



Universidad
Zaragoza

Trabajo Fin de Máster

Simulación de la degeneración del nervio óptico en enfermos esclerosis múltiple. Metodología para la predicción de la evolución de la enfermedad.

Simulation of optic nerve degeneration in multiple sclerosis patients. Methodology for disease evolution prediction.

Autor

Alejandro Martín Gallego

Directora

José Cegoñino Banzo

Máster Universitario en Ingeniería Industrial
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^a. Alejandro Martín Gallego,

con nº de DNI 77132746S 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)

Máster Universitario en Ingeniería Industrial, (Título del Trabajo)

Simulación de la degeneración del nervio óptico en enfermos esclerosis múltiple. Metodología para la predicción de la evolución de la enfermedad.

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

Zaragoza, 20 de Noviembre de 2018

Fdo: Alejandro Martín Gallego

RESUMEN

La esclerosis múltiple (EM) es una enfermedad neurodegenerativa que suele afectar a sujetos adultos jóvenes. Esta enfermedad tiene varios efectos sobre el ser humano entre los que se incluye la desmielinización (la mielina es una sustancia que facilita la transmisión de los impulsos nerviosos) de las fibras del sistema nervioso central (SNC) y la pérdida axonal. Estos efectos están relacionados con el desarrollo de la discapacidad funcional. Además, la atrofia cerebral puede aparecer en fases tempranas de la enfermedad y su avance ser rápido. Por estas razones es de vital importancia tener un mejor conocimiento de la evolución de la patología a la hora de aplicar tratamientos eficaces que intenten frenar la velocidad de avance de la enfermedad, tratando de retardar los efectos neurodegenerativos y discapacitantes de la misma. En este contexto, el Trabajo de Fin de Máster (TFM) se realiza con el fin de contribuir a mejorar el conocimiento actual sobre esta patología, y más concretamente se centra en el estudio de nuevos métodos para predecir la progresión a corto y medio plazo de la enfermedad en cada paciente. 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*). 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.

En segundo lugar, se realizan varios modelos de predicción mediante el entrenamiento y validación de 5 algoritmos de clasificación y aprendizaje supervisado: las redes recurrentes (LSTM, del inglés *Long Short-Term Memory*), las máquinas de soporte vectorial (SVM, del inglés *Support Vector Machine*), el modelo Naïve Bayes, el algoritmo k-vecinos más cercanos (k-NN, del inglés *k-nearest neighbours*) y los árboles de clasificación. 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 algoritmos en la clasificación de los pacientes, 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 como el mejor, siendo éste 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 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, comenzando un tratamiento acorde con su progresión y ayudando a frenar el avance de la enfermedad, a mejorar la calidad de vida de los pacientes y a mejorar la gestión de los recursos del sistema sanitario.

Índice

| | | |
|-------|---|----|
| 1 | INTRODUCCIÓN..... | 3 |
| 1.1 | Motivación..... | 3 |
| 1.2 | Estado del arte..... | 3 |
| 1.3 | Objetivos..... | 4 |
| 2 | PATOLOGÍA..... | 6 |
| 2.1 | Esclerosis Múltiple..... | 6 |
| 2.2 | Hipótesis de trabajo..... | 7 |
| 3 | ANATOMÍA DEL OJO HUMANO..... | 9 |
| 3.1 | Sistema nervioso visual..... | 9 |
| 3.2 | Ojo humano..... | 10 |
| 3.3 | Retina..... | 10 |
| 4 | REDES NEURONALES ARTIFICIALES..... | 13 |
| 4.1 | Inteligencia artificial y aprendizaje automático..... | 13 |
| 4.2 | Redes neuronales artificiales..... | 13 |
| 4.2.1 | Modos de operación..... | 14 |
| 4.3 | Redes neuronales recurrentes (RNN)..... | 14 |
| 4.3.1 | Algoritmo de retropropagación a través del tiempo..... | 16 |
| 4.3.2 | Long Short-Term Memory (LSTM)..... | 17 |
| 4.4 | Algoritmos de clasificación binaria..... | 19 |
| 4.4.1 | Máquinas de soporte vectorial (SVM)..... | 19 |
| 4.4.2 | K vecinos más cercanos (kNN)..... | 22 |
| 4.4.3 | Modelo Naïve Bayes..... | 24 |
| 4.4.4 | Árboles de clasificación..... | 25 |
| 5 | MATERIAL Y MÉTODO..... | 27 |
| 5.1 | Sujetos y protocolo de medida..... | 28 |
| 5.1.1 | Tomografía de coherencia óptica..... | 30 |
| 5.2 | Tratamiento de datos y selección de variables..... | 32 |
| 5.3 | Normalización de los datos..... | 35 |
| 5.4 | Análisis estadístico mediante <i>Machine Learning</i> | 36 |
| 5.4.1 | Validación cruzada..... | 36 |
| 5.4.2 | Curvas ROC..... | 37 |

| | | |
|-------|---|----|
| 6 | EXPERIMENTOS REALIZADOS Y RESULTADOS..... | 38 |
| 6.1 | Predicción mediante redes LSTM. | 38 |
| 6.1.1 | Área bajo las curvas ROC. | 41 |
| 6.2 | Predicción mediante algoritmos de clasificación binaria. | 42 |
| 6.2.1 | Predicción mediante SVM. | 44 |
| 6.2.2 | Predicción mediante kNN. | 46 |
| 6.2.3 | Predicción mediante el modelo NB. | 47 |
| 6.2.4 | Predicción mediante árboles de clasificación. | 48 |
| 6.2.5 | Resultados obtenidos para cada modelo. | 50 |
| 6.2.6 | Área bajo las curvas ROC. | 50 |
| 7 | CONCLUSIONES. | 52 |
| 8 | LÍNEAS DE FUTURO. | 55 |
| 9 | BIBLIOGRAFÍA. | 56 |
| 10 | ÍNDICE DE FIGURAS..... | 60 |
| 11 | ÍNDICE DE TABLAS..... | 62 |

1 INTRODUCCIÓN.

1.1 Motivación.

Las enfermedades neurodegenerativas como la esclerosis múltiple (EM), el Alzheimer o el Parkinson, son enfermedades que afectan el sistema nervioso central y que suponen un deterioro de la salud de los pacientes que la sufren. Todas ellas tienen una consecuencia común, la destrucción de las conexiones neuronales y por tanto la progresiva incapacidad física y mental de los pacientes [1,2]. El estudio del origen y evolución de estas enfermedades se presenta de vital importancia [3,4] ya que no existe cura para ninguna de ellas.

En el presente proyecto nos centraremos en el estudio de la EM, 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 [5,6], 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 [1].

Todo esto hace que tener un conocimiento de la evolución de la patología sea de vital importancia en la aplicación de tratamientos que intenten frenar la velocidad de avance de ésta en los pacientes para ayudar al retardo de los efectos neurodegenerativos y discapacitantes de la misma [7].

Se realiza el presente Trabajo de Fin de Máster (TFM) con la motivación de contribuir a la mejora del conocimiento actual sobre esta patología, y más concretamente nos centramos en el estudio de nuevos métodos para predecir la progresión a corto y largo plazo de la enfermedad en cada paciente.

1.2 Estado del arte.

Actualmente, la esclerosis múltiple solo puede ser diagnosticada con absoluta fiabilidad mediante una autopsia post-mortem o una biopsia, pero existen criterios no invasivos para poder diagnosticarla con cierta certeza [2]. 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) [8]. 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.

Los pacientes suelen ser jóvenes y la progresión de la enfermedad rápida, 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 [7]. Así pues, la búsqueda de un biomarcador no invasivo es una prioridad en la actualidad [4,9,10].

Los síntomas y signos que aparecen pueden ser debidos a múltiples enfermedades diferentes [2], 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 punción 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 (RM) 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) [11].

Se ha intentado utilizar redes neuronales artificiales (RNA) en diversos estudios para predecir la evolución a corto plazo [3] 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 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]. Además, en un estudio previo realizado por nuestro grupo de investigación [16], se han utilizado algoritmos de aprendizaje automático junto a 5 cohortes transversales de la prueba médica conocida como tomografía de coherencia óptica (OCT) para realizar un modelo de diagnóstico. Como conclusión, se obtuvo que estas técnicas puedan resultar de gran ayuda para los especialistas en el diagnóstico de esta enfermedad neurodegenerativa.

Otra línea de investigación se ha centrado en el desarrollo de modelos matemáticos [17] para determinar la heterogeneidad en el daño cerebral en pacientes con EM diagnosticada y así comprender los diferentes subtipos que tiene la enfermedad y finalmente poder definir de forma mucho más precisa los tratamientos para cada sujeto.

1.3 Objetivos.

En el presente trabajo se pretende desarrollar un método que consiga predecir la progresión de la discapacidad y la evolución del daño en el nervio óptico, concretamente en la capa de fibras nerviosas de la retina (RNFL, del inglés Retinal Nerve Fiber Layer), en pacientes que han sido diagnosticados de esclerosis múltiple. Los estudios realizados han demostrado que existe un adelgazamiento de la RNFL mayor en los pacientes afectados de EM al compararlos con controles sanos [18,19].

Este método se desarrollará a partir del análisis de los datos obtenidos por una prueba médica no invasiva llamada OCT (Optic Coherence Tomography) que consiste en el escáner de las capas de la retina. Las ventajas de la OCT frente a la resonancia magnética (RM) son en primer lugar, la rapidez en la exploración con OCT y en segundo lugar, que la ausencia de mielina en los axones de la retina permite determinar de manera clara la muerte axonal, a diferencia de lo que ocurre en la RM donde la aparición de zonas inflamadas por desmielinización puede dificultar la evaluación de la muerte axonal real.

Por un lado, se sabe que el espesor de la capa RNFL está relacionada con el desarrollo de la enfermedad [6] 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 longitudinal de la RNFL en la retina (evolución del espesor en el tiempo) se evaluará la relación del espesor con la progresión de la discapacidad en los enfermos.

Para este estudio se dispone de total de 73 pacientes con datos de OCT y seguimiento a 10 años realizado en el **Servicio de Neuro-Oftalmología del Hospital Miguel Servet de Zaragoza** a través de los cuales se pretende entrenar y validar una serie de modelos desarrollados mediante varios algoritmos de clasificación. En primer lugar, una red neuronal artificial (RNA), concretamente a una red neuronal recurrente (LSTM, del inglés Long Sort-Term Memory), que, a diferencia de otros tipos de algoritmos, introducen el parámetro del tiempo en sus modelos. Por otro lado, una serie de algoritmos de clasificación binaria como son las máquinas de soporte vectorial, el algoritmo “k-vecinos más próximos”, el modelo de Naïve Bayes y los clasificadores basados en árboles de decisión.

Los modelos que se obtendrán como resultado deberán poder predecir la discapacidad de un paciente concreto con el objetivo de adecuar el tratamiento que recibirá el sujeto y tratar así de mitigar la progresión de la enfermedad. Esto resultaría muy útil a nivel de seleccionar un tratamiento más agresivo para pacientes cuya evolución va a ser más rápida, en primer lugar, para tratar de frenar así la progresión de la discapacidad, y, por otra parte, para distribuir mejor los recursos clínicos y económicos del sistema sanitario, ya que estos tratamientos más agresivos resultan ser muy novedosos y más caros.

2 PATOLOGÍA.

2.1 Esclerosis Múltiple.

La Esclerosis Múltiple es una enfermedad neurológica autoinmune, crónica, discapacitante, y de etiología desconocida que afecta principalmente a adultos jóvenes. Se caracteriza por la inflamación y desmielinización de las fibras nerviosas (la mielina es una sustancia grasa que rodea y aísla los nervios, actuando como cobertura de un cable eléctrico y permitiendo que los nervios transmitan sus impulsos más rápidamente), 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 [1].

Hasta el momento no existe una única característica clínica que sirva para predecir el subtipo y/o la progresión de la enfermedad, además, la EM se caracteriza por una alta heterogeneidad, lo que dificulta de manera significativa el estudio y el diagnóstico de esta enfermedad [20]. Existen varios subtipos de EM según la evolución de la enfermedad en cada paciente [21]:

- 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.

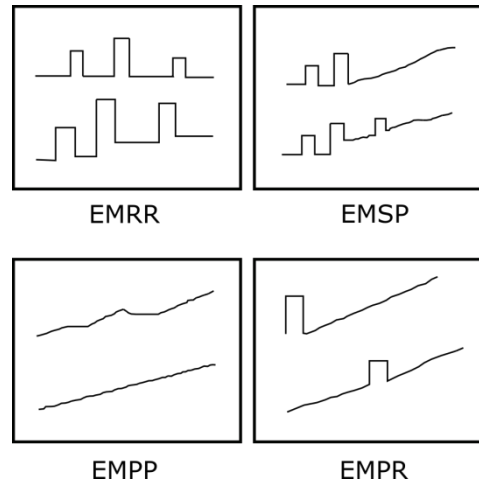


Figura 2.1 Esquema de la evolución temporal de los tipos de EM

La variabilidad en el transcurso de las enfermedades neurodegenerativas se debe a diferentes mecanismos patógenos. En el caso de la EM, estos mecanismos conducen a los efectos como la inflamación crónica, degeneración axonal, desmielinización y remielinización. Estos procesos actúan con diferente severidad para cada subgrupo clínico de la enfermedad, traduciéndose en una evolución diferente de discapacidad en estos subgrupos. Los efectos sobre el ser humano, concretamente en el sistema nervioso central son:

- 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 [6,22].
- 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. Los principales síntomas de la EM son el cansancio, la visión doble o borrosa, problemas del habla, temblor en las manos, debilidad en los miembros, pérdida de fuerza o de sensibilidad en alguna parte del cuerpo, falta de equilibrio, sensación de hormigueo o entumecimiento, problemas de control urinario y dificultad para andar o coordinar movimientos [23].

El seguimiento de los pacientes es imprescindible para conocer la evolución de la enfermedad, identificar la necesidad de tratamiento y valorar la respuesta al mismo, para si procede, modificarlo por otra más eficaz.

2.2 Hipótesis de trabajo.

Para poder realizar una predicción de la progresión en 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 nuestro modelo.

La pérdida axonal en el sistema nervioso central (SNC) [6,22] 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 [24].

Esta reducción del espesor de la RNFL la tomaremos como un biomarcador para la predicción de esclerosis múltiple. El principal problema de esta hipótesis es que todos los seres humanos sufren una pérdida axonal a lo largo de su vida, es decir, que el espesor de la retina se ve disminuido con la edad debido a que las células van muriendo con el paso de los años. Esta reducción del espesor de la retina con la edad (Ver Figura 2.2) complicará de forma considerable la utilización de la RNFL como biomarcador en esta patología.

RETINE FULL LAYERS THICKNESS EVOLUTION WITH AGE

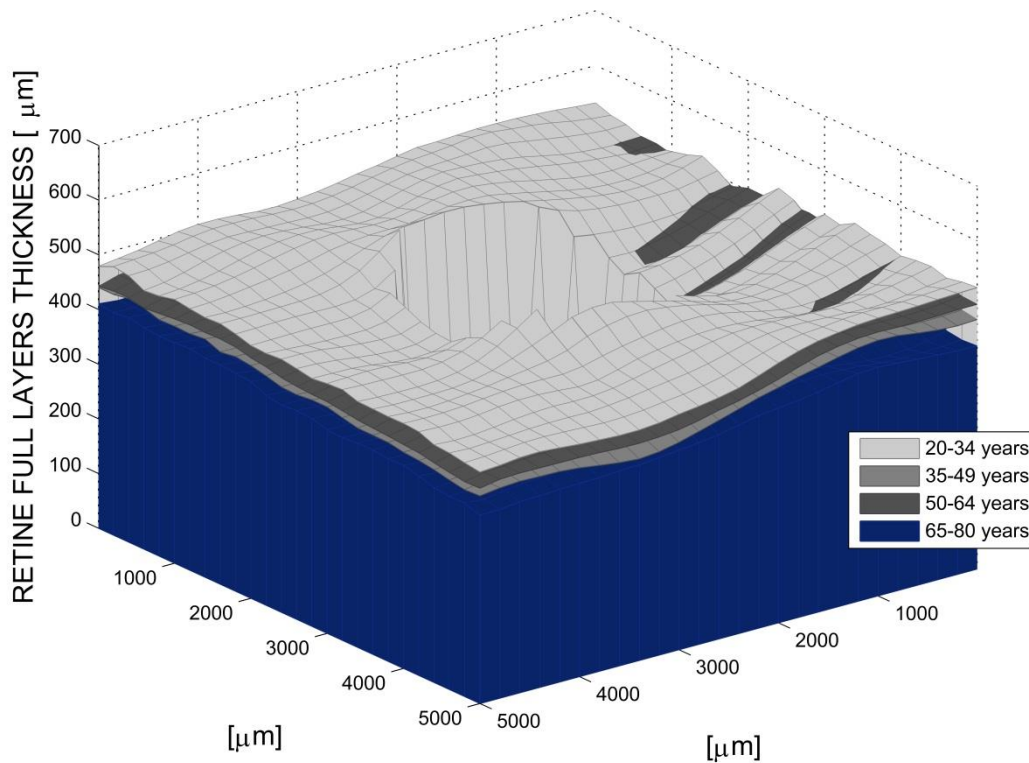


Figura 2.2 Evolución del espesor [μm] de la retina (todas las capas) con la edad de un grupo de pacientes sanos del Hospital Universitario Miguel Servet de Zaragoza.

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 la evolución de la Esclerosis Múltiple.

Además de la evolución del espesor en la retina se van a tener en cuenta otros parámetros que influyen o pueden influir en el desarrollo de la enfermedad, como la edad y el sexo de cada paciente, el subtipo de esclerosis que sufre, la duración de la enfermedad o la edad de diagnóstico, agudeza visual, escala de discapacidad funcional (EDSS), etc.

Con el conjunto de estos parámetros se tratará de realizar una predicción de la progresión de la enfermedad. Para ello se realizará un estudio transversal de los datos de los pacientes diagnosticados de EM procedentes del hospital Miguel Servet, teniendo en cuenta la evolución longitudinal de los parámetros de cada paciente.

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 [25].

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.

Las principales partes del sistema visual [26] son las siguientes:

Nervio óptico: es el nervio sensitivo encargado de transmitir las señales que percibe el ojo hasta el cerebro.

Se origina en la capa de células ganglionares de la retina y está formado por los axones de éstas.

Quiasma óptico: se trata del entrecruzamiento parcial de las fibras axónicas de los nervios ópticos de cada ojo. 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.

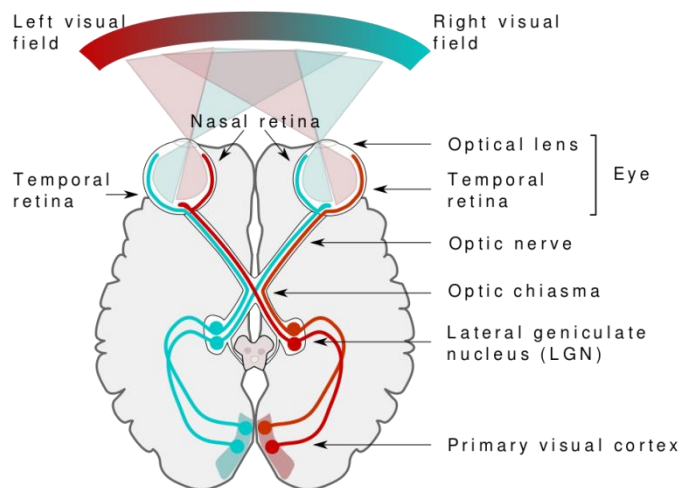


Figura 3.1 Esquema de las principales vías visuales del cerebro.

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 [27].

Las partes del ojo humano son las siguientes [28]:

El iris: membrana circular y 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: estructura transparente del ojo que enfoca el haz de luz en la retina. Tiene forma de lente biconvexa.

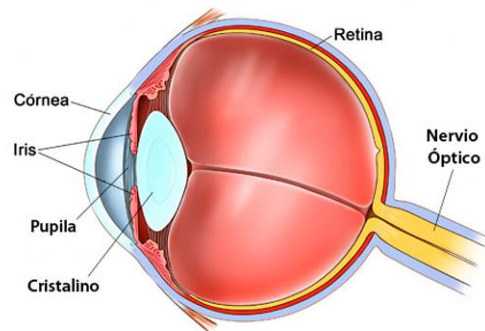


Figura 3.2 Partes del ojo humano [4].

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 capa de tejido del ojo sensible a la luz y se encuentra en la parte posterior del globo ocular. 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.

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 [28]:

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.

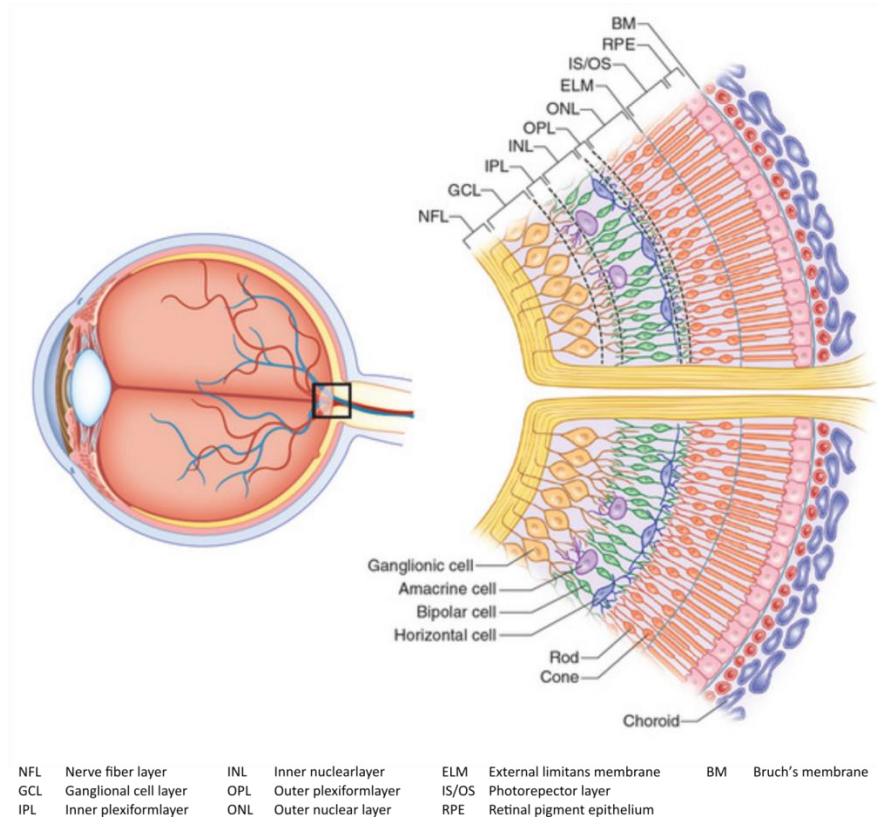


Figura 3.3 Ilustración de las capas de la retina con detalle del nervio óptico [29].

La parte más importante de la retina en el estudio de la predicción de la enfermedad son la capa de células ganglionares (GCL, del inglés Ganglionic Cell Layer) y la capa de fibras nerviosas de la retina (RNFL, del inglés 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 son; el soma (cuerpo de la célula), las dendritas (reciben entradas de otras células) y los axones (que envían salidas a otras células):

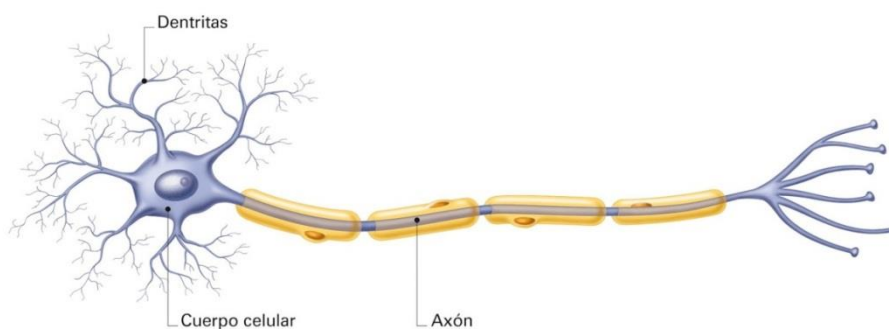


Figura 3.4 Representación de las partes de una célula [30]

La desmielinización y la pérdida axonal son algunos de los efectos de la esclerosis múltiple [6]. 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 como biomarcador 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.

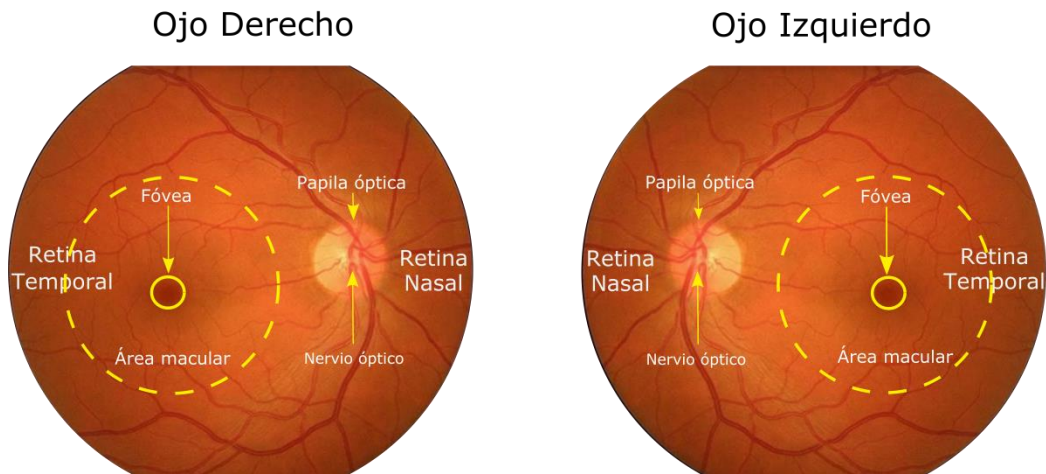


Figura 3.5 Imagen tomada mediante la técnica de OCT (Optical Coherence Tomography) de la retina de un paciente. A la izquierda, imagen de la retina del ojo derecho y a la derecha, imagen de la retina del ojo izquierdo.

En la Figura 3.5 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. En la imagen, también se marca la posición del nervio óptico, la papila, el área macular y la fóvea (centro del área macular), zonas importantes en el estudio de la retina.

4 REDES NEURONALES ARTIFICIALES.

4.1 Inteligencia artificial y aprendizaje automático.

La inteligencia artificial (AI, del inglés *Artificial Intelligence*) 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.

El aprendizaje automático o *Machine Learning* es una aplicación de la AI basada en la idea de dar a las máquinas acceso a los datos y dejarles aprender por sí mismas, es decir, algoritmos que hacen a las máquinas más inteligentes. El *Machine Learning* se centra en el desarrollo de programas informáticos capaces de generalizar comportamientos a partir de los datos recibidos, además de poder cambiar cuando se exponen a nuevos datos [31].

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 lo haría un cerebro humano, y lo que da mayor interés a las redes neuronales es la posibilidad de aprendizaje de una determinada tarea a resolver. Estos modelos son sistemas de aprendizaje basados en ejemplos.

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.

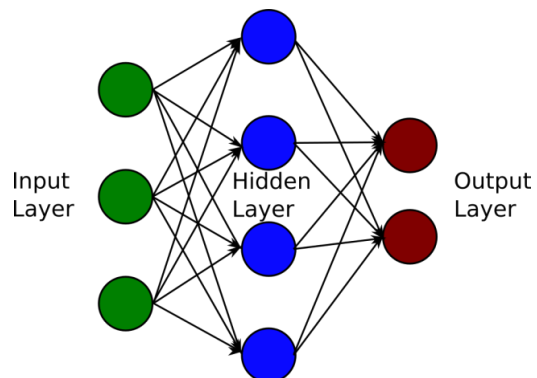


Figura 4.1 Esquema general de una RNA unidireccional de tres capas, de entrada, oculta y de salida.

El primer nivel o capa lo constituyen las células de entrada (Input Layer), que reciben los valores de unos patrones representados como vectores y sirven de entrada a la red. A continuación, se encuentran una serie de capas intermedias denominadas capas ocultas (Hidden Layers), cuyas unidades responde a rasgos particulares que pueden aparecer en los patrones de entrada y su función es “acomodarse” para encontrar la relación entre las entradas y las salidas. Puede haber uno o varios niveles ocultos. El último nivel es el de salida (Output Layer), que servirá respuesta del modelo en cada simulación [32].

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 [33].

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.

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 [34]. 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.

Existen tres criterios de convergencia dependiendo del tipo de red utilizado o del problema a resolver. La finalización del periodo de aprendizaje se puede determinar según un número fijo de ciclos, cuando el error descienda por debajo de una cantidad preestablecida o cuando la modificación de los pesos sinápticos entre ciclo sea irrelevante.

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 Redes neuronales recurrentes (RNN).

Las redes neuronales recurrentes (RNN, del inglés *Recurrent Neural Networks*) son una herramienta muy apropiada para modelar series temporales. Se trata de un tipo de redes que implementa cierta memoria, y por tanto sentido temporal. Esto se consigue implementando algunas neuronas que reciben como entrada la salida de una de las

capas anteriores e inyectan su salida en una de las capas de un nivel anterior a ella. Esta característica de memoria y la posibilidad de modelar series temporales hace que este tipo de redes neuronales sea apropiado para implementar el modelo de predicción que buscamos, ya que la serie de datos longitudinales del seguimiento pueden tratarse como una serie temporal de las características de cada paciente.

Las RNN se caracterizan por crear bucles en las neuronas de la red mediante el uso de las llamadas conexiones recurrentes (Figura 4.2), pudiendo aparecer en la red conexiones de una neurona a ella misma, conexiones entre neuronas de una misma capa o conexiones de las neuronas de una capa a la anterior [32].

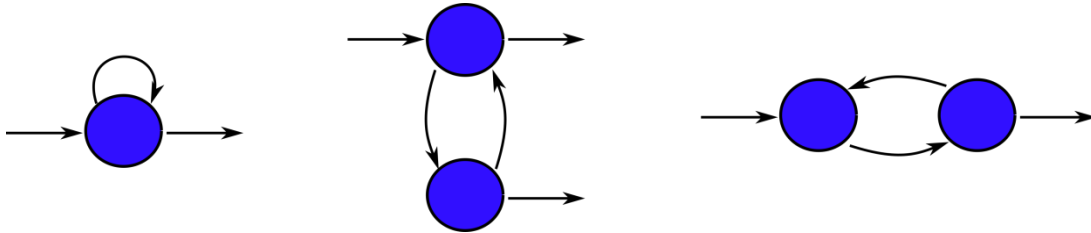


Figura 4.2. Conexiones recurrentes empleadas en las RNN. Izquierda: conexión de una neurona con sí misma. Centro: Conexión entre neuronas de una misma capa. Derecha: Conexión entre dos neuronas de dos capas diferentes.

Al introducir estas conexiones creando bucles, la activación de una neurona ya no depende únicamente de las activaciones de la capa anterior, sino que depende también del estado de cualquier otra neurona conectada a ella. Por tanto, es necesario incluir la variable tiempo en la activación o estado de una neurona dotando a este tipo de redes de un comportamiento temporal o dinámico, lo que facilita el tratamiento de información temporal o patrones dinámicos, es decir, patrones cuyo valor depende de sus valores en instantes anteriores de tiempo. Por tanto, estas redes son las más indicadas para abordar el tipo de problemas temporales [32].

Las redes recurrentes se caracterizan porque sus neuronas reciben como entradas la activación del resto de las neuronas de la red, así como su propia activación. De este modo, si $a_i(t)$ representa la activación de una neurona i de la red en el sistema de tiempo t , su valor viene dado por la siguiente expresión:

$$a_i(t) = f_i \left(\sum_{j \in A \cup B} \omega_{ij} a_j(t-1) \right) \quad (1)$$

donde el índice j varía en el conjunto de todas las neuronas conectadas a la neurona i , f_i es la función de activación, A es el conjunto de neuronas de entrada a la red, B representa el resto de las neuronas de la red y ω_{ij} representa el peso sináptico de la conexión de la neurona j a la neurona i .

En el entrenamiento de este tipo de redes no puede aplicarse el algoritmo de retropropagación (*Back-Propagation*) para entrenar el perceptrón multicapa, sino que se debe llevar a cabo utilizando algoritmos de aprendizaje diferentes como el algoritmo de *retropropagación a través del tiempo* o el *algoritmo de aprendizaje recurrente a tiempo real* [32].

4.3.1 Algoritmo de retropropagación a través del tiempo.

El algoritmo de retropropagación a través del tiempo (*Back-Propagation Through Time*) se basa en la idea de que para cada red recurrente es posible construir una red multicapa con conexiones adelante [35] y con idéntico comportamiento; basta con desarrollar en el tiempo la red recurrente [32,36].

Para explicar esta idea se considera la red recurrente simple de dos neuronas totalmente conectadas (Figura 4.3),

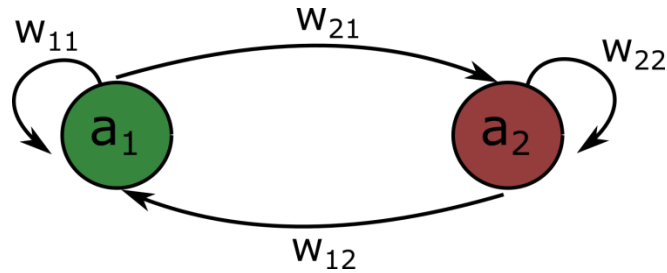


Figura 4.3 Arquitectura de red recurrente con dos neuronas totalmente conectadas.

y la activación de cada una de ellas viene dada por:

$$a_1(t) = f(\omega_{11}a_1(t-1) + \omega_{12}a_2(t-1)) \quad (2)$$

$$a_2(t) = f(\omega_{21}a_1(t-1) + \omega_{22}a_2(t-1)) \quad (3)$$

Desarrollando estas activaciones en el tiempo se obtiene:

$$\begin{aligned} a_i(1) &= f(\omega_{i1}a_1(0) + \omega_{i2}a_2(0)) \\ a_i(2) &= f(\omega_{i1}a_1(1) + \omega_{i2}a_2(1)) \end{aligned} \quad (4)$$

...

$$a_i(n) = f(\omega_{i1}a_1(n-1) + \omega_{i2}a_2(n-1))$$

De este modo, la red recurrente se puede representar mediante una red con conexiones hacia adelante, la cual se obtiene añadiendo una nueva capa por cada unidad de tiempo. Las activaciones hasta el instante de tiempo n se muestran en la Figura 4.4.

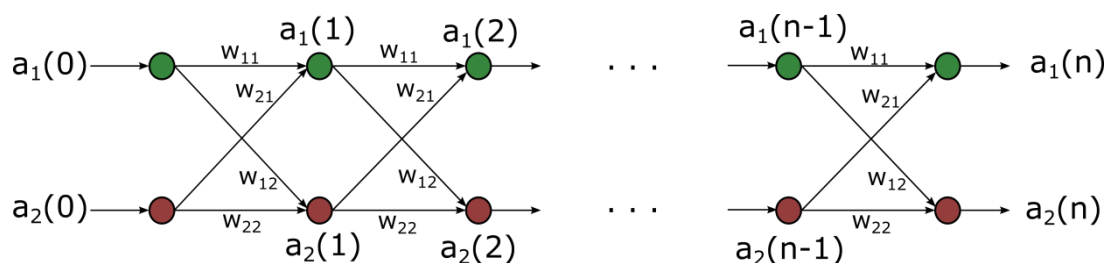


Figura 4.4 Red recurrente desarrollada en el tiempo.

Para aplicar el *algoritmo de retropropagación a través del tiempo*, se debe definir una época o intervalo de tiempo en el cual la red recurrente será desarrollada. Así pues, se define el error cometido por la red en un intervalo de tiempo $[n_0, n_1]$ como:

$$E(n_0, n_1) = \frac{1}{2} \sum_{n=n_0}^{n_1} \sum_{i \in S} e_i^2(n) \quad (5)$$

donde S es el conjunto de índices que representan las neuronas de salida de la red, y $e_i(n)$ es el error que comete la red en el instante n , medido como la diferencia entre la salida de la red y la salida deseada.

Los pesos de la red recurrente se van modificando siguiendo la dirección negativa del gradiente del error, por lo que el cambio del peso ω_{ij} de la neurona j a la i viene dado por:

$$\Delta\omega_{ij} = -\alpha \frac{\partial E(n_0, n_1)}{\partial \omega_{ij}} \quad (6)$$

Este algoritmo de retropropagación en el tiempo presenta ciertas desventajas como un elevado coste computacional y la necesidad de almacenar en memoria el estado de la red desarrollada, lo que limita su utilización en aplicaciones que requieren una adaptación continua de la red recurrente.

En el caso necesitar realizar una aplicación con adaptación a tiempo real de la red recurrente, existe una variante de este algoritmo que, desde el punto de vista computacional es más eficiente, el algoritmo de aprendizaje recurrente a tiempo real [37].

4.3.2 Long Short-Term Memory (LSTM).

Una LSTM es un tipo de red neuronal recurrente que destaca en el aprendizaje, procesamiento y clasificación de datos secuenciales. Las áreas de aplicación más comunes de este tipo de red pasan por el análisis de seguimiento, modelado de lenguaje, reconocimiento de voz y análisis de video [38].

El algoritmo de entrenamiento que normalmente se emplea es el de *retropropagación a lo largo del tiempo* (apartado 4.3.1). No obstante, este algoritmo tiene ciertas limitaciones que pueden generar dependencias a corto plazo disminuyendo la información en los pasos de tiempo anteriores.

La arquitectura típica de un bloque LSTM tiene, además del propio estado de la unidad (c_t), una celda de memoria (g), puerta de entrada (i), puerta de salida (o), una puerta de olvido (f), y además de un estado oculto (h_t). La representación de esta arquitectura podemos verla en la Figura 4.5.

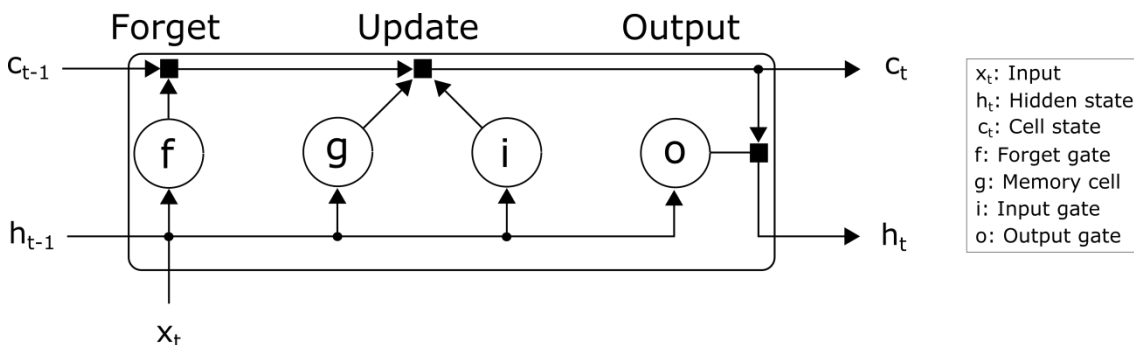


Figura 4.5 Bloque Long Short-Term Memory

Los pesos y los sesgos a la puerta de entrada controlan la medida en la que va a influir el nuevo valor de entrada (x_t) a la celda. De este mismo modo, los pesos y desviaciones de la puerta de olvido y la de salida controlan el grado en que permanece un valor en la celda y el grado en que el valor se usa para calcular la activación de la salida del bloque LSTM.

El siguiente diagrama ilustra el flujo de una serie temporal X con características D de longitud S a través de una capa LSTM.

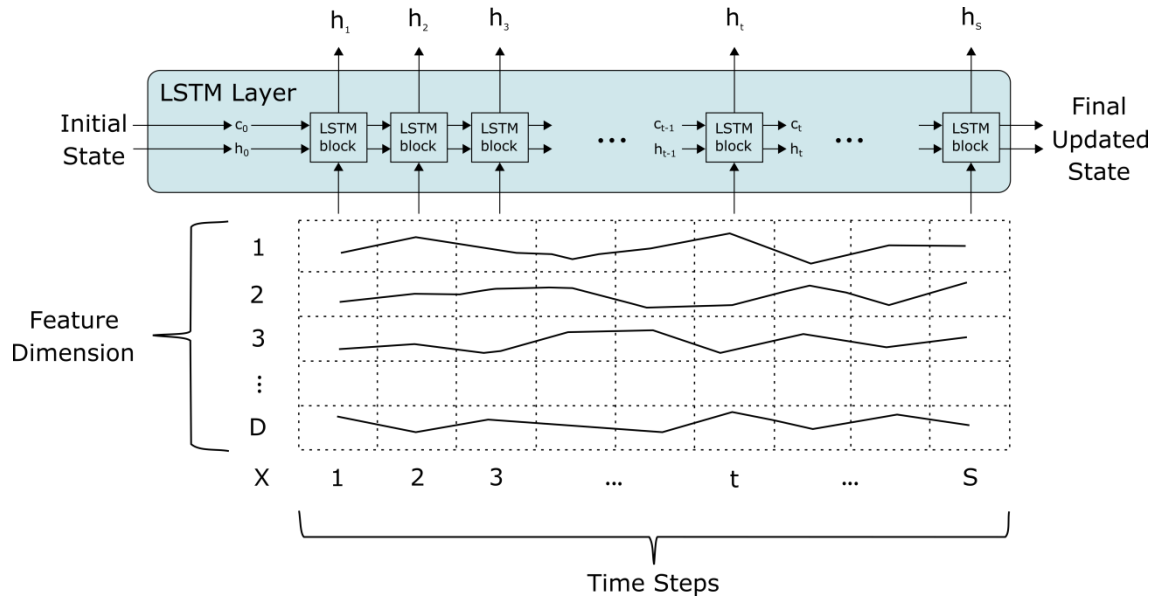


Figura 4.6 Diagrama de flujo de una capa LSTM para una serie temporal X .

En cada paso de tiempo, la capa LSTM agrega o elimina información del estado de la celda, donde la capa controla estas actualizaciones utilizando las puertas y sus pesos y sesgos.

4.4 Algoritmos de clasificación binaria.

En este apartado se van a introducir los algoritmos empleados para el estudio de datos y la clasificación binaria, es decir, para encontrar ciertos patrones en listas de datos y desarrollar la capacidad de clasificar nuevos patrones correctamente. Más concretamente vamos a introducir varios algoritmos empleados en clasificación en dos clases, y algunas modificaciones que permiten realizar la clasificación multiclase.

4.4.1 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 4.7.

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 4.8).

La búsqueda del hiperplano de separación en estos espacios transformados se hará de forma implícita utilizando las funciones *kernel* [39].

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 [40].

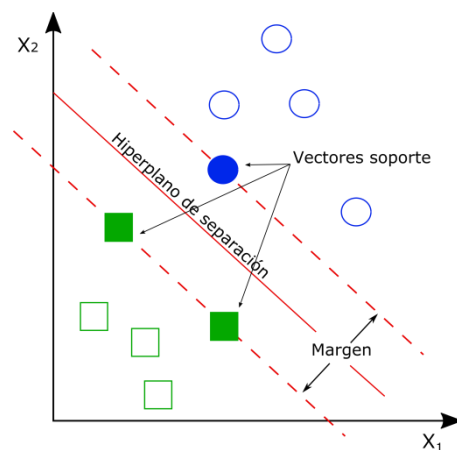


Figura 4.7 Representación de un problema de clasificación 2D resuelto mediante SVM

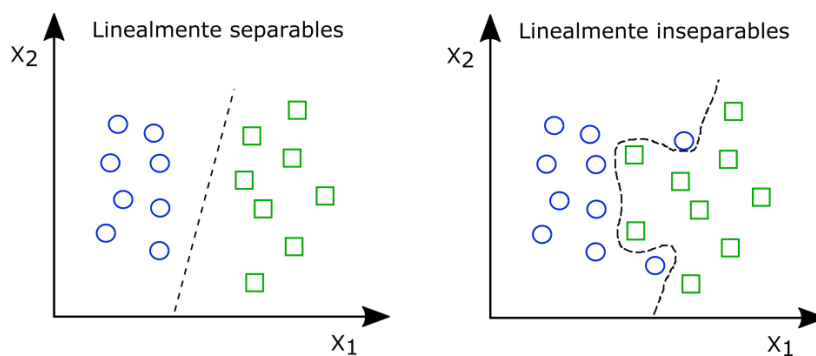


Figura 4.8 Tipos de problemas de clasificación 2D que podemos resolver con SVM; la imagen la izquierda se puede observar un problema de clasificación linealmente separable y en la imagen de la derecha un problema de clasificación inseparable linealmente.

Se va a explicar cómo se resolvería los dos tipos de problemas de clasificación binaria utilizando las SVM [35].

4.4.1.1 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 $x_i \in \mathcal{R}^N$ 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 φ de \mathcal{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 $b \in \mathcal{R}$.

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.

4.4.1.2 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:

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

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

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:

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

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:

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

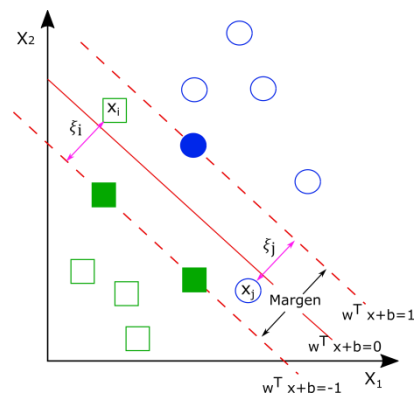


Figura 4.9 Aparición del parámetro ξ_i en el error de clasificación [15]

4.4.1.3 Truco del Kernel para el caso no linealmente separable.

Como no tenemos ningún conocimiento de φ , 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 ;

$$z_i \cdot z_j = \varphi(x_i) \cdot \varphi(x_j) = K(x_i, x_j) \quad (18)$$

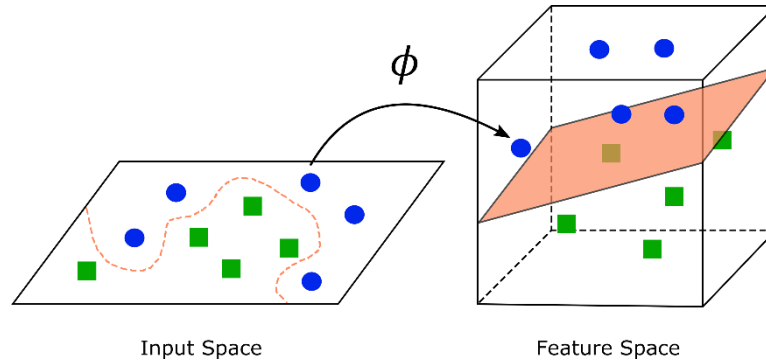


Figura 4.10 Idea del uso de un kernel para la transformación del espacio de los datos [40]

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:

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

4.4.2 K vecinos más cercanos (kNN).

El algoritmo **k-vecinos más cercanos** [41] (**k-NN**, del inglés **k-Nearest Neighbours**) es un algoritmo de aprendizaje supervisado, es decir, que a partir de un conjunto de datos inicial es capaz de clasificar correctamente nuevos patrones. El conjunto de datos típico de este tipo de algoritmos está formado por varios atributos descriptivos y un solo atributo objetivo (también llamado clase).

El algoritmo k-NN clasifica cada dato nuevo en el grupo que corresponda, según tenga k vecinos más cerca de un grupo o de otro. Es decir, calcula la distancia del elemento nuevo a cada uno de los existentes, y ordena dichas distancias de menor a mayor para ir seleccionando el grupo al que pertenecer. Este grupo será, por tanto, el de mayor frecuencia con menores distancias.

En contraste con otros algoritmos de aprendizaje supervisado, k-NN no genera un modelo a partir del aprendizaje con datos de entrenamiento, sino que el aprendizaje sucede en el mismo momento en el que se prueban los datos de test. A este tipo de algoritmos se les llama *lazy learning methods* [42].

Este algoritmo se utiliza en *data mining* (minería de datos), por lo que su uso en *Machine Learning* (aprendizaje automático) es una de sus aplicaciones clave para, por ejemplo, sistemas de recomendación (plataformas de contenido digital, procesos de venta cruzada, etc.).

El funcionamiento del algoritmo se basa en la siguiente descripción:

Sea D el conjunto de datos en el que se distingue la siguiente estructura:

$$D_{m,n} = \begin{bmatrix} d_1 & = & a_{1,1} & \cdots & a_{1,n} & c_1 \\ d_2 & = & a_{2,1} & \ddots & a_{2,n} & c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_m & = & a_{m,1} & \cdots & a_{m,n} & c_m \end{bmatrix} \quad (22)$$

donde d_m indica cada una de las instancias del conjunto de datos, $a_{m,n}$ indica cada uno de los atributos descriptivos y c_m indica la clase a predecir.

Se fija un valor para k , y se calcula una distancia d del conjunto de datos. Tras este proceso, el algoritmo selecciona los patrones de datos de entrenamiento más cercanos y se le asigna la distancia d a la clase más frecuente de entre los patrones seleccionados como más cercanos.

El algoritmo es muy sensible a la variable k , de modo que con valores distintos se puede obtener resultados también muy distintos. En la Figura 4.11 podemos observar un ejemplo de la influencia del valor de k en la clasificación de un mismo conjunto de datos:

- Para $k = 1$, el algoritmo clasificará el nuevo dato (cuadrado rojo) como bola azul.
- Para $k = 2$, el algoritmo no tiene criterio para clasificar el nuevo dato, ya que hay un elemento de cada clase dentro del círculo.
- Para $k = 3$, el algoritmo clasificará el nuevo dato como cuadrado verde.

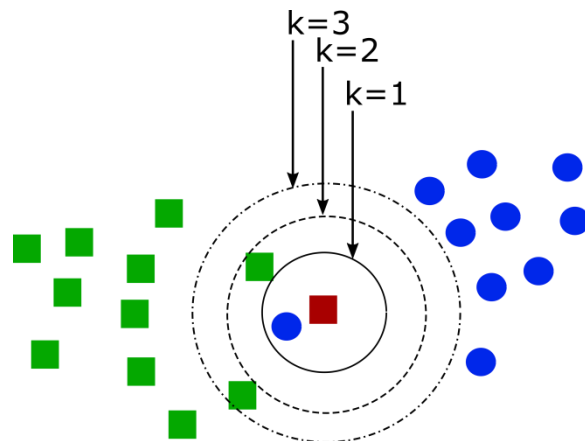


Figura 4.11. Ejemplo de clasificación mediante algoritmo K-NN.

La mayor debilidad de este algoritmo es la lentitud del proceso de clasificación, ya que no se obtiene un modelo de predicción con los datos de entrenamiento, sino que se evalúan el conjunto de datos en el momento que se clasifican los datos de validación.

También puede producirse el llamado efecto de sobre-entrenamiento (*overfitting*), lo que provoca que el modelo que se construye, en lugar de describir de forma genérica las estructuras de los patrones de datos, lo que acaba describiendo sea, exclusivamente, el conjunto de datos de entrenamiento no siendo extrapolable a nuevos patrones.

4.4.3 Modelo Naïve Bayes.

El clasificador “*Bayesiano ingenuo*” o *Naïve Bayes* es un clasificador probabilístico basado en el teorema de Bayes que se resume en la independencia entre las variables predictoras, es decir, que cada una de las características contribuye de manera independiente a la probabilidad de que un patrón de datos corresponda a una clase u otra, independientemente de la presencia o ausencia de las otras variables. Entre las características que poseen los métodos bayesianos se pueden resaltar las siguientes:

- Cada ejemplo observado modifica la probabilidad de que la hipótesis formulada sea correcta, es decir, que la probabilidad estimada para la hipótesis disminuirá o aumentará en función de la más o menos coincidencia con el conjunto de datos.
- Métodos robustos al ruido presente en los ejemplos de entrenamiento y a la posibilidad de tener entre esos ejemplos datos incompletos o erróneos.

En la clasificación de patrones, el problema consiste en encontrar una función $f(x)$ (llamada hipótesis) que clasifique los conjuntos de datos de entrenamiento. La idea de usar el teorema de Bayes en cualquier problema de aprendizaje automático es que se puede estimar las probabilidades a posteriori de cualquier hipótesis consistente con el conjunto de datos de entrenamiento para poder escoger la hipótesis más probable [43]. Entre los algoritmos que utilizan este método destaca el algoritmo *Naïve Bayes*.

Dado un ejemplo x representado por k valores, el clasificador naïve Bayes se basa en encontrar la hipótesis más probable que describa a ese ejemplo [43]. Si la descripción del ejemplo viene dada por los valores $\langle a_1, a_2, \dots, a_n \rangle$, la hipótesis más probable será aquella que cumpla:

$$v_{MAP} = \operatorname{argmax}_{v_j \in V} P(v_j | a_1, \dots, a_n) \quad (23)$$

es decir, la probabilidad de que conocidos los valores que describen a ese ejemplo, éste pertenezca a la clase v_j (donde v_j es el valor de la función de clasificación $f(x)$ en el conjunto finito V). Por el teorema de Bayes [44]:

$$\begin{aligned} v_{MAP} &= \operatorname{argmax}_{v_j \in V} \frac{P(a_1, \dots, a_n | v_j) p(v_j)}{P(a_1, \dots, a_n)} \\ &= \operatorname{argmax}_{v_j \in V} P(a_1, \dots, a_n | v_j) p(v_j) \end{aligned} \quad (24)$$

Se puede estimar $P(v_j)$ contando las veces que aparece el ejemplo v_j en el conjunto de entrenamiento y dividiéndolo por el número total de ejemplos que forman este conjunto. Para estimar el término $P(a_1, \dots, a_n | v_j)$, es decir, las veces en que para cada categoría aparecen los valores del ejemplo x , se debe recorrer todo el conjunto de entrenamiento. Para simplificar el cálculo de este término se recurre a la hipótesis de independencia condicional para factorizar la probabilidad: *Los valores a_j que describen un atributo de un ejemplo cualquiera son independientes entre sí conocido el valor de la categoría a la que pertenecen.* Así la probabilidad a la que pertenecen es el producto de las probabilidades de cada valor por separado:

$$P(a_1, \dots, a_n | v_j) = \prod_i P(a_i | v_j) \quad (25)$$

4.4.4 Árboles de clasificación.

El aprendizaje con este algoritmo utiliza un árbol de decisión como modelo predictivo. El objetivo es crear un modelo que predice el valor de la variable objetivo o clase en función de diversas variables de entrada. En la Figura 4.12 se puede observar un ejemplo de clasificación binaria ($clase \in \{0,1\}$) empleando un árbol de decisión y en él se puede ver como cada uno de los nodos interiores del árbol de decisión corresponde a cada una de las variables de entrada. Estos nodos dividen la característica en dos rangos basados en umbrales (e.g. $x_1 < 3 \Delta x_1 \geq 3$).

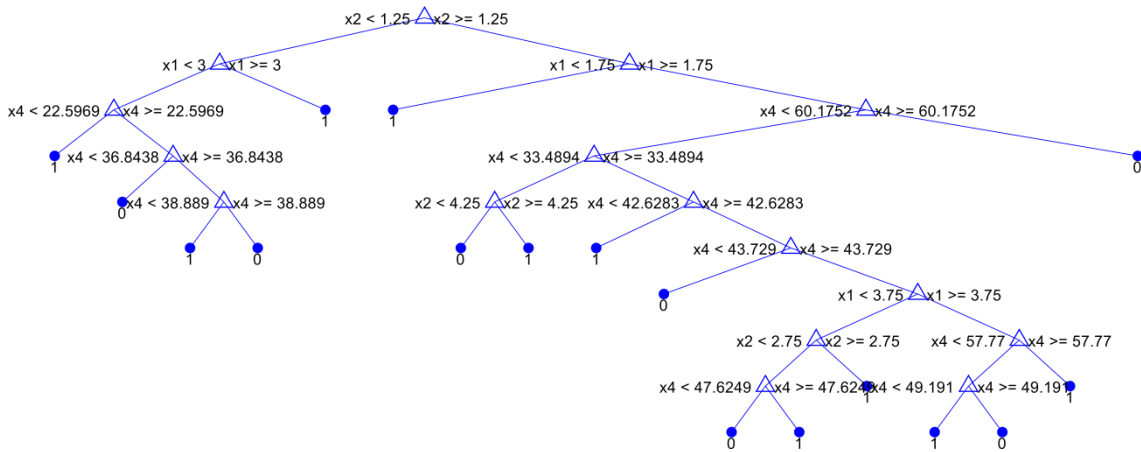


Figura 4.12 Ejemplo de clasificación mediante un árbol de decisión.

El aprendizaje basado en árboles de decisión es una de las técnicas más eficaces para la clasificación supervisada y se basa en la construcción de un árbol de decisión a partir de ejemplos de entrenamiento, cada uno etiquetado con su correspondiente clase. Un árbol de decisión es similar a una estructura de diagrama de flujo, donde cada nodo interno denota una prueba en un atributo, cada rama representa el resultado de una prueba, y cada hoja nodo tiene una etiqueta de clase. El nodo superior en un árbol es el nodo raíz [45].

Los datos de entrenamiento se dan de la forma:

$$(x, Y) = (x_1, x_2, x_3, \dots, x_k, Y) \quad (26)$$

Donde la variable Y es la variable objetivo que estamos tratando de clasificar. El vector x se compone de las variables de entrada $x_1, x_2, x_3, \dots, x_k$ que se utilizan en la clasificación.

Entre las ventajas que ofrecen los árboles de decisión en la minería de datos destacan las siguientes:

- Son muy fáciles de entender e interpretar.
- No requiere mucho tratamiento de datos, como la normalización o eliminación de valores en blanco.
- Capaz de manejar datos numéricos y categorizados.
- Buen funcionamiento con grandes conjuntos de datos, ya que el coste computacional es relativamente bajo.

El problema del aprendizaje con árboles de decisión es que se puede crear un árbol de decisión excesivamente complejo que no generalice bien a partir de los datos de

entrenamiento (sobre-ajuste o sobre-entrenamiento, Figura 4.13). Una hipótesis se dice que sobreajusta los datos de entrenamiento si existe otra hipótesis la cual tiene más error que la primera en los datos de entrenamiento pero menos error en los datos de validación [46].

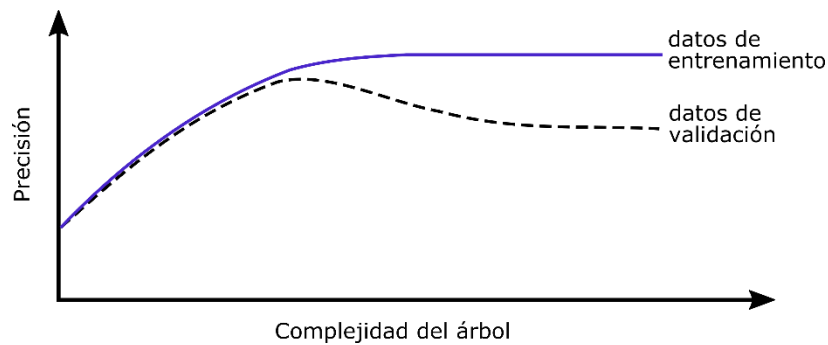


Figura 4.13 Evolución de la precisión en el conjunto de entrenamiento y el de validación en función de la complejidad del árbol de decisión.

En la Figura 4.13 se puede observar como al inicio, al aumentar la complejidad del árbol de decisión, también aumenta la precisión que se alcanza con el algoritmo tanto en los datos de entrenamiento como en los datos de validación, sin embargo, llega un momento donde la precisión en el conjunto de datos de validación comienza a disminuir, es decir, la capacidad de generalización del modelo decrece.

5 MATERIAL Y MÉTODO.

Inicialmente se parte de cinco bases de datos proporcionadas por el servicio de **Neurooftalmología del Hospital Universitario Miguel Servet de Zaragoza** en las que se incluyen una serie de resultados de diferentes pruebas médicas y otros parámetros relacionados con el diagnóstico de la esclerosis múltiple con un seguimiento a 10 años. Las 5 bases de datos se diferencian en el dispositivo y los protocolos de medida utilizados para realizar una prueba médica conocida como tomografía de coherencia óptica (OCT, del inglés Optical Coherence Tomography):

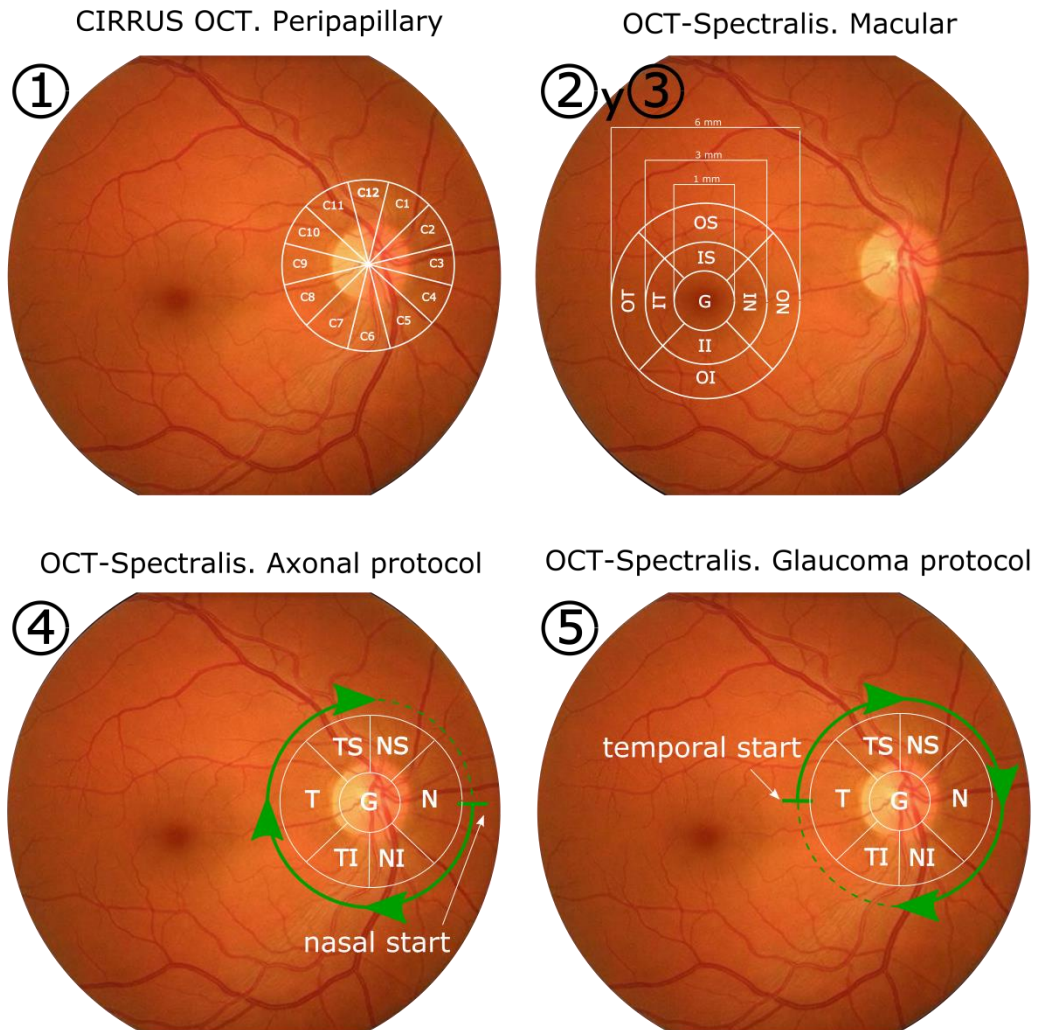


Figura 5.1. Diferencias entre los protocolos de adquisición de OCT para cada una de las bases de datos.

Base datos 1: La primera de las bases de datos corresponde a datos OCT adquiridos mediante el dispositivo CIRRUS-OCT (ver Figura 5.3) e incluye datos del espesor de la RNFL en la zona peripapilar de la retina.

Base datos 2: La segunda base de datos se ha obtenido mediante el equipo OCT-Spectralis (ver Figura i.1). Ésta incluye datos del espesor de la capa de células ganglionares de la retina alrededor de la fóvea, es decir, en la zona macular.

Base datos 3: La tercera, de nuevo se realiza con el equipo OCT-Spectralis, y como la segunda, incluye datos de espesor de la zona macular, pero en este caso de la RNFL.

Base datos 4: En la cuarta base de datos se incluyen datos del espesor de la RNFL en la zona peripapilar obtenidos mediante el equipo OCT-Spectralis. Además, incluye el espesor en una serie de puntos completando una circunferencia alrededor del nervio óptico, adquiridos consecutivamente en sentido de las agujas del reloj e iniciando el barrido desde la zona nasal del nervio óptico. Este protocolo de adquisición se denomina protocolo axonal.

Base datos 5: La quinta de las bases de datos, al igual que la cuarta, posee datos del espesor de la RNFL alrededor de la zona peripapilar, e incluye el barrido de puntos alrededor del nervio óptico, con la diferencia que éste comienza en la zona temporal del nervio óptico y se denomina protocolo glaucoma.

Según se expone en el anexo ii., debido a la gran diferencia entre la frecuencia de adquisición en el seguimiento entre las bases de datos obtenidas mediante el dispositivo OCT-Spectralis, es decir las bases 2, 3, 4 y 5, y la base de datos primera, obtenida mediante el equipo CIRRUS-OCT, será esta última la que se utilice para el desarrollo de los modelos de predicción que se van a plantear.

5.1 Sujetos y protocolo de medida.

Se parte de la **base de datos 1** proporcionada por el Hospital en el que se incluyen los siguientes resultados de diferentes pruebas médicas y otros parámetros relacionados con el diagnóstico de la esclerosis múltiple con un seguimiento a 10 años:

- Edad del paciente en cada una de las visitas.
- Sexo del paciente.
- La duración de la enfermedad en el momento de la visita.
- Fenotipo de EM, es decir, a que subtipo de enfermedad a la que pertenece el paciente. En la base de datos nos encontramos un total de 68 sujetos que desarrollan EM recurrente remitente (EMRR), 4 sujetos que desarrollan EM primaria progresiva (EMPP), y un solo sujeto con EM secundaria progresiva (EMSP).
- Escala expandida del estado de discapacidad (EDSS, del inglés *Expanded Disability Status Scale*), escala que asigna puntuaciones de 0 a 10 con incrementos de 0.5 unidades que representan mayores niveles de discapacidad. Esta escala se basa en la asignación de puntos en ocho sistemas funcionales; piramidal (debilidad o dificultad para mover extremidades), cerebelo (pérdida de coordinación o temblores), bulbo raquídeo (problemas en el habla), sensoriales (entumecimiento o pérdida de sensaciones), función intestinal y vesical, función visual, funciones cerebrales (mentales), otras. La EDSS es uno de los instrumentos que más se utilizan para evaluar la progresión de la enfermedad [47].

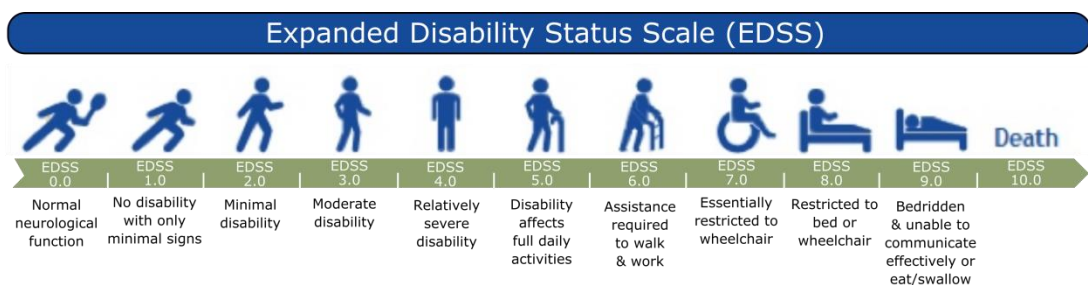


Figura 5.2. Escala expandida del estado de discapacidad.

- EDSS visual: puntuación de discapacidad funcional EDSS para el sistema visual.

- Antecedentes de neuritis óptica: La neuritis óptica es una inflamación del nervio óptico y una manifestación frecuente de la esclerosis múltiple [48].
- Brotes o recaídas en el último año.
- Escala Snellen de agudeza visual: prueba médica que determina la agudeza visual a partir de la identificación por parte del paciente de una serie de letras sobre un fondo blanco.
- Presión intraocular: la presión ocular normal varía entre 12 y 22 mmHg. Una presión intraocular alta es un importante factor de riesgo para la aparición de glaucoma, que es una enfermedad ocular que puede causar un deterioro importante de la capacidad visual.
- Test de Ishihara: test utilizado para el diagnóstico y clasificación de discromastopsias (alteración en la visión de colores, como el daltonismo), aunque es muy útil para conocer el estado del nervio óptico y sus fibras en diferentes patologías.
- Puntuación de la desviación media en el campo visual: es el promedio de los valores obtenidos en el mapa de desviación total. Un valor normal es entre -2 y +2 dB. Cuando más negativo es más anormal.
- Tomografía de coherencia óptica (OCT): procedimiento no invasivo que permite estudiar el espesor de la retina con precisión. En la base de datos que disponemos, se realiza la prueba con el equipo CIRRUS-OCT, cuyo resultado es el espesor medio de la capa de fibras nerviosas de la retina en 4 cuadrantes (superior, nasal, inferior y temporal) alrededor del área peripapilar. Además, en algunos pacientes se incluye el espesor en 12 sectores horarios en la misma zona de la retina. Esta prueba médica la explicaremos más adelante.
- 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. El resultado de esta prueba médica es el valor de latencia de la onda positiva P100 y el valor de la amplitud de la misma.
- 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. Esta prueba médica ofrece como resultado la latencia de la onda positiva P50 y de la onda negativa N95, y los valores de amplitud de las mismas.

Se dispone de datos de un total de 73 pacientes que tienen diagnosticada esclerosis múltiple con un seguimiento de varias visitas a lo largo de unos 10 años (2007 – 2017). Los pacientes a los que corresponden los datos son individuos blancos europeos del área geográfica de Zaragoza.

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 la totalidad de los pacientes han completado todas las pruebas médicas comentadas anteriormente.



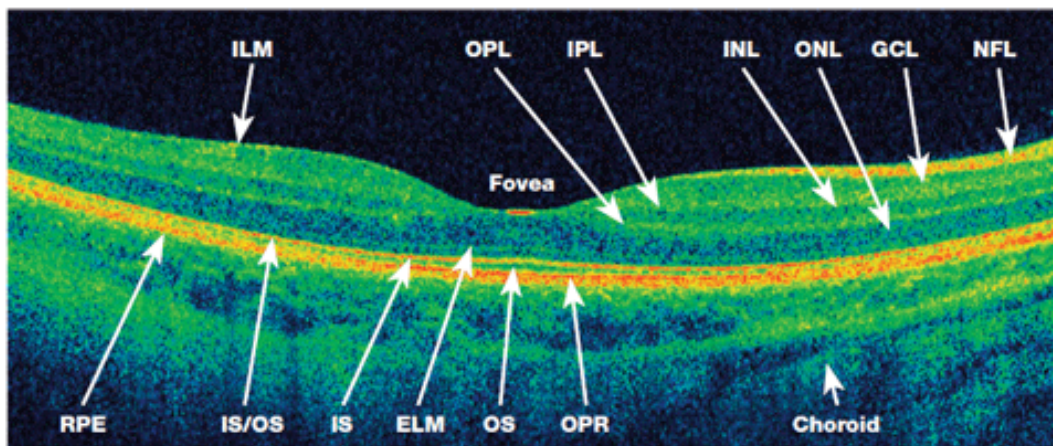
Figura 5.3 Imagen del equipo OCT-CIRRUS en el Hospital Universitario Miguel Servet de Zaragoza.

Tras hacer una búsqueda de bibliografía relacionada con el tema [1,15,49–51], se decide que los datos de espesor de la retina obtenidos mediante la técnica OCT (ver Figura 5.4) van a ser los mejores para realizar un diagnóstico preciso de la progresión de la enfermedad. Por tanto, se descartarán los datos obtenidos mediante polarización láser, potenciales evocados visuales y electroretinograma-pattern.

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).

An HD-OCT scan of a healthy eye



| | | |
|-------------------------------------|--|---|
| NFL: Nerve fiber layer | OPL: Outer plexiform layer | IS/OS: Interface between IS and OS |
| ILM: Inner limiting membrane | ONL: Outer nuclear layer | RPE: Retinal pigment epithelium |
| GCL: Ganglion cell layer | ELM: External limiting membrane | OPR: Outer photoreceptor/ RPE complex |
| IPL: Inner plexiform layer | IS: Photoreceptor inner segment | |
| INL: Inner nuclear layer | OS: Photoreceptor outer segment | |

Figura 5.4 Sección de la retina tomada mediante una tomografía de coherencia óptica [52].

En la Figura 5.4 podemos observar un ejemplo de imagen OCT que representa una sección de la zona macular de un sujeto sano. En esta imagen se puede apreciar en diferentes colores las capas de la retina. A partir de la imagen se puede llegar a obtener el espesor de cada una de estas capas.

La técnica de la OCT se basa en el principio óptico de la interferometría (ver Figura 5.5), 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.

El hospital Miguel Servet de Zaragoza dispone de un equipo de la compañía *Carl Zeiss Meditec AG* para la obtención de los datos OCT,

concretamente un *CIRRUS HD-OCT* que proporciona datos del espesor medio de la RNFL subdividiendo la zona peripapilar de la retina en 4 cuadrantes alrededor del nervio óptico. Estos cuadrantes corresponden a las zonas Temporal (T), Superior (S), Nasal (N) e Inferior (I). Además, el protocolo de adquisición utilizado proporciona datos correspondientes a 12 sectores de 30° formando una circunferencia completa de 360° con disposición horaria como vemos en la Figura 5.6.

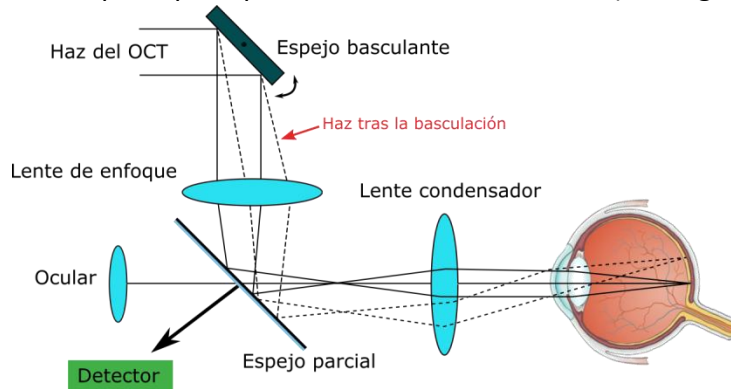


Figura 5.5 Principios de interferometría en los que se basa la OCT.

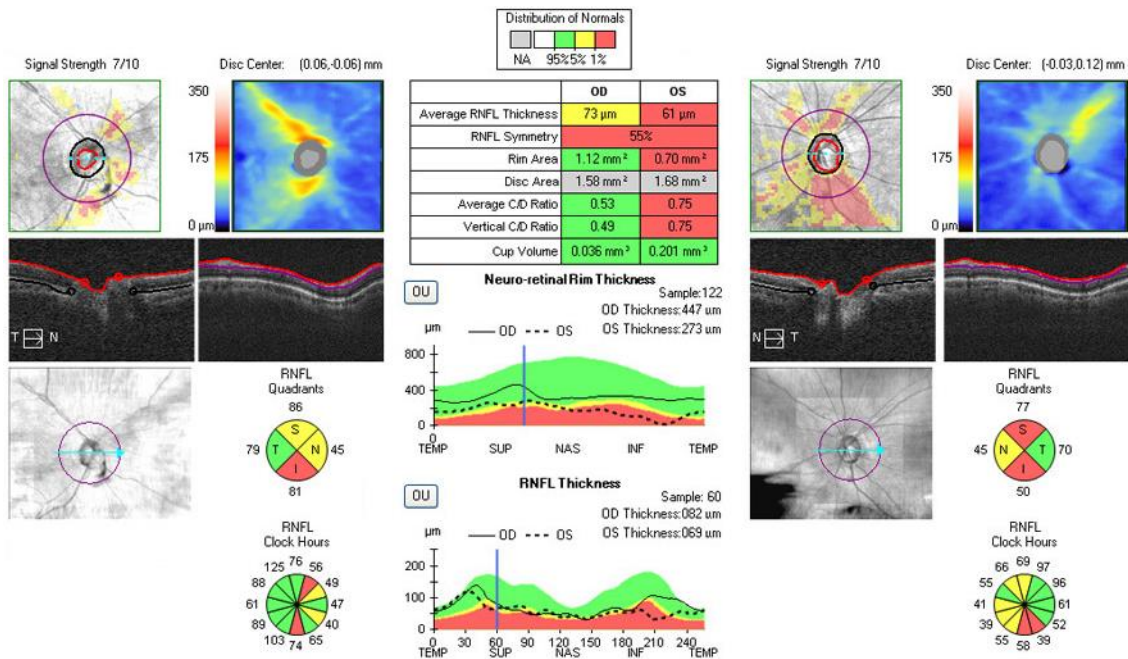


Figura 5.6 Ejemplo de datos que proporciona la OCT-CIRRUS [53].

En la Figura 5.6 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 gráficos 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.

5.2 Tratamiento de datos y selección de variables.

En primer lugar, 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. En la Figura 5.7 se puede observar el año de las visitas de las que se tiene información en el seguimiento a 10 años.

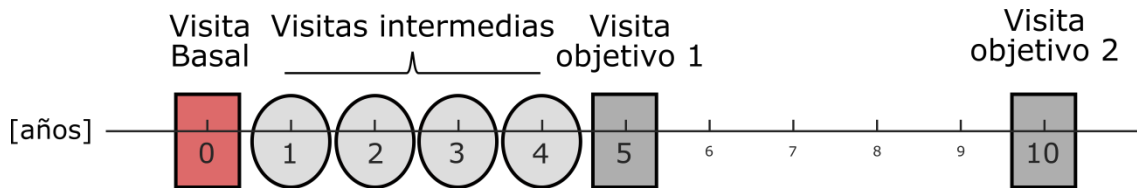


Figura 5.7 Visitas de las que se tiene información de los pacientes.

El siguiente paso es decidir cuántas visitas serán necesarias para el entrenamiento de nuestro modelo de predicción y poder así seleccionar las secuencias de datos válidas.

En la Figura 5.8 se representa la frecuencia del número de visitas disponibles para el conjunto de los pacientes, es decir, el número de pacientes que disponen de la visita basal, el número de pacientes que disponen de la visita 1, la 2,... hasta el número de pacientes que disponen información de la visita 10.

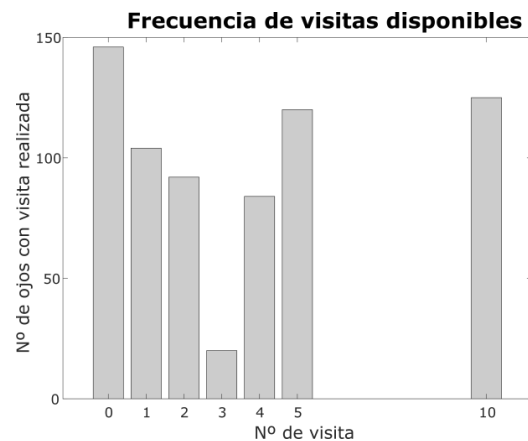


Figura 5.8 Representación de la cantidad de visitas disponibles en el conjunto de los pacientes según la correspondiente visita

En la base de datos, las secuencias disponibles varían en longitud entre 3 y 7 visitas, es decir, hay sujetos con información de 3 visitas (5 pacientes), sujetos con 4 visitas (26 pacientes), otros con 5 visitas (26 pacientes), otros 6 (14 pacientes) y, por último, con 7 visitas (2 pacientes). Además, no siempre son la misma secuencia de visitas, por tanto, será necesario tener en cuenta el número de visita del que estamos tratando para desarrollar el modelo de predicción.

De todos los atributos disponibles en la base de datos vamos a seleccionar los que mayor relevancia tienen o pueden tener en el desarrollo de las enfermedades neurodegenerativas:

- La edad y el sexo del paciente se considera que puede ser relevante en el desarrollo de la esclerosis múltiple, ya que, por un lado, a la mayoría de los pacientes se les diagnostica la enfermedad cuando son jóvenes, pero la edad de diagnóstico en la base de datos varía entre los 13 y los 60 años. Por otro lado, el sexo del paciente puede ser un factor a tener en cuenta, ya que alrededor del 70% de los pacientes son de sexo femenino.

- La duración de la enfermedad en el momento de la visita. Es imprescindible saber en qué momento de la enfermedad se encuentra cada paciente y el porcentaje en su vida durante el cual la enfermedad ha estado progresando.
- Fenotipo de esclerosis múltiple. Este parámetro es muy relevante, ya que como se ha explicado en el apartado 2.1, cada subtipo de enfermedad va a tener una evolución de discapacidad diferente.
- Escala expandida del estado de discapacidad (EDSS). Éste será un parámetro muy importante ya que marca la discapacidad funcional del paciente en el momento de la visita, y por tanto puede ser un buen marcador para saber la progresión de la enfermedad en éste paciente. Además, este parámetro es especialmente relevante a la hora de seleccionar el tipo de tratamiento que se va a proporcionar, por tanto, será un parámetro objetivo idóneo para el modelo de predicción.
- Antecedentes de neuritis óptica. Esto puede llegar a ser relevante, ya que la neuritis óptica es una inflamación del nervio óptico y ésta influye en los resultados de la prueba OCT que tiene como objetivo el espesor de las capas de fibras nerviosas que forman este nervio.
- Brotes o recaídas en el último año. El hecho de sufrir un brote repercute directamente en la evolución de la enfermedad como se ha explicado en el apartado 2.1, por tanto este parámetro se incluye en el modelo.
- Tomografía de coherencia óptica. De los parámetros que esta prueba nos da como resultado, se decide incluir en el modelo de predicción el espesor medio en los 4 cuadrantes (superior, nasal, inferior y temporal, ver Figura 5.6) alrededor del nervio óptico, el espesor medio en la zona peripapilar y el espesor medio en la fovea (zona macular). Con la inclusión de estos parámetros se pretende relacionar la secuencia del espesor de la RNFL con la evolución de la enfermedad en los pacientes, en concreto, con la progresión de la discapacidad funcional de los pacientes.

A continuación, se muestra una tabla con los datos clínicos de los pacientes estudiados, donde aparece la media, la desviación estándar (SD), el valor máximo y el mínimo de cada correspondiente parámetro.

Tabla 1. Datos clínicos de los pacientes estudiados.

| | Media | SD | Max | Min |
|--------------------|--------|-------|-------|-------|
| Edad [años] | 42,08 | 10,52 | 66,05 | 19,69 |
| Sexo (% Mujeres) | 70% | -- | -- | -- |
| Duración EM [años] | 10,59 | 8,33 | 40 | 0,5 |
| Tipo EM (% tipo 1) | 93,15% | -- | -- | -- |
| NO antecedentes | 29% | -- | -- | -- |
| Brotes último año | 23% | -- | -- | -- |
| EDSS Inicial | 2,589 | 1,999 | 7 | 0 |
| EDSS Final | 3,370 | 2,405 | 8,5 | 0 |

Por su parte, en la Tabla 2, aparece los datos clínicos de los pacientes correspondientes a los datos OCT empleados en el desarrollo de los modelos de predicción, es decir, el espesor medio de la zona peripapilar, el espesor medio de la fovea, y los cuatro sectores de espesor alrededor del nervio óptico (superior, nasal, inferior y temporal):

Tabla 2. Datos clínicos de la prueba OCT de los pacientes estudiados.

| Visita | Parámetro | Media [μm] | SD [μm] | Max [μm] | Min [μm] |
|-----------|-----------------|-------------------------|----------------------|-----------------------|-----------------------|
| Visita 0 | Md. Peripapilar | 89,412 | 17,331 | 131,05 | 33,23 |
| | Md. Fóvea | 276,733 | 28,594 | 349 | 137 |
| | Superior | 109,63 | 22,187 | 164 | 27 |
| | Nasal | 74,014 | 20,155 | 139 | 19 |
| | Inferior | 115,548 | 25,626 | 183 | 13 |
| | Temporal | 59,37 | 19,149 | 173 | 26 |
| Visita 1 | Md. Peripapilar | 86,674 | 15,222 | 131,35 | 54,09 |
| | Md. Fóvea | 280,981 | 25,964 | 380 | 194 |
| | Superior | 109,317 | 19,004 | 176 | 65 |
| | Nasal | 71,269 | 19,816 | 124 | 31 |
| | Inferior | 109,471 | 23,783 | 176 | 39 |
| | Temporal | 57,231 | 15,225 | 104 | 27 |
| Visita 2 | Md. Peripapilar | 84,627 | 14,965 | 115 | 42 |
| | Md. Fóvea | 271,663 | 23,235 | 316 | 169 |
| | Superior | 107,88 | 20,325 | 159 | 40 |
| | Nasal | 68,196 | 15,852 | 107 | 37 |
| | Inferior | 106,326 | 23,04 | 154 | 47 |
| | Temporal | 56,342 | 16,337 | 109 | 21 |
| Visita 3 | Md. Peripapilar | 80,569 | 10,946 | 100 | 59 |
| | Md. Fóvea | 263,9 | 25,095 | 319 | 226 |
| | Superior | 102,6 | 14,084 | 133 | 80 |
| | Nasal | 67,7 | 12,757 | 94 | 48 |
| | Inferior | 103,95 | 18,808 | 141 | 66 |
| | Temporal | 48,4 | 13,268 | 75 | 30 |
| Visita 4 | Md. Peripapilar | 82,042 | 14,475 | 120 | 37 |
| | Md. Fóvea | 259,637 | 21,393 | 311 | 212 |
| | Superior | 101,03 | 19,421 | 143 | 35 |
| | Nasal | 67,673 | 13,755 | 101 | 35 |
| | Inferior | 105,851 | 24,334 | 184 | 48 |
| | Temporal | 53,613 | 14,815 | 87 | 28 |
| Visita 5 | Md. Peripapilar | 82,107 | 13,825 | 123 | 53 |
| | Md. Fóvea | 259,025 | 21,794 | 356 | 212 |
| | Superior | 102,208 | 18,235 | 175 | 58 |
| | Nasal | 68,35 | 13,17 | 112 | 42 |
| | Inferior | 107,075 | 22,187 | 180 | 60 |
| | Temporal | 50,946 | 13,425 | 85 | 24 |
| Visita 10 | Md. Peripapilar | 77,656 | 15,478 | 118 | 32 |
| | Md. Fóvea | 255,2 | 26,395 | 320 | 118 |
| | Superior | 94,16 | 19,41 | 138 | 30 |
| | Nasal | 65,04 | 16,196 | 128 | 25 |
| | Inferior | 100,112 | 26,237 | 175 | 32 |
| | Temporal | 50,016 | 16,007 | 140 | 19 |

Una vez seleccionados los parámetros que van a constituir el modelo de predicción se deben ordenar en secuencias de datos aquellos que evolucionan con el tiempo, como son la edad, la duración de la enfermedad, la EDSS, los brotes o recaídas en el último año, y todos los datos de OCT.

En la Figura 5.9 se muestra la evolución de los espesores medios de la tabla anterior a lo largo de las visitas de los pacientes. En ésta se puede observar como la tendencia general es de disminución del espesor de la capa de fibras nerviosas de la retina.

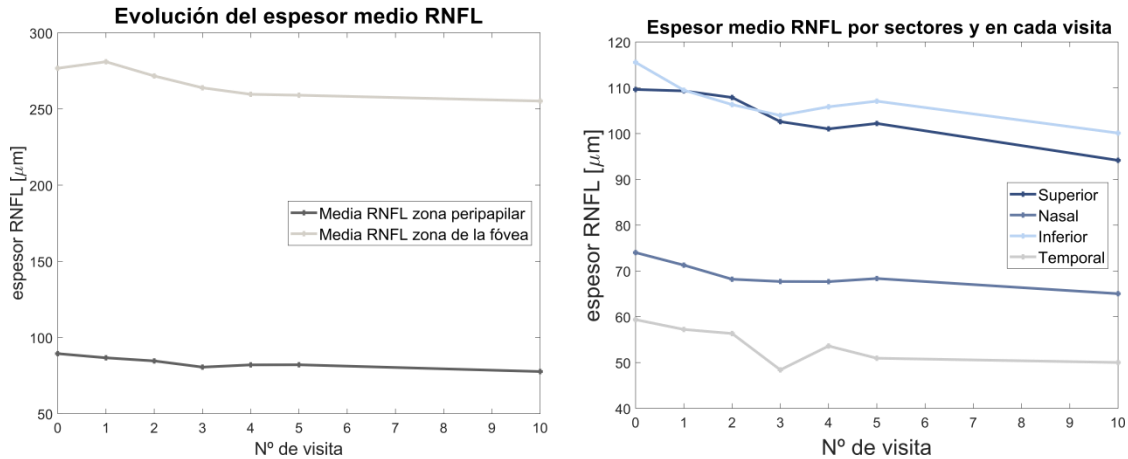


Figura 5.9. Izquierda: Gráfica de la evolución del espesor medio de la RNFL en la zona peripapilar y en la fovea a lo largo del tiempo. Derecha: Gráfica de la evolución del espesor de la RNFL en los cuatro sectores alrededor de la zona peripapilar a lo largo de las 10 visitas.

A partir de estos parámetros y secuencias de parámetros se tratará de encontrar una relación entre la evolución de la capa de fibras nerviosas de la retina y la evolución en la escala de discapacidad funcional EDSS.

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:

$$Z = \frac{X - \bar{X}}{\sigma} \quad (27)$$

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 algoritmos de predicción.

5.4 Análisis estadístico mediante *Machine Learning*.

Se va a evaluar la capacidad de las redes neuronales recurrentes y de algunos algoritmos de clasificación como son las SVM, el algoritmo KNN, los árboles de decisión y el modelo de Naïve Bayes como clasificadores de los datos obtenidos de la OCT para distinguir entre los diferentes grupos de pacientes según vaya a evolucionar la escala de discapacidad funcional EDSS.

- 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 datos correctamente. Para el modelo empleado de **red recurrente** empleamos el algoritmo de *back-propagation through time* en el entrenamiento de la red.
- Las **SVM** 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 [54].
- En contraste con otros algoritmos de aprendizaje supervisado, **k-NN** no genera un modelo a partir del aprendizaje con datos de entrenamiento, sino que el aprendizaje sucede en el mismo momento en el que se prueban los datos de validación.
- El clasificador **NB** por su parte, se basa en el teorema de Bayes, que supone independencia condicional de las variables predictoras dada la clase de cada ejemplo. Este modelo además, nos permitirá conocer cuáles son las variables que más relación tienen con la variable objetivo, ya que la hipótesis generadas por el algoritmo tratará estas variables de forma independiente [55].
- Por último, los clasificadores basados en **árboles de decisión** son fácilmente interpretables que los modelos denominados “caja negra” como pueden ser las redes neuronales artificiales, lo que nos puede ayudar a interpretar el proceso de predicción de los modelos [56].

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.

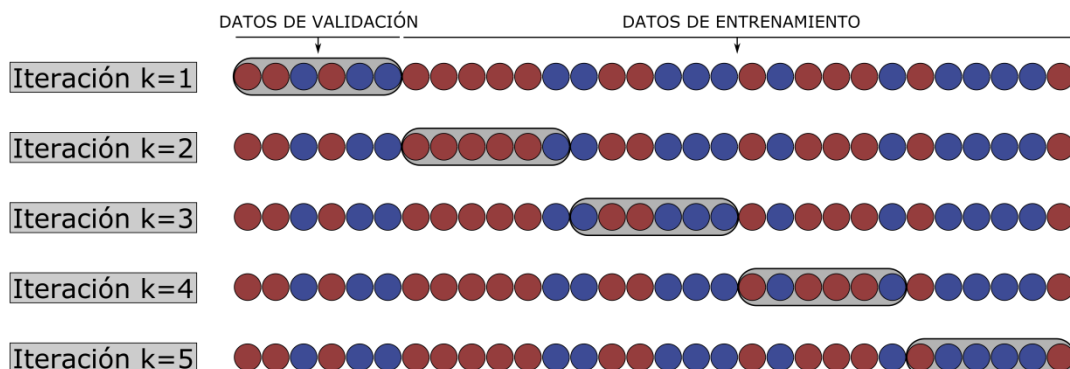


Figura 5.10 Ejemplo de *k-fold cross validation* con $k=5$, en el que se muestra como el grupo de datos de prueba y de datos de entrenamiento varía en cada una de las iteraciones.

En este caso se utiliza *5-fold* y *10-fold cross validation*, que consiste en dividir la muestra inicial de datos en 5 y 10 subconjuntos respectivamente. Uno de los conjuntos se emplea en la validación del modelo y el resto en el entrenamiento del mismo. Esta simulación se repite 5 y 10 veces, variando el conjunto de validación, y obteniendo como resultado final la precisión media de las 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 [51].

$$VPR = \frac{VP}{VP + FN} = \text{sensibilidad o razón de VP} \quad (28)$$

$$FPR = \frac{FP}{FP + VN} = \text{razón de FP} \quad (29)$$

$$ACC = \frac{VP + VN}{P + N} = \text{exactitud} \quad (30)$$

$$SPC = \frac{VN}{N} = \frac{VN}{(FP + VN)} = 1 - FPR = \text{especificidad} \quad (31)$$

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 5.11).

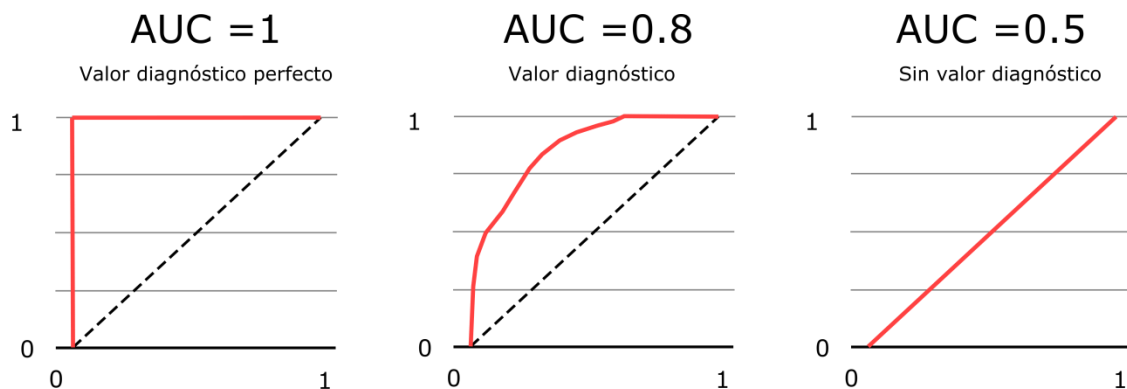


Figura 5.11 Ejemplo de aplicación de las curvas ROC para evaluar la capacidad de una prueba para el diagnóstico de cierta patología en función del ratio de verdaderos positivos y el ratio de falsos positivos VPR-FPR.

6 EXPERIMENTOS REALIZADOS Y RESULTADOS.

En este apartado se describe como se procede a utilizar los algoritmos planteados en la predicción de la evolución de la esclerosis múltiple y los resultados obtenidos de cada uno de los modelos de predicción.

6.1 Predicción mediante redes LSTM.

Como se ha comentado en el apartado 4.3.2, las redes neuronales recurrentes, y especialmente las *Long Short-Term Memory* (LSTM), son apropiadas para tratar secuencias de datos, y por esta razón se han escogido para la predicción de la progresión de la enfermedad objeto del estudio.

Utilizando este tipo de algoritmo se va a plantear dos problemas de predicción diferentes:

- El primero de ellos tratará de predecir la escala de discapacidad funcional del paciente en la visita 5, es decir, 5 años después de la visita basal, a partir de la información proporcionada en la visita basal, y en la visita 1. En este caso se comparará la precisión que puede llegar a alcanzarse utilizando únicamente los datos de la visita basal, y utilizando los datos de la visita basal y la visita 1. Para este modelo de predicción vamos a utilizar los datos de los 73 pacientes disponibles, sumando un total de 146 secuencias de datos (2 ojos por cada paciente).
- El segundo problema propuesto a resolver, tratará de predecir la escala de discapacidad funcional del mismo modo que el primer modelo, con la diferencia de que se tratará de predecir la EDSS en la visita 10 empleando los datos obtenidos en la visita basal, la 1 y la 2. En este caso, se comparará la capacidad de clasificar los pacientes correctamente cuando utilizamos la información de las dos primeras visitas con la precisión alcanzada cuando utilizamos las 3 primeras visitas. Para este modelo se descartan los pacientes que no cuentan con información en la visita 10, quedando 62 sujetos que hacen un total de 124 secuencias disponibles.

Para realizar estos modelos vamos a emplear los algoritmos programados en la librería de *Deep Learning* disponible en Matlab. En concreto, vamos a utilizar la red LSTM bidireccional (*biLstmLayer*), que aprende dependencias bidireccionales entre los pasos de tiempo de las secuencias. Este algoritmo nos permite elegir una serie de parámetros que podemos ajustar para conseguir optimizar la estructura de los algoritmos y maximizar la precisión en la predicción:

- Número de unidades ocultas (*NumHiddenUnits*): especificando el número de unidades o bloques LSTM que formarán la red neuronal podemos ajustar en cierta medida la precisión del modelo, ya que añadir unidades LSTM aportará mayor número de grados de libertad a la red para poder ajustarse al problema, pero a su vez, al añadir más unidades es necesario un mayor número de datos de entrenamiento para hacer un modelo preciso. Por estas razones se debe encontrar un número de bloques ocultos óptimo en cada modelo.
- Función de activación de las unidades LSTM (*StateActivationFunction*): podemos seleccionar entre dos funciones de activación diferentes para actualizar el estado

oculto de las celdas, la función tangente hiperbólica (\tanh) y la función *softsing* ($\text{softsing}(x) = \frac{x}{1+|x|}$).

- Función de activación aplicada a las puertas (*GateActivationFunction*): Esta función se utiliza para el cálculo de las puertas de entrada, salida y olvido. Se puede elegir entre dos funciones de activación: función sigmoide (*sigmoid*) y la función denominada *hard-sigmoid*.
- Estados iniciales de los parámetros: se puede seleccionar el valor en el instante de tiempo 0 tanto en el estado de la celda (*CellState*) como del estado oculto (*HiddenState*) de los bloques LSTM.

Además de los parámetros que podemos variar al crear la red LSTM, hay diferentes opciones de entrenamiento que nos ofrece el software para las redes neuronales de aprendizaje profundo (*deep learning*). En primer lugar, nos da la opción de elegir entre tres algoritmos de optimización diferentes, descenso del gradiente estocástico con momento (SGDM, del inglés *Stochastic Gradient Descent with Momentum*), "RMSProp" (del inglés, *Root Mean Square Propagation*) y "Adam" (derivado del inglés *adaptive moment estimation*). Estos algoritmos utilizan el método del descenso del gradiente estocástico para actualizar los parámetros de la red, que evalúa el gradiente y actualiza los parámetros usando un subconjunto del total de entrenamiento. Este subconjunto se llama mini-lote. Cada evaluación del gradiente usando el mini-lote es una iteración. En cada iteración, el algoritmo da un paso hacia la minimización de la función de error. El paso completo del algoritmo en todo el conjunto de entrenamiento con mini-lotes es una época. Puede especificar el tamaño del mini-lote y el número máximo de épocas usando los argumentos "MiniBatchSize" y "MaxEpochs", respectivamente. Además, podemos especificar otra serie de datos para el entrenamiento como el ratio de entrenamiento o ratio de descenso del gradiente (*GradientDecayFactor*) que puede variar entre 0 y 1 y marcará la velocidad de convergencia, o establecer un límite o umbral para el gradiente (*GradientThreshold*), que marca la máxima reducción de los parámetros de la red cuando el gradiente excede este límite.

Teniendo en cuenta todos estos parámetros, se desarrollan 4 modelos de predicción diferentes (ver Figura 6.1):

1. M1: El primer modelo predecirá la EDSS de los pacientes en el año 5 (o visita 5) con la información de la primera visita en la que el paciente es sometido a la prueba OCT, es decir la visita 0 o basal.
2. M2: El segundo modelo, al igual que el primero, predecirá la EDSS de los pacientes en el año 5, pero ahora con la información de la primera y de la segunda visita, es decir del año 0 (visita basal que tienen todos los pacientes) y la siguiente visita de la que disponga el sujeto (año 1 en la mayoría de los casos, y en su defecto año 2).
3. M3: El tercer modelo tratará de predecir la escala EDSS en el año 10 de cada sujeto utilizando la información de las pruebas realizadas en la primera y segunda visita.
4. M4: Por último, se realizará un modelo que prediga, al igual que el tercero, la escala de discapacidad en el año 10, pero esta vez usando la información de las 3 primeras visitas (0, 1 y 2; 0, 2 y 3; 0, 1 y 3; 0, 1 y 4, dependiendo de las visitas de las que disponga cada sujeto).

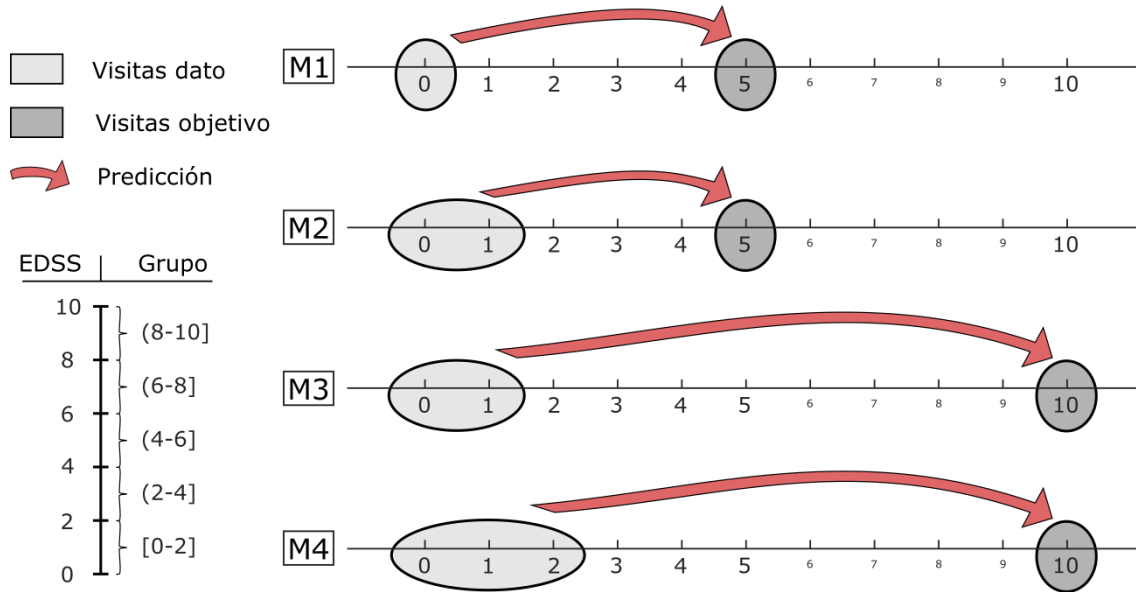


Figura 6.1 Representación de los 4 modelos de predicción desarrollados con LSTM.

Estos modelos serán entrenados con las secuencias de datos de todas las visitas de las que disponga cada sujeto. Además, se le va a asignar a cada sujeto una clase, dependiendo la escala EDSS que tenga en la visita correspondiente a la predicción que deseamos hacer, es decir, en los modelos 1 y 2 la EDSS en la visita 5, y en los modelos 3 y 4 la EDSS en la visita 10. Vamos a clasificar los pacientes en 5 rangos de escala de discapacidad:

Tabla 3. Frecuencia de cada grupo de escala de EDSS para los modelos de clasificación.

| Rango de EDSS | [0-2] | (2-4] | (4-6] | (6-8] | (8-10] | Total |
|-------------------------|-------|-------|-------|-------|--------|-------|
| Nº sujetos en visita 5 | 29 | 19 | 14 | 10 | 1 | 73 |
| Nº sujetos en visita 10 | 21 | 18 | 11 | 11 | 1 | 62 |

A continuación, se van a mostrar los mejores resultados obtenidos para cada uno de los modelos de predicción:

Tabla 4. Tabla de resultados para cada uno de los 4 modelos planteados con las redes LSTM.

| PRUEBA | Tipo RED | NumHid denUnit | Orden | Cross Validation | Normalizado | Mini_batch | Epoch | %ACC |
|-----------------|---------------------|----------------|------------------|------------------|-------------|------------|------------|---------------|
| Modelo 1 | bilstm Layer | 50 | Aleatorio | 10-fold | SI | 10 | 400 | 0,5069 |
| Modelo 2 | bilstm Layer | 20 | Aleatorio | 5-fold | SI | 10 | 400 | 0,6781 |
| Modelo 3 | bilstm Layer | 80 | Aleatorio | 5-fold | NO | 10 | 300 | 0,6320 |
| Modelo 4 | bilstm Layer | 50 | Aleatorio | 10-fold | SI | 10 | 800 | 0,7895 |

Como era de esperar, los modelos que más información aportan son con los que se consigue una mejor predicción del estado de discapacidad del paciente, entre los modelos 1 y 2, el mejor resultado se obtiene utilizando la información de las dos

primeras visitas, y entre los modelos 3 y 4, nuevamente el mejor resultado se obtiene usando la información de las tres primeras visitas.

6.1.1 Área bajo las curvas ROC.

Como se ha comentado en el apartado 5.4.2, las curvas ROC son representaciones gráficas utilizadas para cuantificar la capacidad de un clasificador binario para clasificar correctamente los datos. Sin embargo, en los modelos planteados, la clasificación no es binaria, sino que hay 5 clases diferentes de variable objetivo. Por ese motivo se debe redefinir el concepto de curva ROC para cada uno de éstos modelos. Para poder representar éstas curvas se realiza una clasificación binaria con cada uno de las clases de los modelos, es decir, representaremos una curva ROC para la cada una de las clases, clasificando como correcto o incorrecta la predicción de que un ejemplo pertenezca o no a la clase correspondiente.

A continuación, se realiza la representación de las curvas ROC para cada uno de los modelos que se han creado. Se representan 5 curvas en cada modelo, correspondientes a las 5 subclases de datos en las que se clasifican los pacientes estudiados.

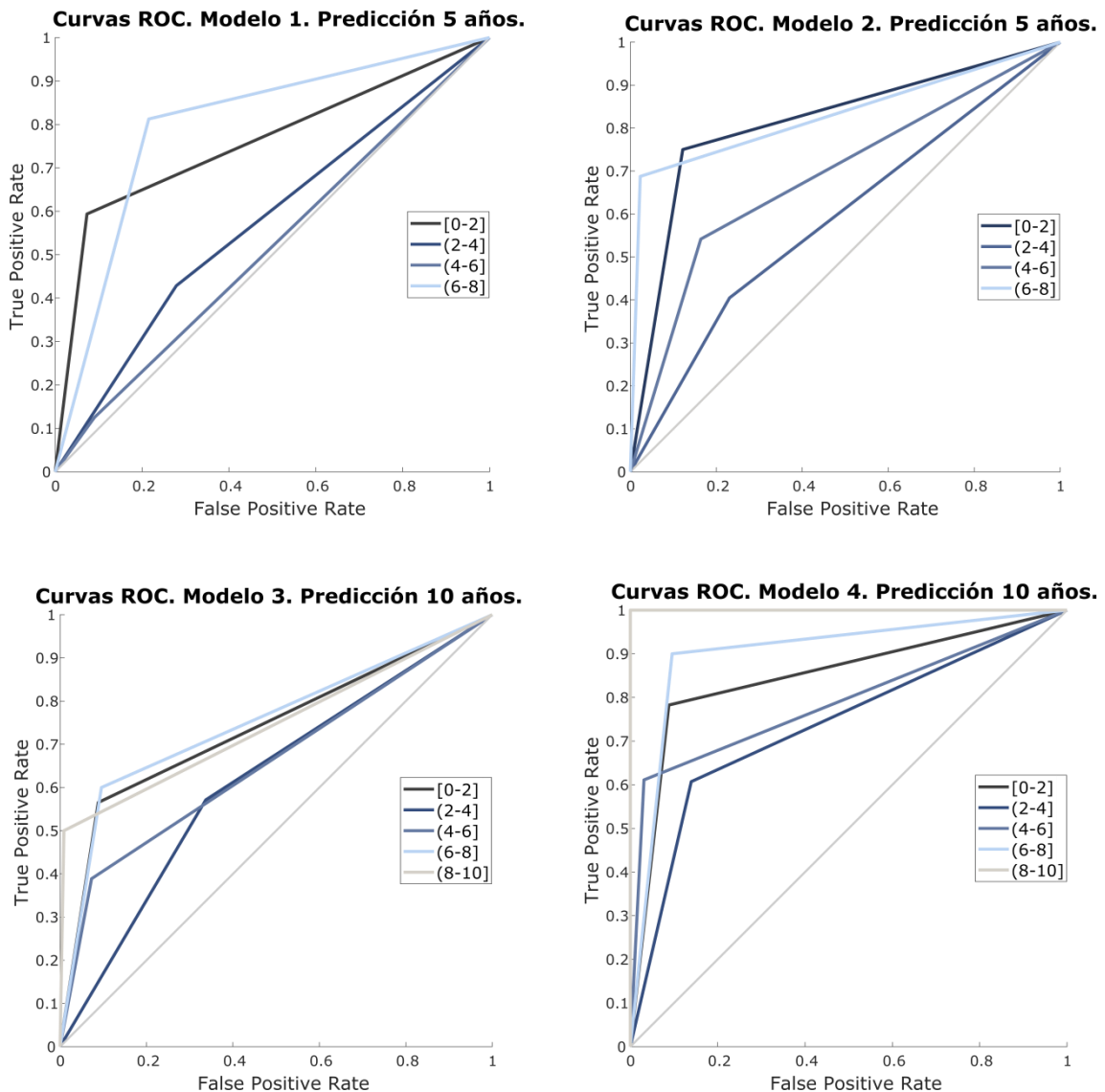


Figura 6.2. Representación de las curvas ROC para los modelos desarrollados con redes LSTM.

Como era de esperar, el modelo con mayor capacidad diagnóstica es el modelo 4, ya que es el modelo que encierra mayor área bajo sus curvas ROC. Éste es el modelo que más información utiliza para predecir la escala de discapacidad de los pacientes.

Si comparamos los primeros modelos, se observa que el modelo 2 tiene mucha mayor capacidad diagnóstica que el 1, ya que el segundo encierra mucha más área bajo sus curvas ROC. Esto es congruente con el hecho de que para el segundo modelo se emplea más información en la predicción de la discapacidad que en el primero. Si observamos los modelos 3 y 4, se puede llegar a la misma conclusión, el modelo 4 tiene mayor capacidad diagnóstica ya que se emplean 3 visitas en la predicción en lugar de 2 visitas como se hace en el tercero.

6.2 Predicción mediante algoritmos de clasificación binaria.

Además de las redes neuronales recurrentes, vamos a emplear otros tipos de algoritmos para intentar predecir la evolución de la enfermedad en los pacientes. En concreto, vamos a emplear clasificadores binarios para saber si la discapacidad en el paciente (EDSS) va a empeorar o no. A diferencia de las redes LSTM, estos clasificadores no utilizan la variable tiempo en sus modelos, por tanto, no reconocen secuencias de datos que se suceden en el tiempo. Por este motivo se debe replantear los modelos y los objetivos que vamos a tratar de predecir.

Se plantean dos problemas de predicción diferentes con estos algoritmos:

- En el primero de ellos tratará de utilizar la información de las dos primeras visitas de cada paciente para predecir si la escala de discapacidad EDSS de los pacientes va a sufrir un empeoramiento o no en la visita 5. Para este modelo de predicción vamos a utilizar los datos de los 73 pacientes disponibles, sumando un total de 146 secuencias de datos (2 ojos por cada paciente).
- El segundo, va a utilizar los datos de las 3 primeras visitas de cada paciente para tratar de predecir si la EDSS va a sufrir un empeoramiento en la visita 10. Para este modelo se descartan los pacientes que no cuentan con información en la visita 10, quedando 62 sujetos que hacen un total de 124 secuencias disponibles.

Para poder resolver los dos problemas planteados, se desarrollan 4 modelos de predicción diferentes (ver Figura 6.3):

1. M1: El primer modelo, tratará de predecir si la escala EDSS va a empeorar en 5 años, utilizando los datos de las dos primeras visitas de cada paciente. En la clasificación de los pacientes, éstos se dividen en dos clases, la primera denominada pacientes que **no-empeoran**, para los cuales la escala EDSS no sufre un aumento mayor a 0,5 punto, y, por otro lado, los pacientes que si **empeoran**, que serán aquellos cuya EDSS sufra un aumento mayor a 0,5 punto en los 5 años.
2. M2: El segundo modelo, al igual que el primero, trata de predecir la evolución de la escala EDSS a los 5 años utilizando los datos de las dos primeras visitas de cada paciente, pero a diferencia del modelo anterior, los pacientes se van a clasificar en 3 clases diferentes; la primera serán los pacientes que sufran una mejoría en cuanto a la EDSS se refiere, es decir, los que tengan una diferencia negativa de la escala EDSS en los 5 años. Por otra parte, estarán los que **no-empeoran**, cuya diferencia de EDSS no supera 1 punto, y por último los pacientes cuya diferencia de EDSS supera 1 punto y se clasifican como pacientes que **empeoran**.

3. M3: El tercer modelo tratará de predecir si la escala EDSS va a empeorar en 10 años utilizando los datos de las tres primeras visitas de cada paciente. Al igual que el modelo primero, los pacientes se clasifican en dos clases diferentes, los que **empeoran** y los que **no-empeoran**.
4. M4: Por último, el modelo cuarto, al igual que el tercero, tratará de predecir si la escala EDSS va a sufrir un empeoramiento a los 10 años utilizando los datos de las 3 primeras visitas. Como en el segundo modelo, se van a clasificar los pacientes en 3 clases de datos, los que **mejoran**, los que **empeoran** y los que **no-empeoran**.

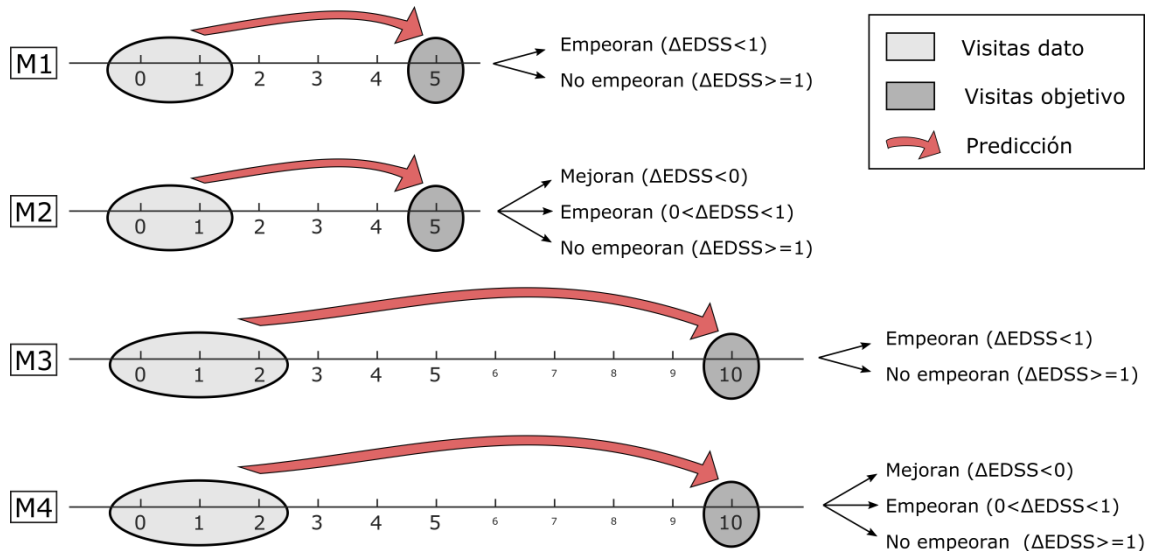


Figura 6.3 Representación de los 4 modelos de predicción desarrollados mediante algoritmos de clasificación.

En la siguiente tabla se muestra la frecuencia de cada una de las clases, tanto para los modelos 1 y 3 que disponen de dos clases de datos (No-empeoran y empeoran) y para los modelos 2 y 4 con 3 clases diferentes de datos (los que mejoran, lo que no empeoran y los que empeoran):

Tabla 5 Frecuencia de cada grupo de paciente respecto a la EDSS.

| Clases | Modelos de 2 clases | | Modelos de 3 clases | | |
|--------------------|---------------------------------------|---------------------------------------|---------------------------------|---|-------------------------------------|
| | No-Empeora ($\Delta EDSS < 0.5$) | Empeora ($\Delta EDSS \geq 0.5$) | Mejora ($\Delta EDSS < 0$) | No-Empeora ($0 < \Delta EDSS < 1$) | Empeora ($\Delta EDSS \geq 1$) |
| Modelo 1 (5 años) | 86 | 58 | -- | -- | -- |
| Modelo 2 (5 años) | -- | -- | 38 | 70 | 36 |
| Modelo 3 (10 años) | 48 | 76 | -- | -- | -- |
| Modelo 4 (10 años) | -- | -- | 10 | 58 | 56 |

Como se ha descrito en el apartado 4.4, para desarrollar los modelos planteados se van a emplear 4 algoritmos de clasificación diferentes: Las máquinas de soporte vectorial (SVM), el algoritmo *k-nearest neighbors* (KNN), el modelo *Naïve Bayes* (NB) y los árboles de decisión (DecisionTrees). A continuación, se va a describir como se tratan de optimizar cada uno de los modelos ajustando ciertos parámetros del

algoritmo para tratar de obtener el mejor resultado en cada uno de ellos. Para el uso de los algoritmos vamos a emplear la librería de *Deep Learning* disponible en Matlab.

6.2.1 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 [57].

Se emplea la función de Matlab *fitcsvm* que genera un modelo SVM de clasificación binaria basado en las funciones kernel. Del mismo modo que en las LSTM se puede ajustar ciertos parámetros del algoritmo para conseguir el modelo óptimo en un problema, en concreto, los parámetros [58] que se van a variar en éste algoritmo son los siguientes:

- 1) Función Kernel (*KernelFunction*): Se pueden utilizar diferentes funciones Kernel para el

Tabla 6. 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 6), para mejorar la capacidad de clasificar de la SVM.

- 3) Algoritmos de optimización empleados (*optimization routine*): los diferentes algoritmos de optimización [59] que se pueden utilizar en la resolución del problema son los siguientes:

Tabla 7. 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 [59]. |
| L1QP | Utiliza programación cuadrática para implementar la minimización. |
| SMO | Optimización mínima secuencial [60]. |

- 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 1000 iteraciones en todas las pruebas que van a realizarse con el algoritmo SVM.

Variando estos parámetros se obtendrá el algoritmo optimizado para cada uno de los modelos.

Como se ha comentado en el apartado 4.4.1, las SVM son algoritmos de clasificación binaria, es decir, a partir de una serie de ejemplos de entrenamiento son capaces de clasificar correctamente un nuevo ejemplo entre dos clases. Sin embargo, es posible utilizar este tipo de algoritmos para realizar una clasificación de ejemplos con más de dos clases. Esto se consigue realizando tantos modelos SVM como tipos de clases estén divididos los ejemplos. Cada uno de los modelos SVM predecirá si un ejemplo pertenece a una clase o no, y este proceso se repetirá para cada una de las clases, obteniendo al final un predictor conjunto que englobe todas las clases.

A continuación, se muestra el diagrama de flujo y un esquema del código empleado en Matlab para utilizar el algoritmo SVM en la predicción multiclase:

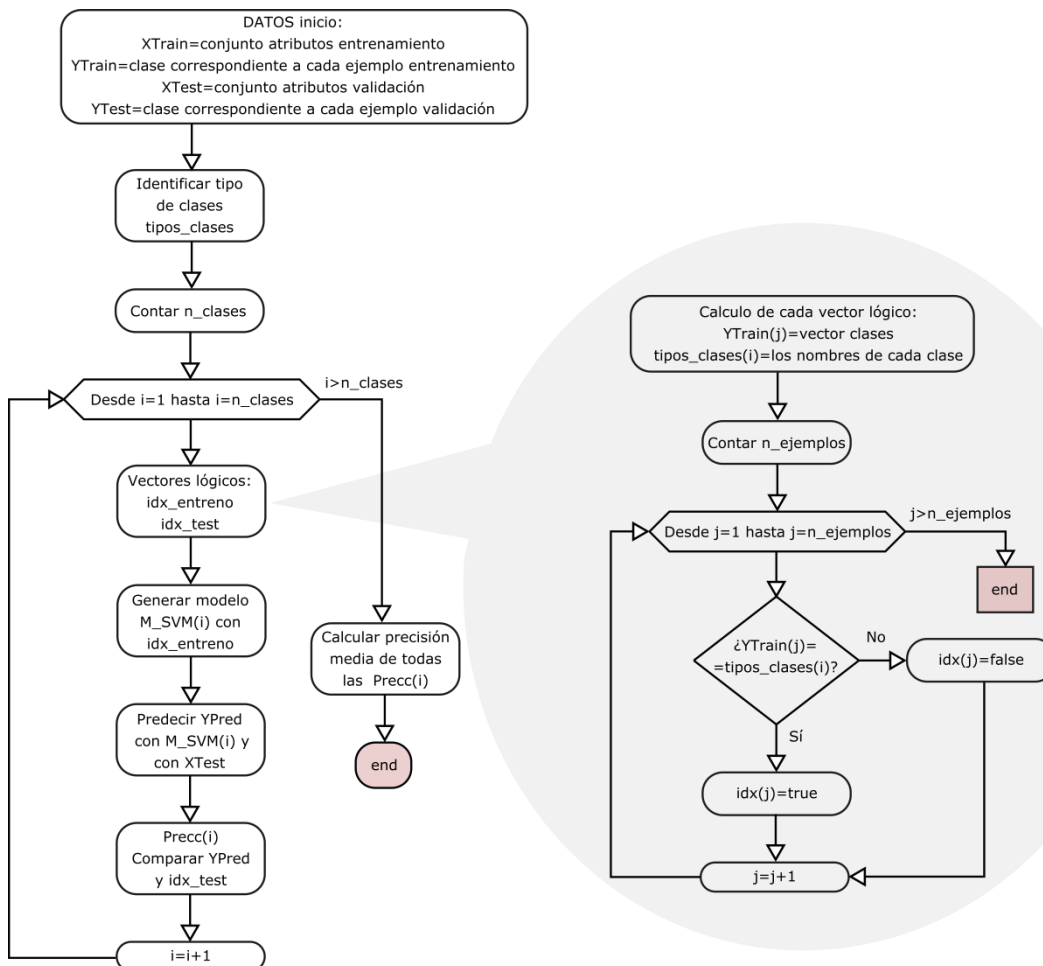


Figura 6.4. Diagrama de flujo del código empleado para utilizar las SVM en la clasificación de más de dos clases.

```

%XTrain --> conjunto atributos de entrenamiento.
%YTrain --> clase de cada ejemplo de entrenamiento.
%XTest --> conjunto atributos de validación.
%YTest --> clase de cada ejemplo de validación.
%Tipos de Clases --> 1.Mejora, 2.No empeora, 3.Empeora
num_clases=3;
for j = 1:num_clases %Se crea un modelo por cada una de las clases a predecir
    indx_entreno = strcmp(YTrain,classes(j)); %Crea para cada uno de los
    modelos un vector lógico con 1 si pertenece a la clase correspondiente o0
    si no (Conjunto entrenamiento).
    indx_test = strcmp(YTest,classes(j)); %Lo mismo pero en el conjunto de
    validación.
    SVMModels{j} = fitcsvm(XTrain,indx_entreno,'ClassNames',[false true]);
    %Crea cada modelo SVM
    YPre=predict(SVMModels{j},XTest); %Predecimos si pertenece o no a la clase
    correspondiente todos los datos del conjunto de validación.
    precision=comp(YTest,YPre); %Comparamos los resultados predichos con los
    de Test para saber la precisión obtenida.
End

```

Figura 6.5 Algoritmo empleado en Matlab para utilizar el algoritmo SVM en clasificación multiclase.

Con este código somos capaces de adaptar el algoritmo SVM para la predicción de conjuntos de datos con más de dos clases, y podemos utilizar este algoritmo en el desarrollo de los modelos 2 y 4.

6.2.2 Predicción mediante kNN.

Para el desarrollo del algoritmo de clasificación *k-nearest neighbour* (KNN) se emplea la función de Matlab *fitcknn*, que genera un modelo de predicción con tantas clases objetivo como clases diferentes tengan los datos de entrenamiento. Los parámetros con lo que vamos a poder optimizar cada modelo son los siguientes:

- 1) Número de “vecinos más cercanos” a encontrar (*NumNeighbours*): Se puede seleccionar el número de vecinos más cercanos que se han de buscar en la clasificación de cada ejemplo a clasificar.
- 2) Métrica de la distancia (*Distance*): en esta función se puede escoger la función que va a encargarse de medir la distancia entre los vecinos. Las funciones entre las que se puede escoger son las siguientes:

Tabla 8. Funciones de métrica de distancia.

| NSMethod | Distance Metric Names |
|-------------------|---|
| <i>Exhaustive</i> | <i>EshastiveSearcher</i> |
| <i>Kdtree</i> | <i>'cityblock', 'chebychev', 'euclidean', 'minkowski'</i> |

- 3) Función de ponderación de la distancia (*DistanceWeight*): Con esta variable se puede especificar si se pondera la distancia entre los vecinos o no, pudiendo elegir entre los parámetros de la siguiente tabla:

Tabla 9. Tipos de función de ponderación de la distancia.

| Función | Descripción |
|-----------------------|---|
| <i>Equal</i> | Sin ponderación de la distancia |
| <i>Inverse</i> | Ponderación de 1/distancia |
| <i>Squaredinverse</i> | Ponderación de 1/distancia ² |
| <i>@fnc</i> | Podemos especificar una función propia |

Además, existe una función que trata de optimizar estos parámetros automáticamente (*OptimizeHyperparameters*) para tratar de minimizar el error cometido por el algoritmo con el conjunto de entrenamiento. Esta función tiene el inconveniente que solo disminuye el error de entrenamiento, sin tener en cuenta el posible sobreentrenamiento o sobreajuste del modelo.

El algoritmo kNN permite predecir conjuntos de datos con más de una clase, es decir, podemos utilizarlo sin hacer ninguna modificación para desarrollar los modelos 2 y 4.

6.2.3 Predicción mediante el modelo NB.

El modelo clasificador Naïve Bayes (NB) trata de relacionar de manera independiente cada una de los atributos de cada ejemplo con la clase a la que pertenece. El entrenamiento del modelo consiste en estimar la función de probabilidad $P(X|Y)$ que prediga la clase Y en función del vector X en cada ejemplo.

Para el desarrollo de los modelos a partir de este algoritmo se emplea la función de Matlab *fitcnb*, que generará un clasificador basado en el teorema de Bayes. Los principales parámetros que podemos utilizar para optimizar un modelo construido a partir de este algoritmo son los siguientes:

- 1) Distribución de los datos (*DistributionNames*): Se puede especificar la distribución que siguen los datos correspondientes a los atributos predictores del conjunto de datos de entrenamiento.

Tabla 10. Tipos de distribuciones correspondientes a los atributos predictores.

| Función | Descripción |
|---------------------------------|---|
| <i>Kernel</i> | Predictores que tienen una distribución continua. Requiere mayor tiempo de cómputo. |
| <i>Multinomial</i> | Apropiada cuando cada observación tiene una distribución aleatoria. |
| <i>Multivariate multinomial</i> | Apropiada cuando las observaciones son categóricas. |
| <i>Normal (Gaussian)</i> | Predictores que tienen una distribución normal o gaussiana. |

- 2) Tipo de suavizado kernel (*Kernel*): En el caso de escoger la distribución kernel se puede elegir el tipo de suavizado con el que se entrena el modelo:

Tabla 11. Tipos de suavizado kernel.

| Kernel | Fórmula |
|-----------------------|--|
| <i>Box (uniforme)</i> | $f(x) = 0.5I\{ x \leq 1\}$ |
| <i>Epanechnikov</i> | $f(x) = 0.75(1 - x^2)I\{ x \leq 1\}$ |
| <i>Gaussiana</i> | $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-0.5x^2)$ |
| <i>Triangular</i> | $f(x) = (1 - x)I\{ x \leq 1\}$ |

Del mismo modo que en el algoritmo KNN, en el algoritmo clasificador NB existe la función de optimización de parámetros que tratará de minimizar el error de entrenamiento cometido por el clasificador en el entrenamiento. De nuevo hay que

tener en cuenta la posibilidad de caer en un sobreajuste del modelo en el conjunto de datos de entrenamiento.

Este algoritmo permite predecir conjuntos de datos con más de una clase, es decir, podemos utilizarlo sin hacer ninguna modificación para desarrollar los modelos 2 y 4.

6.2.4 Predicción mediante árboles de clasificación.

Por último, para el uso del algoritmo de clasificación basado en árboles de decisión se emplea la función de Matlab denominada ***fitctree***. En este caso los parámetros a optimizar son los siguientes:

- 1) Número máximo de división de decisiones (*MaxNumSplits*): Podemos especificar el número máximo de divisiones o nodos de decisión para poder limitar la complejidad del árbol de decisión creado.
- 2) Mínimo número de observaciones por nodo (*MinLeafSize*): También podemos especificar el número mínimo de observaciones por cada uno de los nodos finales.
- 3) Criterio de división (*SplitCriterion*): Podemos elegir entre tres criterios a seguir para definir la división en cada nodo:
 - 'gdi': Índice de diversidad de Gini.
 - 'Twoing': Regla de Twoing.
 - 'desviance': para conseguir la reducción máxima de la desviación.
- 4) Número de predictores para seleccionar al azar para cada división (*NumVariablesToSample*).

De la misma manera que los algoritmos anteriores, se puede seleccionar una función que optimice (*OptimizeHyperparameters*) automáticamente estos parámetros con el objetivo de minimizar el error cometido en el entrenamiento por el modelo.

Los árboles de decisión también pueden utilizarse en la clasificación de conjuntos de datos multiclase, por lo tanto, no será necesario realizar ninguna modificación en el código para desarrollar los modelos con más de dos clases (modelos 2 y 4).

Por otro lado, existe la posibilidad de saber la influencia que tienen cada uno de los atributos o variables de entrada introducidos en el modelo desarrollado mediante árboles de decisión, es decir, saber cuáles de los parámetros de entrada tienen una mayor relación con el resultado del diagnóstico. Esto se consigue utilizando la función de Matlab correspondiente (*oobPermutedPredictorImportance*), que, para un modelo desarrollado con la función ***fitctree***, devuelve un vector con la importancia relativa de todos los atributos de entrada respecto a la predicción de entrenamiento. En este caso concreto el resultado obtenido para la importancia de los atributos del modelo 4 se muestra en la Figura 6.6.

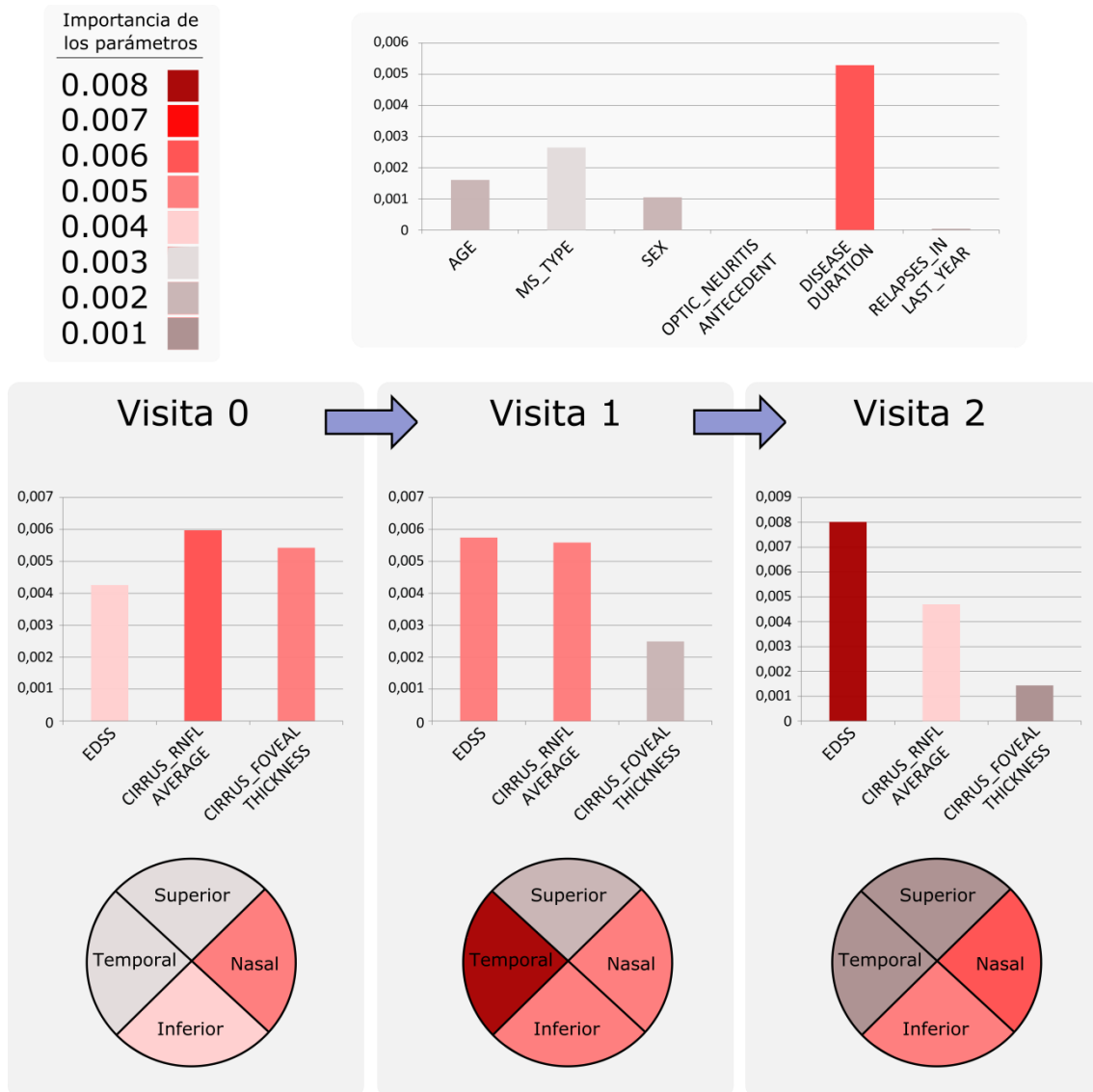


Figura 6.6. Importancia de los atributos de entrada para un modelo construido con la función *fitctree*.

De la figura se puede sacar como conclusión que las variables: *antecedentes de neuritis óptica* y *recaidas en el último año* no tienen apenas importancia en este modelo concreto de predicción desarrollado utilizando los árboles de clasificación.

También podemos observar cuáles son las variables con mayor peso en la predicción. Por un lado, en cuanto a las variables no secuenciales, la que mayor importancia cobra es la *duración de la enfermedad* en el paciente, y por otro lado, para las variables secuenciales, los atributos con mayor importancia son; la *escala de discapacidad EDSS* en las visitas introducidas como dato, y algún dato concreto de espesor de RNFL, como por ejemplo el espesor de la zona temporal de la visita 1 y la zona nasal de las 3 visitas dato.

6.2.5 Resultados obtenidos para cada modelo.

A continuación, se muestran los resultados obtenidos con cada uno de los diferentes algoritmos y modelos planteados.

Tabla 12. Resultados obtenidos para el primer modelo planteado con cada uno de los algoritmos utilizados.

| Modelo | Clasificación | Algoritmo | F. Matlab | k-fold CV | Precisión |
|----------|--------------------------------|----------------------------|----------------|-----------|---------------|
| Modelo 1 | 2 clases Predicción 5 años | SVM (polinomial) | Fitcsvm | 5 | 0,5903 |
| | | SVM (linear) | Fitcsvm | 5 | 0,6458 |
| | | SVM (gaussian) | Fitcsvm | 5 | 0,6181 |
| | | Decision tree | Fitctree | 5 | 0,625 |
| | | K nearest neighbors | Fitcknn | 5 | 0,8958 |
| Modelo 2 | 3 clases Predicción 5 años | Naive Bayes model | Fitcnb | 5 | 0,5625 |
| | | SVM multiclass | Fitcecoc | 5 | 0.5833 |
| | | K nearest neighbors | Fitcknn | 5 | 0,8681 |
| | | Decision tree | Fitctree | 5 | 0,6042 |
| Modelo 3 | 2 clases Predicción 10 años | Naive Bayes model | Fitcnb | 5 | 0.4583 |
| | | SVM (polinomial) | Fitcsvm | 5 | 0,5242 |
| | | SVM (linear) | Fitcsvm | 5 | 0,6935 |
| | | SVM (gaussian) | Fitcsvm | 5 | 0,6855 |
| | | Decision tree | Fitctree | 5 | 0,5726 |
| Modelo 4 | 3 clases Predicción 10 años | K nearest neighbors | Fitcknn | 5 | 0,8629 |
| | | Naive Bayes model | Fitcnb | 5 | 0,5806 |
| | | SVM multiclass | Fitcecoc | 5 | 0.6371 |
| | | K nearest neighbors | Fitcknn | 5 | 0.9113 |
| Modelo 4 | 3 clases Predicción 10 años | Decision tree | Fitctree | 5 | 0,6532 |
| | | Naive Bayes model | Fitcnb | 5 | 0.6371 |

Como puede apreciarse en la tabla de resultados, el algoritmo de clasificación que mejor ha funcionado es el kNN, llegando a alcanzar una precisión de predicción de más del 90% en el último de los modelos. Al igual que sucedía en los modelos planteados con las redes LSTM, el mejor resultado lo obtenemos en el modelo que empleados 3 visitas para predecir la evolución de la escala de discapacidad en la visita 10.

6.2.6 Área bajo las curvas ROC.

En este apartado se van a representar las curvas ROC del mejor resultado de cada modelo, en este caso correspondientes todos a modelos desarrollados con el algoritmo kNN. Para los modelos 1 y 3, que clasifican entre pacientes que empeoran y pacientes que no, se representan una sola curva ROC en cada modelo ya que corresponden a una clasificación binaria. Sin embargo, para los modelos 2 y 4 es necesario representar 3 curvas, una por cada una de las subclases de pacientes en los que se debe clasificar.

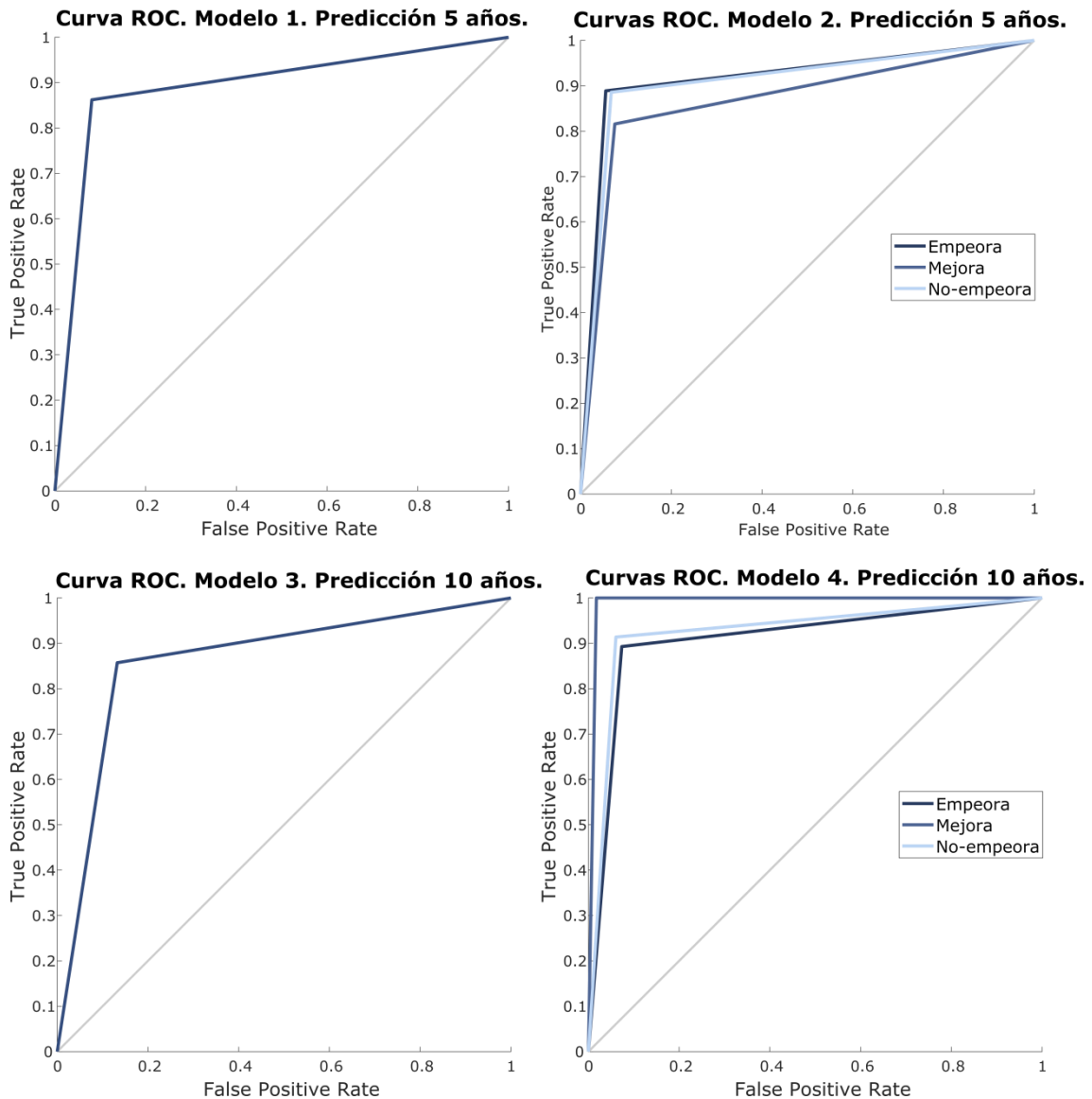


Figura 6.7. Representación de las curvas ROC para los cuatro modelos con algoritmos de clasificación.

Comparando los modelos de clasificación dos a dos, podemos observar como los modelos en los que se introducen 3 clases distintas de variables objetivo predicen mejor que los modelos en los cuales únicamente se distingue entre sujetos que empeoran y sujetos que no lo hacen.

Las curvas ROC correspondientes al modelo 4 son las que más se aproximan al modelo de diagnóstico perfecto. En concreto, la clasificación de los pacientes cuya EDSS mejora en los 10 años de seguimiento es casi perfecta.

7 CONCLUSIONES.

La inteligencia artificial (IA) ha experimentado un gran desarrollo en los últimos años, aplicándose con éxito en numerosos campos científicos. Concretamente, la IA es capaz de mejorar la capacidad de los médicos de predecir el resultado del paciente y el desarrollo de la enfermedad en diferentes patologías. Con la finalidad de ayudar a mejorar el conocimiento de los neurólogos en el desarrollo de la esclerosis múltiple se ha planteado este trabajo, cuyo primer objetivo era conseguir un modelo eficaz que consiga predecir el grado de discapacidad que va a desarrollar un sujeto con el fin de poder adecuar el tratamiento recibido por éste. Además, se pretende comparar la capacidad de los distintos algoritmos de clasificación empleados para el desarrollo de los modelos para saber cuál se adecua mejor al problema concreto tratado.

En primer lugar, viendo los resultados obtenidos para la predicción de la discapacidad funcional de los pacientes (ver apartados 6.1 y 6.2.5), podemos concluir diciendo que, para el problema de predicción de EDSS a partir de las secuencias de datos del espesor de la retina, el algoritmo que mejor se adapta al problema es el algoritmo ***k-nearest neighbours***, llegando a alcanzar un máximo del 91% de precisión en el problema de clasificación en tres clases (Mejora, No-empeora, Empeora) en el seguimiento de 10 años y utilizando la información de las 3 primeras visitas, es decir, con una predicción a 7 años vista.

En los modelos de predicción se pretende relacionar la evolución del espesor de la retina con la evolución de la escala de discapacidad en pacientes afectados por EM. Si observamos la Figura 7.1, se puede ver la evolución que tiene el EDSS medio de todos los pacientes a lo largo del seguimiento a 10 años, y de la misma manera, el cambio que tiene el espesor medio en la zona de la fovea a lo largo del mismo seguimiento. Podemos

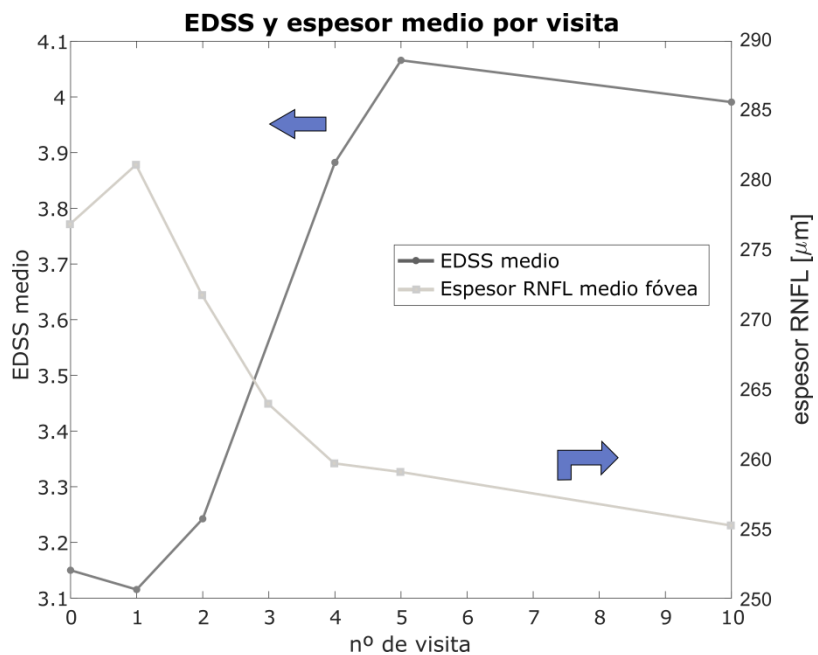


Figura 7.1. EDSS medio y espesor medio de la RNFL en la zona macular a lo largo del seguimiento a 10 años

notar que la escala de discapacidad EDSS sufre un aumento considerable en los primeros 5 años de seguimiento (1 punto EDSS) para después estabilizarse en la visita número 10. Si nos fijamos en el espesor de la RNFL vemos que, de la misma manera, en los primeros 5 años se sufre el mayor descenso de este espesor para más tarde estabilizarse en la visita número 10. Las tendencias de ambos parámetros son inversas, lo que nos hace pensar que puede existir una relación entre los dos parámetros en el progreso de la patología estudiada.

Por otro lado, a la hora de comparar los algoritmos de clasificación, los podemos distinguir en dos grupos, los algoritmos que sufren o pueden sufrir el fenómeno de sobre-entrenamiento o sobre-ajuste y los que no lo sufren. Este fenómeno tiene como consecuencia la disminución de la precisión en el grupo de validación, es decir, una reducción de la capacidad de generalización en los ejemplos que no se han utilizado para el entrenamiento, debido al exceso de ajuste del modelo a los datos utilizados en el entrenamiento. De los algoritmos utilizados; las redes recurrentes LSTM, el algoritmo k-NN, el modelo de Naïve Bayes y los árboles de decisión sufren este fenómeno, por tanto, para todos ellos existe un punto óptimo en el cual las precisiones de entrenamiento y de validación son las mejores. El algoritmo SVM, sin embargo, es un clasificador que se caracteriza por su robustez y por no tener mínimos locales en la función de error, es decir, no sufren el fenómeno de sobre-entrenamiento.

Podemos decir que los modelos que alcanzan una precisión aceptable para poder utilizarlos como prueba diagnóstica son los construidos mediante las redes LSTM (79% de precisión en el modelo 4) y los construidos mediante el algoritmo k-NN (91% de precisión en el modelo 3), por tanto el resto de algoritmos utilizados no son capaces de adaptarse de manera eficaz a la alta no-linealidad del problema, ya sea por su robustez, como las SVM, o por la mayor necesidad de ejemplos para desarrollar un modelo con mayor capacidad de generalización.

Si comparamos las precisiones alcanzadas entre los modelos construidos con las redes neuronales recurrentes y los modelos desarrollados con algoritmos de clasificación binaria, se observa que, por lo general, es mayor la exactitud para estos últimos. Notar la diferencia entre los dos tipos de modelos, ya que, para los primeros, en la tarea de entrenamiento se utiliza todas las visitas de los pacientes, esto es, la secuencia completa de visitas y por tanto toda la información, para el desarrollo del modelo, y en el modo recuerdo, es decir, al predecir el grupo de validación, solo se emplean 1, 2 o 3 de las visitas del paciente. Sin embargo, en los modelos binarios, tanto el modo entrenamiento como el modo recuerdo utilizan el mismo número de visitas de cada paciente, además, no utilizan esta información en forma de secuencia, sino que la emplean en forma de vector. Esta diferencia hace pensar que los modelos desarrollados mediante redes LSTM, al contener la información de toda la secuencia, pueden constituir modelos con más utilidad a la hora de comprender la evolución y el desarrollo de la discapacidad en los pacientes con EM y su relación con la evolución de la capa de fibras nerviosas de la retina. Sin embargo, a la vista de los resultados, se puede decir que para construir un modelo con las redes LSTM que fuera capaz de generalizar mejor sería necesario disponer de un número de secuencias de entrenamiento mucho mayor, lo que supone una limitación en la utilización de estas redes. Por el contrario, utilizando los modelos desarrollados para los modelos binarios se alcanza una predicción considerablemente mayor, sacando la conclusión de la menor necesidad de datos de entrenamiento de éstos algoritmos para una buena generalización, en concreto, para los modelos desarrollados con el algoritmo k-NN.

Para determinar el valor de los modelos desarrollados como biomarcador en la EM, otro factor que se debe tener en cuenta es que todos los pacientes incluidos en las bases de datos proceden del Hospital Miguel Servet de Zaragoza, es decir, de la misma zona geográfica, donde las condiciones ambientales y climáticas son las mismas, además, todos ellos pertenecen a etnia caucásica europea. En diversos estudios se

concluye que éstos parámetros pueden tener una gran relevancia en el desarrollo de las enfermedades neurodegenerativas como la EM [61]. Por ésta razón sería interesante completar las bases de datos con pacientes de diferentes zonas geográficas y poder así estudiar la evolución de la enfermedad de forma global, teniendo en cuenta los parámetros nombrados y ver su influencia en el desarrollo de la patología.

Por último, destacar que la OCT es una técnica sencilla, reproducible, no-invasiva y rápida, que permite medir el grosor y los posibles cambios en éste de las diferentes capas de la retina, pudiendo variar las zonas estudiadas y protocolos de adquisición empleados. Por tanto, además de la inclusión de la prueba en el chequeo del seguimiento de los pacientes, se sugiere la incorporación de los modelos de predicción desarrollados mediante los algoritmos de clasificación con la finalidad de ayudar a los neurólogos en la toma de la decisión de cuál es el tratamiento adecuado para el paciente en función de la evolución de la discapacidad para los próximos años. La aplicación de tratamientos más agresivos a los pacientes cuyo pronóstico sugiera un aumento de discapacidad más pronunciado ayudará a tratar de frenar el avance de esa discapacidad, además, servirá para gestionar de una manera más eficaz y eficiente los recursos de los hospitales, ya que los tratamientos agresivos además de ser más novedosos son notablemente más caros.

8 LÍNEAS DE FUTURO.

En primer lugar, para mejorar los modelos de una manera inmediata se debería intentar completar las bases de datos con un mayor número de sujetos, y poder así realizar el entrenamiento de los modelos con una cantidad mayor de información. En especial, sería interesante llegar a tener unas bases de datos más completas para el equipo OCT-Spectralis ya que el seguimiento de los pacientes en éstas es muy pobre y no se pueden sacar conclusiones de los modelos realizados con la información de este dispositivo (ver anexo iii).

Adicionalmente, puede intentar realizarse un estudio de la sensibilidad de los parámetros utilizados para el desarrollo de los modelos (ver Figura 6.6) y decidir cuáles de éstos tienen una mayor importancia en el desarrollo de la enfermedad. Con esta información se pueden completar los modelos de predicción dando una mayor importancia a los parámetros que estén ligados en mayor medida a la evolución de la enfermedad, incluso eliminar los que no tengan relevancia.

Por otro lado, como se ha comentado anteriormente, sería interesante completar las bases de datos con sujetos de otras zonas geográficas para poder estudiar la influencia de parámetros como las condiciones climáticas en la evolución de la enfermedad.

En la actualidad, el Hospital cuenta con un equipo de OCT más moderno que los que se han empleado en este estudio, concretamente el *DRI OCT Triton*. Este equipo ofrece una serie de protocolos de adquisición más precisos y eficaces, por tanto, se podría intentar conseguir una base de datos completa con el seguimiento de pacientes diagnosticados de la enfermedad, utilizando éste para poder llegar a conseguir una mejoría en los modelos de predicción. Sin embargo, todavía la cantidad de datos que se han obtenido con el equipo son insuficientes para ser analizados.

Por último, vistos los buenos resultados obtenidos en la predicción de la evolución de la esclerosis múltiple, se podría intentar aplicar estas técnicas (prueba OCT junto a algoritmos de clasificación) en la predicción de otras enfermedades neurodegenerativas como el Parkinson y el Alzheimer o en otras enfermedades como el glaucoma o la neuritis óptica, en las cuales el sistema nervioso visual se ve afectado.

9 BIBLIOGRAFÍA.

- [1] Vidal-Jordana Á, Sastre-Garriga J, Montalban X. [Optical coherence tomography in multiple sclerosis]. *Rev Neurol* 2012;54:556–63.
- [2] Fernández Fernández ., Fernández VE. Esclerosis múltiple. Concepto. Etiopatogenia. Fisiopatología. Manifestaciones clínicas. Investigaciones paraclínicas. Diagnóstico. Historia natural. Med - Programa Form Medica Contin Acreditado 2007;9:4867–77. doi:10.1016/S0211-3449(07)75463-2.
- [3] 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.
- [4] 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.
- [5] 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/iops.10-6643.
- [6] 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.
- [7] Pelegrín Molina M aA., Gómez-Conesa A. Esclerosis múltiple: estado actual. *Fisioterapia* 2001;23:144–53. doi:10.1016/S0211-5638(01)72946-X.
- [8] García Férriz P. ESCLEROSIS MÚLTIPLE, Etiopatogenia.Tratamiento n.d.
- [9] 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.
- [10] 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.
- [11] 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.
- [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] Martín-Gallego A. El uso de la retina como biomarcador en enfermedades neurodegenerativa. Simulación numérica. Univ Zaragoza 2018.
- [17] Kotelnikova E, Kiani NA, Abad E, Martinez-Lapiscina EH, Andorra M, Zubizarreta I, et al. Dynamics and heterogeneity of brain damage in multiple sclerosis. *PLoS Comput Biol* 2017;13:1–26. doi:10.1371/journal.pcbi.1005757.
- [18] 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.
- [19] 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.
- [20] Cantó E, Comabella M. Biomarcadores en la esclerosis múltiple: estado actual. *Rev Española Escler Mult* 2012;4:20–9.
- [21] AEDEM-COCEMFE. Tipos de Esclerosis Múltiple n.d.
- [22] 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.
- [23] Porras, M.; Núñez, L.; Plascencia, N.; Quiñones, S.; Sauri S. Esclerosis múltiple. Artículo de revisión. *Rev Mex Neuroci* 2007;8:57–66.
- [24] 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.
- [25] Gonzalez Ruiz V. El Sistema Visual Humano 2014. https://w3.ual.es/~vruiz/Docencia/Apuntes/Perception/Sistema_Visual/index.html.
- [26] Aznar Casanova JA. Psicología de la percepción visual n.d. http://www.ub.edu/pa1/node/area_visual.
- [27] Instituto de la visión n.d. <http://www.institutodelavision.org/webs/ivn/index.php/es/56-partes-del-ojo>.
- [28] Alañón Fernández FJ. Anatomía Y Fisiología Del Aparato Ocular. *Oftalmol En Atención Primaria* n.d.:11–46.
- [29] The retina reference n.d. <http://www.retinareference.com>.

- [30] Nelly. Esquema Sistema Nervioso. Educa y Crea 2018.
- [31] Contreras FA. Introducción a Machine Learning. vol. 1. 2016.
- [32] Viñuela PI, León IMG. Redes Neuronales Artificiales Un Enfoque Practico 2008:229.
- [33] Martín del Brío B, Sanz Molina A. REDES NEURONALES Y SISTEMAS BORROSOS 1997:387.
- [34] Paulasen DF. Histology & Cell Biology: Examination & Board Review. n.d.
- [35] Haykin S. Neural Networks and Learning Machines. vol. 3. 2008. doi:978-0131471399.
- [36] Werbos PJ. Backpropagation Through Time: What It Does and How to Do It. Proc IEEE 1990;78:1550–60. doi:10.1109/5.58337.
- [37] Williams RJ, Peng J. An Efficient Gradient-Based Algorithm for On-Line Training of Recurrent Network Trajectories 1 Introduction 2 Formal Assumptions and Definitions. Time 1990:1–12.
- [38] Hochreiter S, Jürgen Schmidhuber JJ. Long short term memory. Neural computation. Mem Neural Comput 1997;9:1735–80.
- [39] Heidelberg-Engineering. Spectralis OCT. Heidelberg, Ger 2018. <https://business-lounge.heidelbergengineering.com/es/en/products/>.
- [40] Betancour G. Las máquinas de soporte vectorial (SVMs). Sci Tech 2005:67–72. doi:10.22517/23447214.6895.
- [41] García S. El algoritmo K-NN y su importancia en el modelado de datos. AnaliticaWeb 2018. <https://www.analiticaweb.es/algoritmo-knn-modelado-datos/>.
- [42] Zhang ML, Zhou ZH. ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognit 2007;40:2038–48. doi:10.1016/j.patcog.2006.12.019.
- [43] Luque CM. El algoritmo Naïve Bayes Resumen 2003:1–3.
- [44] Morales E. Aprendizaje Bayesiano Contenido 2012;2012:1–86.
- [45] Berk. Classification and Regression Trees. Data Min with Ratt R Art Excav Data Knowl Discov Use R 2009:36–350. doi:10.1007/s00038-011-0315-z.
- [46] Quintero M, Amézquita E. Árboles de Decisión. Internet 2008:45.
- [47] Cadavid D, Tang Y, O'Neill G. Sensibilidad de la escala ampliada del estado de discapacidad (EDSS) a la progresión de la enfermedad y la intervención terapéutica en las formas progresivas de la esclerosis múltiple. Rev Neurol 2010;51:321–9.
- [48] Ayuso Blanco T, Aliseda D, Ajuria I, Zandío B, Mayor S, Navarro MC. Neuritis óptica inflamatoria. An Sist Sanit Navar 2009;32:249–63.
- [49] 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/iops.07-0834.

- [50] Rut M^a López Lizcano. Valor del grosor de la capa de fibras nerviosas de la retina medido mediante tomografía de coherencia óptica como marcador de daño neuroaxonal en la esclerosis múltiple. 2012.
- [51] 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.
- [52] OCT Retinal Scan. Noel Templeton Optometrists. 2016. <https://noeltempleton.com>.
- [53] CIRRUS HD-OCT. Carl Zeiss AG n.d.
- [54] Carmona Suárez EJ. Máquinas de Vectores Soporte (SVM) 2014:1–25.
- [55] Larrañaga P. Clasificadores Bayesianos 2002:1–17. doi:10.3390/nano6050092.
- [56] Larrañaga, Pedro and Inza, Inaki and Moujahid A. Árboles de Clasificación 2013:1–6.
- [57] Bottou L, Lin C. Support vector machine solvers. *Large Scale Kernel Mach* 2007:1–27. doi:10.1.1.127.511.
- [58] MathWorks. MathWorks. Train binary support vector machine classifier. n.d. https://es.mathworks.com/help/stats/fitcsvm.html?s_tid=srchtitle#namevaluepairarguments.
- [59] Kecman V, Huang T, Vogt M. Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets : n.d.;274:255–74.
- [60] 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.
- [61] 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 | 6 |
| Figura 2.2 Evolución del espesor [μm] de la retina (todas las capas) con la edad de un grupo de pacientes sanos del Hospital Universitario Miguel Servet de Zaragoza. | 8 |
| Figura 3.1 Esquema de las principales vías visuales del cerebro. | 9 |
| Figura 3.2 Partes del ojo humano [4]. | 10 |
| Figura 3.3 Ilustración de las capas de la retina con detalle del nervio óptico [29]. | 11 |
| Figura 3.4 Representación de las partes de una célula [30] | 11 |
| Figura 3.5 Imagen tomada mediante la técnica de OCT (Optical Coherence Tomography) de la retina de un paciente. A la izquierda, imagen de la retina del ojo derecho y a la derecha, imagen de la retina del ojo izquierdo. | 12 |
| Figura 4.1 Esquema general de una RNA unidireccional de tres capas, de entrada, oculta y de salida. | 13 |
| Figura 4.2. Conexiones recurrentes empleadas en las RNN. Izquierda: conexión de una neurona con sí misma. Centro: Conexión entre neuronas de una misma capa. Derecha: Conexión entre dos neuronas de dos capas diferentes. | 15 |
| Figura 4.3 Arquitectura de red recurrente con dos neuronas totalmente conectadas. | 16 |
| Figura 4.4 Red recurrente desarrollada en el tiempo. | 16 |
| Figura 4.5 Bloque Long Short-Term Memory | 17 |
| Figura 4.6 Diagrama de flujo de una capa LSTM para una serie temporal X. | 18 |
| Figura 4.7 Representación de un problema de clasificación 2D resuelto mediante SVM | 19 |
| Figura 4.8 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. | 19 |
| Figura 4.9 Aparición del parámetro ξ_i en el error de clasificación [15] | 21 |
| Figura 4.10 Idea del uso de un kernel para la transformación del espacio de los datos [40]. | 22 |
| Figura 4.11. Ejemplo de clasificación mediante algoritmo K-NN. | 23 |
| Figura 4.12 Ejemplo de clasificación mediante un árbol de decisión. | 25 |
| Figura 4.13 Evolución de la precisión en el conjunto de entrenamiento y el de validación en función de la complejidad del árbol de decisión. | 26 |
| Figura 5.1. Diferencias entre los protocolos de adquisición de OCT para cada una de las bases de datos. | 27 |
| Figura 5.2. Escala expandida del estado de discapacidad. | 28 |

| | |
|---|----|
| Figura 5.3 Imagen del equipo OCT-CIRRUS en el Hospital Universitario Miguel Servet de Zaragoza..... | 30 |
| Figura 5.4 Sección de la retina tomada mediante una tomografía de coherencia óptica [52]..... | 30 |
| Figura 5.5 Principios de interferometría en los que se basa la OCT..... | 31 |
| Figura 5.6 Ejemplo de datos que proporciona la OCT-CIRRUS [53]. | 31 |
| Figura 5.7 Visitas de las que se tiene información de los pacientes. | 32 |
| Figura 5.8 Representación de la cantidad de visitas disponibles en el conjunto de los pacientes según la correspondiente visita | 32 |
| Figura 5.9. Izquierda: Gráfica de la evolución del espesor medio de la RNFL en la zona peripapilar y en la fovea a lo largo del tiempo. Derecha: Gráfica de la evolución del espesor de la RNFL en los cuatro sectores alrededor de la zona peripapilar a lo largo de las 10 visitas..... | 35 |
| Figura 5.10 Ejemplo de k-fold cross validation con k=5, en el que se muestra como el grupo de datos de prueba y de datos de entrenamiento varía en cada una de las iteraciones. | 36 |
| Figura 5.11 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. | 37 |
| Figura 6.1 Representación de los 4 modelos de predicción desarrollados con LSTM. ... | 40 |
| Figura 6.2. Representación de las curvas ROC para los modelos desarrollados con redes LSTM. | 41 |
| Figura 6.3 Representación de los 4 modelos de predicción desarrollados mediante algoritmos de clasificación. | 43 |
| Figura 6.4. Diagrama de flujo del código empleado para utilizar las SVM en la clasificación de más de dos clases..... | 45 |
| Figura 6.5 Algoritmo empleado en Matlab para utilizar el algoritmo SVM en clasificación multiclase. | 46 |
| Figura 6.6. Importancia de los atributos de entrada para un modelo construido con la función <i>fitctree</i> | 49 |
| Figura 6.7. Representación de las curvas ROC para los cuatro modelos con algoritmos de clasificación..... | 51 |
| Figura 7.1. EDSS medio y espesor medio de la RNFL en la zona macular a lo largo del seguimiento a 10 años..... | 52 |

11 ÍNDICE DE TABLAS.

| | |
|---|----|
| Tabla 1. Datos clínicos de los pacientes estudiados..... | 33 |
| Tabla 2. Datos clínicos de la prueba OCT de los pacientes estudiados..... | 34 |
| Tabla 3. Frecuencia de cada grupo de escala de EDSS para los modelos de clasificación. | 40 |
| Tabla 4. Tabla de resultados para cada uno de los 4 modelos planteados con las redes LSTM. | 40 |
| Tabla 5 Frecuencia de cada grupo de paciente respecto a la EDSS. | 43 |
| Tabla 6. Tipos de función kernel para poder entrenar la SVM. | 44 |
| Tabla 7. Algoritmos de optimización empleados en la resolución del problema de clasificación mediante máquinas de soporte vectorial. | 44 |
| Tabla 8. Funciones de métrica de distancia. | 46 |
| Tabla 9. Tipos de función de ponderación de la distancia. | 46 |
| Tabla 10. Tipos de distribuciones correspondientes a los atributos predictores. | 47 |
| Tabla 11. Tipos de suavizado kernel..... | 47 |
| Tabla 12. Resultados obtenidos para el primer modelo planteado con cada uno de los algoritmos utilizados. | 50 |



Universidad
Zaragoza

Trabajo Fin de Máster

ANEXOS

Simulación de la degeneración del nervio óptico en enfermos esclerosis múltiple. Metodología para la predicción de la evolución de la enfermedad.

Simulation of optic nerve degeneration in multiple sclerosis patients. Methodology for disease evolution prediction.

Autor

Alejandro Martín Gallego

Directora

José Cegoñino Banzo

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

Índice Anexos

| | | |
|-------|--|----|
| i. | MATERIAL Y MÉTODO BASE DATOS OCT-SPECTRALIS. | 65 |
| i.1 | OCT-Spectralis..... | 65 |
| ii. | FRECUENCIA DE ADQUISICIÓN DE VISITAS EN LAS 5 BASES DE DATOS..... | 69 |
| ii.1 | Base de datos 1: Datos del espesor de la RNFL en la zona peripapilar. Equipo CIRRUS-OCT..... | 69 |
| ii.2 | Base de datos 2: Datos espesor de la GCL en la zona macular..... | 71 |
| ii.3 | Base de datos 3: Datos espesor de la RNFL en la zona macular..... | 72 |
| ii.4 | Base de datos 4: Datos espesor de la RNFL en la zona peripapilar. Protocolo axonal..... | 73 |
| ii.5 | Base de datos 5: Datos espesor de la RNFL en la zona peripapilar. Protocolo glaucoma..... | 74 |
| iii. | EXPERIMENTOS REALIZADOS Y RESULTADOS BASE DATOS OCT-SPECTRALIS. | 75 |
| iii.1 | Modelos de predicción con las bases de datos OCT-Spectralis..... | 75 |
| iii.2 | Resultados obtenidos. | 76 |
| iii.3 | Curvas ROC..... | 78 |
| iv. | ÍNDICE DE FIGURAS ANEXOS. | 79 |
| v. | ÍNDICE DE TABLAS ANEXOS..... | 80 |

i. MATERIAL Y MÉTODO BASE DATOS OCT-SPECTRALIS.

Como se ha explicado en el apartado 5, se dispone de 5 bases de datos en total, pero la que mayor frecuencia de adquisición posee es la perteneciente al equipo CIRRUS-OCT. En este apartado se explica los datos que incluyen cada una de las otras 4 bases de datos, que han sido obtenidas mediante el equipo OCT-Spectralis.

i.1 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.



Figura i.1. Equipo OCT-Spectralis en el Hospital Universitario Miguel Servet de Zaragoza.

En la Figura i.2 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 comprar 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).

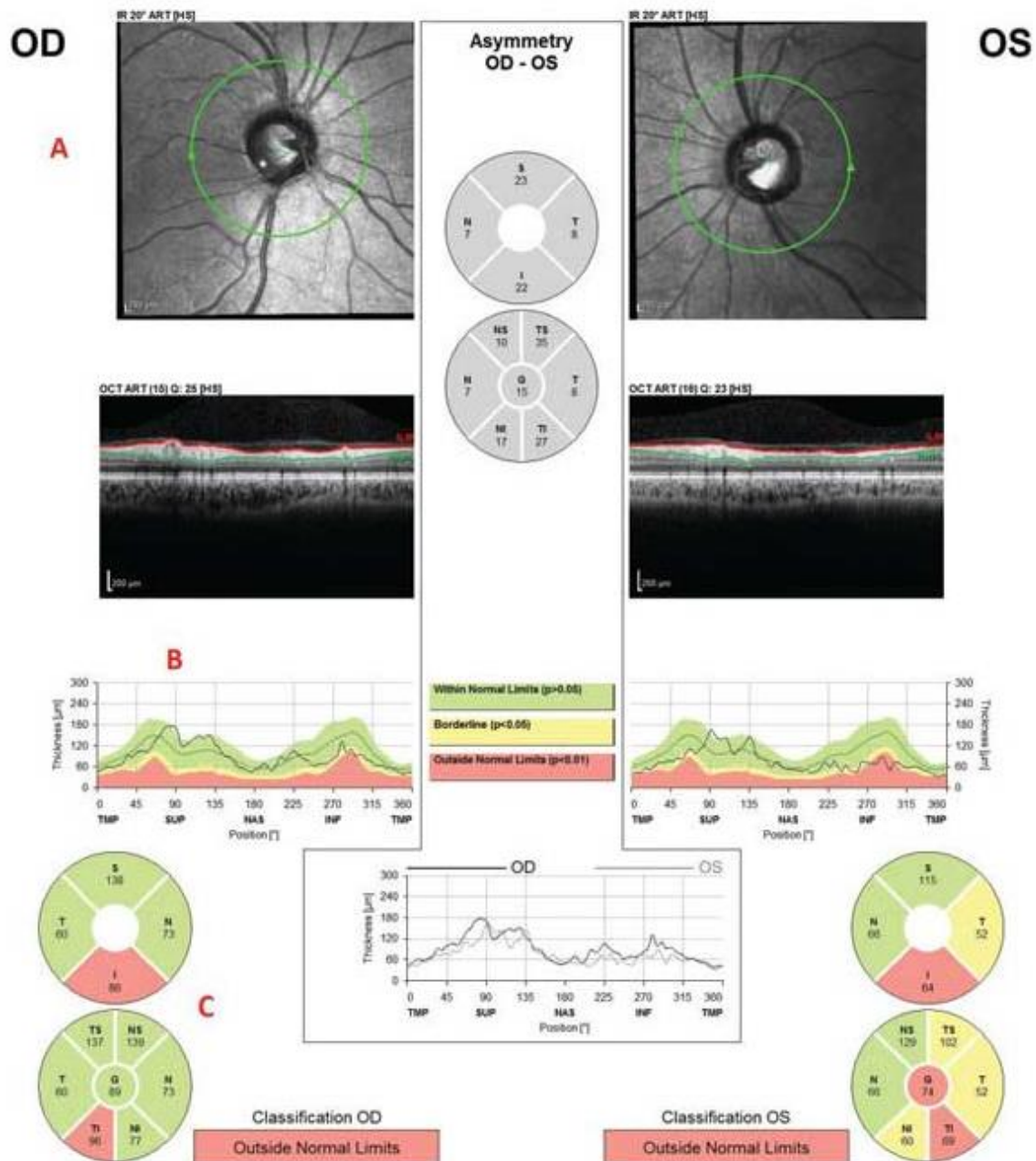


Figura i.2 Ejemplo de datos que proporciona OCT-Spectralis

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**.

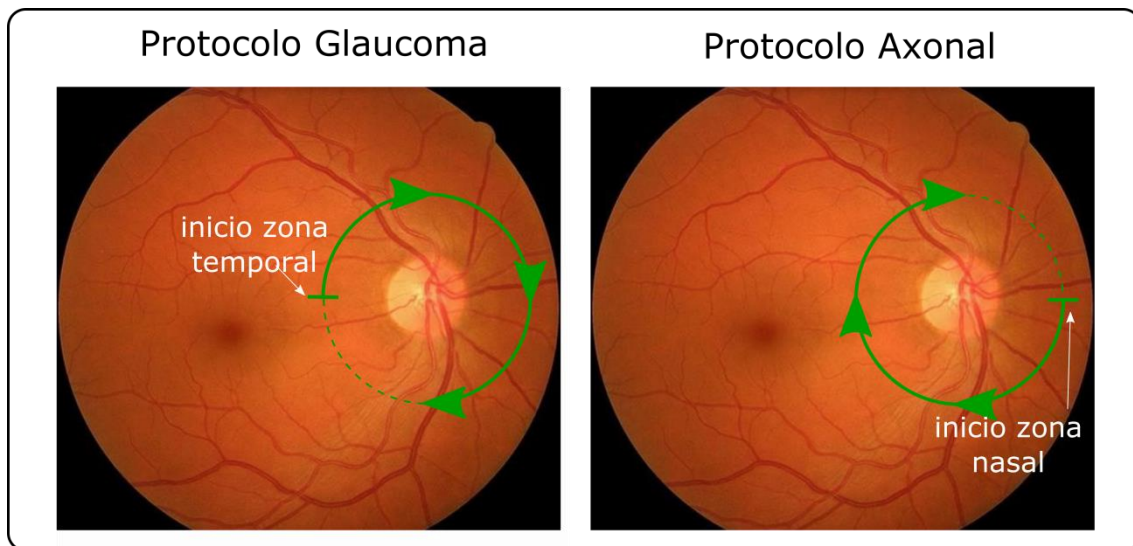


Figura i.3 A la izquierda patrón de barrido realizado por el protocolo glaucoma (de zona temporal a zona temporal) en un ojo derecho, a la derecha patrón de adquisición realizado por el protocolo axonal (de zona nasal a zona nasal) en el mismo ojo derecho.

En la Figura i.3 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 fóvea (Ver Figura i.4).

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.

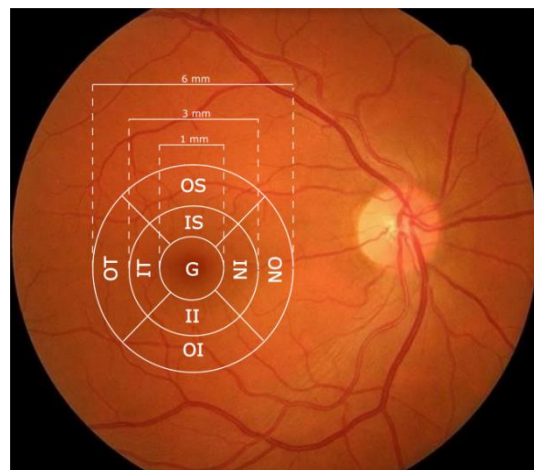


Figura i.4 Representación de la adquisición de datos alrededor de la zona macular.

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.

Para todos los protocolos de los que se dispone información, en concreto para las 4 bases de datos, se ha recogido los siguientes parámetros en el seguimiento de los pacientes:

- Edad del paciente en cada una de las visitas.
- Sexo del paciente.
- La duración de la enfermedad en el momento de la visita.
- Fenotipo de EM, es decir, a que subtipo de enfermedad a la que pertenece el paciente.
- Escala expandida del estado de discapacidad (EDSS, del inglés *Expanded Disability Status Scale*).

Además, en cada una de las bases de datos se ha recogido el espesor de la retina con diferentes parámetros.

Base datos 2: Datos del espesor medio de la capa de células ganglionares (GCL, del inglés Ganglionic Cell Layer) de la retina en los segmentos del área macular. En este protocolo se obtienen datos del espesor de la GCL en 9 sectores diferentes de la zona macular de la retina como se puede observar en la Figura i.4.

Base datos 3: 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.

Base datos 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 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 (Ver Figura i.3).

Base datos 5: 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 (Ver Figura i.3).

ii. FRECUENCIA DE ADQUISICIÓN DE VISITAS EN LAS 5 BASES DE DATOS.

Este apartado trata de comprar la frecuencia de adquisición de datos en cada paciente en las 5 bases de datos disponibles para justificar la elección de la primera en el desarrollo de los mejores modelos de predicción.

Se van a representar las mismas gráficas para las 5 bases de datos y poder así compararlas. Puede observarse que, tanto la frecuencia de adquisición como el número de secuencias de datos disponibles de la base de datos correspondiente al equipo CIRRUS-OCT es mucho mayor que el resto.

Por esa razón se emplearán los datos de la primera de las bases de datos para el desarrollo de los modelos de predicción, ya que se dispone de unas series temporales de datos más largas y completas.

ii.1 Base de datos 1: Datos del espesor de la RNFL en la zona peripapilar. Equipo CIRRUS-OCT.

En la Figura ii.1 se puede observar, por una parte, el número de ojos totales que disponen de cada una de las visitas, es decir, cuántos ojos disponemos con la visita basal, cuántos con la visita 1, la visita 2, ... hasta la visita 10. Por otra parte, el número de pacientes que disponen de una secuencia con 3 visitas, con 4 visitas,... hasta con 7 visitas:

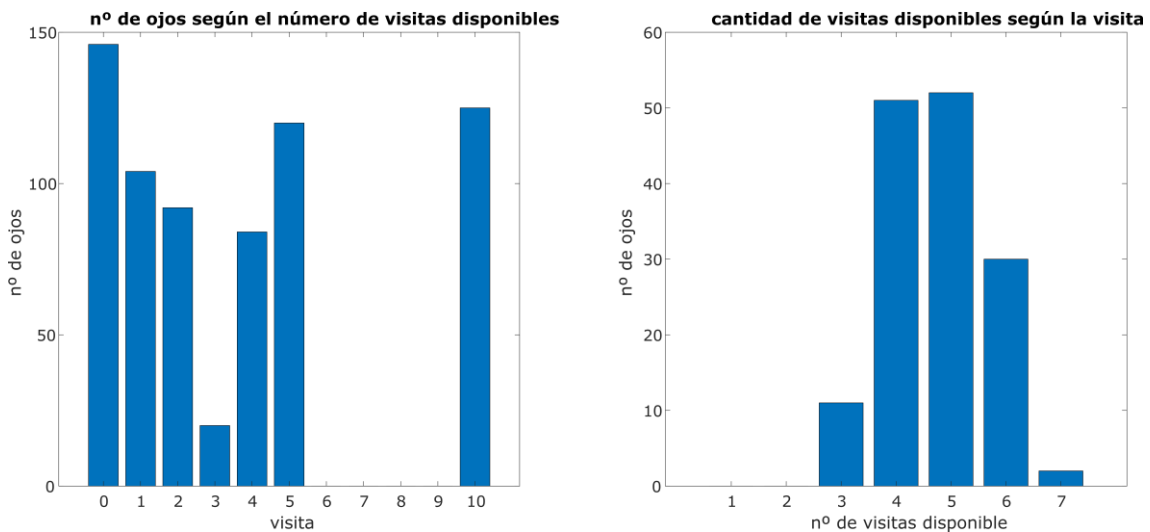


Figura ii.1. Izquierda: Frecuencia de disponibilidad de datos en cada una de las visitas en número de ojos.
Derecha: Frecuencia de disponibilidad de pacientes según la cantidad de visitas que dispone.

Para ver de una forma más gráfica la frecuencia de adquisición de cada uno de los pacientes disponibles en esta base de datos se representa la siguiente gráfica. Ésta representa para cada uno de los pacientes disponibles (cada fila representa a un paciente), cuáles son las visitas de las que dispone información (azul oscuro) en la base de datos.

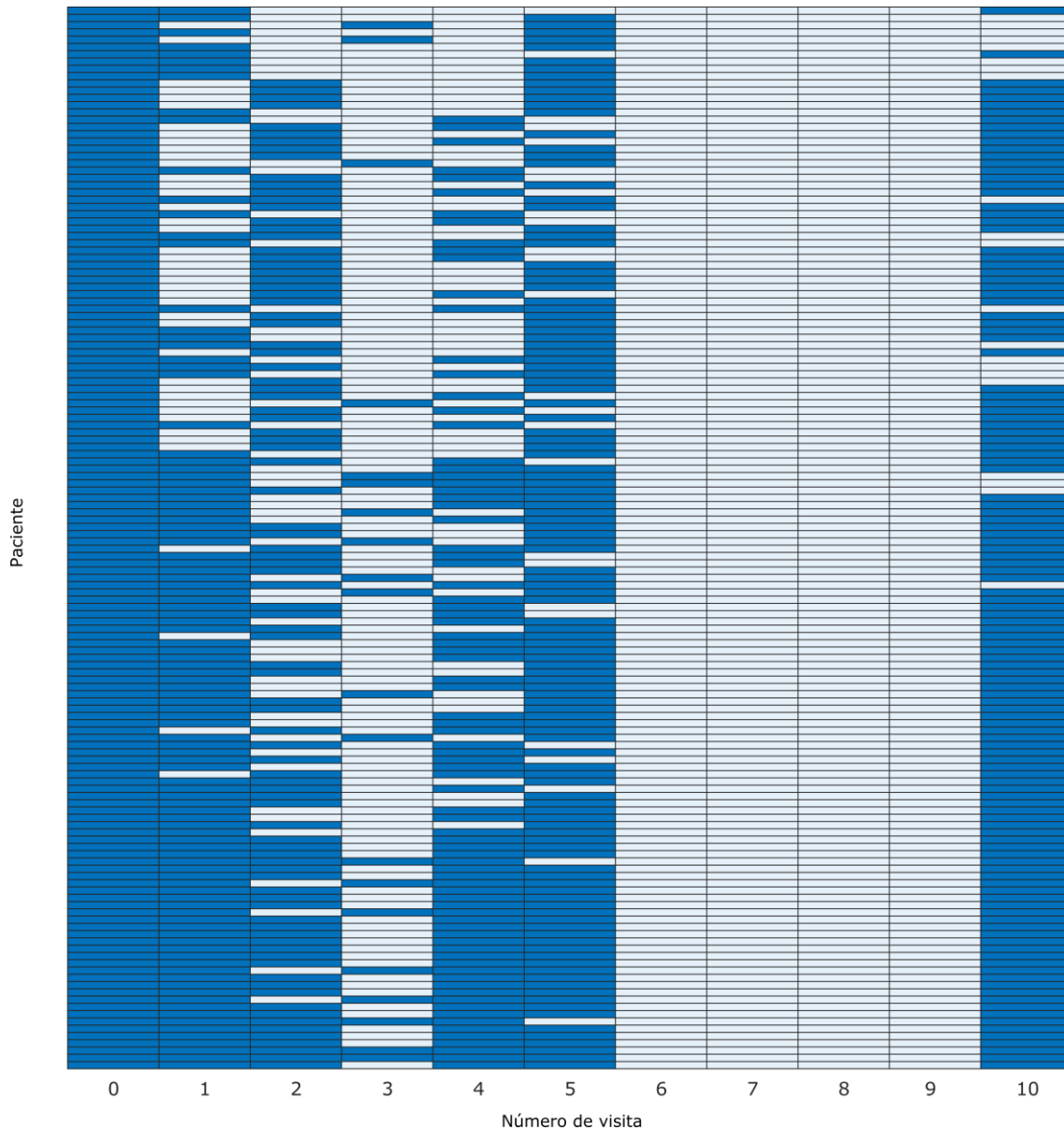


Figura ii.2 Número de visitas de las que dispone cada uno de los pacientes (144 ojos en total) utilizados en los modelos de predicción desarrollados con la base de datos 1.

Para esta base de datos, se puede observar que la totalidad de los pacientes tienen disponible la visita basal, y la gran mayoría de ellos disponen de la visita 5 y 10. Cabe destacar que los pacientes que no disponen de la visita número 10 no podrán ser utilizados en los modelos de predicción a 10 años.

ii.2 Base de datos 2: Datos espesor de la GCL en la zona macular.

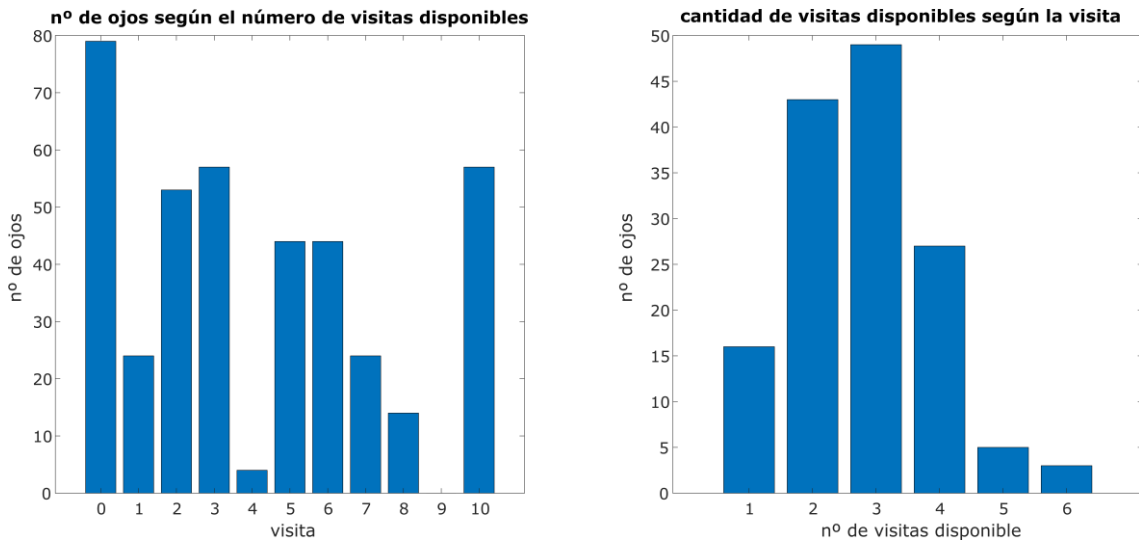


Figura ii.3. Izquierda: Frecuencia de disponibilidad de datos en cada una de las visitas en número de ojos. Derecha: Frecuencia de disponibilidad de pacientes según la cantidad de visitas que dispone.

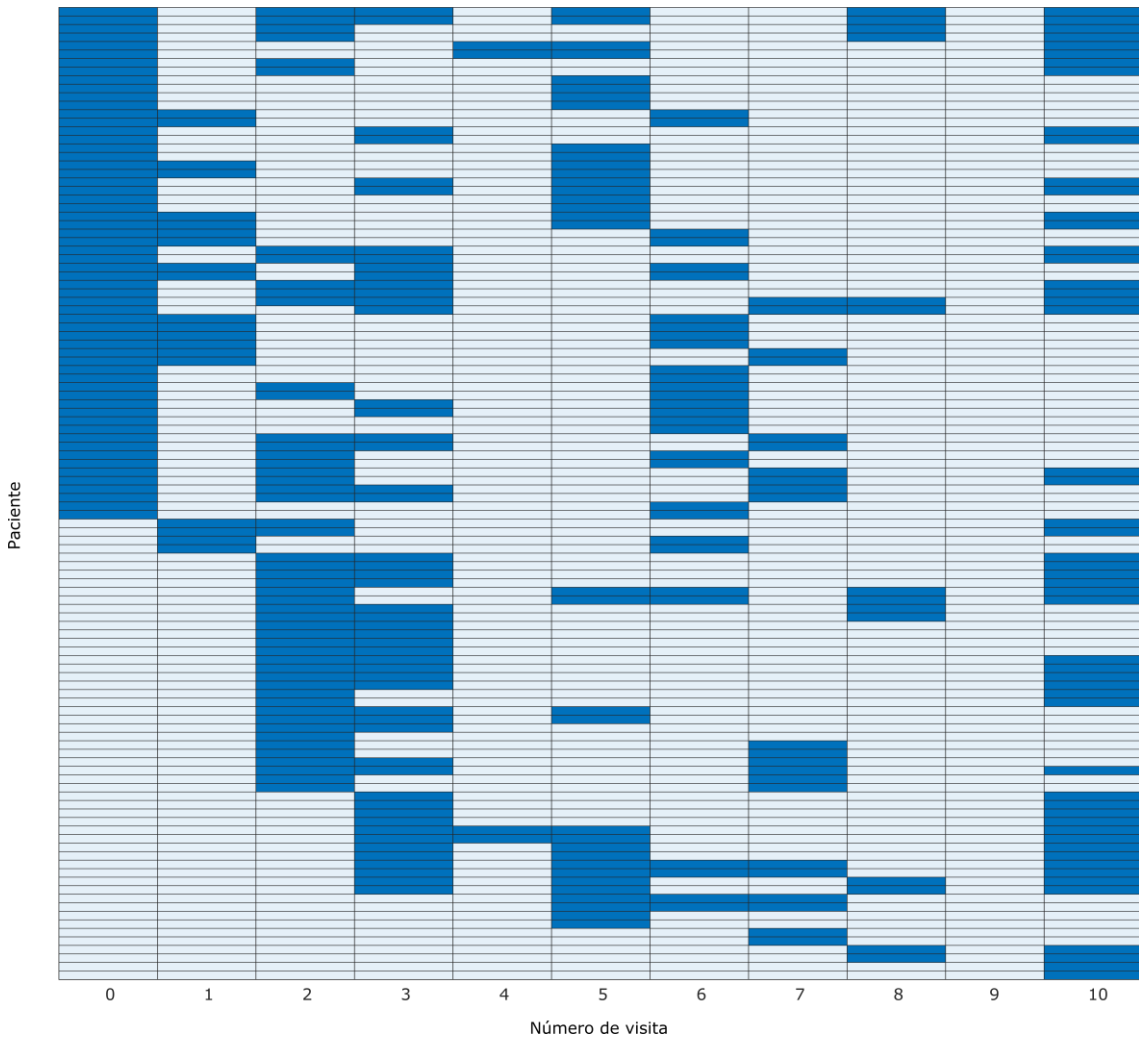


Figura ii.4. Número de visitas de las que dispone cada uno de los pacientes (114 ojos en total) utilizados en los modelos de predicción desarrollados con la base de datos 2.

ii.3 Base de datos 3: Datos espesor de la RNFL en la zona macular.

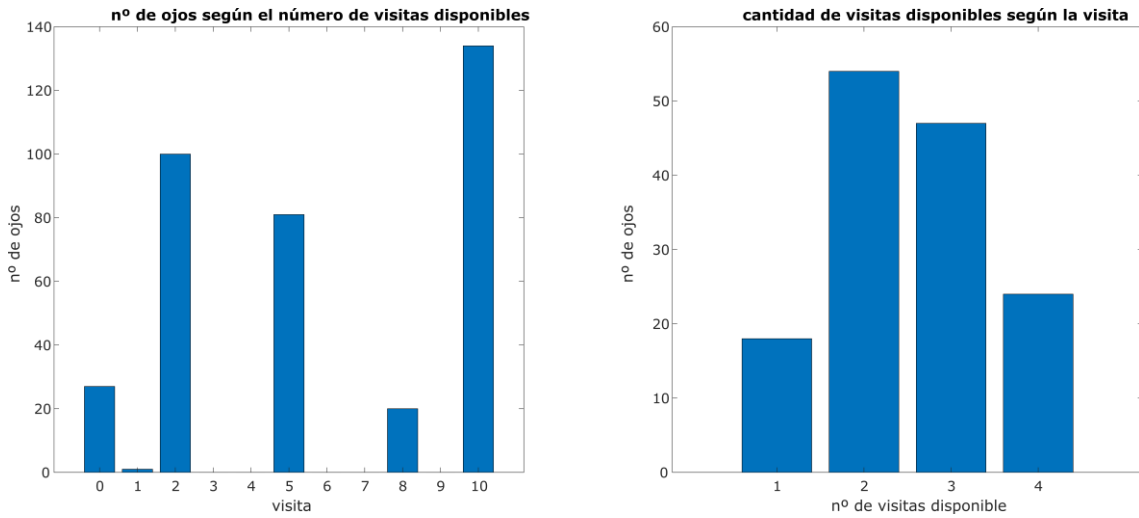


Figura ii.5. Izquierda: Frecuencia de disponibilidad de datos en cada una de las visitas en número de ojos. Derecha: Frecuencia de disponibilidad de pacientes según la cantidad de visitas que dispone.

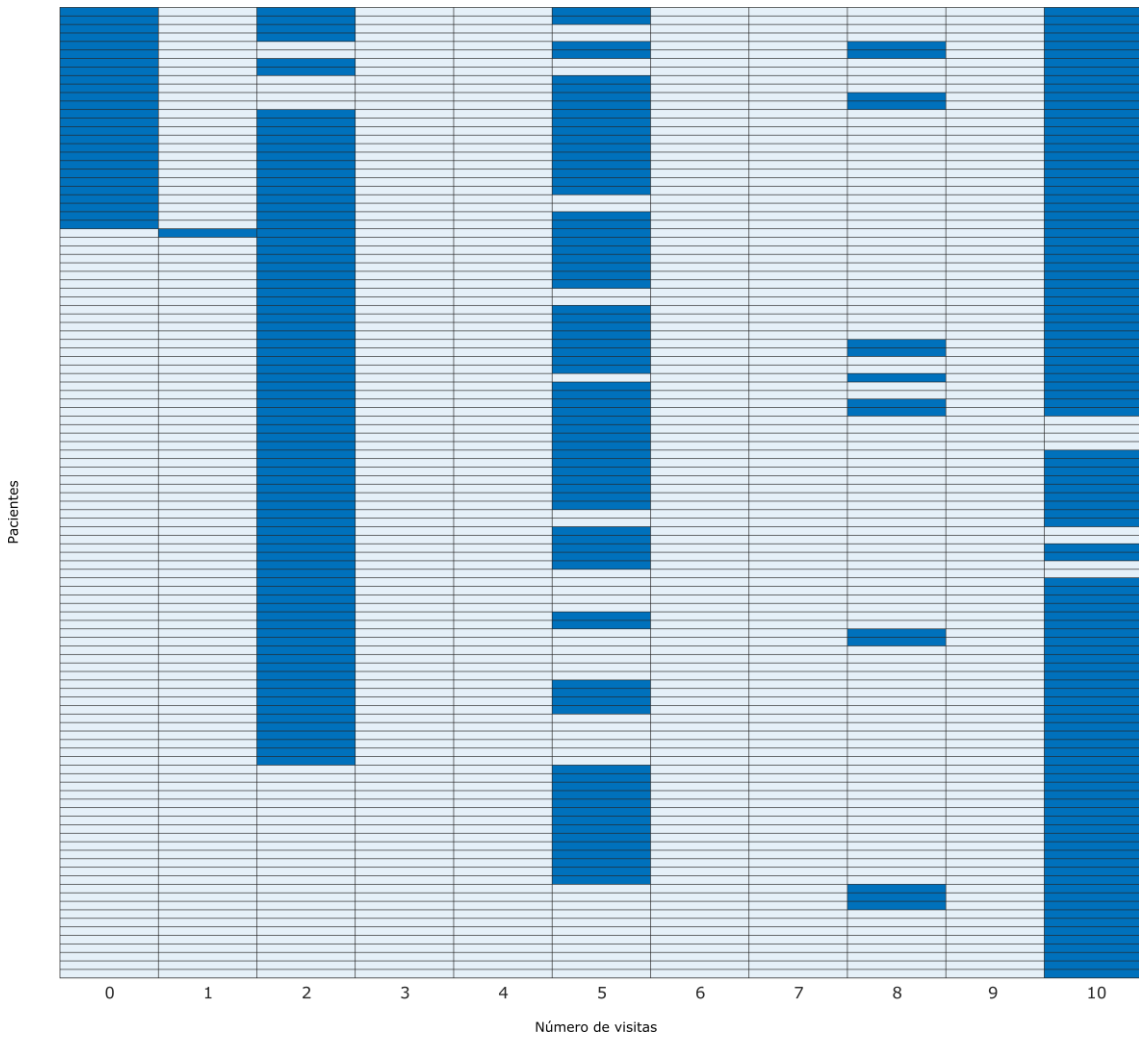


Figura ii.6. Número de visitas de las que dispone cada uno de los pacientes (114 ojos en total) utilizados en los modelos de predicción desarrollados con la base de datos 3.

ii.4 Base de datos 4: Datos espesor de la RNFL en la zona peripapilar. Protocolo axonal.

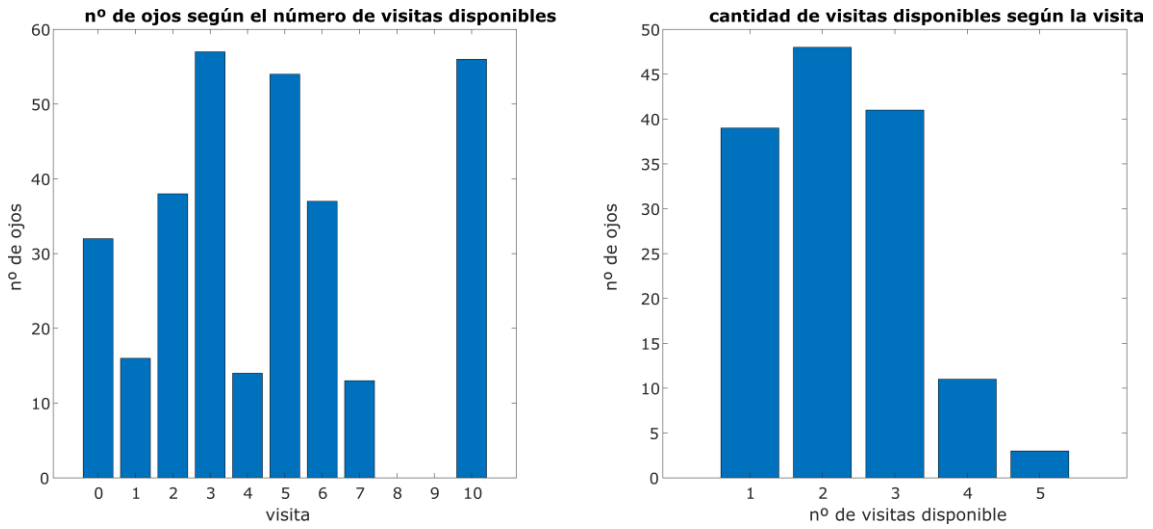


Figura ii.7. Izquierda: Frecuencia de disponibilidad de datos en cada una de las visitas en número de ojos.
 Derecha: Frecuencia de disponibilidad de pacientes según la cantidad de visitas que dispone.

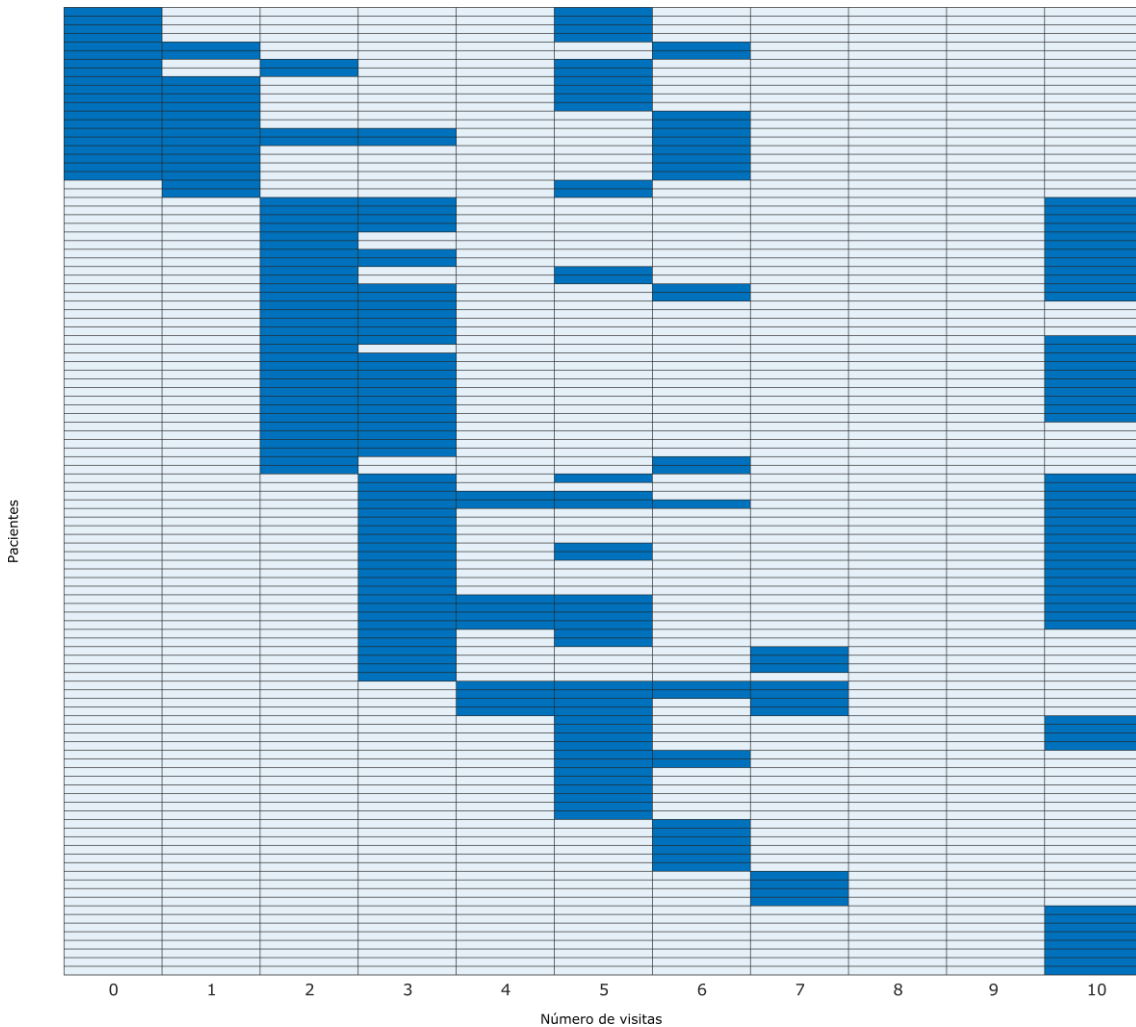


Figura ii.8. Número de visitas de las que dispone cada uno de los pacientes (112 ojos en total) utilizados en los modelos de predicción desarrollados con la base de datos 4.

ii.5 Base de datos 5: Datos espesor de la RNFL en la zona peripapilar. Protocolo glaucoma.

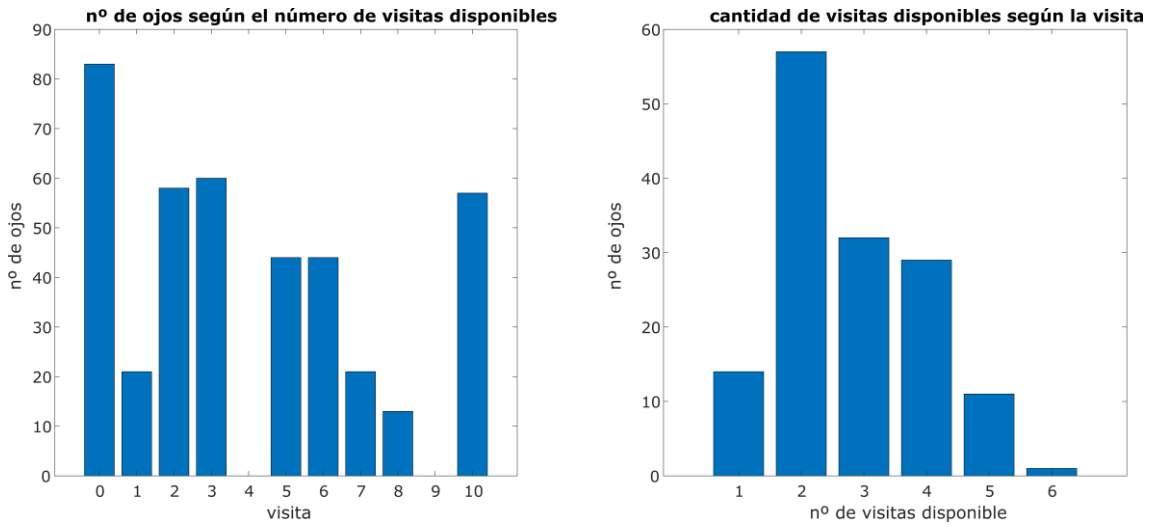


Figura ii.9. Izquierda: Frecuencia de disponibilidad de datos en cada una de las visitas en número de ojos. Derecha: Frecuencia de disponibilidad de pacientes según la cantidad de visitas que dispone.

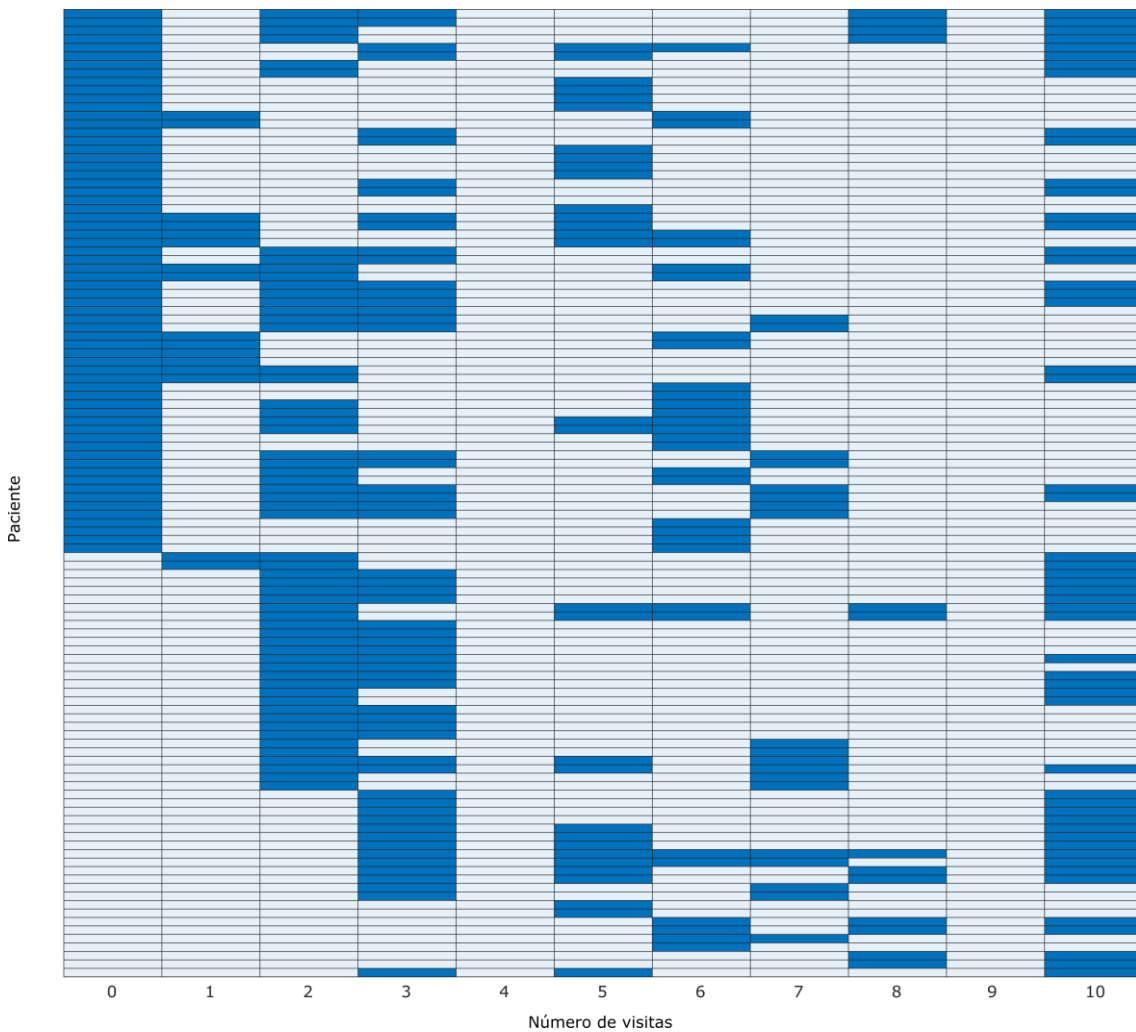


Figura ii.10. Número de visitas de las que dispone cada uno de los pacientes (114 ojos en total) utilizados en los modelos de predicción desarrollados con la base de datos 5.

iii. EXPERIMENTOS REALIZADOS Y RESULTADOS BASE DATOS OCT-SPECTRALIS.

Como se ha explicado en el apartado ii, las bases de datos que provienen del equipo OCT-Spectralis, no disponen de la misma frecuencia de adquisición que la base de datos del equipo CIRRUS-OCT. Sin embargo, Spectralis dispone de una serie de protocolos de adquisición que puede hacer interesante desarrollar modelos de predicción utilizando la información de estas bases.

En este apartado se describe como se procede a utilizar los algoritmos planteados en la predicción de la evolución de la esclerosis múltiple y los resultados obtenidos de cada uno de los modelos de predicción desarrollados mediante la información incluida en las bases de datos de OCT-Spectralis.

iii.1 Modelos de predicción con las bases de datos OCT-Spectralis.

Para estos modelos de predicción vamos a utilizar la información de los pacientes que tengan, al menos, **3 visitas disponibles** con el objetivo de predecir la evolución de la enfermedad en la última de ellas, utilizando la información de las dos primeras. El número de pacientes disponibles para la realización de estos modelos de cada base de datos se puede ver en la Tabla 1 y en la Tabla 2.

Para generar los modelos vamos a utilizar los mismos algoritmos que se han empleado en los apartados 6.1 y 6.2, tratando de optimizarlos de la misma forma.

Así pues, vamos a plantear dos modelos de predicción para cada una de estas bases de datos:

1. M1: Modelo de predicción desarrollado con las redes neuronales recurrentes LSTM, para lo cual se emplearán todas las secuencias disponibles de los pacientes para el entrenamiento del algoritmo. En el modo recuerdo, el modelo tratará de predecir la EDSS de los pacientes en la última visita de la que disponga cada sujeto con la información de la primera y de la segunda.

Tabla 1. Frecuencia de datos de cada una de las 5 clases en las que se clasifican los ojos utilizados para el modelo de predicción M1.

| Rango de EDSS | [0-2] | (2-4] | (4-6] | (6-8] | (8-10] | Total |
|---------------|-------|-------|-------|-------|--------|-------|
| Base datos 2 | 28 | 22 | 8 | 14 | 2 | 74 |
| Base datos 3 | 28 | 17 | 8 | 8 | 2 | 63 |
| Base datos 4 | 20 | 16 | 4 | 8 | 0 | 48 |
| Base datos 5 | 28 | 15 | 8 | 14 | 2 | 67 |

2. M2: El segundo modelo se desarrollará con los algoritmos de clasificación binaria, y tratará de predecir si la escala EDSS va a empeorar, utilizando los datos de las dos primeras visitas de cada paciente. En la clasificación de los pacientes, al igual que en los modelos desarrollados para la base de datos CIRRUS, éstos se dividen en dos clases, la primera denominada pacientes que **no-empeoran**, para los cuales la escala EDSS no sufre un aumento mayor a 1 punto, y por otro lado, los pacientes que si **empeoran**, que serán aquellos cuya EDSS sufra un aumento mayor a 1 punto en los 5 años.
3. M3: El tercer modelo, al igual que el segundo, se desarrollará con los algoritmos de clasificación binaria, y tratará de predecir si la escala EDSS va a empeorar, utilizando los datos de las dos primeras visitas de cada paciente. En la clasificación

de los pacientes, éstos se dividen en tres clases, la primera serán los pacientes que sufran un **mejoría** en cuanto a la EDSS se refiere, es decir, los que tengan una diferencia negativa de la escala EDSS en los 5 años. Por otra parte estarán los que **no-empeoran**, cuya diferencia de EDSS no supera 1 punto, y por último los pacientes cuya diferencia de EDSS supera 1 punto y se clasifican como pacientes que **empeoran**.

Tabla 2. Frecuencia de datos de cada clase en las que se clasifican los ojos utilizados para los modelos de predicción M2 y M3.

| Clases | M2: modelos 2 clases | | | M3: modelos 3 clases | | Total |
|--------------|-------------------------------------|-------------------------------------|---------------------------------|---|-------------------------------------|-------|
| | No-Empeora ($\Delta EDSS < 1$) | Empeora ($\Delta EDSS \geq 1$) | Mejora ($\Delta EDSS < 0$) | No-Empeora ($0 < \Delta EDSS < 1$) | Empeora ($\Delta EDSS \geq 1$) | |
| Base datos 2 | 38 | 36 | -- | -- | -- | 74 |
| | -- | -- | 4 | 34 | 36 | 74 |
| Base datos 3 | 31 | 32 | -- | -- | -- | 63 |
| | -- | -- | 4 | 27 | 32 | 63 |
| Base datos 4 | 22 | 26 | -- | -- | -- | 48 |
| | -- | -- | 2 | 20 | 26 | 48 |
| Base datos 5 | 34 | 33 | -- | -- | -- | 67 |
| | -- | -- | 4 | 30 | 33 | 67 |

Basándonos en estos planteamientos se van a desarrollar los modelos con cada uno de los algoritmos que se han empleado anteriormente en los apartados 6.1 y 6.2.

iii.2 Resultados obtenidos.

A continuación se muestran los resultados obtenidos para cada una de las 4 bases de datos disponibles a partir de la información del equipo OCT-Spectralis:

Tabla 3. Resultados por cada modelo en la base de datos correspondiente a la capa de células ganglionares en la zona macular.

| Modelo | Clasificación | Algoritmo | F. Matlab | k-fold CV | Precisión |
|----------|---------------|----------------------------|----------------|-----------|---------------|
| Modelo 1 | 5 clases | LSTM | bilstmLayer | 5 | 0.6629 |
| | | SVM (polinomial) | Fitcsvm | 5 | 0.5135 |
| Modelo 2 | 2 clases | SVM (linear) | Fitcsvm | 5 | 0.6216 |
| | | SVM (gaussian) | Fitcsvm | 5 | 0.6486 |
| | | Decision tree | Fitctree | 5 | 0.5541 |
| | | K nearest neighbors | Fitcknn | 5 | 0.7027 |
| | | Naive Bayes model | Fitcnb | 5 | 0.6081 |
| Modelo 3 | 3 clases | SVM multiclass | Fitcecoc | 5 | 0.6892 |
| | | K nearest neighbors | Fitcknn | 5 | 0.7838 |
| | | Decision tree | Fitctree | 5 | 0.6216 |
| | | Naive Bayes model | Fitcnb | 5 | 0.4595 |

Para la base de datos 2, el modelo con el que se ha conseguido llegar a una mayor precisión es el modelo 3, que trata de predecir la evolución de los pacientes en 3 clases diferentes (Mejora, No-empeoran, Empeoran) utilizando el algoritmo de predicción k-NN.

Tabla 4. Resultados por cada modelo en la base de datos correspondiente a la capa de fibras nerviosas de la retina en la zona macular.

| Modelo | Clasificación | Algoritmo | F. Matlab | k-fold CV | Precisión |
|----------|---------------|----------------------------|----------------|-----------|---------------|
| Modelo 1 | 5 clases | LSTM | bilstmLayer | 5 | 0.619 |
| Modelo 2 | 2 clases | SVM (polinomial) | Fitcsvm | 5 | 0.444 |
| | | SVM (linear) | Fitcsvm | 5 | 0.6508 |
| | | SVM (gaussian) | Fitcsvm | 5 | 0.6508 |
| | | Decision tree | Fitctree | 5 | 0.5873 |
| | | K nearest neighbors | Fitcknn | 5 | 0.6349 |
| | | Naive Bayes model | Fitcnb | 5 | 0.619 |
| Modelo 3 | 3 clases | SVM multiclass | Fitcecoc | 5 | 0.5714 |
| | | K nearest neighbors | Fitcknn | 5 | 0.7143 |
| | | Decision tree | Fitctree | 5 | 0.5238 |
| | | Naive Bayes model | Fitcnb | 5 | -- |

Para la base de datos 3, del mismo modo que para la base de datos 2, el mejor modelo de predicción resulta ser el tercero, y de nuevo con el algoritmo de predicción k-NN.

Tabla 5. Resultados por cada modelo en la base de datos correspondiente a la capa de fibras nerviosas de la retina en la zona peripapilar con protocolo de adquisición axonal.

| Modelo | Clasificación | Algoritmo | F. Matlab | k-fold CV | Precisión |
|----------|---------------|----------------------------|----------------|-----------|---------------|
| Modelo 1 | 5 clases | LSTM | bilstmLayer | 5 | 0.4375 |
| Modelo 2 | 2 clases | SVM (polinomial) | Fitcsvm | 5 | 0.3542 |
| | | SVM (linear) | Fitcsvm | 5 | 0.6667 |
| | | SVM (gaussian) | Fitcsvm | 5 | 0.7083 |
| | | Decision tree | Fitctree | 5 | 0.5417 |
| | | K nearest neighbors | Fitcknn | 5 | 0.75 |
| | | Naive Bayes model | Fitcnb | 5 | 0.625 |
| Modelo 3 | 3 clases | SVM multiclass | Fitcecoc | 5 | 0.6042 |
| | | K nearest neighbors | Fitcknn | 5 | 0.7708 |
| | | Decision tree | Fitctree | 5 | 0.7083 |
| | | Naive Bayes model | Fitcnb | 5 | -- |

Para la cuarta base de datos, al igual que para las anteriores el mejor modelo de predicción es el tercero, desarrollado con el algoritmo de clasificación k-NN.

Tabla 6. Resultados por cada modelo en la base de datos correspondiente a la capa de fibras nerviosas de la retina en la zona peripapilar con protocolo de adquisición glaucoma.

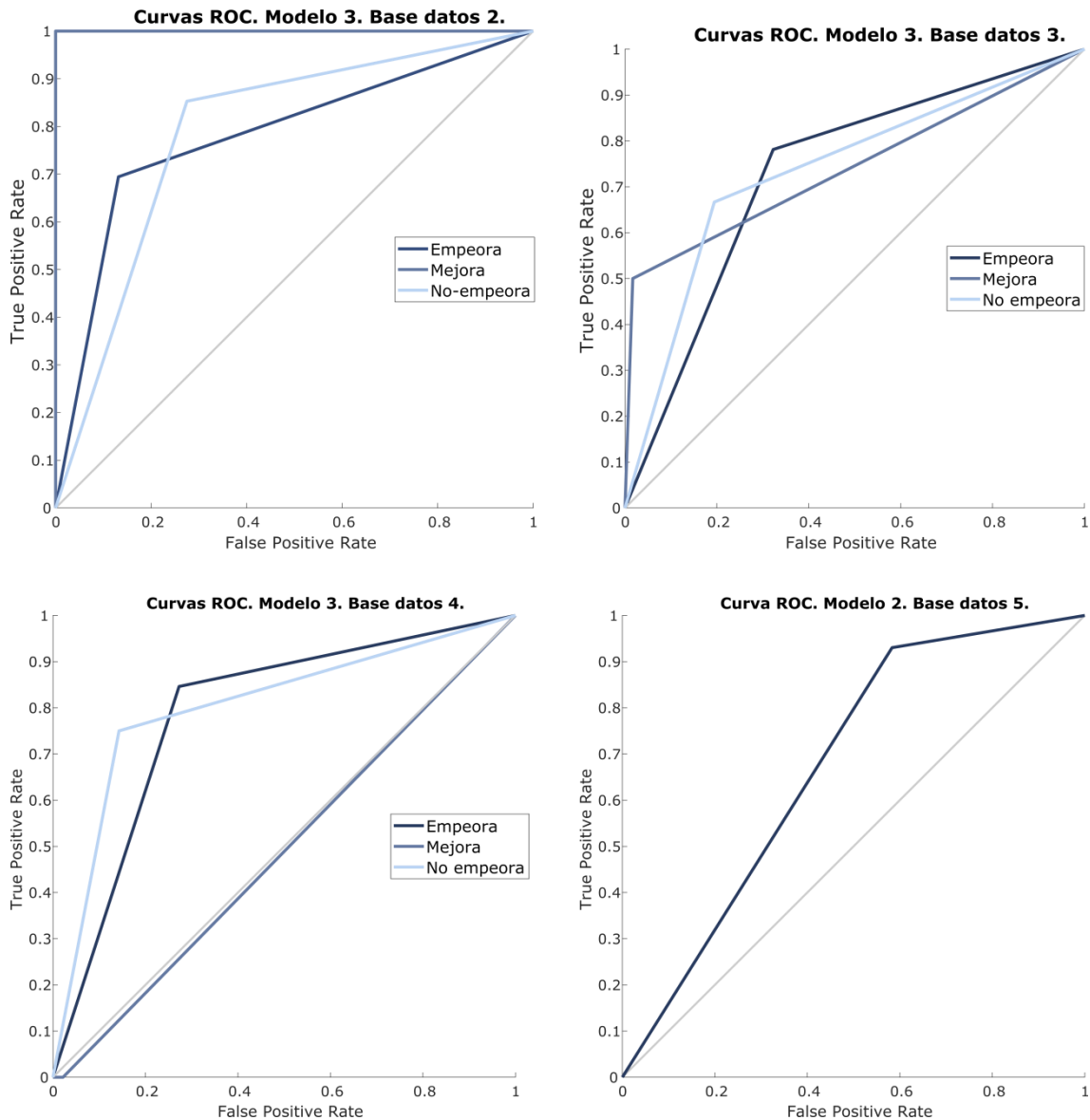
| Modelo | Clasificación | Algoritmo | F. Matlab | k-fold CV | Precisión |
|----------|---------------|--------------------------|---------------|-----------|---------------|
| Modelo 1 | 5 clases | LSTM | bilstmLayer | 5 | 0.5373 |
| Modelo 2 | 2 clases | SVM (polinomial) | Fitcsvm | 5 | 0.6119 |
| | | SVM (linear) | Fitcsvm | 5 | 0.6866 |
| | | SVM (gaussian) | Fitcsvm | 5 | 0.5373 |
| | | Decision tree | Fitctree | 5 | 0.6119 |
| | | K nearest neighbors | Fitcknn | 5 | 0.6567 |
| | | Naive Bayes model | Fitcnb | 5 | 0.7463 |

| | | | | | |
|----------|----------|----------------------------|----------------|----------|---------------|
| Modelo 3 | 3 clases | SVM multiclass | Fitcecoc | 5 | 0.5821 |
| | | K nearest neighbors | Fitcknn | 5 | 0.7015 |
| | | Decision tree | Fitctree | 5 | 0.5821 |
| | | Naive Bayes model | Fitcnb | 5 | -- |

Por último, para la base de datos cuarta, el mejor modelo de predicción es el segundo, además, a diferencia de las anteriores, el algoritmo utilizado para su desarrollo es modelo de Naïve-Bayes.

iii.3 Curvas ROC.

En este apartado se representan las curvas ROC para el mejor resultado de predicción obtenido en cada base de datos, que como se ha comentado anteriormente se obtiene en todas ellas para el modelo 3, excepto para la base quinta que se obtiene para el modelo 2.



iv. ÍNDICE DE FIGURAS ANEXOS.

| | |
|--|----|
| Figura i.1. Equipo OCT-Spectralis en el Hospital Universitario Miguel Servet de Zaragoza..... | 65 |
| Figura i.2 Ejemplo de datos que proporciona OCT-Spectralis..... | 66 |
| Figura i.3 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..... | 67 |
| Figura i.4 Representación de la adquisición de datos alrededor de la zona macular.... | 67 |
| Figura ii.1. Izquierda: Frecuencia de disponibilidad de datos en cada una de las visitas en número de ojos. Derecha: Frecuencia de disponibilidad de pacientes según la cantidad de visitas que dispone. | 69 |
| Figura ii.2 Número de visitas de las que dispone cada uno de los pacientes (144 ojos en total) utilizados en los modelos de predicción desarrollados con la base de datos 1. ... | 70 |
| Figura ii.3. Izquierda: Frecuencia de disponibilidad de datos en cada una de las visitas en número de ojos. Derecha: Frecuencia de disponibilidad de pacientes según la cantidad de visitas que dispone. | 71 |
| Figura ii.4. Número de visitas de las que dispone cada uno de los pacientes (114 ojos en total) utilizados en los modelos de predicción desarrollados con la base de datos 2. | 71 |
| Figura ii.5. Izquierda: Frecuencia de disponibilidad de datos en cada una de las visitas en número de ojos. Derecha: Frecuencia de disponibilidad de pacientes según la cantidad de visitas que dispone. | 72 |
| Figura ii.6. Número de visitas de las que dispone cada uno de los pacientes (114 ojos en total) utilizados en los modelos de predicción desarrollados con la base de datos 3. | 72 |
| Figura ii.7. Izquierda: Frecuencia de disponibilidad de datos en cada una de las visitas en número de ojos. Derecha: Frecuencia de disponibilidad de pacientes según la cantidad de visitas que dispone. | 73 |
| Figura ii.8. Número de visitas de las que dispone cada uno de los pacientes (112 ojos en total) utilizados en los modelos de predicción desarrollados con la base de datos 4. | 73 |
| Figura ii.9. Izquierda: Frecuencia de disponibilidad de datos en cada una de las visitas en número de ojos. Derecha: Frecuencia de disponibilidad de pacientes según la cantidad de visitas que dispone. | 74 |
| Figura ii.10. Número de visitas de las que dispone cada uno de los pacientes (114 ojos en total) utilizados en los modelos de predicción desarrollados con la base de datos 5. | 74 |

v. ÍNDICE DE TABLAS ANEXOS.

| | |
|---|----|
| Tabla 1. Frecuencia de datos de cada una de las 5 clases en las que se clasifican los ojos utilizados para el modelo de predicción M1. | 75 |
| Tabla 2. Frecuencia de datos de cada clase en las que se clasifican los ojos utilizados para los modelos de predicción M2 y M3. | 76 |
| Tabla 3. Resultados por cada modelo en la base de datos correspondiente a la capa de células ganglionares en la zona macular. | 76 |
| Tabla 4. Resultados por cada modelo en la base de datos correspondiente a la capa de fibras nerviosas de la retina en la zona macular. | 77 |
| Tabla 5. Resultados por cada modelo en la base de datos correspondiente a la capa de fibras nerviosas de la retina en la zona peripapilar con protocolo de adquisición axonal. | 77 |
| Tabla 6. Resultados por cada modelo en la base de datos correspondiente a la capa de fibras nerviosas de la retina en la zona peripapilar con protocolo de adquisición glaucoma. | 77 |