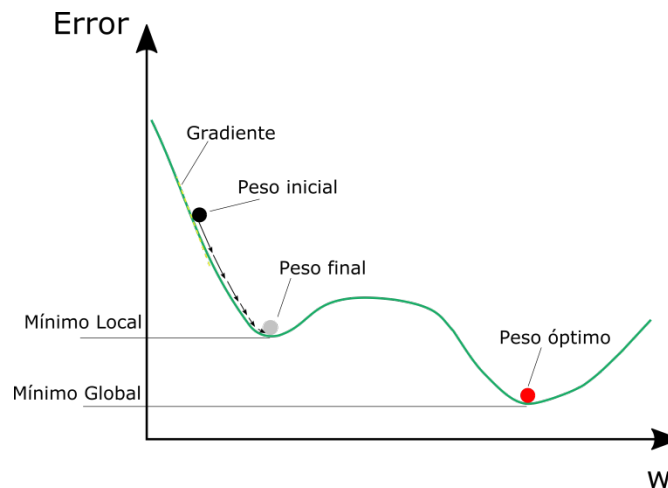


Figura 4.4 Representación de un mínimo local y el mínimo global en función de los pesos sinápticos.

#### 4.4 Máquinas de soporte vectorial (SVM).

Las máquinas de soporte vectorial o máquinas de vectores soporte (SVM, del inglés Support Vector Machines) son un conjunto de algoritmos de aprendizaje supervisado, que, a partir de un conjunto de ejemplos de entrenamiento, puede construir un modelo que prediga la clase de un nuevo conjunto de datos.

Su funcionamiento se basa en encontrar el mejor hiperplano que separa todos los puntos de una clase con los de otra, maximizando el margen entre las dos clases. Los vectores soporte son los puntos que marcan la separación del hiperplano. Ver Figura 4.5.

Las SVM pertenecen a la categoría de los clasificadores lineales, puesto que inducen separadores lineales o hiperplanos, ya sea en el espacio original de los ejemplos de entrada, si éstos son separables, o en el espacio transformado, si los ejemplos no son linealmente separables en el espacio original (ver Figura 4.6).

La búsqueda del hiperplano de separación en estos espacios transformados se hará de forma implícita utilizando las funciones *kernel* [29].

Las fortalezas que caracterizan las SVM son que el entrenamiento es relativamente fácil, no hay óptimo local como en las redes neuronales y se pueden usar datos no tradicionales como cadenas de caracteres como entrada. Por su parte, la debilidad es que se necesita una buena función *kernel*, es decir, se necesitan metodologías eficientes para sintonizar los parámetros de inicialización de la SVM [30].

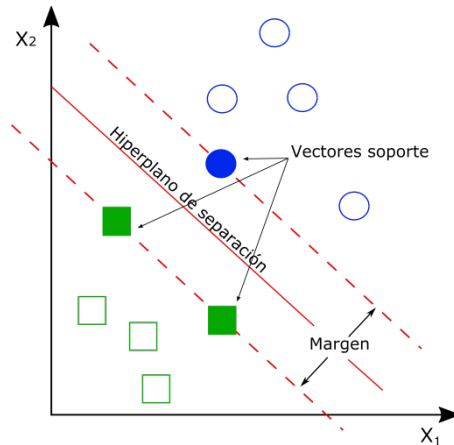


Figura 4.5 Representación de un problema de clasificación 2D resuelto mediante SVM

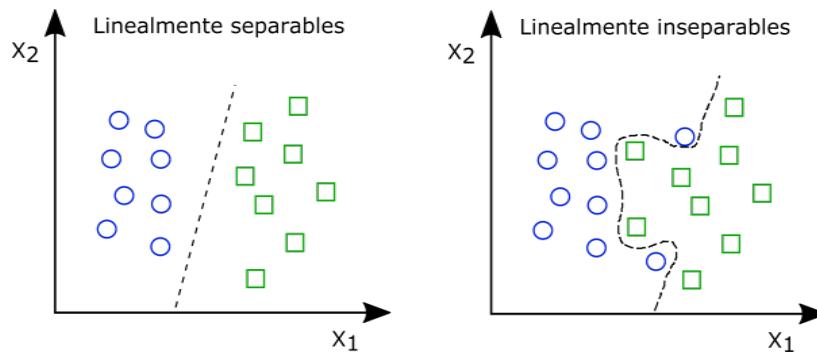


Figura 4.6 Tipos de problemas de clasificación 2D que podemos resolver con SVM; la imagen la izquierda se puede observar un problema de clasificación linealmente separable y en la imagen de la derecha un problema de clasificación inseparable linealmente.

Se va a explicar cómo se resolvería los dos tipos de problemas de clasificación binaria utilizando las SVM [31].

##### 4.4.1 SVM para clasificación binaria de ejemplos separables linealmente.

Dado un conjunto  $S$  de puntos etiquetados para entrenamiento:

$$(y_1, x_1), \dots, (y_l, x_l) \quad (7)$$

Y cada punto de entrenamiento  $x_i \in \mathbb{R}^N$  perteneciente a alguna de dos clases  $y_i \in \{-1, 1\}$  para  $i = 1, \dots, l$ .

En la mayoría de los casos, la búsqueda del hiperplano adecuado en un espacio de entrada es demasiado restrictivo por lo que se mapea el espacio de entrada en un espacio de características de una dimensión mayor y se busca el hiperplano óptimo allí.

Sea  $z = \varphi(x)$  el vector en el espacio de características con un mapeo  $\varphi$  de  $\mathfrak{R}^N$  a un espacio de características  $Z$ , se desea encontrar el hiperplano:

$$w \cdot z + b = 0 \quad (8)$$

Definido por  $w$  y  $b$ , tal que podamos separar el punto  $x_i$  de acuerdo a la función:

$$f(x_i) = \text{sign}(w \cdot z_i + b) = \begin{cases} 1 & y_i = 1 \\ -1 & y_i = -1 \end{cases} \quad (9)$$

Donde  $w \in Z$  y  $b \in \mathfrak{R}$ .

Se resume diciendo que el conjunto  $S$  es linealmente separable si existe  $(w, b)$  que cumple las siguientes inecuaciones:

$$\begin{cases} (w \cdot z_i + b) \geq 1, & y_i = 1 \\ (w \cdot z_i + b) \leq -1, & y_i = -1 \end{cases} \quad i = 1, \dots, l \quad (10)$$

Podemos encontrar así un único hiperplano óptimo, para el cual, el margen entre las proyecciones de los puntos de entrenamiento es máximo.

#### 4.4.2 SVM para clasificación binaria de ejemplos no linealmente separables.

Si el conjunto  $S$  no es linealmente separable, se debe admitir errores en la clasificación en la formulación de las SVM, generalizando el caso anterior introduciendo algunas variables  $\xi_i \geq 0$  de tal modo que se modifica la ecuación (10) de la siguiente manera:

$$y_i(w \cdot z_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l \quad (11)$$

Los  $\xi_i \neq 0$  en la ecuación (11) son aquellos para los cuales el punto  $x_i$  no satisface ecuación (10). Entonces el término  $\sum_{i=1}^l \xi_i$  puede servir como medida del error en la clasificación.

El problema del hiperplano óptimo es entonces redefinido como la solución al problema:

$$\begin{aligned} \min & \left\{ \frac{1}{2} w \cdot w + C \sum_{i=1}^l \xi_i \right\} \\ & y_i(w \cdot z_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l \\ & \xi_i \geq 0, \quad i = 1, \dots, l \end{aligned} \quad (12)$$

Donde  $C$  es una constante definida como un parámetro de regularización. Éste es el único parámetro libre de ser ajustado en la formulación de la SVM. El ajuste de éste parámetro puede hacer un balance entre la maximización del margen y el error en la clasificación.

El problema de búsqueda del hiperplano óptimo es un problema de programación cuadrática que puede ser resuelto construyendo un Lagrangiano y transformándolo en dual:

$$\begin{aligned} \text{Max } W(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \cdot \alpha_j \cdot y_i \cdot y_j \cdot z_i \cdot z_j \\ & \sum_{i=1}^l y_i \cdot \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l \end{aligned} \quad (13)$$

Donde  $\alpha = (\alpha_1, \dots, \alpha_l)$  es un vector de multiplicadores de Lagrange positivos asociados con las constantes en la ecuación (11).

Según el teorema de Kuhn-Tucker, la solución  $\bar{\alpha}_i$  de la ecuación (12), satisface:

$$\bar{\alpha}_i(y_i(\bar{w} \cdot z_i + \bar{b}) - 1 + \bar{\xi}_i) = 0, \quad i = 1, \dots, l \tag{14}$$

$$(C - \bar{\alpha}_i)\bar{\xi}_i = 0, \quad i = 1, \dots, l \tag{15}$$

El punto  $x_i$  correspondiente con  $\bar{\alpha}_i > 0$  es llamado *vector de soporte*. Pero hay dos tipos de vectores soporte en un caso no separable. En el caso de  $0 < \bar{\alpha}_i < C$ , el correspondiente vector de soporte  $x_i$  satisface las igualdades  $y_i(\bar{w} \cdot z_i + \bar{b}) = 1$  y  $\bar{\xi}_i = 0$ . En el caso de  $\bar{\alpha}_i = C$ , el correspondiente  $\bar{\xi}_i$  es diferente de cero y el correspondiente vector soporte  $x_i$  no satisface la ecuación (10), estos vectores soporte son los llamados errores.

Para construir el hiperplano óptimo  $\bar{w} \cdot z + \bar{b}$ , se utiliza:

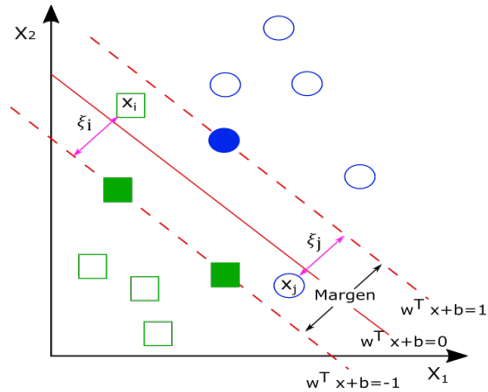


Figura 4.7 Aparición del parámetro  $\xi_i$  en el error de clasificación [15]

$$\bar{w} = \sum_{i=1}^l \bar{\alpha}_i y_i z_i \tag{16}$$

Y el escalar  $b$  puede ser determinado de las condiciones de Kuhn-Tucker (15).

La función de decisión generalizada de (9) y (16) es tal que:

$$f(x) = \text{sign}(w \cdot z + b) = \text{sign}\left(\sum_{i=1}^l \alpha_i y_i z_i \cdot z + b\right) \tag{17}$$

#### 4.4.3 Truco del Kernel para el caso no linealmente separable.

Como no tenemos ningún conocimiento de  $\varphi$ , el cálculo de las ecuaciones (13) y (17) es imposible. Para poder resolverlo se necesita la función *Kernel*  $K(\cdot, \cdot)$  que calcule el producto de los puntos de entrada en el espacio de características  $Z$ ;

$$z_i \cdot z_j = \varphi(x_i) \cdot \varphi(x_j) = K(x_i, x_j) \tag{18}$$

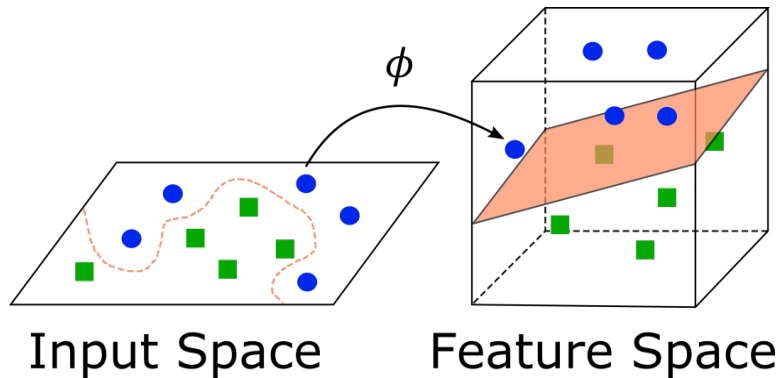


Figura 4.8 Idea del uso de un kernel para la transformación del espacio de los datos [30]

Se puede utilizar el kernel polinomial de grado  $d$  para construir un clasificador SVM:

$$K(x_i, x_j) = (1 + x_i \cdot x_j)^d \quad (19)$$

Entonces el hiperplano no lineal de separación puede ser la solución de:

$$\begin{aligned} \text{Max } W(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \sum_{i=1}^l y_i \alpha_i &= 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l \end{aligned} \quad (20)$$

Y la función de decisión es:

$$f(x) = \text{sign}(w \cdot z + b) = \text{sign} \left( \sum_{i=1}^l \alpha_i y_i K(x_i, x_j) + b \right) \quad (21)$$

## 5 MATERIAL Y MÉTODO.

### 5.1 Sujetos y protocolo de medida.

Se parte de una base de datos proporcionada por el Hospital Universitario Miguel Servet de Zaragoza en el que se incluyen una serie de resultados de diferentes pruebas médicas relacionadas con el diagnóstico de la esclerosis múltiple:

- Tomografía de coherencia óptica (OCT): procedimiento no invasivo que permite estudiar el espesor de la retina con precisión.
- Polarización láser: es un procedimiento parecido al OCT, pero con menor resolución y más específico del glaucoma.
- Potenciales evocados visuales (PEV): mide los impulsos nerviosos que viajan a través del nervio óptico y son provocados por un estímulo visual, y permiten estudiar la recepción cerebral de estímulos.
- Electroretinograma-pattern (PERG): mide la funcionalidad de las capas superficiales de la retina (células ganglionares) a través del estudio de los impulsos nerviosos provocados por un estímulo conocido como pattern o damero.

Se dispone de datos de un total de 180 pacientes que tienen diagnosticada esclerosis múltiple con un seguimiento de varias visitas a lo largo de unos 6 años (2006 – 2012) y datos de 85 pacientes sanos (controles) con mucha menor frecuencia de obtención de datos. Los pacientes a los que corresponden los datos son individuos blancos europeos.

Hay que destacar que la base de datos no es completa, es decir, no se dispone del mismo número de visitas para cada paciente. Tampoco todos los pacientes han completado todas las pruebas médicas comentadas anteriormente.

Tras hacer una búsqueda de bibliografía relacionada con el tema [3,15,32–34], se decide que los datos de espesor de la retina obtenidos mediante la técnica OCT (ver Figura 5.2) van a ser los mejores para realizar un diagnóstico preciso de la enfermedad.

La base de datos proporcionada por el Hospital, dispone de 5 grupos distintos de datos para los pacientes de la prueba médica denominada OCT. El primero de ellos corresponde a la prueba realizada con el equipo CIRRUS HD-OCT, mientras que los otros 4 grupos de datos se obtienen con un equipo OCT-Spectralis. Más adelante se explicará las diferencias entre los cinco grupos de datos.



**Figura 5.1** A la izquierda, imagen del equipo de OCT-Spectralis, a la derecha imagen del equipo OCT-CIRRUS en el Hospital Universitario Miguel Servet de Zaragoza.

### 5.1.1 Tomografía de coherencia óptica.

La tomografía de coherencia óptica (OCT) es una técnica de imagen no invasiva que nos permite visualizar in-vivo las diferentes capas de la retina. Entre ellas se encuentra la capa de fibras nerviosas de la retina (RNFL).

La técnica de la OCT se basa en el principio óptico de la interferometría, que consiste en combinar las ondas de luz provenientes de diferentes receptores o superficies, en este caso de tejidos, para generar una imagen de gran resolución.

An HD-OCT scan of a healthy eye

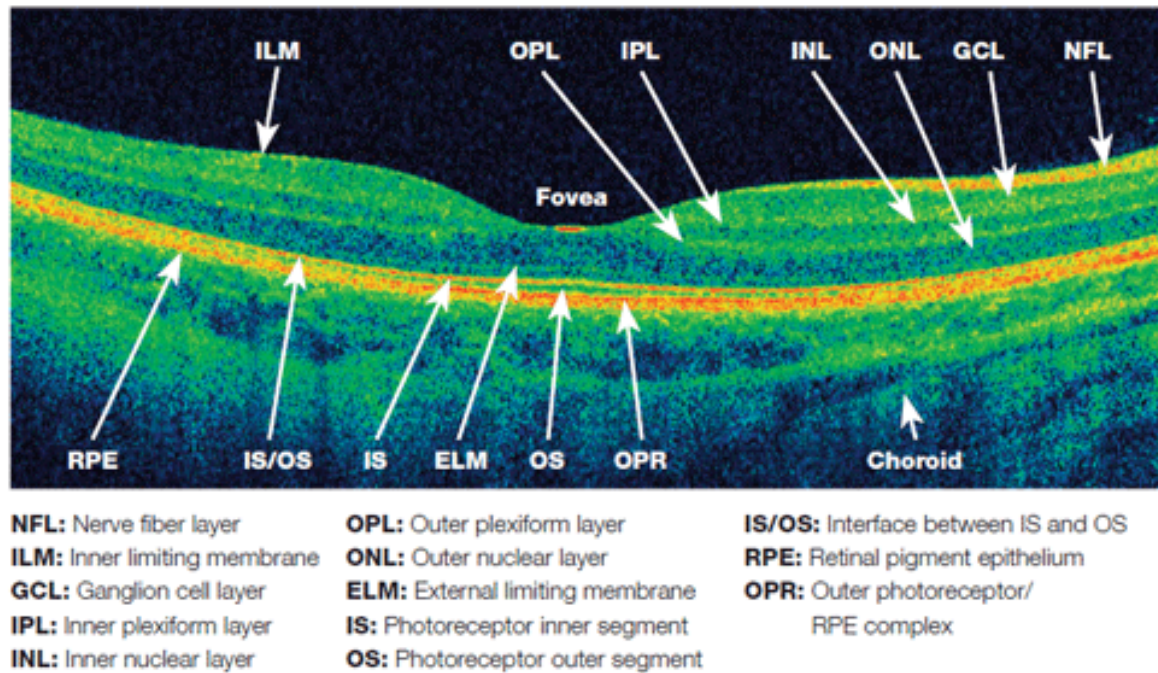


Figura 5.2 Sección de la retina tomada mediante una tomografía de coherencia óptica.

Se dispone de dos equipos para la obtención de los datos OCT;

- El primero de la compañía *Carl Zeiss Meditec AG*, concretamente un *CIRRUS HD-OCT* que proporciona datos del espesor medio de la RNFL subdividiendo la zona peripapilar de la retina en 12 cuadrantes de 30° formando una circunferencia completa de 360° como vemos en la Figura 5.3.
- El segundo de la compañía *Heidelberg-Engineering*, concretamente un *OCT-Spectralis* genera un mapa con la medida del espesor de la RNFL en seis sectores en la zona peripapilar de la retina (superonasal, nasal, inferonasal, inferotemporal, temporal, superotemporal en el sentido de las agujas del reloj para el ojo derecho, y contrario a las agujas del reloj en el ojo izquierdo) que podemos observar en la Figura 5.11. El Spectralis nos proporciona además la medida del espesor de la RNFL en 768 puntos formando un círculo peripapilar.

### 5.2 Tratamiento de datos.

En primer lugar, para cada uno de los grupos de los que se dispone se debe realizar una selección de datos, eliminando los que están incompletos y ordenando cada paciente en orden cronológico según la fecha de la visita.

Para realizar el diagnóstico de la enfermedad se va a utilizar únicamente los datos de la primera visita de cada paciente, ya que en la primera consulta será cuando se debe diagnosticar si dicho individuo padece la enfermedad o por el contrario está sano.

Para seleccionar los datos que se van a utilizar en la predicción de la enfermedad se debe realizar un análisis de la importancia de los parámetros y así poder utilizar los que mayor relevancia tengan. Éstos se utilizarán como entradas en el desarrollo de la estrategia de predicción.

### 5.2.1 Selección de variables.

Hemos seleccionado la OCT como prueba médica para el diagnóstico, por tanto, los datos que proporciona esta prueba los introduciremos como entrada para la red neuronal diseñada.

Como disponemos de varias series de datos según procedan de un equipo u otro, o según el protocolo seguido en la adquisición de los datos, se va a seleccionar diferentes entradas de datos con el objetivo de realizar distintos entrenamientos de redes neuronales y comparar así los resultados obtenidos para cada predicción.

#### 5.2.1.1 CIRRUS HD-OCT.

En la base de datos se dispone de los resultados que proporciona la OCT-CIRRUS, entre los cuales se escogen como más representativos los datos del espesor medio de la capa de fibras nerviosas de la retina (RNFL) para doce sectores correspondientes a la zona peripapilar alrededor del nervio óptico. También utilizamos el espesor medio de la zona peripapilar completa y de la zona de la mácula.

Además de los espesores de la RNFL se cree que la edad del paciente y el sexo del mismo pueden ser determinantes en la predicción.

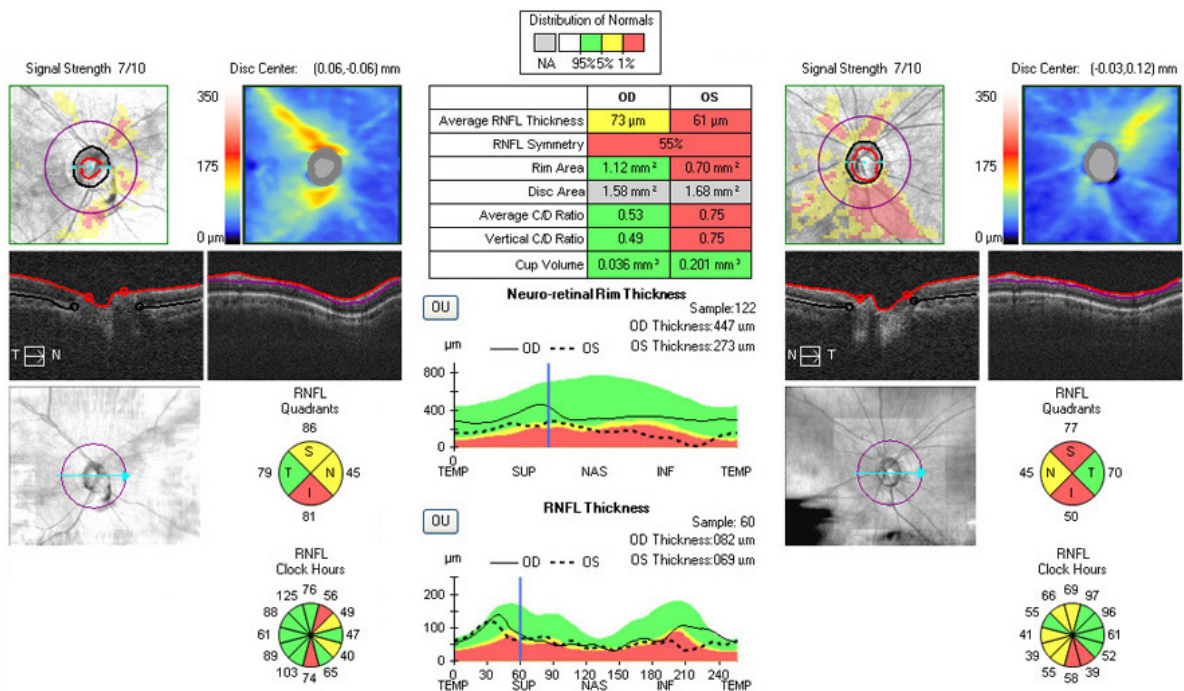


Figura 5.3 Ejemplo de datos que proporciona la OCT-CIRRUS

En la Figura 5.3 puede observarse un ejemplo de los datos que muestra una prueba OCT realizada con el equipo CIRRUS HD-OCT para un paciente. En la imagen podemos ver en la parte central superior un resumen, tanto para el ojo derecho (OD) como para el ojo izquierdo (OS), de los principales parámetros de espesor, área y volumen en la zona peripapilar de la retina. También observamos dos gráficas que representan el espesor del borde del disco óptico (gráfica superior) y de la capa de las fibras nerviosas de la retina (gráfica inferior).

Para cada ojo nos muestra una representación de la zona peripapilar dividida en cuatro cuadrantes, superior (S), nasal (N), inferior (I) y temporal (T), para cada ojo y dividida en 12



sectores, de C1 hasta C12, correspondientes a las horas del reloj, en los cuales muestran el espesor medio en cada uno de los cuadrantes de la RNFL.

Por último, para la selección de las variables de entrada se realiza un análisis de sensibilidad de éstas mediante una regresión lineal y ayudado del software Weka [35], obteniendo la importancia de las zonas más representativas de la adquisición peripapilar mediante el equipo CIRRUS OCT:

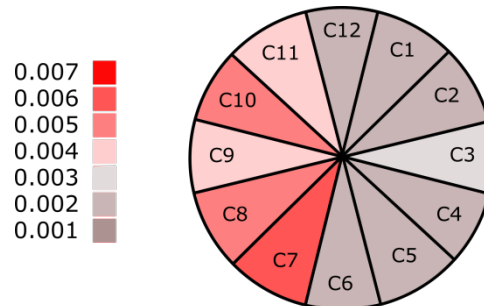


Figura 5.4 Importancia para los sectores horarios de la zona peripapilar

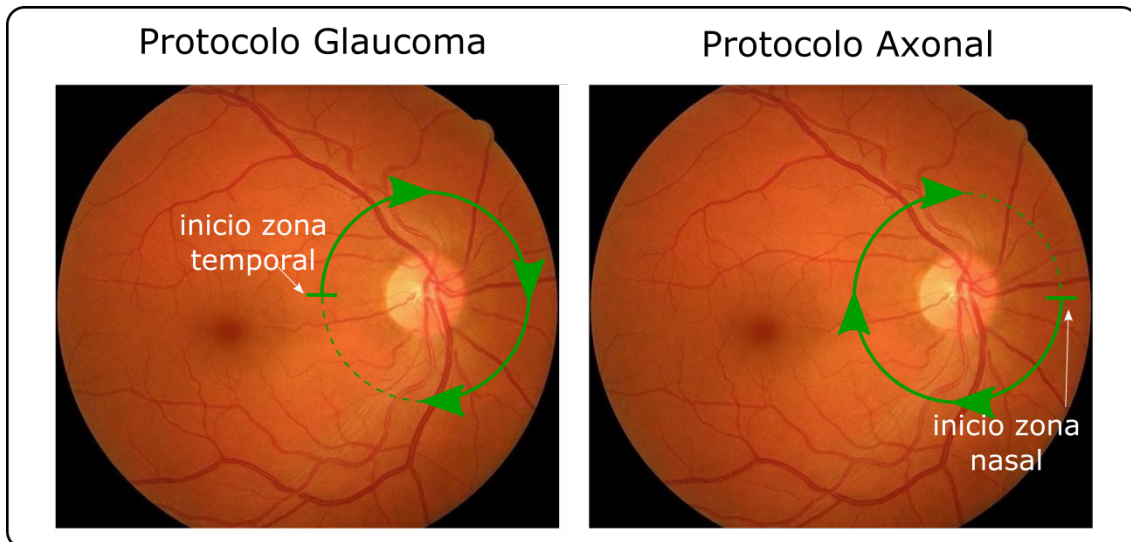
En este caso se puede apreciar que las zonas más representativas en la predicción de la enfermedad son las zonas horarias C7, C8 y C10, correspondiendo las dos primeras a la zona temporal inferior de la zona peripapilar y la tercera a la zona temporal superior.

### 5.2.1.2 OCT-Spectralis.

El equipo de OCT de la compañía *Heidelberg-Engineering*, ofrece una amplia versatilidad de exploraciones, incluyendo varios protocolos de adquisición y de análisis de los cuales, varios están diseñados específicamente para la detección y seguimiento del daño glaucomatoso y que se centran en el análisis del nervio óptico y de la capa de las fibras nerviosas de la retina.

En la actualidad, el software del equipo dispone de un protocolo diseñado específicamente para realizar un análisis axonal de las células ganglionares de la retina, llamado **patrón RNFL o protocolo axonal**, en el que la adquisición de los datos durante el barrido se hace partiendo y finalizando en el sector nasal de la cabeza del nervio óptico, de forma que en el sector temporal se mejoran la reproducibilidad y la fiabilidad de las mediciones. La importancia de esto se debe a que el sector temporal es el que presenta más alteraciones de forma más precoz en las enfermedades neurodegenerativas.

Los software previos, comenzaban a captar la imagen sobre la zona temporal del nervio óptico, recorriendo 360° y terminando de nuevo en la zona temporal. De esta manera la fiabilidad de los datos obtenidos era elevada en la zona nasal. Este **protocolo se le denomina adquisición glaucoma**.



**Figura 5.5** A la izquierda patrón de barrido realizado por el protocolo glaucoma (de zona temporal a zona temporal) en un ojo derecho, a la derecha patrón de adquisición realizado por el protocolo axonal (de zona nasal a zona nasal) en el mismo ojo derecho.

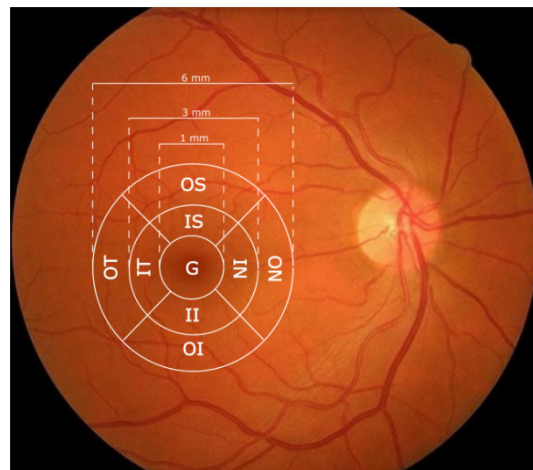
En la Figura 5.5 se puede observar la diferencia entre el protocolo de adquisición de glaucoma, que comienza el barrido de la cabeza del nervio óptico en la zona temporal recorriendo 360° hasta volver al punto de partida y el protocolo de adquisición axonal, donde al contrario que el de glaucoma, comienza el barrido en la zona nasal.

Distinguiendo entre estos dos protocolos de adquisición alrededor del nervio óptico, se puede separar los estudios clínicos de los pacientes, y se crea una base de datos para cada uno de los protocolos.

Además, el equipo OCT-Spectralis cuenta con un protocolo de barrido en el área macular realizando una adquisición de datos en forma circular centrada en la fovea (Ver Figura 5.6).

Esta adquisición se realiza tanto para la capa de las fibras nerviosas de la retina (RNFL) como en la capa de las células ganglionares, obteniendo así otras dos bases de datos, distinguiendo los datos obtenidos para la capa de células ganglionares y la RNFL.

Con todo esto, se van a conformar cuatro bases de datos a partir de las pruebas realizadas con el equipo OCT-Spectralis. La primera corresponde con la adquisición de datos en la capa de células ganglionares en el área macular, la segunda corresponde a la adquisición de datos en la RNFL en el área macular, la tercera al protocolo axonal de barrido del área peripapilar, y la cuarta y última a la adquisición mediante el protocolo glaucoma.

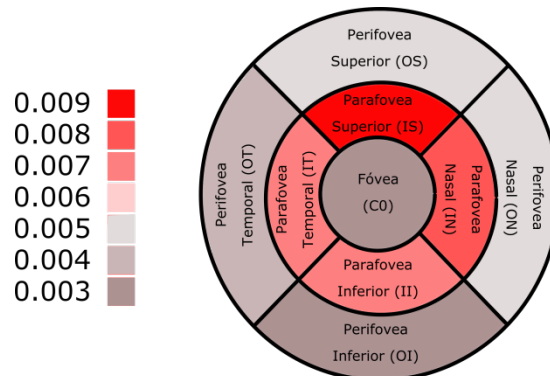


**Figura 5.6** Representación de la adquisición de datos alrededor de la zona macular.

De la misma forma que en el apartado anterior se realiza el análisis de sensibilidad de las entradas mediante una regresión lineal para cada uno de los 4 grupos de datos que se obtienen de éste equipo:

1. Datos del espesor medio de la capa de células ganglionares de la retina en los segmentos del área macular.

En este protocolo se obtienen datos de 9 sectores diferentes de la zona macular de la retina. Se estudiará la relación de cada uno de estos sectores con la enfermedad.

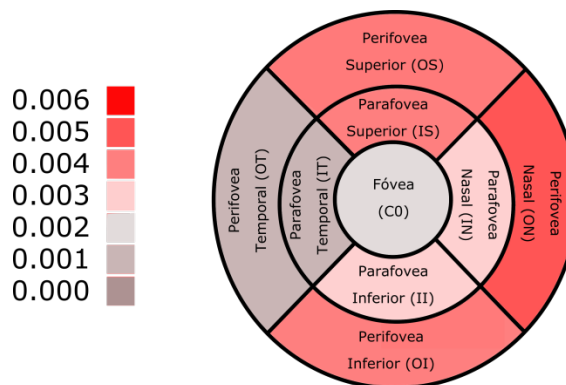


**Figura 5.7 Relevancia de los sectores de barrido de la zona macular en la capa de las células ganglionares**

Podemos observar que, en este caso, las zonas con mayor peso de los atributos la cogen las zonas de la parafovea (zona interior del barrido macular) cobrando especial importancia la zona de la parafovea superior.

2. Datos del espesor medio de la RNFL en los segmentos del área macular.

Al igual que el caso anterior, este protocolo diferencia en 9 sectores la zona macular de la retina, pero en este caso la medida corresponde a la capa de fibras nerviosas de la retina.



**Figura 5.8 Relevancia de los sectores de barrido de la zona macular en la capa de las células ganglionares**

En el caso de la zona macular para la capa de fibras nerviosas de la retina se observa que la zona perifóvea nasal es la que mayor peso tiene en la regresión lineal. Observando la Figura 5.6 se puede ver que esta zona nasal de la adquisición macular coincide con la zona temporal en la adquisición peripapilar, deduciendo la especial relevancia de la zona entre la papila y la mácula en la predicción de la esclerosis múltiple.

3. Datos del espesor medio de la RNFL en los segmentos de la zona peripapilar y 768 medidas de la circunferencia peripapilar mediante el protocolo axonal.

En este caso además de los sectores de la zona peripapilar, se van a incluir en la regresión lineal los 768 puntos de barrido peripapilar. Para representar la importancia de los puntos de barrido se decide agrupar éstos en franjas de 24 puntos.

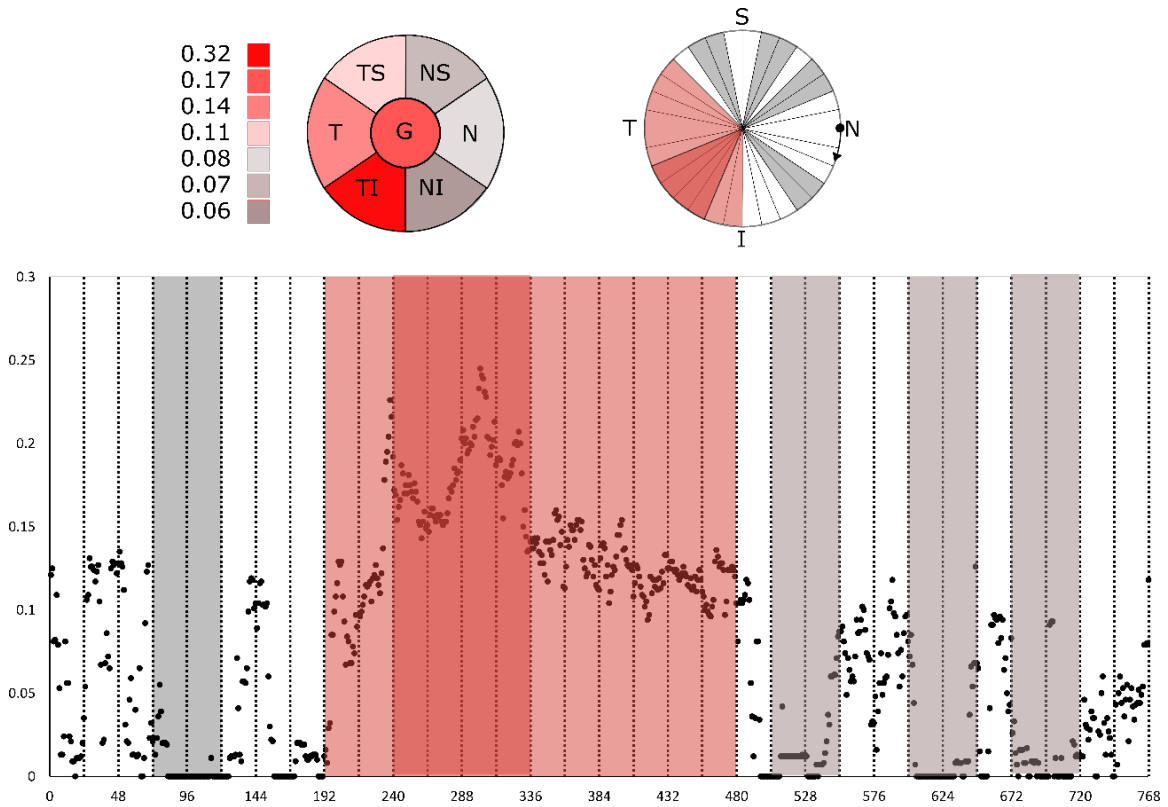


Figura 5.9 Importancia para el barrido peripapilar de la RNFL con el protocolo axonal.

En la Figura 5.9 se observa que el sector con mayor importancia es la zona temporal inferior. Esto además se ratifica con la representación de los puntos de barrido, cobrando mayor relevancia en la regresión los puntos 240-336, que coinciden con dicha zona.

4. Datos del espesor medio de la RNFL en los segmentos de la zona peripapilar y 768 medidas de la circunferencia peripapilar mediante el protocolo glaucoma.

Del mismo modo que en el protocolo anterior, se realiza una regresión lineal para todos los puntos de barrido y los sectores de adquisición. Los puntos de barrido comienzan en este caso en la zona temporal.

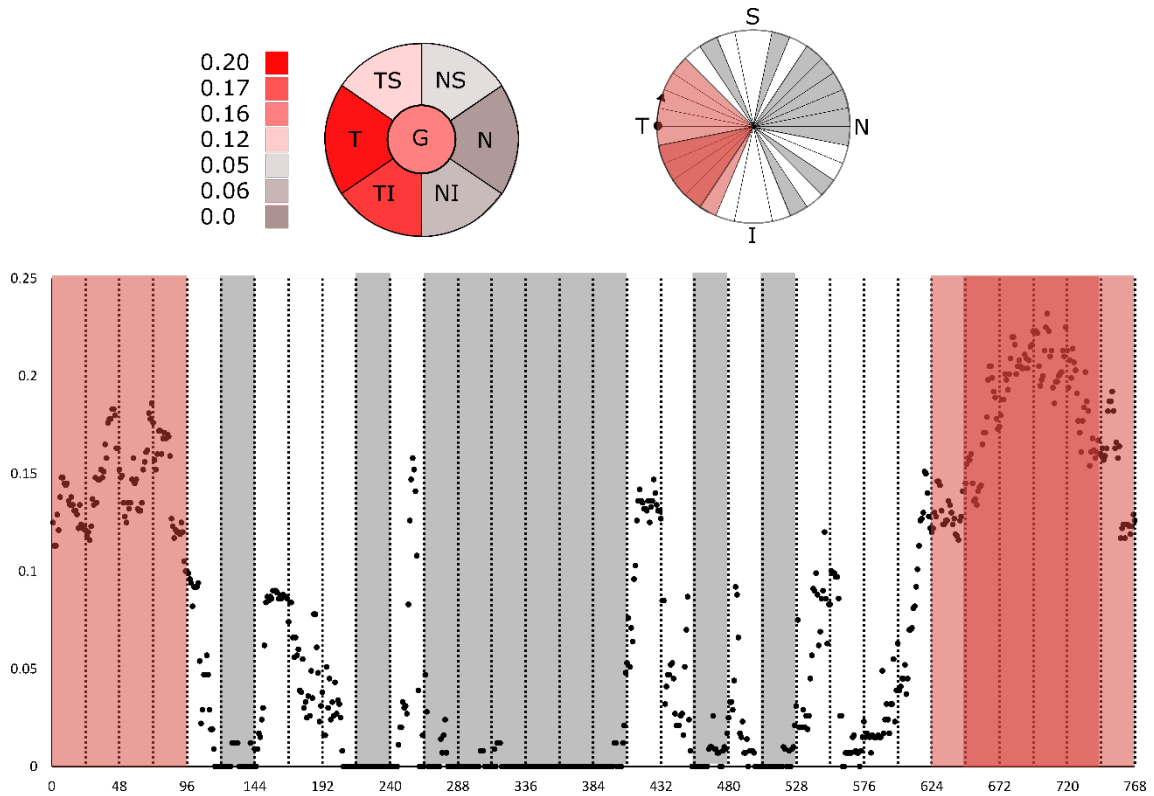


Figura 5.10 Importancia de los puntos para el barrido peripapilar de la RNFL con el protocolo glaucoma.

De nuevo, la zona con mayor peso en la regresión es la zona temporal. Los puntos de barrido con mayor importancia en esta ocasión van entre 656-752, que al ser el protocolo glaucoma y comenzar la adquisición en la zona temporal, de nuevo coinciden en la zona temporal inferior de la papila.

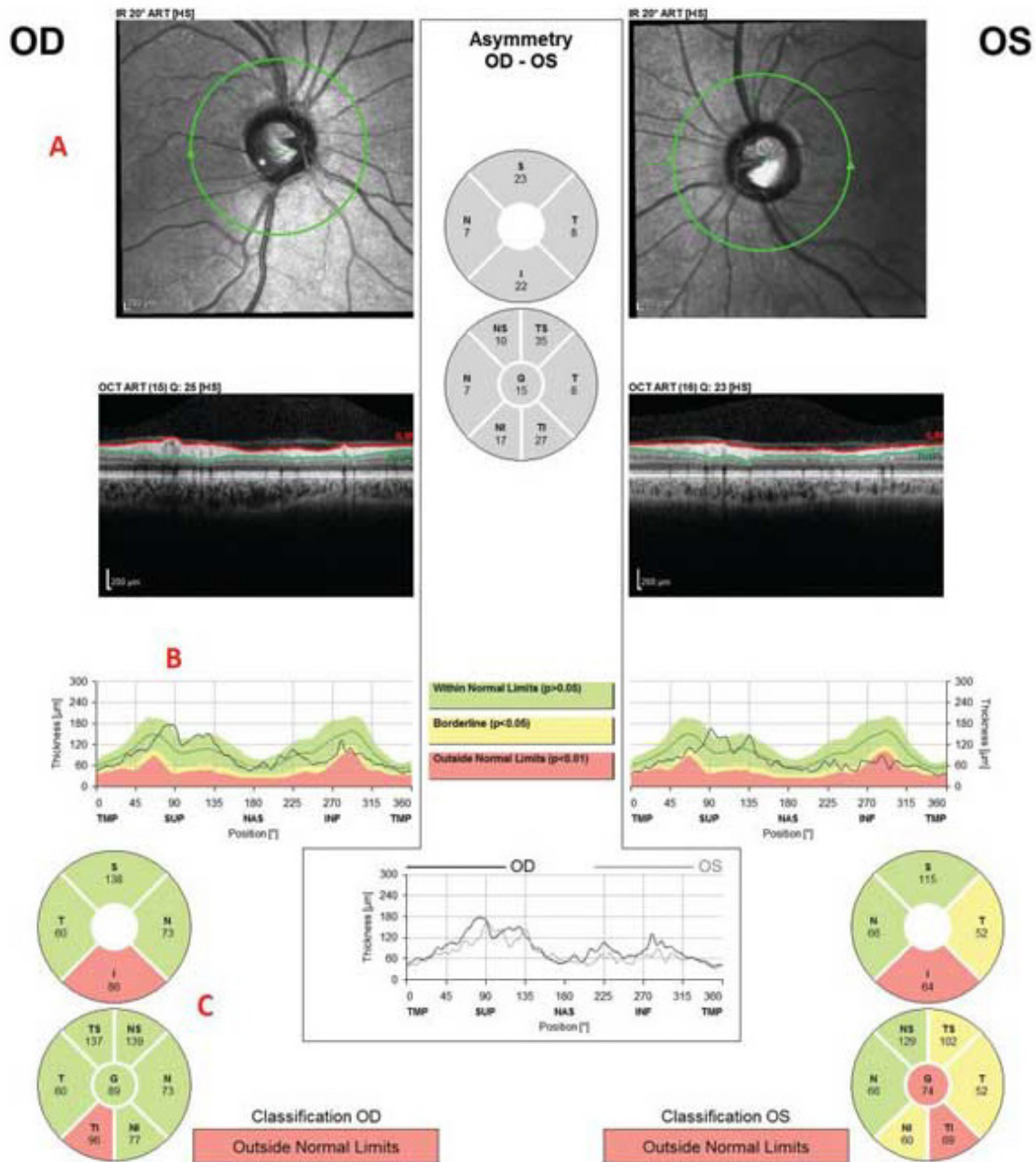


Figura 5.11 Ejemplo de datos que proporciona OCT-Spectralis

En la Figura 5.11 podemos observar un ejemplo del interfaz que ofrece el equipo durante la realización de la prueba a un paciente. Esta prueba concretamente corresponde al protocolo de adquisición glaucoma para ambos ojos.

En las imágenes superiores podemos observar el recorrido de adquisición en la cabeza del nervio óptico, comenzando en la zona temporal (flecha verde), y terminando igualmente en la zona temporal. Las imágenes inmediatamente inferiores corresponden al gráfico de datos del espesor de las capas de la retina en la zona de barrido.

Las gráficas que se encuentran debajo representan el espesor de la RNFL de la misma sección de barrido desarrollada y la comparo con los valores normales para sujetos sanos. Por último, aparece una representación de los sectores de la zona peripapilar, arriba dividida en 4, nasal (N),

inferior (I), temporal (T) y superior (S), y debajo dividida en 7, central (G), temporal-superior (TS), temporal (T), temporal-inferior (TI), nasal-inferior (NI), nasal (N), nasal-superior (NS).

### 5.3 Normalización de los datos.

La normalización de la base de datos ayuda a evitar efectos negativos como la ralentización de los procesos y la ineficiencia de las operaciones, minimizar la cantidad de datos duplicados y poder organizar los datos en grupos lógicos. En el caso de las redes neuronales es un paso obligatorio para el correcto funcionamiento del modelo.

La normalización empleada en el modelo es la normalización de una distribución caracterizada por los argumentos valor medio y desviación estándar. La ecuación del valor normalizado es la siguiente:

$$Z = \frac{X - \bar{X}}{\sigma} \quad (22)$$

Siendo:  $Z = \text{valor normalizado}$   
 $X = \text{valor sin normalizar}$   
 $\bar{X} = \text{valor medio}$   
 $\sigma = \text{desviación estándar}$

Así pues, tenemos listos los datos para introducir en los modelos creados de redes neuronales.

### 5.4 Análisis estadístico mediante el método de las redes neuronales.

Se va a evaluar la capacidad del Perceptrón multicapa y de las SVM como clasificadores de los datos obtenidos de la OCT para distinguir entre pacientes sanos y pacientes con EM.

Las RNAs son herramientas que se usan fundamentalmente para buscar patrones en conjuntos de datos de entrenamiento, y tras el proceso de entrenamiento aprender esos patrones y desarrollar la capacidad de clasificar nuevos patrones correctamente. Para el modelo empleado de perceptrón multicapa empleamos el algoritmo de *back-propagation* en el entrenamiento de la red.

Las SVM por su parte, están reconocidos como unos de los mejores clasificadores binarios porque siempre encuentran la mejor división posible entre las dos clases de datos, sin finalizar en mínimos locales como si puede suceder en el perceptrón multicapa [36].

#### 5.4.1 Validación cruzada.

La validación cruzada (en inglés, k-fold cross validation) es una técnica utilizada para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición entre los datos de entrenamiento y prueba. Consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones.

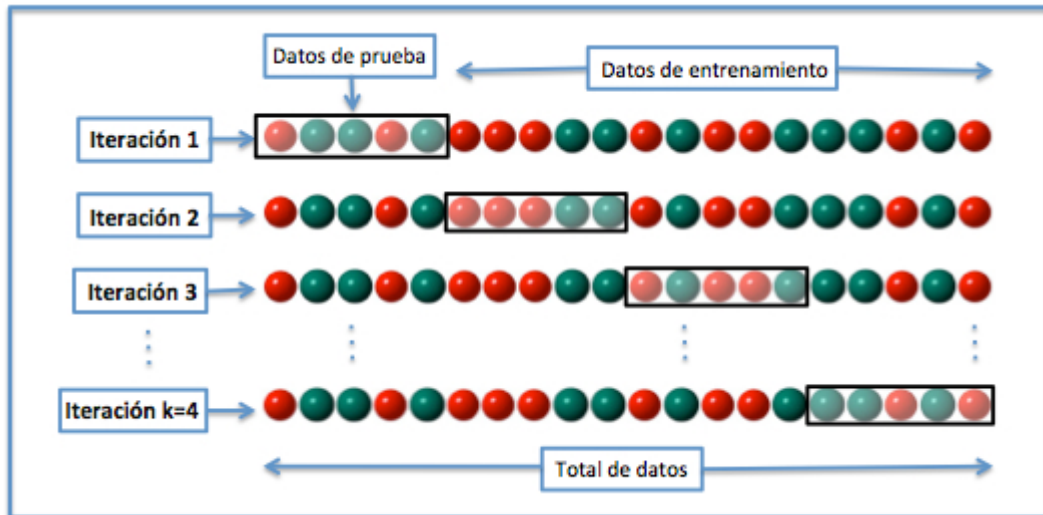


Figura 5.12 Ejemplo de k-fold cross validation con k=4, en el que se muestra como el grupo de datos de prueba y de datos de entrenamiento varía en cada una de las iteraciones.

En este caso se utiliza *5-fold cross validation*, que consiste en dividir la muestra inicial de datos en 5 subconjuntos. Uno de los conjuntos se emplea en la validación de la red neuronal y los otros cuatro en el entrenamiento de la misma. Esta simulación se repite 5 veces, variando el conjunto de validación, y obteniendo como resultado final la precisión media de las 5 iteraciones como resultado final para esa arquitectura de red concreta.

#### 5.4.2 Curvas ROC.

Una curva ROC es una representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario según se varía el umbral de discriminación. Otra interpretación de este gráfico es la representación de la razón o ratio de verdaderos positivos (VPR) frente a la razón de falsos positivos (FPR) también según se varía el umbral de discriminación (valor a partir del cual consideramos que un caso es un positivo). Esta curva ROC nos va a permitir seleccionar los modelos óptimos de diagnóstico [34].

$$VPR = \frac{VP}{VP + FN} = \text{sensibilidad o razón de VP} \quad (23)$$

$$FPR = \frac{FP}{FP + VN} = \text{razón de FP} \quad (24)$$

$$ACC = \frac{VP + VN}{P + N} = \text{exactitud} \quad (25)$$

$$SPC = \frac{VN}{N} = \frac{VN}{(FP + VN)} = 1 - FRP = \text{especificidad} \quad (26)$$

Siendo:  $VP = \text{verdaderos positivos}$

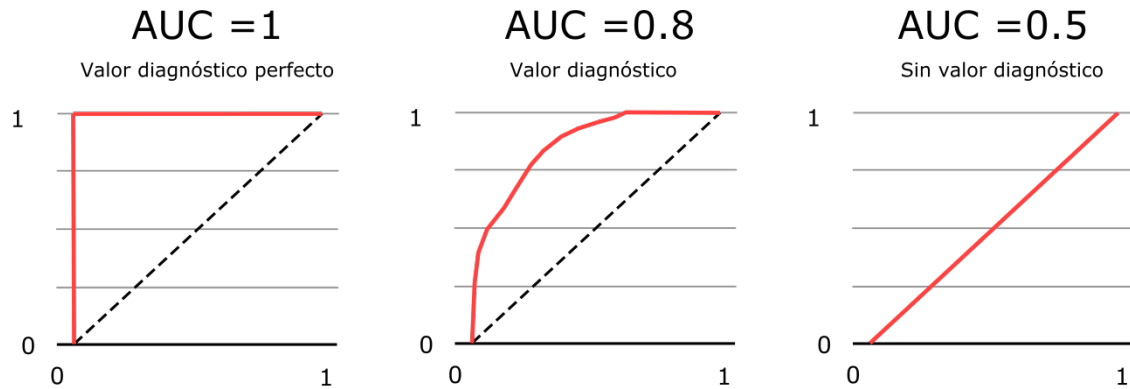
$VN = \text{verdaderos negativos}$



$FP = \text{falsos positivos}$

$FN = \text{falsos negativos}$

Para la elección entre dos pruebas de diagnóstico se realiza la comparación del área bajo la curva AUC (Area Under Curve) de ambas pruebas. Esta área tiene un valor entre 0.5 y 1, siendo 1 una prueba diagnóstica perfecta y 0.5 una muy mala. Así pues se elegirá la prueba diagnóstica que tenga un mayor valor de AUC (ver Figura 5.13).



**Figura 5.13** Ejemplo de aplicación de las curvas ROC para evaluar la capacidad de una prueba para el diagnóstico de cierta patología en función del ratio de verdaderos positivos y el ratio de falsos positivos VPR-FPR.

## 6 EXPERIMENTOS REALIZADOS Y RESULTADOS.

Como se ha comentado en el apartado 5.2.1, se van a realizar 5 modelos de predicción basados en el perceptrón multicapa y otros 5 modelos basados en las SVM, correspondiendo a cada una de las bases de datos en las que se ha dividido los resultados de las pruebas médicas de OCT:

- 1) Modelo 1: base de datos obtenida mediante el equipo CIRRUS HD-OCT.
- 2) Modelo 2: base de datos correspondiente a la capa de células ganglionares de la retina en la zona macular, obtenida con el equipo OCT-Spectralis.
- 3) Modelo 3: base de datos correspondiente a la RNFL en la zona macular, obtenida con el equipo OCT-Spectralis.
- 4) Modelo 4: base de datos correspondiente a la RNFL obtenida con el protocolo axonal del equipo OCT-Spectralis.
- 5) Modelo 5: base de datos que corresponde a la RNFL obtenida con el protocolo glaucoma del equipo OCT-Spectralis.

Se realiza una serie de pruebas con los modelos de predicción para conseguir el diagnóstico más fiable en cada uno de los dichos modelos y poder evaluar cuál es el protocolo que tiene mayor capacidad de diagnóstico de la esclerosis múltiple, así como poder comparar la capacidad como clasificadores del perceptrón multicapa y de las SVM en este problema concreto.

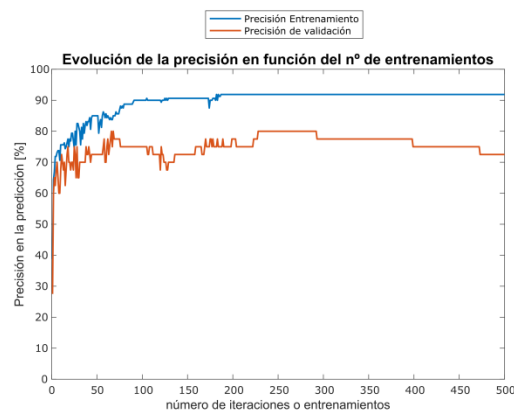
### 6.1 Predicción mediante el Perceptrón Multicapa.

La utilización del perceptrón para el diagnóstico se debe fundamentalmente en su capacidad para generalización a partir de ejemplos, es decir, proporcionar una respuesta correcta ante patrones que no han sido empleados en su entrenamiento.

En el proceso de entrenamiento se debe considerar por una parte el error de aprendizaje, calculado como el error cometido por la red con los patrones de entrenamiento, y por otra parte el error de generalización, calculado como el error cometido por el algoritmo con el conjunto de patrones reservados para la validación [26]. Podemos minimizar el error de entrenamiento sin más que aumentar las iteraciones (número de veces que introducimos los parámetros de entrenamiento en el algoritmo de retropropagación), pero la disminución excesiva de este error suele conllevar la degradación de la capacidad de generalización. Este fenómeno se conoce como **sobre-aprendizaje** o sobre-entrenamiento, y se debe tener en cuenta para la consecución de un modelo de diagnosis eficaz.

Un ejemplo de este fenómeno se puede observar en uno de los experimentos realizados con el modelo 1 (ver ANEXO III.1), donde puede apreciarse como la precisión en los datos de entrenamiento aumentan con el número de iteraciones hasta llegar aproximadamente a 90%, manteniéndose constante a partir de entonces. Sin embargo, en los datos de validación se aprecia un aumento de precisión hasta alrededor de 250 ciclos de entrenamiento donde comienza a descender ligeramente.

Por otro lado, la convergencia de la red depende de la velocidad de cambio de los pesos de la red, siendo proporcionales al gradiente del error como se refleja en las ecuaciones (5) y (6). Esta proporcionalidad viene determinada por la **razón o tasa de aprendizaje ( $\alpha$ )**, marcando así la velocidad de convergencia de la red. El valor de este parámetro suele estar comprendido entre 0 y 1. Se debe obtener la tasa de aprendizaje óptima para el problema a tratar sabiendo que



**Figura 6.1 Ejemplo de sobre-aprendizaje para una arquitectura de red concreta de 1 capa oculta y 5 neuronas en ésta.**

valores cercanos a 1 ayudarán a una rápida convergencia, pero pueden tener efectos negativos como que el método salte un mínimo o incluso oscile alrededor del mínimo.

En el desarrollo práctico de estas redes el principal problema es no disponer de suficientes patrones para entrenar una determinada arquitectura. Esto es debido al fenómeno denominado como **maldición de la dimensionalidad** (*curse of dimensionality*) [26], que consiste en que el número de datos necesarios para el correcto funcionamiento de la clasificación crece exponencialmente con la dimensión del espacio de entrada, lo que se agrava en los problemas con elevada dimensión de entrada, como el que se está tratando, disponer de un escaso número de patrones de aprendizaje. Para, en la medida de lo posible, evitar el fenómeno comentado se debe minimizar el número de pesos en la arquitectura de la red reduciendo así el número de patrones necesarios en el aprendizaje. Esto se consigue disminuyendo el número de entradas en el modelo y reduciendo la cantidad de neuronas ocultas hasta el mínimo de éstas que consigan una generalización razonable del problema, en este caso, la clasificación de los pacientes en sanos y enfermos. Por esta razón se intenta conseguir el menor número de capas ocultas en todos los modelos estudiados.

Otro factor a tener en cuenta es la posible **agrupación de datos** de cada modelo. Para ello se realiza una prueba de predicción con el conjunto completo de datos y otra separando los datos por sexo, es decir, un modelo para mujeres y otro para hombres. Esto es debido a que se observa un mayor número de enfermos de sexo femenino, alrededor del 70%, y se cree poder llegar a alcanzar un diagnóstico con mayor precisión. Tras realizar dicha prueba se concluye que no mejora la predicción de diagnóstico, coincidiendo además con el resultado del estudio de la sensibilidad de parámetros que minimizaba la importancia del sexo del paciente como dato de entrada en la predicción.

Para cada base de datos se realizan las siguientes pruebas con el fin de conseguir la mejor arquitectura de red en cada caso:

- Generalización frente a memorización: se estudia el número de iteraciones óptimo para evitar el efecto de sobre-aprendizaje.
- Selección de razón de aprendizaje: se selecciona la razón de aprendizaje óptima que marcará la velocidad de convergencia de la red.
- Selección de la arquitectura de red: se estudia el número de capas y neuronas ocultas que ofrece una mayor precisión en la predicción de la enfermedad.
- En los modelos 4 y 5, además, se debe seleccionar el número de entradas a la red, ya que existe la posibilidad de agrupar los puntos de barrido peripapilar para disminuir la dimensión de entrada a la misma y evitar el fenómeno de la maldición de la dimensionalidad.

Tras realizar las pruebas de los 5 modelos (ANEXO III) se exponen a continuación el mejor resultado obtenido para cada uno.

**Tabla 2 Resultados obtenidos por cada modelo de predicción para el perceptrón multicapa.**

Modelo	Datos	Capas Ocultas	Nº Neuronas ocultas por capa	Nº iteraciones	Razón de aprendizaje $\alpha$	Precisión de diagnóstico [%]	AUC
Base 1	Todos	2	5; 2	200	0,5	81	0.8001
Base 2	Todos	2	8; 4	200	0,5	79.87	0.8498
Base 3	Todos	2	1; 4	200	0,5	82.43	0.7953
Base 4	Todos	1	8	200	0,5	92.3	0.9446
Base 5	Todos	1	8	200	0,5	84.42	0.8561

## 6.2 Predicción mediante SVM.

El algoritmo *Support Vector Machine* (SVM) es probablemente el algoritmo de aprendizaje más empleado debido a que consigue patrones con un robusto rendimiento utilizando conceptos de la teoría de optimización [37].

Del mismo modo que en el perceptrón multicapa se puede ajustar ciertos parámetros del algoritmo para conseguir el óptimo en un problema concreto. Los parámetros [38] que se van a variar en éste algoritmo son los siguientes:

- 1) Función Kernel (*KernelFunction*): Se pueden utilizar diferentes funciones Kernel para el entrenamiento de la red. Las opciones que se contemplan aparecen en la siguiente tabla:

**Tabla 3. Tipos de función kernel para poder entrenar la SVM.**

Función Kernel	Descripción	Fórmula
Gaussiana	Función kernel Gaussiana o radial. Suele emplearse en aprendizaje con una sola clase.	$K(x_j, x_k) = e^{(-\ x_j - x_k\ ^2)}$
Lineal	Función kernel Lineal. Suele utilizarse por defecto en el aprendizaje con dos clases.	$K(x_j, x_k) = x_j' x_k$
Polynomial	Función kernel Polinomial. Se usa pudiendo especificar el orden del polinomio empleado ( $q$ ).	$K(x_j, x_k) = (1 + x_j' x_k)^q$

- 2) Orden del polinomio (*PolynomialOrder*): cuando se emplea la función kernel tipo polinomial, se puede especificar el orden con el que se quiera calcular la matriz K (ver Tabla 3 y ecuación (19)), para mejorar la capacidad de clasificar de la SVM.
- 3) Algoritmos de optimización empleados (*optimization routine*): los diferentes algoritmos de optimización [39] que se pueden utilizar en la resolución del problema son los siguientes:

**Tabla 4. Algoritmos de optimización empleados en la resolución del problema de clasificación mediante máquinas de soporte vectorial.**

Algoritmos de Optimización	Descripción
ISDA	Iterative Single Data Algorithm [39].
L1QP	Utiliza programación cuadrática para implementar la minimización.
SMO	Optimización mínima secuencial [40].

- 4) Número de iteraciones (*CacheSize*): al igual que en el perceptrón multicapa podemos elegir el número de iteraciones en el aprendizaje. Tras realizar un pequeño estudio sobre la influencia del número de iteraciones en la precisión de clasificación se observa que este parámetro no influye en tanta medida como en el perceptrón multicapa. Se va a emplear a partir de ahora 1000 iteraciones en todas las pruebas que van a realizarse con el algoritmo SVM.

Variando éstos parámetros se tratará de llegar a una configuración de SVM óptima para cada modelo:

**Tabla 5 Resultados obtenidos por cada modelo de predicción para las máquinas de soporte virtual.**

Modelo	Datos	Función kernel	Algoritmo Optimización	Nº entradas	Precisión de diagnóstico [%]	AUC
Base 1	Todos	Lineal	SMO	16	68	0.6581
Base 2	Todos	Polinomial	SMO	12	76.76	0.6963
Base 3	Todos	Polinomial	SMO	12	74	0.6712
Base 4	Todos	Lineal	SMO	25	78.97	0.7307
Base 5	Todos	Lineal	SMO	25	76.82	0.7037

Como puede apreciarse en la Tabla 5, el algoritmo de optimización empleado es el de optimización de mínima secuencia en todos los modelos. Esto es debido a que se realiza un análisis (ANEXO III.2) en el que se compara la precisión en el diagnóstico obtenida con los diferentes algoritmos de optimización para el modelo 1, que corresponde a la base de datos del equipo CIRRUS-OCT. El resultado de éste análisis tiene como conclusión la utilización del algoritmo SMO para el resto de los modelos.

### 6.3 Precisiones alcanzadas.

Las mejores precisiones alcanzadas para el perceptrón (MLP) y para las máquinas de soporte vectorial (SVM) para cada una de las bases de datos se muestran en la Tabla 6.

**Tabla 6. Máximas precisiones de diagnóstico alcanzadas con cada modelo realizado, tanto con el perceptrón multicapa como con las máquinas de soporte vectorial.**

Tipo de Red	Modelo	Datos	Estructura de red	Precisión de diagnóstico [%]	AUC
MLP	Base 1	Todos	16-5-2-2	81	0.8001
	Base 2	Todos	12-8-4-2	79.87	0.8498
	Base 3	Todos	12-1-4-2	82.43	0.7953
	Base 4	Todos	777-8-2	<b>92.3</b>	<b>0.9446</b>
	Base 5	Todos	777-8-2	84.42	0.8561
SVM	Base 1	Todos	Lineal (16 entradas)	68	0.6581
	Base 2	Todos	Polinomial (12 entradas)	76.76	0.6963
	Base 3	Todos	Polinomial (12 entradas)	74	0.6712
	Base 4	Todos	Lineal (25 entradas)	78.97	0.7307
	Base 5	Todos	Lineal (25 entradas)	76.82	0.7037

Destacar en el caso del perceptrón multicapa la precisión en el diagnóstico obtenida para la base de datos 4, correspondiente al OCT-Spectralis, área peripapilar con barrido de la cabeza del nervio óptico mediante el protocolo axonal, para la cual se obtiene un 92,3% de precisión.

En el caso de las SVM, destaca la precisión de 78.97% para el mismo modelo 4.

### 6.4 Área bajo las curvas ROC.

En este apartado se realiza la representación de las curvas ROC y se calcula el área encerrada por la curva para poder comparar los modelos y su valor diagnóstico.

En primer lugar, en la Figura 6.2 se muestran las curvas ROC y el área bajo las curvas (AUC) de la mejor arquitectura de red conseguida según la precisión en el diagnóstico para cada uno de los 5 modelos realizados con el perceptrón multicapa. En segundo lugar, en la Figura 6.3 se muestran las 5 curvas ROC y su respectivo AUC para los 5 modelos, pero realizados con las máquinas de soporte vectorial.

En el caso del perceptrón multicapa se obtiene un  $AUC = 0.9442$  en el modelo 4, lo que puede considerarse como una prueba con muy buen valor diagnóstico, sin embargo, en el caso de las SVM, en el mejor de los modelos se obtiene un  $AUC = 0.7307$ , no se consideraría una prueba con suficiente valor diagnóstico.

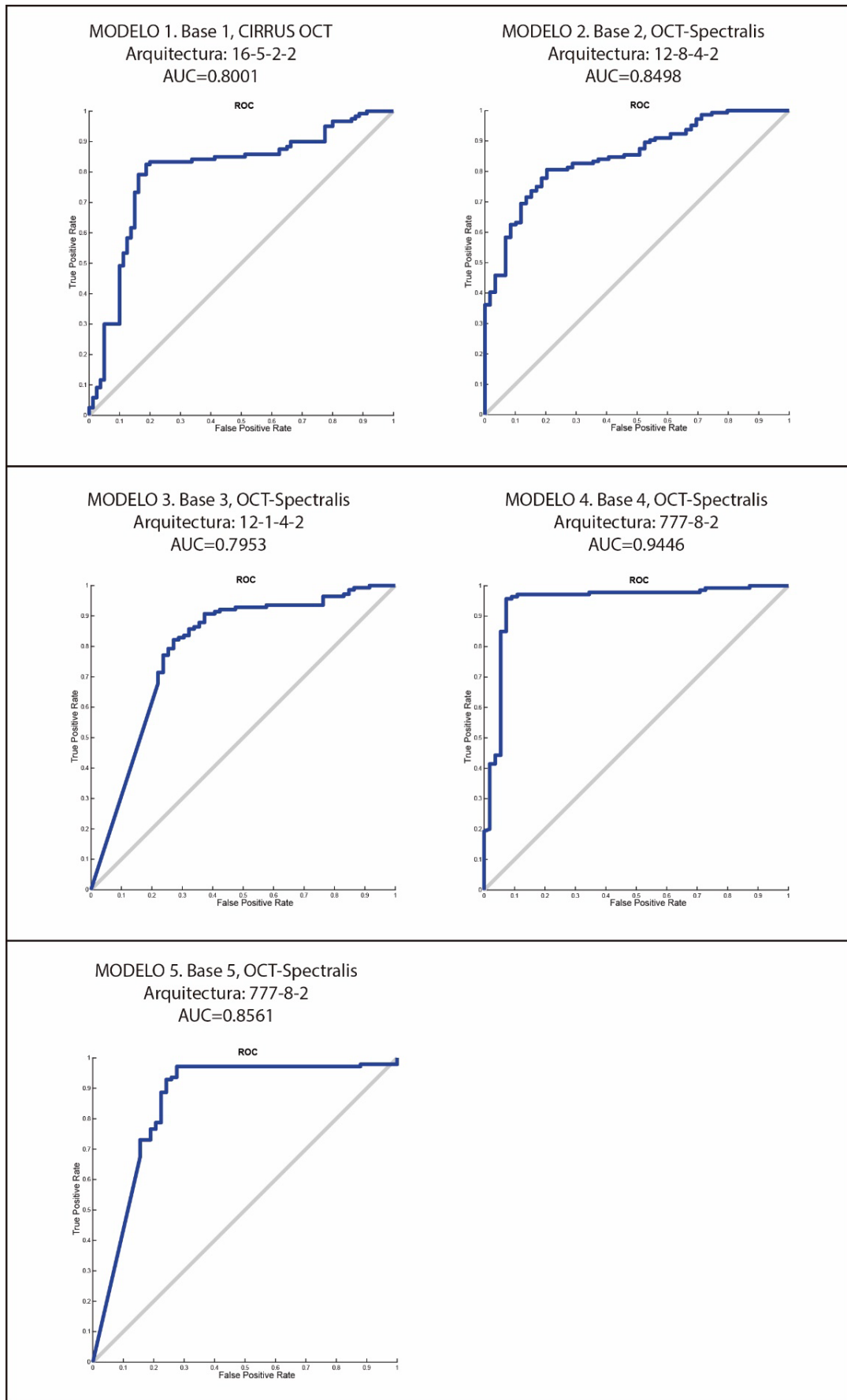
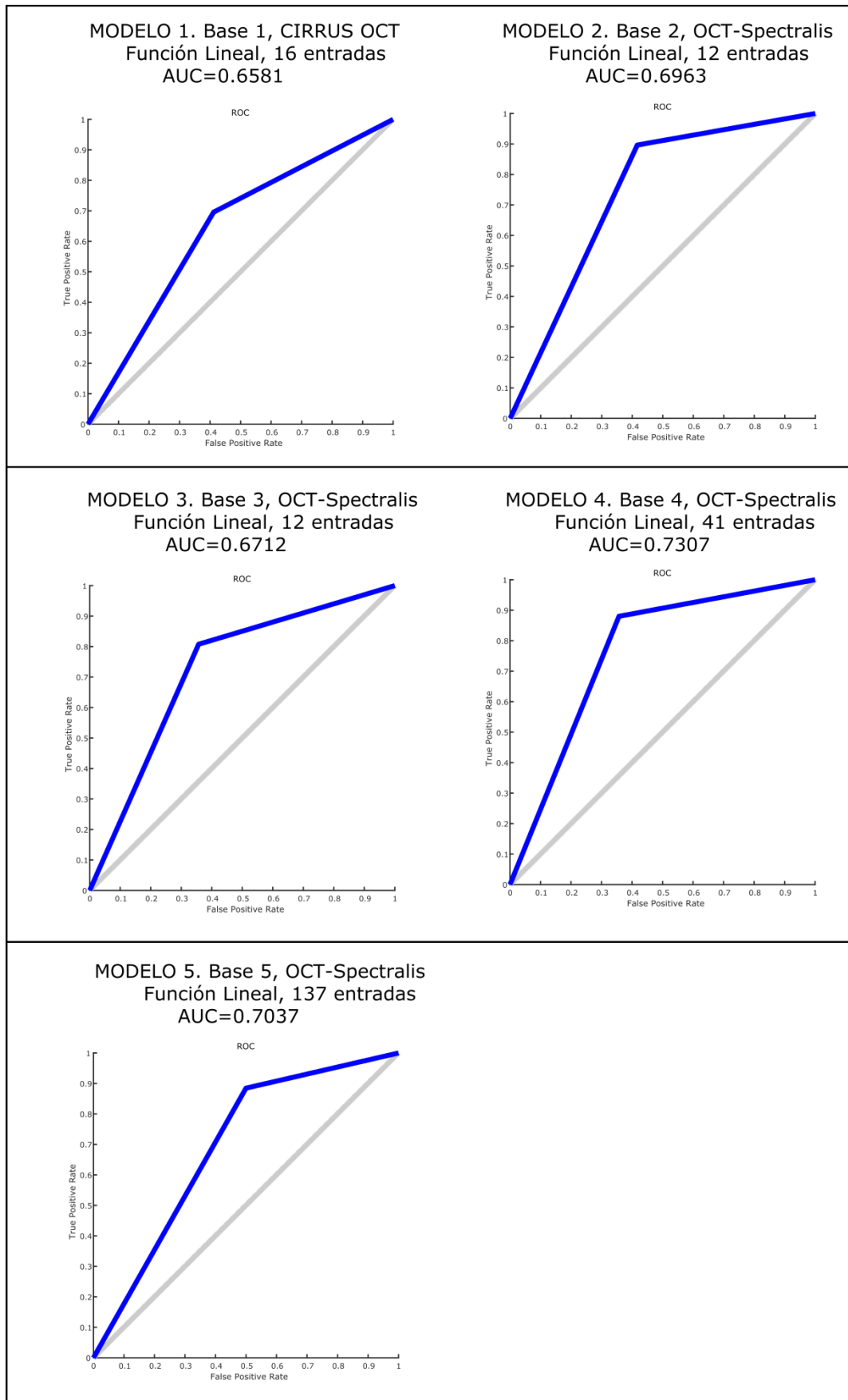


Figura 6.2 Curvas ROC para los 5 modelos realizados con el perceptrón multicapa.



**Figura 6.3** Curvas ROC para los 5 modelos realizados con las SVM.



## 7 CONCLUSIONES.

Las redes neuronales artificiales han experimentado un gran desarrollo en los últimos años, aplicándose con éxito en numerosos campos científicos y mejorando la capacidad de los médicos de predecir el resultado del paciente en diferentes patologías. El primer objetivo era comparar la capacidad de dos tipos de redes neuronales artificiales (MLP y SVM) para detectar diferencias entre la retina de pacientes con esclerosis múltiple y pacientes sanos. El segundo objetivo era comparar el valor como biomarcador de los diferentes equipos de OCT y protocolos de medida de éstos, para conseguir finalmente un modelo que sea capaz de diagnosticar con fiabilidad la enfermedad.

En primer lugar, viendo los resultados obtenidos de precisión en el diagnóstico (ver apartado 6.3) mediante las pruebas realizadas a los diferentes modelos construidos con el perceptrón multicapa y las máquinas de soporte vectorial podemos concluir que, para el problema concreto que se ha tratado, las SVM constituyen un algoritmo de clasificación extremadamente robusto, por esta razón, no son capaces de ajustarse a alta no-linealidad del problema y la precisión conseguida mediante éste algoritmo apenas alcanza el 80%. Sin embargo, el perceptrón alcanza un 92% de precisión, adaptándose mejor a través de la rectificación de los pesos que dan forma al algoritmo. Se debe destacar que la mayoría de los pacientes mal clasificados por los algoritmos son pacientes sanos. Esto puede deberse a la menor disponibilidad de datos de controles sanos que sujetos con EM.

Las principales diferencias entre los dos clasificadores utilizados quedan recogidas en la siguiente tabla:

**Tabla 7. Comparación de prestaciones entre el perceptrón multicapa y las máquinas de soporte vectorial.**

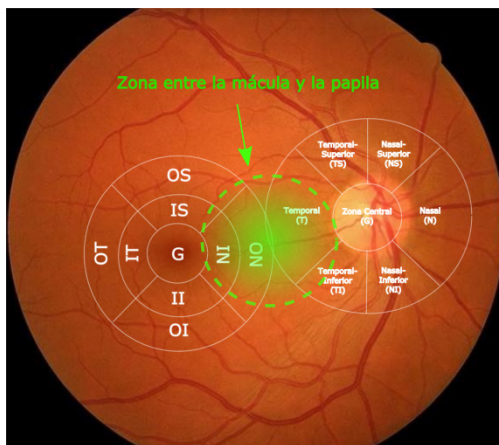
MLP	SVM
<ul style="list-style-type: none"> <li>• Capas ocultas transforman a espacios de cualquier dimensión.</li> <li>• El espacio de búsqueda tienen múltiples mínimos locales.</li> <li>• El entrenamiento es costoso.</li> <li>• La clasificación es muy eficiente.</li> <li>• Se diseña el número de capas ocultas y neuronas.</li> <li>• Muy buen funcionamiento en problemas típicos.</li> </ul>	<ul style="list-style-type: none"> <li>• La función Kernel transforma a espacios de dimensión superior.</li> <li>• El espacio de búsqueda tienen sólo un mínimo global.</li> <li>• El entrenamiento es muy eficiente.</li> <li>• La clasificación es muy eficiente.</li> <li>• Se diseña la función kernel.</li> <li>• Muy buen funcionamiento en problemas típicos.</li> <li>• Extremadamente robusto para generalización, menos necesidad de heurísticos para entrenamiento.</li> </ul>

Se puede decir que una de las razones principales de que la precisión con el MLP haya sido mayor que con las SVM es que ofrece una mayor flexibilidad con sus diferentes estructuras, a diferencia de las máquinas de soporte vectorial caracterizadas por su robustez, entrando en juego el desequilibrio entre el número de pacientes con EM diagnosticada y controles sanos. Sin embargo, éste último algoritmo tiene una menor necesidad de heurísticos para una buena generalización que con el MLP, pudiendo estas últimas sufrir la llamada maldición de dimensionalidad (apartado 6.1), necesitando mayor cantidad de datos históricos para una buena generalización al aumentar la dimensión de la entrada del problema.

La predicción con un 92.3% de precisión y una AUC=0.9446 resulta una prueba con gran valor diagnóstico, no obstante, ésta alta precisión hay que tomarla con precaución, ya que sería necesario obtener más pacientes y controles para aumentar los datos de validación y realizar un test final que validara completamente el modelo, o por el contrario demostrara cierto sobreaprendizaje en el perceptrón.

A la hora de comparar los diferentes modelos, según a la base de datos que pertenezca, un importante parámetro que determinará el valor del OCT como biomarcador en las enfermedades neurodegenerativas será la calidad de las imágenes obtenidas mediante esta técnica. Otro, será la repetibilidad del aparato, entrando en juego la habilidad del especialista en cuestión que realiza la prueba a cada paciente y las posibles diferencias que puede haber en la misma medida realizada por dos especialistas diferentes. Puede ser éste uno de los motivos de que los dos modelos realizados con la base de datos obtenida a partir del aparato CIRRUS OCT tengan menor valor diagnóstico que el resto de los modelos provenientes del OCT-Spectralis, ya que, éste último es más moderno y preciso que el anterior.

Las mejores precisiones y AUC se consiguen con la base de datos 4, correspondiente a la adquisición de la zona peripapilar con barrido de la cabeza del nervio óptico con el protocolo axonal (sentido nasal-nasal), llegando a un 92.3% de precisión en el diagnóstico y un  $AUC = 0.9446$  que equivale a un diagnóstico casi-perfecto. La mejor predicción coincide con el modelo 4, cuyo protocolo de adquisición axonal mejora la precisión y reproducibilidad del cuadrante temporal de la zona peripapilar, lo que hace que los datos obtenidos para esa zona sean más fiables que los del protocolo glaucoma (modelo 5). Este aumento de la precisión coincide con algunos estudios [15] que sugieren que la zona temporal es la que puede verse más afectada por la esclerosis múltiple además de coincidir con el análisis de sensibilidad de atributos realizado para estas bases de datos en las que se muestran una mayor importancia de la zona temporal del área peripapilar.



**Figura 7.1 Ojo derecho. Zona destacada entre la papila y la mácula en la predicción de la esclerosis múltiple.**

A partir de los datos obtenidos en el análisis de sensibilidad de los atributos, se ve como en la adquisición macular de la RNFL las zonas más relevantes son las nasales, mientras que en las adquisiciones peripapilares, las zonas con mayor importancia son las temporales. Observando una imagen de la retina con los protocolos de adquisición (Figura 7.1) se puede ver que la zona temporal de la papila corresponde con la zona nasal de la mácula. A partir de esto, se puede deducir que el área comprendida entre la mácula y la papila tendrá especial importancia en la evolución de la esclerosis múltiple y será un dato relevante para considerar la retina como un biomarcador en el diagnóstico de la enfermedad.

Otro factor a tener en cuenta es que todos los pacientes incluidos en las bases de datos son procedentes del Hospital Universitario Miguel Servet de Zaragoza, y todos ellos pertenecen a pacientes de etnia caucásica europea y de la misma zona geográfica, con las mismas condiciones medioambientales. Se considera en diversos estudios que éstos parámetros [41] pueden tener cierta relevancia en el desarrollo de la esclerosis múltiple. Por ésta razón, para llegar a obtener un modelo de diagnóstico general debería obtenerse una base de datos que englobara pacientes y controles de otras zonas geográficas.

Por último, concluir diciendo que la OCT es una técnica sencilla, reproducible, rápida y no invasiva, que permite medir el grosor de las diferentes capas de la retina y los posibles cambios que puedan producirse en ella. Por tanto, sugerir que la inclusión de la retina como biomarcador en el diagnóstico, a través de los datos obtenidos con la OCT y con el correspondiente algoritmo de predicción, puede ayudar mucho a los neurólogos a clasificar a un paciente con esclerosis múltiple y poder llegar a conseguir un diagnóstico precoz y eficiente, alargando y mejorando la calidad de vida de los enfermos, pudiendo comenzar así un tratamiento que consiga frenar la progresión de la enfermedad y el daño neuronal.

## 8 LÍNEAS DE FUTURO.

En primer lugar, para mejorar el modelo de predicción se debería comenzar por conseguir un mayor número de sujetos con los que poder entrenar el algoritmo. A su vez, es importante conseguir más controles sanos para intentar corregir la deficiencia de éstos en la base actual, consiguiendo llegar a un equilibrio entre pacientes sanos y enfermos.

Por otra parte, se puede intentar utilizar otros algoritmos de aprendizaje para repetir las pruebas realizadas y comparar los resultados. En concreto algunos de los algoritmos que pueden encajar en la naturaleza del problema tratado son los árboles de decisión o las regresiones lineal y logarítmica. Otra forma de mejorar los modelos obtenidos sería añadir el algoritmo conocido como “*Boosting*”, que es un meta-algoritmo que reduce el sesgo y la varianza en un contexto de aprendizaje supervisado.

Actualmente el Hospital cuenta con un equipo de OCT más moderno que el OCT-Spectralis, concretamente el DRI OCT Triton. Por tanto, podrían emplearse nuevas bases de datos más precisas obtenidas con este nuevo aparato para llegar a obtener una mejora en el diagnóstico, pero todavía la cantidad de datos que se han obtenido no es suficientemente elevada.

Respecto al valor de la retina como biomarcador; a partir de las pruebas realizadas con los algoritmos, el análisis de sensibilidad de los atributos y apoyándose en otros estudios [15] se llega a la conclusión de que algunas zonas de la RNFL se ven mucho más afectadas que otras en el desarrollo de la EM. Así pues, se podría seguir trabajando en encontrar puntos de la retina (zona entre la mácula y la papila) a los que dar más peso como entrada en los clasificadores, incluso eliminar las zonas menos significativas con el objetivo de mejorar la precisión en el diagnóstico.

Visto los buenos resultados obtenidos en el diagnóstico de la EM, se podría intentar aplicar estos algoritmos, a partir de los resultados de la OCT, a la predicción y diagnóstico de otras enfermedades neurodegenerativas como el Parkinson u otro tipo de enfermedades como el glaucoma o la neuritis óptica, que afectan a las zonas del sistema nervioso visual.

Por último, se podría aplicar la inteligencia artificial a la predicción de la evolución longitudinal en los pacientes que han sido diagnosticados con la EM con el objetivo de saber la evolución que va a tener la enfermedad en el paciente (Figura 2.1) y poder realizar un mejor tratamiento individual a cada paciente. Algunos de los algoritmos que podrían emplearse en esta predicción de evolución serían las redes neuronales recurrentes LSTM (Long Short-Term Memory), que introducen la variable tiempo y constituyen herramientas útiles en series temporales.

## 9 BIBLIOGRAFÍA.

- [1] Garcia-Martin E, Pueyo V, Pinilla I, Ara JR, Martin J, Fernandez J. Fourier-Domain OCT in multiple sclerosis patients: Reproducibility and ability to detect retinal nerve fiber layer atrophy. *Investig Ophthalmol Vis Sci* 2011;52:4124–31. doi:10.1167/iovs.10-6643.
- [2] Ferguson B, Matyszak MK, Esiri MM, Perry VH. Axonal damage in acute multiple sclerosis lesions. *Brain* 1997;120:393–9. doi:10.1093/brain/120.3.393.
- [3] Vidal-Jordana Á, Sastre-Garriga J, Montalban X. [Optical coherence tomography in multiple sclerosis]. *Rev Neurol* 2012;54:556–63.
- [4] Pelegrín Molina M aA., Gómez-Conesa A. Esclerosis múltiple: estado actual. *Fisioterapia* 2001;23:144–53. doi:10.1016/S0211-5638(01)72946-X.
- [5] Fernández Fernández ., Fernández VE. Esclerosis múltiple. Concepto. Etiopatogenia. Fisiopatología. Manifestaciones clínicas. Investigaciones paraclínicas. Diagnóstico. Historia natural. *Med - Programa Form Médica Contin Acreditado* 2007;9:4867–77. doi:10.1016/S0211-3449(07)75463-2.
- [6] García Ferriz P. ESCLEROSIS MÚLTIPLE, Etiopatogenia. Tratamiento n.d.
- [7] Fernández-Domínguez J, García-Rodríguez R, Mateos V. [Transorbital echography for assessment of optical nerve atrophy in demyelinating diseases: a pilot study]. *Rev Neurol* 2012;54:587–92.
- [8] Sergott RC, Frohman E, Glanzman R, AL-Sabbagh A. The role of optical coherence tomography in multiple sclerosis: Expert panel consensus. *J Neurol Sci* 2007;263:3–14. doi:10.1016/j.jns.2007.05.024.
- [9] Frohman E, Costello F, Zivadinov R, Stuve O, Conger A, Winslow H, et al. Optical coherence tomography in multiple sclerosis. *Lancet Neurol* 2006;5:853–63. doi:10.1016/S1474-4422(06)70573-7.
- [10] Frohman E, Costello F, Zivadinov R, Stuve O, Conger A, Winslow H, et al. Optical coherence tomography in multiple sclerosis. *Lancet Neurol* 2006;5:853–63. doi:10.1016/S1474-4422(06)70573-7.
- [11] Bejarano B, Bianco M, Gonzalez-Moron D, Sepulcre J, Goñi J, Arcocha J, et al. Computational classifiers for predicting the short-term course of Multiple sclerosis. *BMC Neurol* 2011;11. doi:10.1186/1471-2377-11-67.
- [12] Zhao Y, Healy BC, Rotstein D, Guttmann CRG, Bakshi R, Weiner HL, et al. Exploration of machine learning techniques in predicting multiple sclerosis disease course. *PLoS One* 2017;12:e0174866. doi:10.1371/journal.pone.0174866.
- [13] Garcia-Martin E, Pablo LE, Herrero R, Satue M, Polo V, Larrosa JM, et al. Diagnostic ability of a linear discriminant function for spectral-domain optical coherence tomography in patients with multiple sclerosis. *Ophthalmology* 2012;119:1705–11. doi:10.1016/j.ophtha.2012.01.046.
- [14] Garcia-Martin E, Polo V, Larrosa JM, Marques ML, Herrero R, Martin J, et al. Retinal layer segmentation in patients with multiple sclerosis using spectral domain optical coherence tomography. *Ophthalmology* 2014;121:573–9. doi:10.1016/j.ophtha.2013.09.035.
- [15] Garcia-Martin E, Herrero R, Bambo MP, Ara JR, Martin J, Polo V, et al. Artificial neural network techniques to improve the ability of optical coherence tomography to detect optic neuritis. *Semin Ophthalmol* 2015;30:11–9. doi:10.3109/08820538.2013.810277.

- [16] Britze J, Pihl-Jensen G, Frederiksen JL. Retinal ganglion cell analysis in multiple sclerosis and optic neuritis: a systematic review and meta-analysis. *J Neurol* 2017;264:1837–53. doi:10.1007/s00415-017-8531-y.
- [17] Petzold A, de Boer JF, Schippling S, Vermersch P, Kardon R, Green A, et al. Optical coherence tomography in multiple sclerosis: a systematic review and meta-analysis. *Lancet Neurol* 2010;9:921–32. doi:https://doi.org/10.1016/S1474-4422(10)70168-X.
- [18] AEDEM-COCEMFE. Tipos de Esclerosis Múltiple n.d.
- [19] Trapp BD, Peterson J, Ransohoff RM, Rudick R, Mörk S, Bö L. Axonal Transection in the Lesions of Multiple Sclerosis. *N Engl J Med* 1998;338:278–85. doi:10.1056/NEJM199801293380502.
- [20] Soler García A, Padilla Parrado F, Figueroa-Ortiz LC, González Gómez A, García-Ben A, García-Ben E, et al. Análisis del grosor macular y de capa de fibras nerviosas en pacientes con esclerosis múltiple en relación con su nivel de gravedad y antecedentes previos de neuritis óptica. *Neurología* 2016;31:379–88. doi:10.1016/j.nrl.2014.09.003.
- [21] Gonzalez Ruiz V. El Sistema Visual Humano 2014. [https://w3.ual.es/~vruiz/Docencia/Apuntes/Perception/Sistema\\_Visual/index.html](https://w3.ual.es/~vruiz/Docencia/Apuntes/Perception/Sistema_Visual/index.html).
- [22] Aznar Casanova JA. Psicología de la percepción visual n.d. [http://www.ub.edu/pa1/node/area\\_visual](http://www.ub.edu/pa1/node/area_visual).
- [23] Instituto de la visión n.d. <http://www.institutodelavision.org/webs/ivn/index.php/es/56-partes-del-ojo>.
- [24] Alañón Fernández FJ. Anatomía Y Fisiología Del Aparato Ocular. *Oftalmol En Atención Primaria* n.d.:11–46.
- [25] Paulasen DF. *Histology & Cell Biology: Examination & Board Review*. n.d.
- [26] Martín del Brío B, Sanz Molina A. REDES NEURONALES Y SISTEMAS BORROSOS 1997:387.
- [27] Perceptron E. Perceptron multicapa 1969:15–32.
- [28] Viñuela PI, León IMG. *Redes Neuronales Artificiales Un Enfoque Practico* 2008:229.
- [29] Heidelberg-Engineering. Spectralis OCT. Heidelberg, Ger 2018. <https://business-lounge.heidelbergengineering.com/es/en/products/>.
- [30] Betancour G. Las máquinas de soporte vectorial (SVMs). *Sci Tech* 2005:67–72. doi:10.22517/23447214.6895.
- [31] Haykin S. *Neural Networks and Learning Machines*. vol. 3. 2008. doi:978-0131471399.
- [32] Gundogan FC, Demirkaya S, Sobaci G. Is optical coherence tomography really a new biomarker candidate in multiple sclerosis? - A structural and functional evaluation. *Investig Ophthalmol Vis Sci* 2007;48:5773–81. doi:10.1167/iovs.07-0834.
- [33] Rut M<sup>a</sup> López Lizcano. VALOR DEL GROSOR DE LA CAPA DE FIBRAS NERVIOSAS DE LA RETINA MEDIDO MEDIANTE TOMOGRAFIA DE COHERENCIA ÓPTICA COMO MARCADOR DE DAÑO NEUROAXONAL EN LA ESCLEROSIS MÚLTIPLE 2012.
- [34] Garcia-Martin E, Pablo LE, Herrero R, Ara JR, Martin J, Larrosa JM, et al. Neural networks to identify multiple sclerosis with optical coherence tomography. *Acta Ophthalmol* 2013;91:1–7. doi:10.1111/aos.12156.
- [35] H. Witten , Eibe Frank , Len Trigg , Mark Hall , Geoffrey Holmes SJC. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Fr Wittgenstein, L 1999.

- [36] Carmona Suárez EJ. Máquinas de Vectores Soporte (SVM) 2014:1–25.
- [37] Bottou L, Lin C. Support vector machine solvers. Large Scale Kernel Mach 2007:1–27. doi:10.1.1.127.511.
- [38] MathWorks. MathWorks. Train binary support vector machine classifier. n.d. [https://es.mathworks.com/help/stats/fitcsvm.html?s\\_tid=srchtitle#namevaluepairarguments](https://es.mathworks.com/help/stats/fitcsvm.html?s_tid=srchtitle#namevaluepairarguments).
- [39] Kecman V, Huang T, Vogt M. Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets : n.d.;274:255–74.
- [40] Fan, Rong-En And Chen, Pai-Hsuen And Lin C-J. Working Set Selection Using Second Order Information for Training Support Vector Machines. J Mach Learn Res 2005;6:1889--1918.
- [41] Compston A, Coles A. Multiple sclerosis. Lancet (London, England) 2008;372:1502–17. doi:10.1016/S0140-6736(08)61620-7.

## 10 ÍNDICE DE FIGURAS.

Figura 2.1 Esquema de la evolución temporal de los tipos de EM .....	5
Figura 2.2 Evolución del espesor [ $\mu\text{m}$ ] de la retina (todas las capas) con la edad de un grupo de pacientes sanos del Hospital Universitario Miguel Servet de Zaragoza. ....	6
Figura 3.1 Esquema de las principales vías visuales del cerebro .....	7
Figura 3.2 Partes del ojo humano. [4] .....	8
Figura 3.3 Partes del ojo humano y detalle de la retina [25] .....	8
Figura 3.4 Ilustración de las capas de la retina de Ramón y Cajal .....	9
Figura 3.5 Representación de las partes de una célula .....	10
Figura 3.6 Imagen tomada mediante la técnica de OCT (Optical Coherence Tomography) de la retina de un ser humano. A la izquierda, imagen de la retina del ojo derecho y a la derecha, imagen de la retina del ojo izquierdo. ....	10
Figura 4.1 Esquema general de una RNA unidireccional de tres capas, de entrada, oculta y de salida. ....	11
Figura 4.2 Ejemplo de un perceptrón multicapa con 7 entradas, 2 capas ocultas de neuronas y 2 salidas. ....	12
Figura 4.3 Espacio de los pesos $w$ y el descenso por el gradiente hacia un mínimo. ....	14
Figura 4.4 Representación de un mínimo local y el mínimo global de la función de error. ....	14
Figura 4.5 Representación de un problema de clasificación 2D resuelto mediante SVM .....	15
Figura 4.6 Tipos de problemas de clasificación 2D que podemos resolver con SVM; la imagen a la izquierda se puede observar un problema de clasificación linealmente separable y en la imagen de la derecha un problema de clasificación inseparable linealmente. ....	15
Figura 4.7 Aparición del parámetro $\xi$ en el error de clasificación [15] .....	17
Figura 4.8 Idea del uso de un kernel para la transformación del espacio de los datos [30] .....	17
Figura 5.1 A la izquierda, imagen del equipo de OCT-Spectralis, a la derecha imagen del equipo OCT-CIRRUS en el Hospital Universitario Miguel Servet de Zaragoza.....	19
Figura 5.2 Sección de la retina tomada mediante una tomografía de coherencia óptica.....	20
Figura 5.3 Ejemplo de datos que proporciona la OCT-CIRRUS .....	21
Figura 5.4 Importancia para los sectores horarios de la zona peripapilar .....	22
Figura 5.5 A la izquierda patrón de barrido realizado por el protocolo glaucoma (de zona temporal a zona temporal) en un ojo derecho, a la derecha patrón de adquisición realizado por el protocolo axonal (de zona nasal a zona nasal) en el mismo ojo derecho. ....	23
Figura 5.6 Representación de la adquisición de datos alrededor de la zona macular.....	23
Figura 5.7 Relevancia de los sectores de barrido de la zona macular en la capa de las células ganglionares .....	24
Figura 5.8 Relevancia de los sectores de barrido de la zona macular en la capa de las células ganglionares .....	24
Figura 5.9 Importancia para el barrido peripapilar de la RNFL con el protocolo axonal. ....	25
Figura 5.10 Importancia de los puntos para el barrido peripapilar de la RNFL con el protocolo glaucoma. ....	26
Figura 5.11 Ejemplo de datos que proporciona OCT-Spectralis.....	27

Figura 5.12 Ejemplo de k-fold cross validation con $k=4$ , en el que se muestra como el grupo de datos de prueba y de datos de entrenamiento varía en cada una de las iteraciones.....	29
Figura 5.13 Ejemplo de aplicación de las curvas ROC para evaluar la capacidad de una prueba para el diagnóstico de cierta patología en función del ratio de verdaderos positivos y el ratio de falsos positivos VPR-FPR.....	30
Figura 6.1 Ejemplo de sobre-aprendizaje para una arquitectura de red concreta de 1 capa oculta y 5 neuronas en ésta. ....	31
Figura 6.2 Curvas ROC para los 5 modelos realizados con el perceptrón multicapa. ....	36
Figura 6.3 Curvas ROC para los 5 modelos realizados con las SVM.....	37
Figura 7.1 Ojo derecho. Zona destacada entre la papila y la mácula en la predicción de la esclerosis múltiple.....	39



## 11 ÍNDICE DE TABLAS.

Tabla 1 Funciones de activación para el perceptrón multicapa.....	13
Tabla 2 Resultados obtenidos por cada modelo de predicción para el perceptrón multicapa. ....	32
Tabla 3. Tipos de función kernel para poder entrenar la SVM. ....	33
Tabla 4. Algoritmos de optimización empleados en la resolución del problema de clasificación mediante máquinas de soporte vectorial. ....	33
Tabla 5 Resultados obtenidos por cada modelo de predicción para las máquinas de soporte virtual. ....	34
Tabla 6. Máximas precisiones de diagnóstico alcanzadas con cada modelo realizado, tanto con el perceptrón multicapa como con las máquinas de soporte vectorial. ....	35
Tabla 7. Comparación de prestaciones entre el perceptrón multicapa y las máquinas de soporte vectorial.....	38

## Trabajo Fin de Máster

# ANEXOS

El uso de la retina como biomarcador en enfermedades neurodegenerativas. Simulación numérica.

Use of retina as a biomarker in neurodegenerative diseases.  
Numerical simulation.

Autor

**Alejandro Martín Gallego**

Directora

**Amaya Pérez del Palomar Aldea**

Máster Universitario en Ingeniería Mecánica  
Mecánica de Medios Continuos y Teoría de Estructuras  
Escuela de Ingeniería y Arquitectura de la Universidad de Zaragoza  
Curso 2017-2018

# Índice de Anexos

I.	CARACTERÍSTICAS CLÍNICAS.....	2
I.1	CIRRUS HD-OCT.....	2
I.1.1	Modelo 1: base de datos obtenida mediante el equipo CIRRUS HD-OCT.....	2
I.2	OCT-SPECTRALIS.....	3
I.2.1	Modelo 2: base de datos correspondiente a la capa de células ganglionares de la retina en la zona macular, obtenida con el equipo OCT-Spectralis.....	4
I.2.2	Modelo 3: base de datos correspondiente a la RNFL en la zona macular, obtenida con el equipo OCT-Spectralis.....	4
I.2.3	Modelo 4: base de datos correspondiente a la RNFL obtenida con el protocolo axonal del equipo OCT-Spectralis.....	5
I.2.4	Modelo 5: base de datos que corresponde a la RNFL obtenida con el protocolo glaucoma del equipo OCT-Spectralis.....	5
II.	FORMULACIÓN DEL PERCEPTRÓN MULTICAPA.....	7
II.1	Cálculo del valor en cada neurona.....	7
II.2	BackPropagation (descenso del gradiente).....	8
II.3	Cálculo de las derivadas del error respecto al peso en cada neurona.....	8
III.	EXPERIMENTOS REALIZADOS Y RESULTADOS.....	10
III.1	Predicción mediante el Perceptrón Multicapa.....	10
III.1.1	Modelo 1: base de datos obtenida mediante el equipo CIRRUS HD-OCT.....	11
III.1.2	Modelo 2: base de datos correspondiente a la capa de células ganglionares de la retina en la zona macular, obtenida con el equipo OCT-Spectralis.....	15
III.1.3	Modelo 3: base de datos correspondiente a la RNFL en la zona macular, obtenida con el equipo OCT-Spectralis.....	17
III.1.4	Modelo 4: base de datos correspondiente a la RNFL obtenida con el protocolo axonal del equipo OCT-Spectralis.....	18
III.1.5	Modelo 5: base de datos que corresponde a la RNFL obtenida con el protocolo glaucoma del equipo OCT-Spectralis.....	19
III.2	Predicción mediante SVM.....	21
III.2.1	Modelo 1: base de datos obtenida mediante el equipo CIRRUS HD-OCT.....	22
III.2.2	Modelo 2: base de datos correspondiente a la capa de células ganglionares de la retina en la zona macular, obtenida con el equipo OCT-Spectralis.....	22
III.2.3	Modelo 3: base de datos correspondiente a la RNFL en la zona macular, obtenida con el equipo OCT-Spectralis.....	23
III.2.4	Modelo 4: base de datos correspondiente a la RNFL obtenida con el protocolo axonal del equipo OCT-Spectralis.....	23
III.2.5	Modelo 5: base de datos que corresponde a la RNFL obtenida con el protocolo glaucoma del equipo OCT-Spectralis.....	24

## I. CARACTERÍSTICAS CLÍNICAS.

En este anexo se presentan las características clínicas de los pacientes y controles para cada una de las bases de datos tratadas.

### I.1 CIRRUS HD-OCT.

#### I.1.1 Modelo 1: base de datos obtenida mediante el equipo CIRRUS HD-OCT.

Los datos clínicos de los sujetos que disponen de la prueba médica realizada mediante el equipo CIRRUS se muestran en la Tabla 1. En ésta se puede observar los valores medios, el rango de valor mínimo y máximo y la desviación estándar de la población de individuos de cada uno de los parámetros que se han seleccionado como relevantes en la predicción, diferenciando los datos de los sujetos con esclerosis múltiple diagnosticada y los controles.

**Tabla 1. Características clínicas y medidas de espesor medio de la RNFL que proporciona el OCT\_CIRRUS en la población de sujetos sanos y sujetos enfermos.**

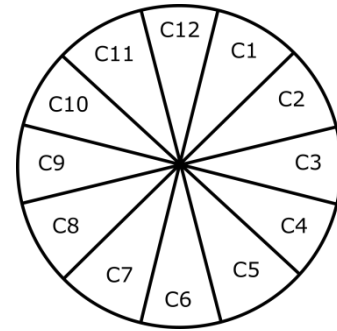
Parámetro	Sujetos Sanos (n=162 ojos)		Sujetos con EM (n=238 ojos)	
	Media (rango)	SD	Media (rango)	SD
Edad (años)	38.67 (19.85-63.59)	13.65	43.65 (5.32-69.29)	11.16
Mujeres (%)	69.75		72.5	
Espesor medio RNFL zona peripapilar	96.44 (68-123)	9.89	84 (33.23-120)	14.35
Espesor medio RNFL zona mácula	270.77 (224-360)	24.05	259.64 (212-383)	21.72
Espesor medio zona 1	109 (43-175)	26.4	96.8 (26-189)	22.2
Espesor medio zona 2	89.5 (24-142)	19.6	83.4 (6-167)	21.3
Espesor medio zona 3	57.2 (29-102)	10.7	55.5 (19-94)	11.3
Espesor medio zona 4	65.8 (37-117)	14.2	63.8 (25-99)	15.1
Espesor medio zona 5	103 (50-170)	25.6	93.4 (22-186)	26.5
Espesor medio zona 6	136 (11-216)	30	122 (20-224)	32.9
Espesor medio zona 7	141 (69-225)	26.1	114 (14-172)	29.5
Espesor medio zona 8	71.3 (43-148)	18	54.3 (10-149)	18.5
Espesor medio zona 9	53 (34-109)	10.8	43.9 (10-97)	12.7
Espesor medio zona 10	81.4 (44-162)	19	63.9 (19-174)	20.3
Espesor medio zona 11	131 (66-196)	24.3	112 (36-173)	25.6
Espesor medio zona 12	119 (42-190)	27.2	106 (11-186)	27.8

Para el grupo de datos de CIRRUS-OCT se dispone de un total de 81 sujetos clínicamente sanos (162 ojos) y 119 sujetos con EM diagnosticada (238 ojos).

Así pues, se define el modelo de predicción con 16 parámetros de entrada:

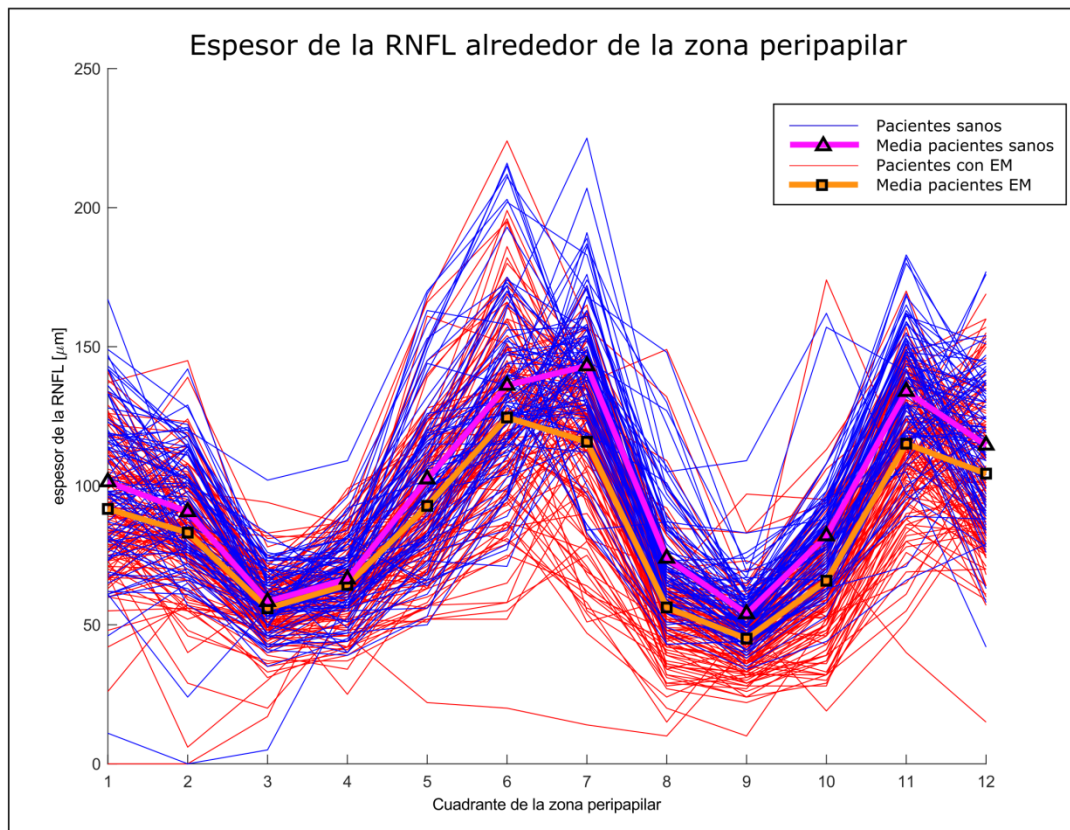
1. Edad del paciente
2. Sexo
3. Espesor medio de la RNFL en la zona peripapilar.
4. Espesor medio de la RNFL en la zona de la mácula.
5. C1: espesor medio correspondiente a la zona 1.
6. C2: espesor medio correspondiente a la zona 2.
7. C3: espesor medio correspondiente a la zona 3.
8. C4: espesor medio correspondiente a la zona 4.

9. C5: espesor medio correspondiente a la zona 5.
10. C6: espesor medio correspondiente a la zona 6.
11. C7: espesor medio correspondiente a la zona 7.
12. C8: espesor medio correspondiente a la zona 8.
13. C9: espesor medio correspondiente a la zona 9.
14. C10: espesor medio correspondiente a la zona 10.
15. C11: espesor medio correspondiente a la zona 11.
16. C12: espesor medio correspondiente a la zona 12.



**Figura I.1 Representación de los cuadrantes en la zona peripapilar**

Para saber si hay alguna diferencia apreciable entre el espesor de la RNFL de los pacientes con EM y los pacientes sanos se representa una gráfica del espesor en el mapa de cuadrantes horarios que nos proporciona la OCT-CIRRUS (Figura I.2).



**Figura I.2 Representación del espesor medio de la RNFL para los datos del ojo en cada cuadrante horario del mapa peripapilar para los pacientes con EM (líneas rojas) y sanos (líneas azules), el promedio de los pacientes con EM (línea naranja con cuadrados negros) y de los pacientes sanos (línea fucsia con triángulos negros).**

Como se muestra en la Figura I.2 a priori no parece que haya ninguna relación clara entre el espesor de la capa de fibras nerviosas de la retina con la enfermedad, únicamente puede apreciarse como el espesor medio de los pacientes sanos (línea fucsia con triángulos negros) es algo mayor que el espesor medio de los pacientes con EM (línea naranja con cuadrados negros).

## I.2 OCT-SPECTRALIS.

Distinguiendo los cuatro tipos de datos que proporciona el equipo Spectralis, se muestran los valores clínicos de los sujetos sometidos a cada una de las pruebas:

1. Datos del espesor medio de la capa de células ganglionares de la retina en los segmentos del área macular → Modelo 2.
2. Datos del espesor medio de la RNFL en los segmentos del área macular → Modelo 3.
3. Datos del espesor medio de la RNFL en los segmentos de la zona peripapilar y 768 medidas de la circunferencia peripapilar mediante el protocolo axonal → Modelo 4.
4. Datos del espesor medio de la RNFL en los segmentos de la zona peripapilar y 768 medidas de la circunferencia peripapilar mediante el protocolo glaucoma → Modelo 5.

### I.2.1 Modelo 2: base de datos correspondiente a la capa de células ganglionares de la retina en la zona macular, obtenida con el equipo OCT-Spectralis.

Tabla 2 Características clínicas y medidas de espesor medio de la capa de células ganglionares en la zona macular que proporciona el OCT-Spectralis en la población de sujetos sanos y sujetos enfermos.

Parámetro	Sujetos Sano (n=124 ojos)		Sujetos con EM (n=253 ojos)	
	Media (rango)	SD	Media (rango)	SD
Edad (años)	41.73 (26.22-72.79)	13.96	45.51 (24.74-72.04)	10.52
Mujeres (%)	76		74.6	
Espesor medio C0 ( $\mu\text{m}$ )	15.02 (9-24)	3.17	15.39 (6-51)	6.63
Espesor medio IN ( $\mu\text{m}$ )	52.25 (32-63)	4.62	41.62 (14-61)	10.26
Espesor medio ON ( $\mu\text{m}$ )	37.37 (30-47)	3.81	32.67 (17-52)	5.13
Espesor medio IS ( $\mu\text{m}$ )	52.90 (33-63)	5.06	43.71 (17-61)	9.11
Espesor medio OS ( $\mu\text{m}$ )	37.12 (28-45)	3.74	32.83 (19-46)	4.87
Espesor medio IT ( $\mu\text{m}$ )	46.99 (32-58)	4.41	38.18 (14-55)	9.43
Espesor medio OT ( $\mu\text{m}$ )	35.45 (26-47)	3.93	30.05 (13-45)	5.42
Espesor medio II ( $\mu\text{m}$ )	52.94 (41-62)	4.28	43.02 (14-60)	10.01
Espesor medio OI ( $\mu\text{m}$ )	34.86 (21-42)	4.06	31.63 (15-43)	4.36
Espesor medio Centro ( $\mu\text{m}$ )	3.424 (0-14)	2.38	5.24 (0-50)	8.2

### I.2.2 Modelo 3: base de datos correspondiente a la RNFL en la zona macular, obtenida con el equipo OCT-Spectralis.

Tabla 3 Características clínicas y medidas de espesor medio de la RNFL en la zona macular que proporciona el OCT-Spectralis en la población de sujetos sanos y sujetos enfermos.

Parámetro	Sujetos Sano (n=139 ojos)		Sujetos con EM (n=275 ojos)	
	Media (rango)	SD	Media (rango)	SD
Edad (años)	42.14 (26.22-72.78)	13.76	44.26 (24.74-72.03)	10.96
Mujeres (%)	70.5		76	
Espesor medio C0 ( $\mu\text{m}$ )	12.30 (8-16)	1.61	12.42 (5-27)	2.71
Espesor medio IN ( $\mu\text{m}$ )	21.49 (17-29)	2.28	19.42 (12-48)	3.48
Espesor medio ON ( $\mu\text{m}$ )	52.13 (40-67)	5.55	41.54 (16-67)	11.16
Espesor medio IS ( $\mu\text{m}$ )	23.96 (15-31)	2.31	20.63 (11-33)	3.59
Espesor medio OS ( $\mu\text{m}$ )	36.81 (29-47)	4.33	30.53 (16-54)	6.96
Espesor medio IT ( $\mu\text{m}$ )	17.33 (14-21)	1.36	17.49 (13-27)	1.95
Espesor medio OT ( $\mu\text{m}$ )	18.70 (15-24)	1.7	18.28 (13.56)	2.82
Espesor medio II ( $\mu\text{m}$ )	25.88 (20-36)	3.09	22.42 (14-40)	4.17
Espesor medio OI ( $\mu\text{m}$ )	38.63 (25-58)	7.56	32.21 (16-50)	7.5
Espesor medio Centro ( $\mu\text{m}$ )	1.22 (2.41)	2.41	3.08 (0-24)	4.9

### I.2.3 Modelo 4: base de datos correspondiente a la RNFL obtenida con el protocolo axonal del equipo OCT-Spectralis.

Tabla 4 Características clínicas y medidas de espesor medio de la RNFL en la zona peripapilar que proporciona el OCT-Spectralis en la población de sujetos sanos y sujetos enfermos, con protocolo de adquisición axonal.

Parámetro	Sujetos Sano (n=113 ojos)		Sujetos con EM (n=279 ojos)	
	Media (rango)	SD	Media (rango)	SD
Edad (años)	52.1 (27.15-72.78)	13.3	45.54 (24.74-72.36)	10.81
Mujeres (%)	93		72.4	
Espesor medio G ( $\mu\text{m}$ )	98.34 (69-121)	10.99	85.68 (40-159)	16.6
Espesor medio NS ( $\mu\text{m}$ )	106.68 (32-161)	25.21	93.62 (23-240)	23.17
Espesor medio N ( $\mu\text{m}$ )	76.79 (49-119)	13.47	69.60 (10-168)	19.19
Espesor medio NI ( $\mu\text{m}$ )	116.83 (68-188)	24.21	105.14 (28-190)	28.81
Espesor medio TI ( $\mu\text{m}$ )	146.07 (83-185)	18.49	115.32 (46-217)	26.57
Espesor medio T ( $\mu\text{m}$ )	69.23 (42-99)	12.17	58.58 (23-139)	16.72
Espesor medio TS ( $\mu\text{m}$ )	131.77 (79-174)	19.2	114.84 (46-220)	24.98

### I.2.4 Modelo 5: base de datos que corresponde a la RNFL obtenida con el protocolo glaucoma del equipo OCT-Spectralis.

Tabla 5 Características clínicas y medidas de espesor medio de la RNFL en la zona peripapilar que proporciona el OCT-Spectralis en la población de sujetos sanos y sujetos enfermos, con protocolo de adquisición de glaucoma.

Parámetro	Sujetos Sano (n=235 ojos)		Sujetos con EM (n=378 ojos)	
	Media (rango)	SD	Media (rango)	SD
Edad (años)	42.09 (23.25-72.78)	15.04	44.78 (24.74-72.04)	10.38
Mujeres (%)	80		73.35	
Espesor medio G ( $\mu\text{m}$ )	99.12 (65-211)	12.38	88.78 (34-273)	24.21
Espesor medio NS ( $\mu\text{m}$ )	72.13 (47-135)	13.14	61.48 (28-699)	38.01
Espesor medio N ( $\mu\text{m}$ )	133.2 (72-183)	19.12	120.26 (38-422)	35.28
Espesor medio NI ( $\mu\text{m}$ )	147.48 (71-190)	19.79	121.02 (7-488)	38.62
Espesor medio TI ( $\mu\text{m}$ )	75.95 (33-540)	33.16	71.46 (11-293)	26.04
Espesor medio T ( $\mu\text{m}$ )	106.13 (23-170)	24.25	98.07 (18-370)	29.82
Espesor medio TS ( $\mu\text{m}$ )	111.89 (59.183)	20.42	104.93 (0-330)	36.12

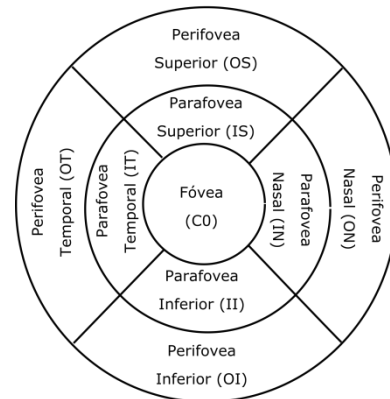
Con los cuatro tipos de datos que podemos obtener de este equipo, podemos plantear cuatro modelos de predicción según los datos que se obtienen de cada tipo:

Para el **primer y segundo tipo de datos**, que tenemos el espesor medio de la RNFL y de la capa de células ganglionares en el área macular planteamos una red neuronal con las siguientes entradas:

Se define el modelo de predicción con 12 parámetros de entrada:

1. Edad.
2. Sexo.
3. Mean C0: sector circular central o fovea.

4. Mean IN: sector nasal-interior o parafovea nasal.
5. Mean ON: sector nasal-exterior o perifovea nasal.
6. Mean IS: sector superior-interior o parafovea superior.
7. Mean OS: sector superior-exterior o perifovea superior.
8. Mean IT: sector temporal-interior o parafovea temporal.
9. Mean OT: sector temporal-exterior o perifovea temporal.
10. Mean II: sector inferior-interior o parafovea inferior.
11. Mean OI: sector inferior-exterior o perifovea inferior.
12. Center: Espesor del punto central.



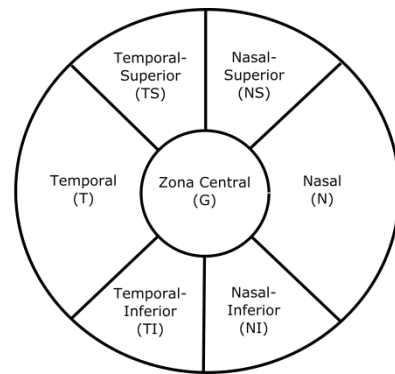
**Figura I.3 Representación de los cuadrantes en el área macular**

La representación gráfica de los cuadrantes o sectores que corresponden con las entradas para el modelo de predicción de las dos primeras bases de datos del equipo OCT-Spectralis se muestran en la Figura I.3.

Para el **tercer y cuarto tipo de datos** planteamos una red neuronal con las siguientes entradas:

Se define el modelo de predicción con  $(9 + i)$  parámetros de entrada:

1. Edad.
2. Sexo.
3. Mean G: espesor medio en el centro de la zona peripapilar.
4. Mean NS: espesor medio de la zona nasal-superior.
5. Mean N: espesor medio de la zona nasal.
6. Mean NI: espesor medio de la zona nasal-inferior.
7. Mean TI: espesor medio de la zona temporal-inferior.
8. Mean T: espesor medio de la zona temporal.
9. Mean TS: espesor medio de la zona temporal-superior.
10. Se realizarán diferentes pruebas agrupando los 768 puntos que se obtienen en el barrido del exterior de la zona peripapilar en grupos de diferente tamaño, realizando modelos con diferente número de entradas según los grupos que se realizan.



**Figura I.4 Representación de los segmentos en la zona peripapilar**

La representación de los sectores en los que se divide la zona peripapilar se muestran en la Figura I.4 y el círculo de barrido de los 768 puntos de adquisición del espesor se muestra con un círculo verde en la Figura 5.5.



## II. FORMULACIÓN DEL PERCEPTRÓN MULTICAPA.

En este anexo II se presenta la formulación necesaria para la programación completa del perceptrón multicapa empleado.

### II.1 Cálculo del valor en cada neurona.

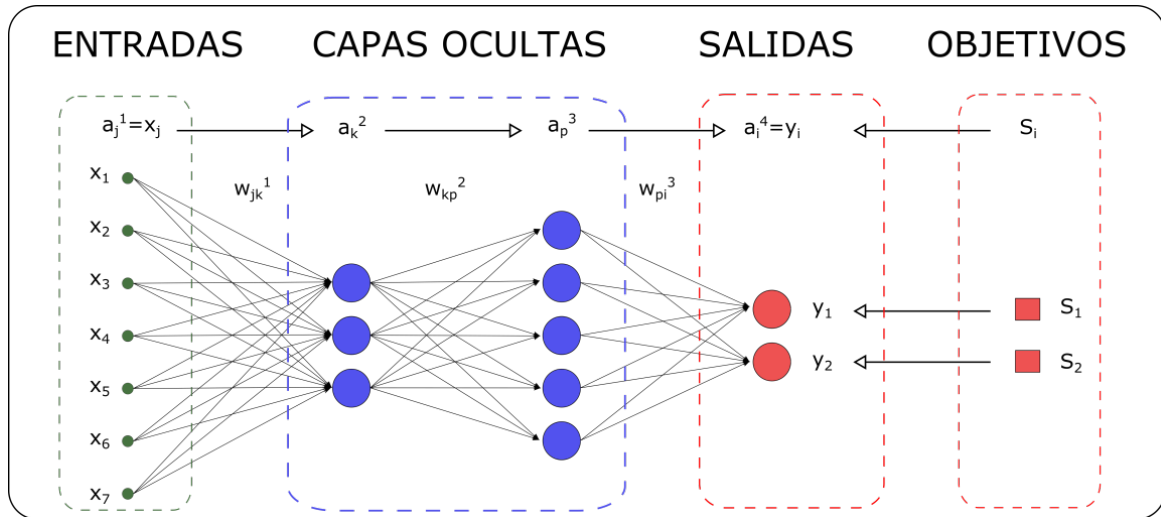


Figura II.1 Representación de perceptrón multicapa y terminología.

$a_i^1 = x_i \rightarrow$  neuronas capa 1 son igual a las entradas

$a_i^4 = y_i \rightarrow$  neuronas de la capa 4 son igual a las salidas

$$a_i^k = f\left(u_i^k + \sum_{j=1}^{n_{k-1}} a_j^{k-1} \cdot \omega_{ji}^{k-1}\right) \rightarrow \text{valor de las neuronas en cada capa}$$

Dónde:

$$\left[ \begin{array}{l} i = 1, \dots, n_k \rightarrow n^\circ \text{ de neurona en capa } k \\ j = 1, \dots, n_{k+1} \rightarrow n^\circ \text{ de neurona en capa } k + 1 \\ k = 1, \dots, 4 \rightarrow n^\circ \text{ de capas} \end{array} \right]$$

Y la función sigmoide:

$$f(x) = \frac{1}{1 + e^{-x}} \rightarrow f'(x) = f(x) \cdot (1 - f(x))$$

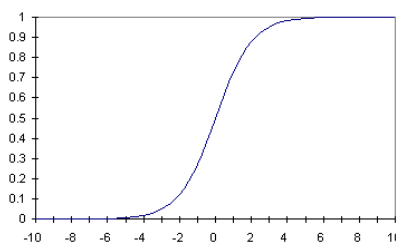


Figura II.2 Representación de la función sigmoide.

## II.2 BackPropagation (descenso del gradiente).

$$\begin{cases} \omega_{ji}^k = \omega_{ji}^k - \alpha \cdot \frac{\partial e}{\partial \omega_{ji}^k} \\ u_i^k = u_i^k - \alpha \cdot \frac{\partial e}{\partial u_i^k} \end{cases}$$

Dónde:

$\omega_{ji}^k \rightarrow$  actualización del valor del peso entre la neurona  $j$  e  $i$

$u_i^k \rightarrow$  actualización del umbral de activación de la neurona  $i$

## II.3 Cálculo de las derivadas del error respecto al peso en cada neurona.

$$\begin{cases} j = 1, \dots, n_1 \rightarrow \text{número de entradas} \\ k = 1, \dots, n_2 \rightarrow \text{número neuronas capa 2} \\ p = 1, \dots, n_3 \rightarrow \text{número neuronas capa 3} \\ i = 1, \dots, n_4 \rightarrow \text{número de salidas} \end{cases}$$

**Salida en función de los pesos entre capas**

$$\left[ \begin{aligned} \frac{\partial y_i}{\partial \omega_{jk}^1} &= x_j \cdot a_k^2 \cdot (1 - a_k^2) \cdot \left[ \sum_{p=1}^{n_3} (\omega_{kp}^2 \cdot a_p^3 \cdot (1 - a_p^3) \cdot \omega_{pi}^3) \right] \cdot y_i \cdot (1 - y_i) \\ \frac{\partial y_i}{\partial \omega_{kp}^2} &= a_k^2 \cdot a_p^3 \cdot (1 - a_p^3) \cdot \omega_{pi}^3 \cdot y_i \cdot (1 - y_i) \\ \frac{\partial y_i}{\partial \omega_{pi}^3} &= a_p^3 \cdot y_i \cdot (1 - y_i) \end{aligned} \right]$$

**Salida en función de los umbrales de cada neurona**

$$\left[ \begin{aligned} \frac{\partial y_i}{\partial u^2} &= a_k^2 \cdot (1 - a_k^2) \cdot \left[ \sum_{p=1}^{n_3} (\omega_{kp}^2 \cdot a_p^3 \cdot (1 - a_p^3) \cdot \omega_{pi}^3) \right] \cdot y_i \cdot (1 - y_i) \\ \frac{\partial y_i}{\partial u^3} &= a_p^3 \cdot (1 - a_p^3) \cdot \omega_{pi}^3 \cdot y_i \cdot (1 - y_i) \\ \frac{\partial y_i}{\partial u^4} &= y_i \cdot (1 - y_i) \end{aligned} \right]$$

$$\frac{\partial e}{\partial X} = \frac{\partial e}{\partial y_i} \cdot \frac{\partial y_i}{\partial X} = -(s_i - y_i) \cdot \frac{\partial y_i}{\partial X}$$

**Error en función de los pesos entre capas**

$$\left[ \begin{array}{l} \frac{\partial e}{\partial \omega_{jk}^1} = x_j \cdot a_k^2 \cdot (1 - a_k^2) \cdot \sum_{p=1}^{n_3} \left( \omega_{kp}^2 \cdot a_p^3 \cdot (1 - a_p^3) \cdot \sum_{i=1}^{n_4} (\omega_{pi}^3 \cdot y_i \cdot (1 - y_i) \cdot (-(s_i - y_i))) \right) \\ \frac{\partial e}{\partial \omega_{kp}^2} = a_k^2 \cdot a_p^3 \cdot (1 - a_p^3) \cdot \sum_{i=1}^{n_4} (\omega_{pi}^3 \cdot y_i \cdot (1 - y_i) \cdot (-(s_i - y_i))) \\ \frac{\partial e}{\partial \omega_{pi}^3} = a_p^3 \cdot y_i \cdot (1 - y_i) \cdot (-(s_i - y_i)) \end{array} \right]$$

**Error en función de los umbrales de cada neurona**

$$\left[ \begin{array}{l} \frac{\partial e}{\partial u_k^2} = a_k^2 \cdot (1 - a_k^2) \cdot \sum_{p=1}^{n_3} \left( \omega_{kp}^2 \cdot a_p^3 \cdot (1 - a_p^3) \cdot \sum_{i=1}^{n_4} (\omega_{pi}^3 \cdot y_i \cdot (1 - y_i) \cdot (-(s_i - y_i))) \right) \\ \frac{\partial e}{\partial u_p^3} = a_p^3 \cdot (1 - a_p^3) \cdot \sum_{i=1}^{n_4} (\omega_{pi}^3 \cdot y_i \cdot (1 - y_i) \cdot (-(s_i - y_i))) \\ \frac{\partial e}{\partial u_i^4} = y_i \cdot (1 - y_i) \cdot (-(s_i - y_i)) \end{array} \right]$$

**Error en función de los pesos entre capas**

$$\left[ \begin{array}{l} \frac{\partial e}{\partial \omega_{jk}^1} = x_j \cdot \delta_k^2 \\ \frac{\partial e}{\partial \omega_{kp}^2} = a_k^2 \cdot \delta_p^3 \\ \frac{\partial e}{\partial \omega_{pi}^3} = a_p^3 \cdot \delta_i^4 \end{array} \right]$$

**Error en función de los umbrales de cada neurona**

$$\left[ \begin{array}{l} \frac{\partial e}{\partial u_k^2} = \delta_k^2 \\ \frac{\partial e}{\partial u_p^3} = \delta_p^3 \\ \frac{\partial e}{\partial u_i^4} = \delta_i^4 \end{array} \right]$$

Dónde:

$$\left\{ \begin{array}{l} \delta_k^2 = a_k^2 \cdot (1 - a_k^2) \cdot \sum_{p=1}^{n_3} \left( \omega_{kp}^2 \cdot a_p^3 \cdot (1 - a_p^3) \cdot \sum_{i=1}^{n_4} (\omega_{pi}^3 \cdot y_i \cdot (1 - y_i) \cdot (-(s_i - y_i))) \right) \\ \delta_p^3 = a_p^3 \cdot (1 - a_p^3) \cdot \sum_{i=1}^{n_4} (\omega_{pi}^3 \cdot y_i \cdot (1 - y_i) \cdot (-(s_i - y_i))) = a_p^3 \cdot (1 - a_p^3) \cdot \sum_{i=1}^{n_4} (\omega_{pi}^3 \cdot \delta_i^4) \\ \delta_i^4 = y_i \cdot (1 - y_i) \cdot (-(s_i - y_i)) \end{array} \right\}$$

### III. EXPERIMENTOS REALIZADOS Y RESULTADOS.

Como se ha comentado en apartado 5.2.1, se van a realizar 5 modelos de predicción basados en el perceptrón multicapa y otros 5 modelos basados en las SVM, correspondiendo a cada una de las bases de datos en las que se ha dividido los resultados de las pruebas médicas de OCT:

- 1) Modelo 1: base de datos obtenida mediante el equipo CIRRUS HD-OCT.
- 2) Modelo 2: base de datos correspondiente a la capa de células ganglionares de la retina en la zona macular, obtenida con el equipo OCT-Spectralis.
- 3) Modelo 3: base de datos correspondiente a la RNFL en la zona macular, obtenida con el equipo OCT-Spectralis.
- 4) Modelo 4: base de datos correspondiente a la RNFL obtenida con el protocolo axonal del equipo OCT-Spectralis.
- 5) Modelo 5: base de datos que corresponde a la RNFL obtenida con el protocolo glaucoma del equipo OCT-Spectralis.

Se van a realizar una serie de pruebas o experimentos con los modelos de predicción para conseguir el diagnóstico más fiable en cada uno de los dichos modelos y poder evaluar cuál es el protocolo que tiene mayor capacidad de diagnóstico de la esclerosis múltiple, así como poder comparar la capacidad como clasificadores del perceptrón multicapa y de las SVM en este problema concreto.

#### III.1 Predicción mediante el Perceptrón Multicapa.

En primer lugar, se debe decidir sobre la agrupación o no de datos. Se decide probar a separar los pacientes por sexo, ya que se observa una mayor probabilidad de que un enfermo de EM sea mujer que hombre, alrededor del 70%, y como consecuencia se puede llegar a alcanzar una predicción con mayor precisión. Esta prueba se va a realizar con los datos de la primera base de datos, es decir, con los datos obtenidos con el equipo CIRRUS HD-OCT.

- 1) Con todos los datos disponibles: Número de sujetos totales es de 200, de los cuales 160 se emplearán para entrenar la red neuronal y 40 para su validación.
- 2) Datos del sexo femenino. El número de mujeres totales es de 143, de los cuales 116 se emplearán para el entrenamiento y 27 para la validación.
- 3) Datos del sexo masculino: El número de hombres es de 57, de los cuales 44 se emplearán en el entrenamiento y 13 para la validación de los modelos.

Se realiza una la prueba con una misma arquitectura de red para comprar los resultados:

**Tabla 6. Tabla de precisión en el diagnóstico utilizando todos los datos disponibles o separando los datos por sexo para una arquitectura de red de una sola capa oculta, y 100 ciclos de aprendizaje.**

Grupo	Razón de aprendizaje	nº neuronas capa oculta	Precisión de entrenamiento	Precisión de validación
<b>Ambos sexos</b>	<b>0,9</b>	4	83	67
		<b>5</b>	<b>81</b>	<b>80</b>
		6	83	67
Mujeres	0,9	4	90	63
		5	78	72
		6	91	63
Hombres	0,9	4	76	63
		5	76	53
		6	68	46

Como se puede observar en la Tabla 6, el mejor modelo de diagnóstico será el que se obtenga utilizando el conjunto de datos completo como patrones para el entrenamiento de la red neuronal. Así pues, se decide **no separar los datos por sexo** en ningún modelo realizado con el perceptrón multicapa.

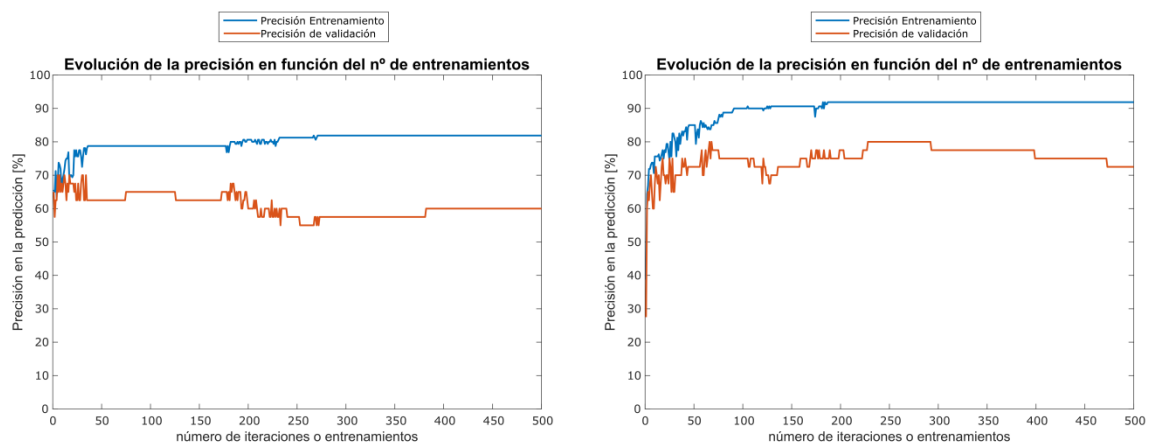
Se pasa a estudiar cada modelo por separado siguiendo el mismo procedimiento con todas y cada una de las bases de datos disponibles para llegar a obtener la mejor estructura de perceptrón en cada una de ellas y así comparar el valor como biomarcador de cada una de las diferentes pruebas realizadas con la OCT.

### III.1.1 Modelo 1: base de datos obtenida mediante el equipo CIRRUS HD-OCT.

En primer lugar, se realiza una serie de pruebas cambiando los parámetros y la arquitectura del perceptrón multicapa para llegar a obtener la mejor predicción posible en el diagnóstico con los datos proporcionados con la primera de las bases de datos que vamos a emplear.

#### III.1.1.1 Generalización frente a memorización.

Para evitar el fenómeno de sobre-aprendizaje comentado anteriormente, se debe obtener el número de iteraciones óptima para cada uno de los modelos que se van a realizar. Esto conlleva que se debe estudiar la evolución de la precisión en el diagnóstico con el número de iteraciones como un parámetro fundamental en la consecución de un modelo preciso. Para ver cuando se produce el fenómeno de sobre-aprendizaje se muestra a continuación dos casos en los que se observa la evolución de la predicción de entrenamiento y la de validación con el número de ciclos de aprendizaje:



**Figura III.1 Evolución de la precisión de entrenamiento (línea azul) y la precisión de validación o generalización (línea naranja) frente al número de iteraciones para la base de datos perteneciente al equipo CIRRUS HD-OCT (modelo 1) en dos casos distintos.**

La Figura III.1 muestra la aparición del fenómeno de sobre-aprendizaje para dos estructuras de perceptrón multicapa diferentes, a la derecha, modelo con una capa oculta y 5 neuronas en ésta, 16 parámetros de entrada y una razón de aprendizaje  $\alpha = 0.9$ . A la izquierda, modelo con 4 neuronas ocultas, 16 parámetros de entrada y una razón de aprendizaje  $\alpha = 0.5$ . Se puede observar que comienza a disminuir la precisión de validación a partir de unas 200 iteraciones, aunque aumente la precisión de entrenamiento.

La explicación de este fenómeno es que al principio la red se adapta progresivamente al conjunto de entrenamiento, acomodándose al problema y mejorando la generalización, sin embargo, en un momento dado el sistema se ajusta demasiado a los patrones utilizados en el entrenamiento, incluso al ruido presente en ellos, creciendo el error de generalización [23].

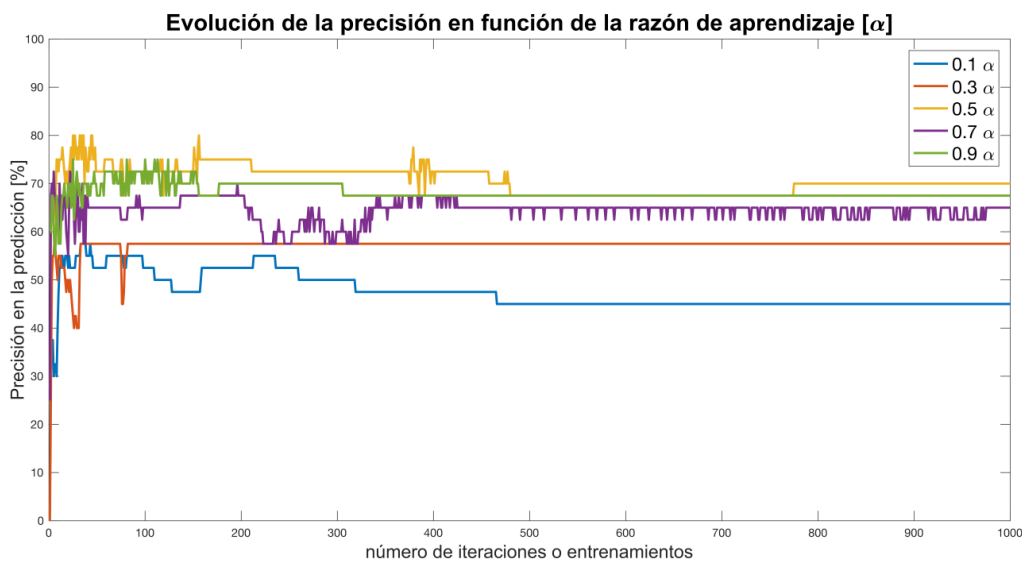
Para evitar este fenómeno indeseado en el proceso de aprendizaje debemos obtener el número de iteraciones óptimo para la arquitectura de red empleada y para el problema tratado. Este punto óptimo dependerá fuertemente de la naturaleza del problema a tratar y no queda más

remedio que estudiar el fenómeno para cada modelo que se va a implementar mediante el perceptrón multicapa.

### III.1.1.2 Selección de la razón de aprendizaje.

El cambio de un peso en la red es proporcional al gradiente del error como se refleja en las ecuaciones (5) y (6). Esta proporcionalidad viene determinada por la razón de aprendizaje ( $\alpha$ ), marcando así la velocidad de convergencia de la red. El valor de este parámetro suele estar comprendido entre 0-1 y se debe obtener la tasa de aprendizaje óptima para el problema a tratar sabiendo que valores cercanos a 1 ayudarán a una rápida convergencia, pero pueden tener efectos negativos como que el método salte un mínimo o incluso oscile alrededor del mínimo.

Para seleccionar una razón de aprendizaje óptima se realiza un estudio para la arquitectura con una sola capa oculta variando la tasa de aprendizaje  $\alpha$  y comparando las precisiones alcanzadas.



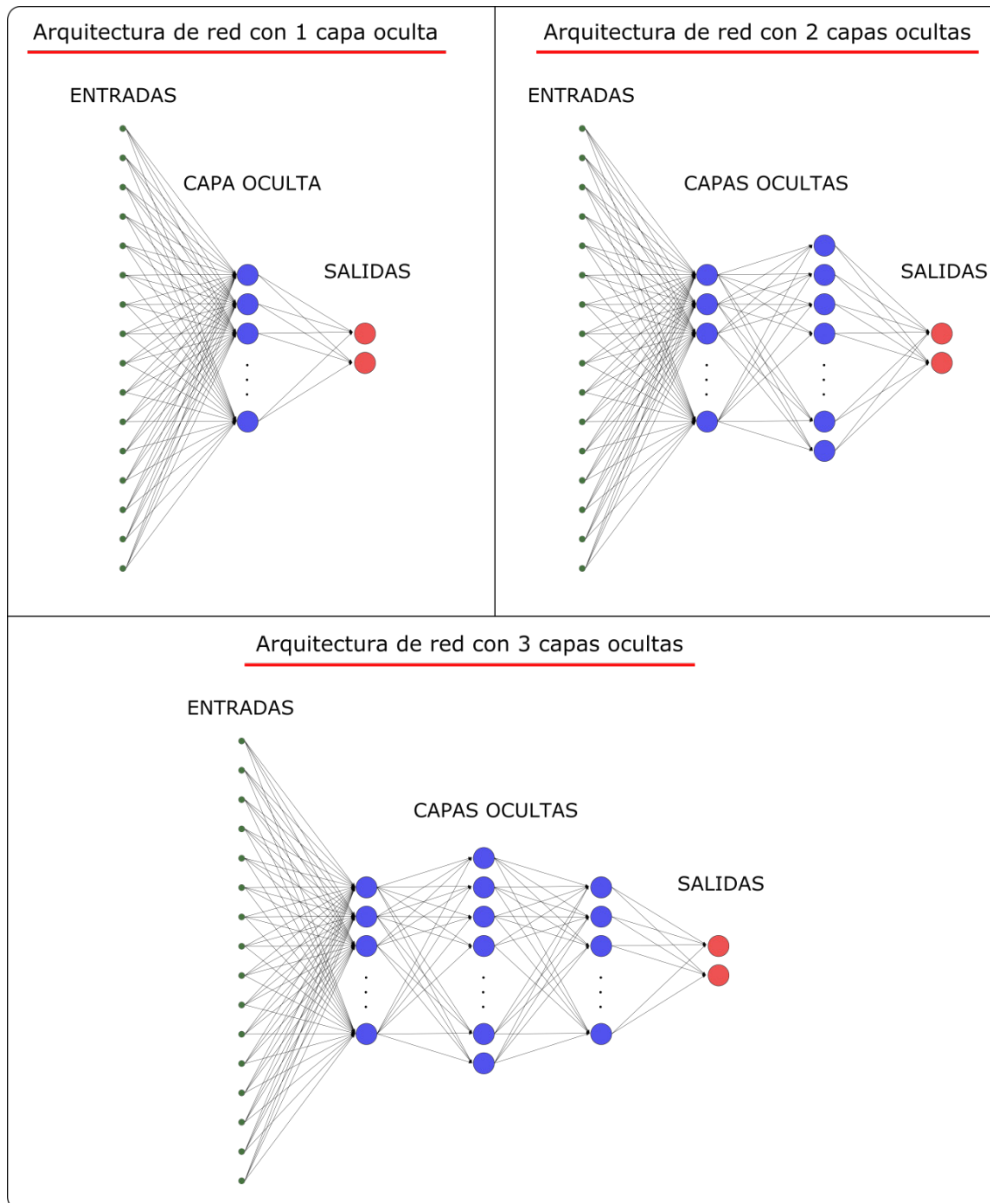
**Figura III.2** Gráfica de la evolución en la precisión de validación alcanzada para un perceptrón multicapa con una única capa oculta, con 5 neuronas en la misma, para diferentes razones de aprendizaje (ver leyenda), en función del número de entrenamientos realizados.

En la Figura III.2 se muestra la evolución de los errores de validación para las diferentes razones de aprendizaje (0.1-0.9). Se observa que a tasas pequeñas se retrasa la convergencia, mientras que a razones de aprendizaje altas se acelera la convergencia del algoritmo aunque puede volverse algo inestable (como en el caso de  $\alpha = 0.7$ ). Para la resolución del problema que se está tratando se determina la **tasa de aprendizaje  $\alpha = 0.5$**  como la mejor, ya que se alcanza la mayor predicción (línea amarilla), teniendo en cuenta que aproximadamente a los 200 ciclos de aprendizaje se produce el fenómeno anteriormente comentado de sobre-entrenamiento.

### III.1.1.3 Arquitectura del Perceptrón multicapa.

Se debe encontrar la arquitectura de perceptrón multicapa que mejor se ajusta al problema que deseamos resolver. Para ello se van a realizar una serie de pruebas con diferentes arquitecturas de perceptrón, variando el número de capas ocultas y el número de neuronas en cada una de éstas capas ocultas. La estructura de la red óptima (número de capas y neuronas ocultas) depende de la complejidad del problema a resolver, siendo recomendable conseguir la arquitectura más simple que resuelve el problema cumpliendo los objetivos propuestos [33].

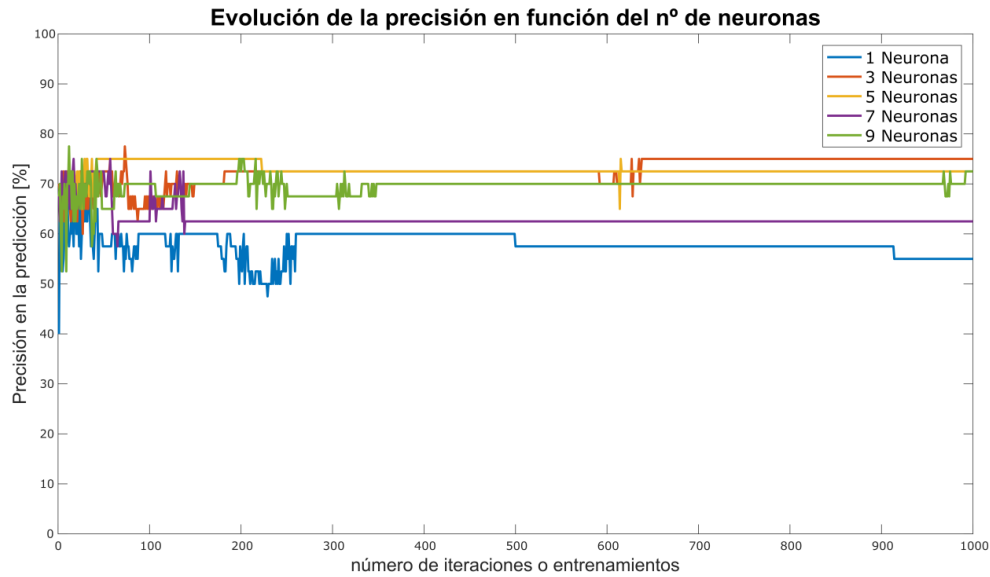
En la Figura III.3 se muestran las diferentes arquitecturas de perceptrón multicapa que se van a estudiar en el caso que se está tratando para el diagnóstico de la EM.



**Figura III.3** Arquitecturas de perceptrón multicapa que se van a emplear en el diagnóstico de la enfermedad. La primera imagen corresponde a la arquitectura más simple, con una sola capa de neuronas ocultas, la segunda corresponde a una estructura de red con dos capas ocultas y la tercera a tres.

### III.1.1.3.1 Una sola capa oculta.

En primer lugar, se utiliza la arquitectura de red más simple, la de una sola capa de neuronas ocultas. Para evaluar cómo afecta añadir o quitar neuronas de la capa oculta a la capacidad de diagnóstico se realiza una serie de simulaciones con diferente número de éstas y se comprara la precisión en el diagnóstico.



**Figura III.4** Gráfica de la evolución en la precisión de validación alcanzada para un perceptrón multicapa con una única capa oculta, con razón de aprendizaje 0.5, variando el número de neuronas en la capa oculta (ver leyenda), en función del número de entrenamientos realizados.

Se puede apreciar en la Figura III.4 que la arquitectura con 5 neuronas en la capa oculta es en este caso la que mejores resultados de diagnóstico ofrece, ya que con poco más de 200 ciclos de aprendizaje obtiene cerca de un 80% de precisión en la predicción. No se puede sacar una conclusión general acerca del efecto de añadir neuronas en la capa oculta, ya que el número de neuronas adecuado depende del problema a resolver [25].

### III.1.1.3.2 Dos capas de neuronas ocultas.

Del mismo modo que en el apartado anterior se debe determinar el número óptimo de neuronas en cada una de las dos capas ocultas para mejorar la precisión en el diagnóstico. Además, para esta arquitectura se estudia el número de iteraciones necesarias para conseguir el mejor diagnóstico y evitar el fenómeno de sobreaprendizaje.

En la Figura III.5 se observa que, igual que en el caso anterior, será suficiente con realizar 200 iteraciones en el proceso de entrenamiento, evitando así el fenómeno comentado, ya que éste aparece alrededor del ciclo 300.

En segundo lugar, se realiza un estudio variando en número de neuronas en las dos capas ocultas entre 1 y 8. La razón de aprendizaje utilizada es igual a 0.5, el número de ciclos de entrenamiento 200.



**Figura III.5** Evolución de la precisión de diagnóstico para una arquitectura de dos capas de neuronas con 7 neuronas en la primera capa oculta, 1 neurona en la segunda y una razón de aprendizaje  $\alpha = 0.5$ .



**Tabla 7. Precisiones en la predicción alcanzadas con una arquitectura de red de dos capas ocultas, variando el número de neuronas en cada capa.**

		nº de neuronas en la 2ª capa oculta							
		1	2	3	4	5	6	7	8
nº de neuronas en la 1ª capa oculta	1	52,50	67,50	57,50	62,50	70,00	72,50	67,50	67,50
	2	75,00	72,50	67,50	70,00	77,50	70,00	72,50	67,50
	3	77,50	67,50	65,00	75,00	67,50	72,50	62,50	62,50
	4	72,50	75,00	70,00	75,00	75,00	75,00	65,00	70,00
	5	67,50	82,50	67,50	80,00	60,00	75,00	75,00	65,00
	6	70,00	67,50	67,50	72,50	67,50	72,50	70,00	65,00
	7	62,50	67,50	65,00	70,00	80,00	62,50	65,00	65,00
	8	67,50	70,00	65,00	72,50	72,50	70,00	75,00	77,50

Tras realizar el estudio con una arquitectura de red de dos capas ocultas (Tabla 7) se puede observar que se ha conseguido mejorar en cierto punto la precisión de diagnóstico. Más concretamente, con un perceptrón que disponga de 5 neuronas en la primera capa oculta y 2 neuronas en la segunda aumentan la precisión hasta un 82.5%.

### III.1.1.3.3 Tres capas de neuronas ocultas.

Por último, se decide realizar una serie de simulaciones con una estructura de red con 3 capas ocultas con el objetivo de ver si esto puede mejorar la precisión de diagnóstico. Del mismo modo que en los apartados anteriores se va a variar el número de neuronas en cada una de las capas del perceptrón multicapa para ver la evolución en el diagnóstico.

**Tabla 8 Precisiones alcanzadas con una arquitectura de red de 3 capas ocultas. Tabla izquierda con 1 neurona en la primera capa, tabla central con 2 neuronas y tabla derecha con 3.**

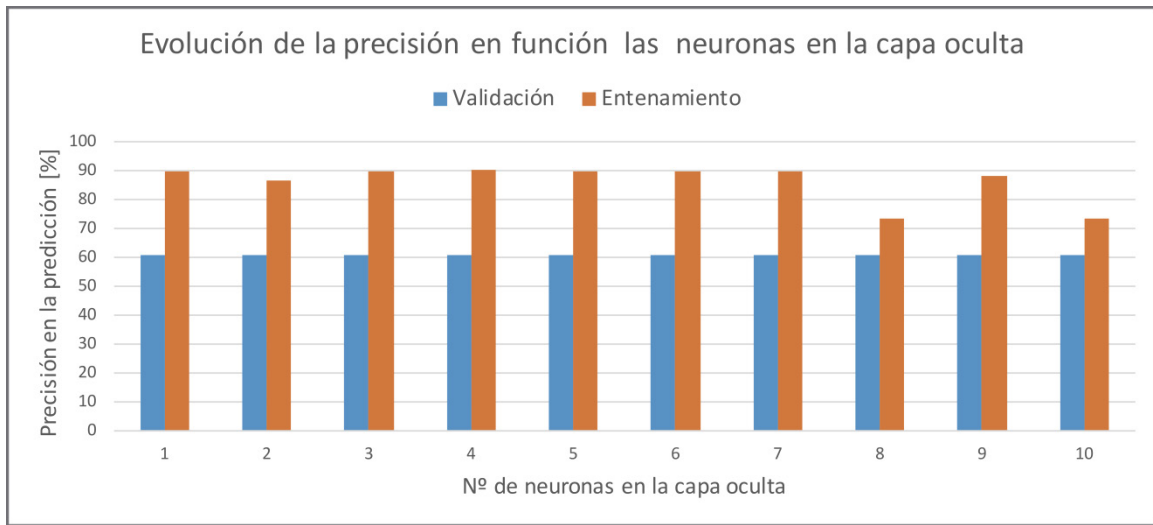
		1ª capa oculta = 1 neurona						1ª capa oculta = 2 neuronas						1ª capa oculta = 3 neuronas					
		nº de neuronas en la 2ª capa oculta						nº de neuronas en la 2ª capa oculta						nº de neuronas en la 2ª capa oculta					
		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
nº de neuronas en la 1ª capa oculta	1	55,0	65,0	67,5	55,0	50,0	72,5	72,5	67,5	75,0	72,5	77,5	72,5	57,5	67,5	75,0	67,5	70,0	72,5
	2	70,0	70,0	70,0	57,5	72,5	67,5	57,5	67,5	67,5	67,5	75,0	62,5	67,5	70,0	67,5	67,5	75,0	60,0
	3	12,5	52,5	60,0	55,0	52,5	60,0	62,5	70,0	72,5	62,5	67,5	67,5	72,5	70,0	50,0	52,5	0,0	70,0
	4	57,5	70,0	70,0	65,0	0,0	70,0	67,5	75,0	67,5	72,5	75,0	70,0	67,5	67,5	70,0	70,0	67,5	65,0
	5	70,0	62,5	75,0	75,0	72,5	67,5	50,0	55,0	65,0	70,0	60,0	70,0	62,5	67,5	60,0	67,5	65,0	75,0
	6	55,0	75,0	52,5	17,5	0,0	67,5	67,5	70,0	50,0	72,5	0,0	72,5	77,5	62,5	67,5	70,0	70,0	65,0

En estas simulaciones (Tabla 8) se observa que la precisión en el diagnóstico se reduce algo respecto a la arquitectura empleada en el apartado anterior con dos capas ocultas. La máxima precisión obtenida en la prueba realizada es del 77,5% frente al 82,5% obtenido en el apartado anterior.

### III.1.2 Modelo 2: base de datos correspondiente a la capa de células ganglionares de la retina en la zona macular, obtenida con el equipo OCT-Spectralis.

Tras estudiar el número de ciclos de entrenamiento óptimo se llega a la misma conclusión que en el modelo 1, con 200 ciclos de aprendizaje es suficiente para llegar a la mejor precisión de diagnóstico evitando el fenómeno de sobre-aprendizaje en la red.

En primer lugar se va a realizar un estudio únicamente con la arquitectura de una sola capa de neuronas ocultas variando el número de éstas (Figura III.6).



**Figura III.6 Evolución de la precisión de entrenamiento y validación en función del número de neuronas en la capa oculta para una arquitectura de red de una sola capa oculta, razón de aprendizaje  $\alpha = 0.5$  y número de ciclos de entrenamiento igual a 200.**

Con los datos que se obtienen de la segunda base de datos se consigue poco más del 60% de precisión de diagnóstico con el conjunto de datos de validación, es decir, apenas tiene valor de diagnóstico el modelo 2 utilizando el perceptrón multicapa con una sola capa oculta.

Se realiza a continuación el mismo estudio variando el número de neuronas pero con la arquitectura de red de dos capas ocultas (Tabla 9).

**Tabla 9 Precisión de validación para la arquitectura de dos capas ocultas en el modelo 2, razón de aprendizaje  $\alpha = 0.5$  y 200 ciclos de entrenamiento.**

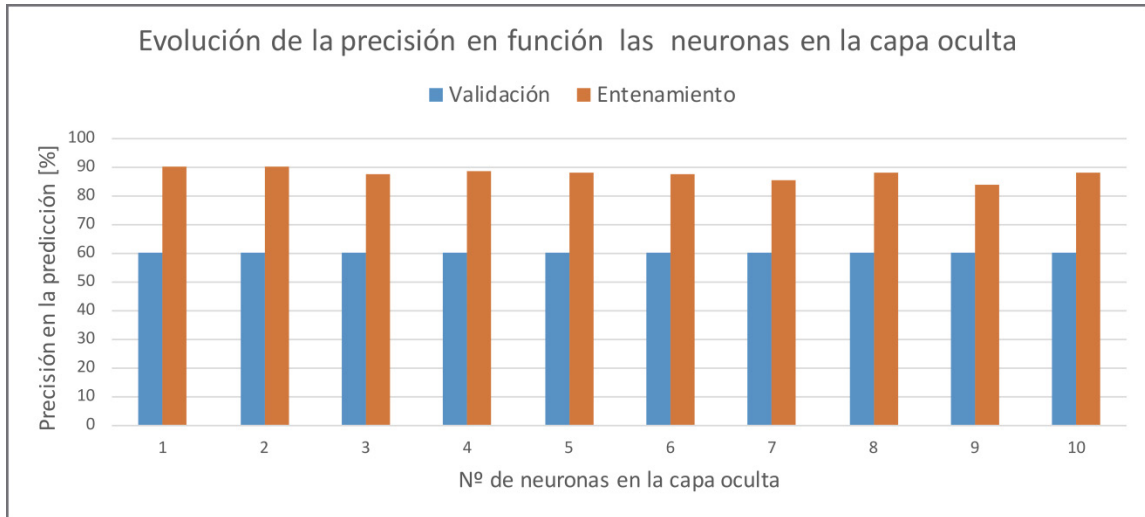
		nº de neuronas en la 2ª capa oculta							
		1	2	3	4	5	6	7	8
nº de neuronas en la 1ª capa oculta	1	63,41	70,73	78,05	78,05	85,37	82,93	70,73	87,80
	2	78,05	87,80	78,05	78,05	78,05	75,61	78,05	78,05
	3	80,49	78,05	87,80	78,05	87,80	78,05	90,24	78,05
	4	85,37	78,05	85,37	75,61	80,49	80,49	82,93	78,05
	5	78,05	85,37	75,61	78,05	78,05	80,49	78,05	60,98
	6	75,61	78,05	85,37	87,80	85,37	78,05	87,80	87,80
	7	85,37	80,49	75,61	85,37	85,37	87,80	73,17	80,49
	8	82,93	75,61	85,37	90,24	85,37	80,49	85,37	85,37

Añadiendo una segunda capa de neuronas ocultas se puede apreciar que la precisión de validación aumenta considerablemente respecto a una sola capa de neuronas ocultas. La mejor precisión obtenida es de un 90% tanto para la arquitectura de 3 neuronas en la primera capa oculta y 7 en la segunda como para la arquitectura de 8 neuronas en la primera capa y 4 en la segunda.

### III.1.3 Modelo 3: base de datos correspondiente a la RNFL en la zona macular, obtenida con el equipo OCT-Spectralis.

De la misma manera que para la base de datos anterior, se estudia el número de ciclos de aprendizaje óptimos obteniendo igualmente un número de ciclo igual a 200.

Se realiza la misma prueba variando el número de neuronas para la arquitectura de una sola capa oculta (Figura III.7) obteniendo una precisión del 60% sin importar el número de neuronas en la capa oculta. Se puede sacar la misma conclusión a cerca de la escasa capacidad de predicción del modelo 3 con el perceptrón multicapa.



**Figura III.7 Evolución de la precisión de entrenamiento y validación en función del número de neuronas en la capa oculta para una arquitectura de red de una sola capa oculta, razón de aprendizaje  $\alpha = 0.5$  y número de ciclos de entrenamiento igual a 200.**

Se realiza a continuación las simulaciones correspondientes para comprobar la precisión que se puede llegar a alcanzar con la arquitectura de dos capas ocultas:

**Tabla 10 Precisión de validación para la arquitectura de dos capas ocultas en el modelo 3, razón de aprendizaje  $\alpha = 0.5$  y 200 ciclos de entrenamiento.**

		nº de neuronas en la 2ª capa oculta							
		1	2	3	4	5	6	7	8
nº de neuronas en la 1ª capa oculta	1	85,00	82,50	90,00	95,00	77,50	67,50	90,00	72,50
	2	82,50	70,00	70,00	82,50	72,50	52,50	72,50	42,50
	3	70,00	90,00	65,00	75,00	67,50	65,00	67,50	70,00
	4	72,50	72,50	82,50	72,50	70,00	72,50	57,50	65,00
	5	65,00	70,00	67,50	70,00	72,50	72,50	72,50	72,50
	6	67,50	72,50	65,00	70,00	65,00	70,00	67,50	65,00
	7	87,50	72,50	72,50	57,50	75,00	72,50	67,50	62,50
	8	75,00	70,00	62,50	90,00	72,50	72,50	70,00	52,50

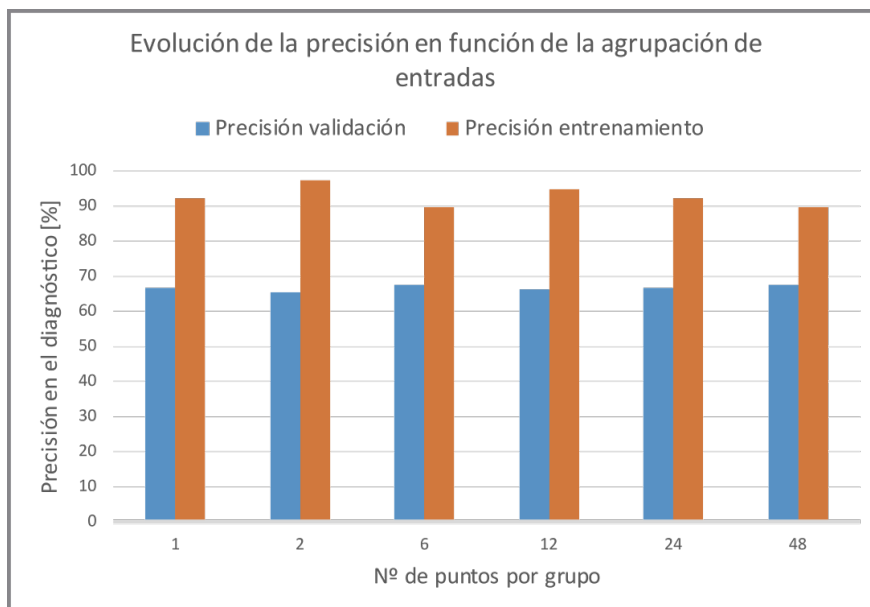
Igual que sucede en el modelo anterior, la precisión en el diagnóstico crece mucho al introducir una nueva capa de neuronas oculta. Más concretamente la precisión más alta que se alcanza en este modelo con dos capas ocultas es para una arquitectura con una neurona en la primera capa y cuatro en la segunda, alcanzando un 95% de precisión.

### III.1.4 Modelo 4: base de datos correspondiente a la RNFL obtenida con el protocolo axonal del equipo OCT-Spectralis.

A diferencia de los modelos anteriores, en el modelo 4 y 5 se introduce una nueva variable en la estructura de la red neuronal. Esta nueva variable es el número de datos que se introducirán como entrada en la red neuronal. Como se ha comentado en el apartado 5.2.1, el número de entradas dependerán del número de puntos del barrido peripapilar que incluyamos en cada grupo. Se realizan 6 pruebas, agrupando un número de puntos diferente en cada una de ellas (ver 5.2.1.2):

- 1) Grupos de 1 punto, es decir, introducimos todos los puntos como entrada a la red neuronal ( $9 + 768 = 777$  entradas).
- 2) Grupos de 2 puntos, es decir, agrupamos de dos en dos los puntos del barrido peripapilar, introduciendo cada grupo como entrada en el modelo ( $9 + \frac{768}{2} = 393$  entradas).
- 3) Grupos de 6 puntos, ( $9 + \frac{768}{6} = 137$  entradas).
- 4) Grupos de 12 puntos, ( $9 + \frac{768}{12} = 73$  entradas).
- 5) Grupos de 24 puntos, ( $9 + \frac{768}{24} = 41$  entradas).
- 6) Grupos de 48 puntos, ( $9 + \frac{768}{48} = 25$  entradas).

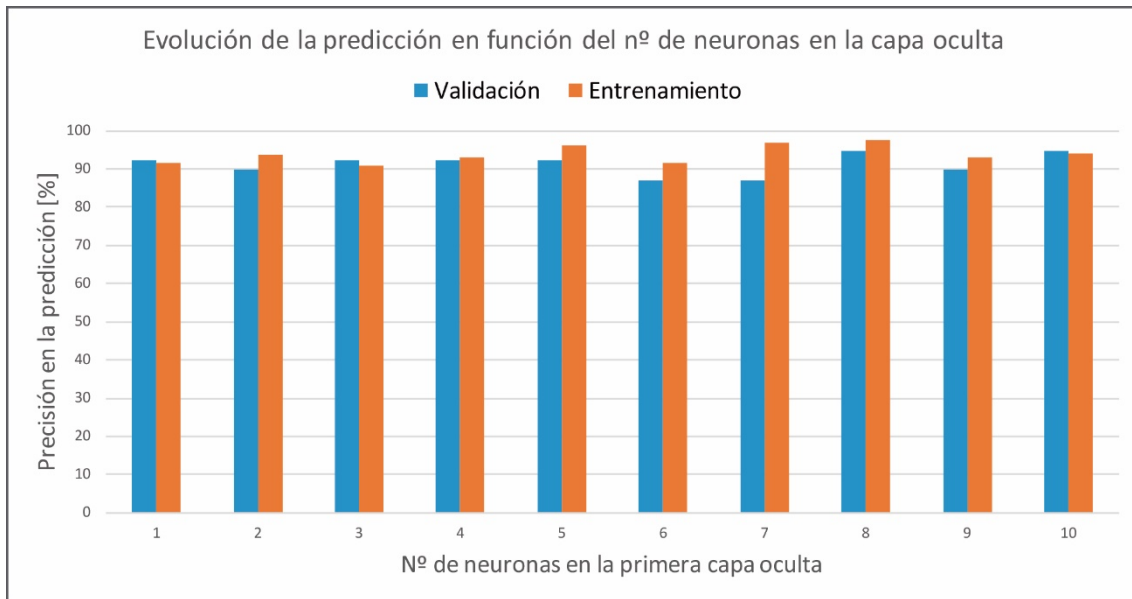
Estas pruebas se realizan con una arquitectura de red de una sola capa oculta, 5 neuronas en ésta y una razón de aprendizaje igual a 0,5. Se representa a continuación la evolución de la precisión en el diagnóstico de la EM según el número de puntos en cada grupo del barrido peripapilar:



**Figura III.8** Evolución de la precisión de entrenamiento y la precisión de validación frente al número de entradas para la base de datos perteneciente al equipo OCT-Spectralis (modelo 4) y una arquitectura de red de una sola capa oculta, con 5 neuronas en ésta, variando el número de puntos por grupo (1, 2, 6, 12, 24 y 48), una razón de aprendizaje  $\alpha = 0.5$  y un número de iteraciones igual a 500.

Como puede observarse en la Figura III.8, tanto las precisiones de entrenamiento como las de generalización que se alcanzan son muy parecidas, alrededor de 90% y 67% respectivamente, variando el número de entradas utilizado. Se decide **utilizar todos los puntos disponibles en el barrido peripapilar como entradas** y de esta manera no perder nada de información al realizar los valores medios. Como inconveniente es el mayor gasto computacional que tiene la simulación el gran número de entradas (777) en el perceptrón multicapa.

Una vez se ha seleccionado el número de entradas se varía el número de neuronas en la capa oculta del mismo modo que en los apartados anteriores:

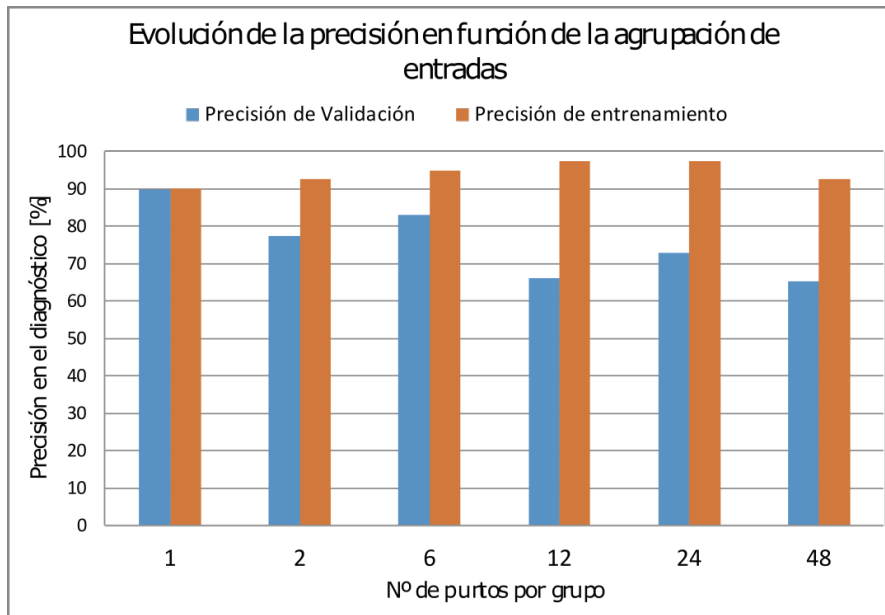


**Figura III.9** Evolución de la precisión de entrenamiento frente al número de entradas para la base de datos perteneciente a la OCT-Spectralis (modelo 4) con una arquitectura de red de una sola capa oculta y variando el número de neuronas en ésta

Tras realizar esta prueba se obtiene la máxima precisión en la predicción para los datos de validación con la arquitectura de red de 8 neuronas en la capa oculta, llegando a alcanzar una precisión de 94,87%. Además, la precisión en el grupo de entrenamiento también es la más alta con un 97,43%.

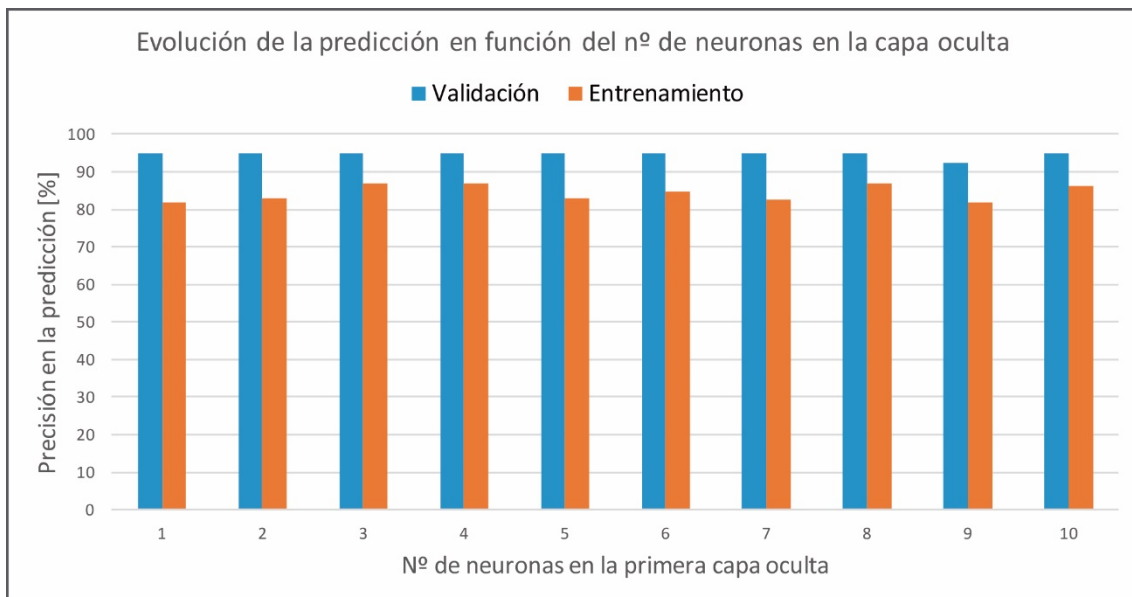
### **III.1.5 Modelo 5: base de datos que corresponde a la RNFL obtenida con el protocolo glaucoma del equipo OCT-Spectralis.**

Exactamente del mismo modo que en el modelo 4, lo primero que se ha de seleccionar es el número de puntos que se van a incluir como parámetros de entrada para el diagnóstico de la EM mediante la base de datos 5. Las pruebas realizadas (Figura III.10) muestran que utilizando todos los puntos de barrido peripapilar se obtienen precisiones de entrenamiento y validación de cerca del 90%, siendo la de validación mucho mayor que en los otros casos simulados. Por esta razón se decide realizar las siguientes pruebas con **todos los puntos disponibles** para el entrenamiento de la red neuronal.



**Figura III.10** Evolución de la precisión de entrenamiento y la precisión de validación frente al número de entradas para la base de datos perteneciente al equipo OCT-Spectralis (modelo 5) y una arquitectura de red de una sola capa oculta, con 5 neuronas en ésta, variando el número de puntos por grupo (1, 2, 6, 12, 24 y 48), una razón de aprendizaje  $\alpha = 0.5$  y un número de iteraciones igual a 200.

Se realiza el mismo estudio que en el apartado anterior, variando el número de neuronas en la capa oculta para ver la evolución en la precisión de predicción:



**Figura III.11** Evolución de la precisión de entrenamiento frente al número de entradas para la base de datos perteneciente a la OCT-Spectralis (modelo 5) con una arquitectura de red de una sola capa oculta y variando el número de neuronas en ésta.

Para esta base de datos, al igual que en la anterior, la máxima precisión alcanzada se da para la arquitectura de red de 8 neuronas en la capa oculta, llegando a un 95% de precisión en el grupo de validación, pero únicamente se alcanza un 86.79% en el grupo de entrenamiento.

### III.2 Predicción mediante SVM.

En primer lugar, hay que decidir si se separan los datos por sexo o por el contrario se realiza cada modelo con el conjunto completo de los pacientes.

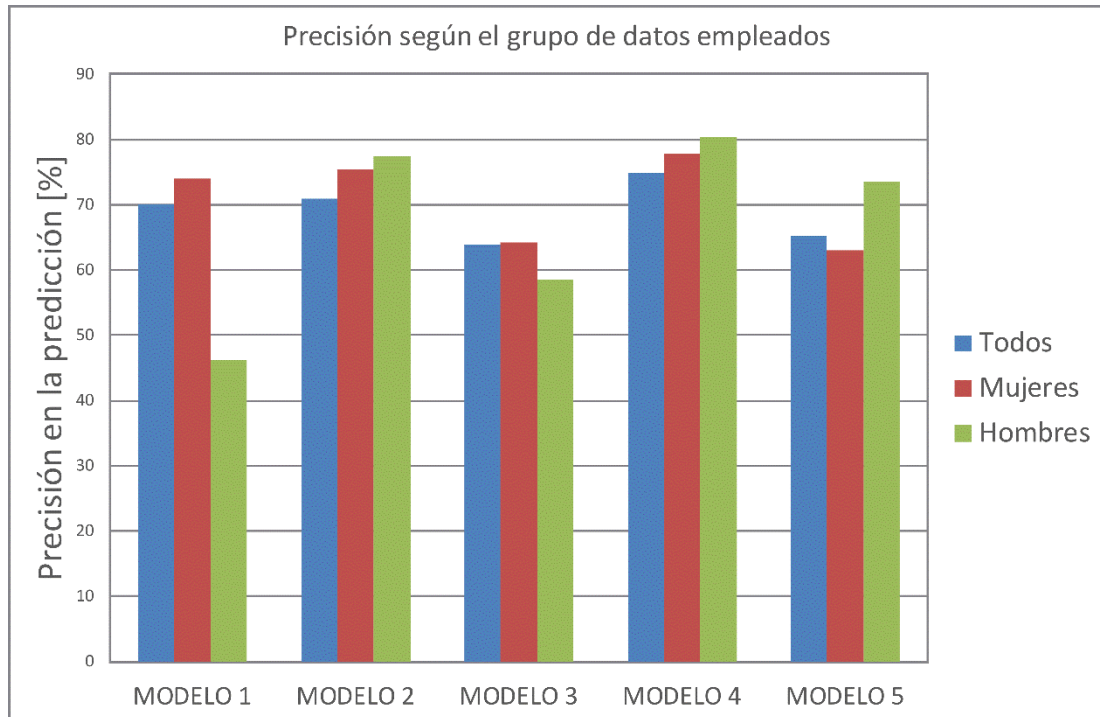


Figura III.12 Precisión en el diagnóstico alcanzada para las 5 bases de datos disponibles según se utilicen los pacientes de ambos sexos juntos (barra azul), solo los pacientes de sexo femenino (barra roja) o los del masculino (barra verde).

En la Figura III.12 se puede observar que no hay una tendencia general de la precisión separando o no los datos por sexo de los pacientes. Por este motivo será necesario estudiar todas las posibilidades en cada base de datos para encontrar la arquitectura adecuada en cada una de ellas.

Se debe definir también el algoritmo de optimización y la función kernel que se va a emplear en este caso concreto. Para ello se realiza un estudio para evaluar la influencia de estos dos parámetros en la precisión de diagnóstico. Esta prueba la realizamos con los resultados obtenidos con el equipo CIRRUS, es decir, con el modelo 1.

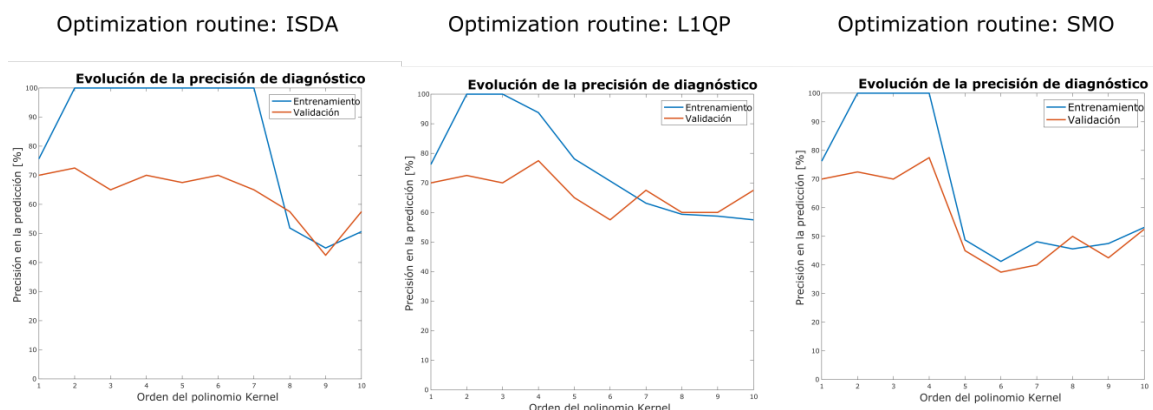


Figura III.13 Evolución de la precisión en la predicción de las redes SVM en función del orden del polinomio kernel empleado. A la izquierda utilizando la rutina de optimización ISDA, en medio utilizando la optimización L1QP, a la derecha empleando la optimización SMO.

En la Figura III.13 se muestran los resultados de precisión de entrenamiento y de validación en diferentes pruebas realizadas. En la primera figura, a la izquierda, se muestra la evolución de la precisión del algoritmo que ha sido entrenado empleando una rutina de optimización ISDA y una función kernel tipo polinomial variando el orden  $q$  de esta entre 1 y 10. La imagen del centro y de la derecha representan la misma prueba, pero utilizando la rutina de optimización LIQP y SMO respectivamente. Puede observarse que la máxima predicción se obtiene para el algoritmo de optimización SMO con el orden polinomial  $q=4$  de la función kernel. De esta manera se decide realizar los modelos de predicción con **el modelo de optimización SMO**, que además es el recomendado por la librería de Matlab utilizada [35].

### III.2.1 Modelo 1: base de datos obtenida mediante el equipo CIRRUS HD-OCT.

Para la primera base de datos, CIRRUS HD-OCT, se realizan las pruebas oportunas para poder observar la influencia de la función kernel y la separación o no de los datos por sexo. Como se ha comentado en el apartado anterior, se ha escogido la función de optimización SMO.

**Tabla 11. Precisiones alcanzadas según la función kernel empleada en el algoritmo en la base de datos 1.**

Función Kernel	Datos	Nº entradas	Nº Pacientes totales	Precisión Predicción [%]
	<b>Todos</b>		<b>160</b>	<b>68</b>
<b>Lineal</b>	Mujeres	<b>16</b>	116	67.22
	Hombres		44	54.71
Polinomial (grado 4)	Todos	16	160	64
	Mujeres		116	63.79
	Hombres		44	52.27
Gaussiana	Todos	16	160	60
	Mujeres		116	62.07
	Hombres		44	54.54

En la Tabla 11 se puede ver que la máxima precisión de validación alcanzada por el algoritmo para esta base de datos es 68% de precisión con todos los datos juntos (mujeres + hombres), para una función kernel lineal. La precisión separando los datos según el sexo es inferior y por tanto será preferible no separarlos al igual que en las pruebas realizadas con el perceptrón multicapa.

### III.2.2 Modelo 2: base de datos correspondiente a la capa de células ganglionares de la retina en la zona macular, obtenida con el equipo OCT-Spectralis.

En la segunda base de datos, correspondiente a la adquisición macular de la capa de células ganglionares de la retina, se realizan la misma prueba que en la base anterior, con la función de optimización SMO.

**Tabla 12. Precisiones alcanzadas según la función kernel empleada en el algoritmo en la base de datos 2.**

Función Kernel	Datos	Nº entradas	Nº Pacientes totales	Precisión Predicción [%]
Lineal	Todos	12	162	70.94
	Mujeres		124	75.48
	Hombres		38	74
<b>Polinomial (grado 4)</b>	<b>Todos</b>	<b>12</b>	<b>162</b>	<b>76.76</b>
	Mujeres		124	79.35
	Hombres		38	76
Gaussiana	Todos	12	162	77.37
	Mujeres		124	74.83
	Hombres		38	75



La Tabla 12 muestra las precisiones alcanzadas para esta base en concreto, y pueden apreciarse que son ligeramente más altas que en las anteriores. La configuración del algoritmo con la que se consigue las mayores precisiones en este caso es la polinomial, que llega a un 76.76% de precisión de diagnóstico.

### III.2.3 Modelo 3: base de datos correspondiente a la RNFL en la zona macular, obtenida con el equipo OCT-Spectralis.

De nuevo se realiza la prueba de variar la función kernel empleada en el algoritmo, esta vez para la tercera base de datos, adquisición de la zona macular correspondiente a la fibra de capas nerviosas de la retina.

**Tabla 13. Precisiones alcanzadas según la función kernel empleada en el algoritmo en la base de datos 3.**

Función Kernel	Datos	Nº entradas	Nº Pacientes totales	Precisión Predicción [%]
Lineal	Todos		159	62.9
	Mujeres	12	121	62.76
	Hombres		38	64
<b>Polinomial (grado 4)</b>	<b>Todos</b>		<b>159</b>	<b>74</b>
	Mujeres	<b>12</b>	121	66.86
	Hombres		38	70
Gaussiana	Todos		159	63.5
	Mujeres	12	121	63.33
	Hombres		38	64

Se puede observar en la Tabla 13 que al igual que en el modelo anterior la configuración que mayor precisión ofrece es la polinomial, llegando en este caso a una precisión de 74% con la función kernel polinomial. Con respecto a la base de datos anterior, hay que destacar que las precisiones generales alcanzadas han descendido ligeramente.

### III.2.4 Modelo 4: base de datos correspondiente a la RNFL obtenida con el protocolo axonal del equipo OCT-Spectralis.

El modelo 4 emplea la base de datos correspondiente a la adquisición peripapilar con barrido de la circunferencia alrededor del nervio óptico en sentido nasal-nasal, es decir con protocolo de adquisición axonal (ver Figura 5.6). En esta adquisición axonal se obtienen 768 puntos que, al igual que se ha hecho anteriormente, se estudia el agrupamiento de los mismos de manera que se intenta reducir el número de entradas en el clasificador.

En primer lugar, se selecciona la función kernel que vamos a utilizar para la base de datos de la misma manera que en los casos anteriores, introduciendo todos los puntos de barrido como entrada en el algoritmo.

**Tabla 14. Precisiones alcanzadas según la función kernel empleada en el algoritmo en la base de datos 4.**

Función Kernel	Datos	Nº entradas	Nº Pacientes totales	Precisión Predicción [%]
<b>Lineal</b>	<b>Todos</b>		<b>195</b>	<b>76.41</b>
	Mujeres	<b>777</b>	149	76.48
	Hombres		46	76.66
Polinomial (grado 4)	Todos		195	71.79
	Mujeres	777	149	68.38
	Hombres		46	74.66
Gaussiana	Todos		195	71.79
	Mujeres	777	149	70.83
	Hombres		46	78.33

La mayor precisión alcanzada para esta prueba corresponde a la función kernel lineal y sin separar los datos por sexo nuevamente, llegando a un 76.41% de precisión. Así pues, se selecciona la función lineal para variar el número de entradas del algoritmo, procediendo a realizar un estudio del número de puntos óptimo en cada grupo:

**Tabla 15. Precisiones alcanzadas en la base 4 según el agrupamiento de datos del barrido peripapilar, para la función kernel lineal.**

Datos	Nº entradas	Nº Pacientes totales	Precisión Predicción [%]
<b>Todos</b>	777	<b>195</b>	76.41
	393		74.35
	137		74.87
	73		69.23
	<b>41</b>		71.28
	<b>25</b>		<b>78.97</b>

Tras realizar las pruebas agrupando los puntos en grupos de 1, 2, 6, 12, 24 y 48 puntos, se obtiene en la Tabla 15 las predicciones más elevadas. Se llega a conseguir de esta manera un 78.97% de precisión con la función kernel lineal y un total de 25 entradas, es decir, agrupando los puntos de barrido de 48 en 48.

### III.2.5 Modelo 5: base de datos que corresponde a la RNFL obtenida con el protocolo glaucoma del equipo OCT-Spectralis.

Por último, se estudia la base de datos 5, que corresponde a la adquisición peripapilar con barrido alrededor del nervio óptico con sentido temporal-temporal (ver Figura 5.6). Del mismo modo que en el apartado anterior se debe escoger en primer lugar la función kernel con la que se va a realizar el entrenamiento de la SVM.

**Tabla 16. Precisiones alcanzadas según la función kernel empleada en el algoritmo en la base de datos 5.**

Función Kernel	Datos	Nº entradas	Nº Pacientes totales	Precisión Predicción [%]
<b>Lineal</b>	<b>Todos</b>		<b>195</b>	<b>69.2</b>
	Mujeres	<b>777</b>	149	66.32
	Hombres		46	73
Polinomial (grado 4)	Todos		195	68.74
	Mujeres	777	149	67.8
	Hombres		46	66
Gaussiana	Todos		195	68.54
	Mujeres	777	149	68.4
	Hombres		46	66

Puede observarse en la Tabla 16 que las precisiones más altas se obtienen para la función kernel polinomial, alcanzando esta un 69.2% de precisión. Por ello se procede a estudiar el número de entradas óptimo con la función lineal.

**Tabla 17. Precisiones alcanzadas en la base 4 según el agrupamiento de datos del barrido peripapilar, para la función kernel lineal.**

Datos	Nº entradas	Nº Pacientes totales	Precisión Predicción [%]
	777		69.2
	393		65.71
<b>Todos</b>	137	<b>195</b>	69.26
	73		63.50
	41		66.76
	<b>25</b>		<b>76.82</b>

Las precisiones de diagnóstico para los resultados de la prueba médica correspondiente a la base de datos 5 alcanzan un 76.82%, estando los puntos del barrido peripapilar agrupados de 48 en 48, es decir, con 25 entradas al algoritmo.