

Universidad Autónoma de Querétaro

Facultad de Informática

Algoritmo de aprendizaje automático basado en información genética  
para predecir la respuesta de pacientes con esclerosis múltiple al  
tratamiento de interferón beta

Tesis

Que como parte de los requisitos para obtener el Grado de

Doctor en Ciencias de la Computación

Presenta

Edgar Rafael Ponce de León Sánchez

Dirigido por:

Dr. Jorge Domingo Mendiola Santibañez

Co-Director:

Dr. Omar Arturo Domínguez Ramírez

Querétaro, Qro. a 04 de diciembre de 2024

La presente obra está bajo la licencia:  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

### Usted es libre de:

**Compartir** — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

### Bajo los siguientes términos:



**Atribución** — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



**NoComercial** — Usted no puede hacer uso del material con [propósitos comerciales](#).



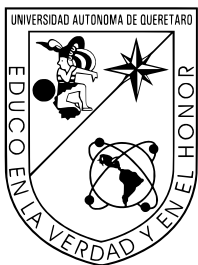
**SinDerivadas** — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

**No hay restricciones adicionales** — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

### Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



Universidad Autónoma de Querétaro

Facultad de Informática

Doctorado en Ciencias de la Computación

Algoritmo de aprendizaje automático basado en información genética  
para predecir la respuesta de pacientes con esclerosis múltiple al  
tratamiento de interferón beta

Tesis

Que como parte de los requisitos para obtener el Grado de

Doctor en Ciencias de la Computación

Presenta

Edgar Rafael Ponce de León Sánchez

Dirigido por:

Dr. Jorge Domingo Mendiola Santibañez

Co-Director:

Dr. Omar Arturo Domínguez Ramírez

Dra. Ana Marcela Herrera Navarro

Vocal

Dr. Alberto Vásquez Cervantes

Suplente

Dr. Hugo Jiménez Hernández

Suplente

Centro Universitario, Querétaro, Qro.

04 de diciembre de 2024

México

## **DEDICATORIA**

Dedico esta Tesis a mi director, Dr. Jorge Mendiola, a mi co-director, Dr. Omar Domínguez, y a mis sinodales, Dra. Ana Marcela Herrera, Dr. Alberto Vásquez, y Dr. Hugo Jiménez, por enseñarme, por compartir su experiencia, por su paciencia y por su apoyo constante, que fueron fundamentales para la realización de este trabajo. Su guía me ayudo en mi formación académica y me motivó en momentos de incertidumbre. Su confianza en mí me impulsó a seguir adelante y superar los desafíos que enfrenté.

A mi esposa, por su amor incondicional y su apoyo moral y espiritual, que fueron muy importantes en los momentos difíciles. A mis padres, hermanos, tíos y primos, por sus palabras de aliento, consejos y oraciones. A mis abuelos, que aunque no están físicamente presentes, sus enseñanzas siguen guiándome día a día.

A la Universidad Autónoma del Estado de Querétaro, por brindarme la oportunidad de crecer académica y profesionalmente. A la Facultad de Informática, en especial al M.S.I. José Alejandro Vargas y al Dr. Mauricio Ibarra, por su apoyo constante y disposición, por su confianza en mi trabajo y el ambiente de aprendizaje que me ofrecieron, que fueron esenciales para la culminación de esta investigación.

A mis amigos y compañeros, por su compañía y apoyo en los momentos de adversidad. Su amistad me ayudó a mantener el ánimo para continuar en este camino.

A todos los colegas y colaboradores que participaron en esta investigación. Su ayuda en la revisión de mi trabajo y valiosos comentarios enriquecieron este proyecto. Ésta Tesis es el resultado de un esfuerzo colectivo.

## **AGRADECIMIENTOS**

Agradezco principalmente a Dios por concederme la vida, el entendimiento y la fortaleza para llegar hasta este momento. Sin Su ayuda, todo este proyecto no habría sido posible.

"A ti, oh Dios de mis padres, te doy gracias y te alabo, porque me has dado sabiduría y fuerza, y ahora me has revelado lo que te pedimos; ..." Daniel 2:23a  
RVR1960

## TABLA DE CONTENIDOS

<b>Dedicatoria .....</b>	<b>I</b>
<b>Agradecimientos .....</b>	<b>II</b>
<b>Contenido .....</b>	<b>III</b>
<b>Índice de Figuras .....</b>	<b>VI</b>
<b>Índice de Tablas .....</b>	<b>X</b>
<b>Abreviaturas y siglas .....</b>	<b>XII</b>
<b>Resumen.....</b>	<b>XVII</b>
<b>1 Introducción.....</b>	<b>19</b>
1.1 DESCRIPCIÓN DEL PROBLEMA.....	21
1.2 JUSTIFICACIÓN.....	23
1.3 ESTRUCTURA DE LA TESIS.....	25
<b>2 Antecedentes.....</b>	<b>26</b>
<b>3 Hipótesis.....</b>	<b>30</b>
3.1 SUPUESTOS DE INVESTIGACIÓN .....	30
<b>4 Objetivos.....</b>	<b>31</b>
4.1 OBJETIVO GENERAL .....	31
4.2 OBJETIVOS ESPECÍFICOS.....	31
<b>5 Material y Métodos.....</b>	<b>32</b>
5.1 IMPORTACIÓN DE DATOS .....	32
5.2 METODOLOGÍA .....	32
5.3 ASIGNACIÓN DE ETIQUETAS DE CLASE.....	33
5.3.1 <i>Preprocesamiento de datos</i> .....	33
5.3.2 <i>Sistema experto de lógica difusa</i> .....	36

5.3.3	<i>Agrupamiento jerárquico</i> .....	38
5.4	TÉCNICAS CONVENCIONALES DE ML .....	40
5.4.1	<i>K-Vecinos</i> .....	41
5.4.2	<i>Modelos lineales</i> .....	42
5.4.3	<i>Support Vector Machine</i> .....	43
5.4.4	<i>Decision Tree</i> .....	43
5.5	MODELOS DE <i>Deep Learning</i> .....	44
5.5.1	<i>Redes Neuronales Artificiales</i> .....	45
5.5.2	<i>Perceptron Multi-capas</i> .....	46
5.5.3	<i>Esquema de optimización</i> .....	46
5.5.4	<i>Optimización de hiperparámetros con Algoritmo Genético (GA)</i> ..	48
5.5.5	<i>Métricas de rendimiento</i> .....	49
<b>6</b>	<b>Resultados y Discusión</b> .....	<b>58</b>
6.1	DISTRIBUCIÓN DE LOS DATOS .....	58
6.2	SISTEMA EXPERTO DIFUSO .....	59
6.3	OPTIMIZACIÓN DE HIPERPARÁMETROS CON GA .....	61
6.4	EVALUACIÓN DE RENDIMIENTO .....	62
6.5	PUBLICACIONES .....	62
<b>7</b>	<b>Conclusiones</b> .....	<b>71</b>
	<b>Referencias</b> .....	<b>73</b>
<b>8</b>	<b>Apéndice (Aplicación de un algoritmo de ML en el estudio del progre- so de la MS)</b> .....	<b>84</b>
8.1	INTRODUCCIÓN .....	84
8.2	MATERIALES Y MÉTODOS .....	85
8.2.1	<i>Base de datos</i> .....	85
8.2.2	<i>Algoritmo propuesto</i> .....	85
8.2.3	<i>Morfología matemática</i> .....	86
8.2.4	<i>Transformaciones geodésicas</i> .....	88
8.2.5	<i>Reconstrucción morfológica</i> .....	89
8.2.6	<i>Apertura y cierre por reconstrucción</i> .....	90
8.2.7	<i>Medidas de imagen</i> .....	91

8.3	RESULTADOS.....	92
8.3.1	<i>Algoritmo (Etapa 1)</i> .....	93
8.3.2	<i>Algoritmo (Etapa 2)</i> .....	95
8.4	DISCUSIÓN .....	95



## ÍNDICE DE FIGURAS

5.1	Citocina IL-2. Valores de expresión de 25 pacientes con RRMS correspondientes a cinco dosis: antes de la primer inyección de IFN- $\beta - 1b$ , segunda, y los meses 1, 12, 24. ....	33
5.2	Citocina IL-4. Los valores de expresión parecen más dispersos que los de la citocina IL-2. ....	34
5.3	Metodología propuesta: 1) Se recopilan las características demográficas, clínicas, y genéticas de pacientes con RRMS tratados con IFN- $\beta - 1b$ durante 2 años y se preprocesan los datos aplicando la técnica de estandarización y la prueba de KS. 2) Se ingresan las variables demográficas y clínicas a un sistema experto difuso para obtener las etiquetas de clase no conocidas. 3) Se ingresan las variables genéticas con las salidas etiquetadas a un GA que busque los hiperparámetros de configuración óptimos para entrenar a un modelo ANN que realice estimaciones de la respuesta al IFN- $\beta - 1b$ . ....	35
5.4	Estructura de un sistema difuso, que consiste de tres etapas: 1) Fuzzyficador, 2) Razonamiento aproximado y 3) Defuzzyficador.....	37
5.5	Conjunto de valores lingüísticos correspondientes a tres etiquetas que describen la variable de entrada "Edad" (conjunto difuso $A_1^{(1)}$ ). ....	39
5.6	Conjunto de valores lingüísticos correspondientes a tres etiquetas que describen la variable de entrada "EDSS después de dos años" (conjunto difuso $A_4^{(1)}$ ). ....	40
5.7	Conjunto de valores lingüísticos correspondientes a tres etiquetas que describen la variable de salida "Respuesta al IFN- $\beta$ " (conjunto difuso $B(y)$ ). ....	41
5.8	Arquitectura de la ANN propuesta. La capa lineal aplica una transformación lineal a los datos de entrada: $y = xA^T + b$ , donde $A$ es la matriz de pesos y $b$ es el vector de compensación. ....	50
5.9	Diagrama de matriz de confusión para las etiquetas de clase: Respondedor alto (0), respondedor bajo (1) y respondedor medio (2) al IFN- $\beta 1b$ .....	51

6.1	Gráfico de las funciones de distribución acumulativa empírica de las variables demográficas y clínicas. ....	59
6.2	Gráfico de las funciones de distribución acumulativa empírica de las variables genéticas (primera dosis).....	61
6.3	Gráfico de inferencia de reglas en la aplicación <i>Fuzzy Logic Designer</i> , que muestra los resultados de la combinación de los valores de inferencia de las reglas 13 y 15. ....	62
6.4	Gráfico de resultados de inferencia de la muestra 7, que incluye los valores de las etiquetas lingüísticas: bajo, medio y alto. ....	63
6.5	Resultados de la función de pérdida del modelo ANN. Los parámetros de entrenamiento fueron configurados por el GA implementado: 'num_neuronas_ocultas' = 64, 'num_capas_ocultas' = 2, 'func_activacion' = 'relu', 'optimizador' = 'adamax'. ....	64
6.6	Resultados de la métrica de precisión del modelo ANN. ....	65
8.1	Procedimiento de la primer etapa del algoritmo propuesto: Transformaciones morfológicas apertura en MRI de cerebros sanos y con MS, y mediciones de granulometría para estimar el diagnóstico de MS.....	86
8.2	Procedimiento de la segunda etapa del algoritmo propuesto: Transformaciones morfológicas de cierre en MRI de cerebros de pacientes con MS y mediciones de granulometría para estimar el tamaño de las lesiones de la enfermedad. ....	87
8.3	SE isótropos planos elementales para grafos octagonales y de disco: (a) Diamante. (b) Octágono. (c) Disco. El origen de cada SE está en su centro. ....	87
8.4	Dilatación geodésica de una imagen binaria o un conjunto $Y$ dentro de una máscara geodésica $X$ . El conjunto marcador se dilata primero mediante el SE isótropo elemental y luego se intersecta con la máscara geodésica: $\delta_X^{(1)}(Y) = \delta^{(1)}(Y) \cap X$ . ....	89
8.5	Erosión geodésica de una señal marcador $f$ 1D con respecto a una señal máscara $g$ . Debido al operador máximo puntual, todos los píxeles de la erosión elemental $f$ con valores inferiores a $g$ se establecen en el valor de $g$ . ....	89

8.6	Reconstrucción morfológica autodual $R^v$ de una imagen máscara $g$ a partir de una imagen marcador $f$ . ....	91
8.7	Implementación de transformaciones morfológicas de apertura: (a) Imagen de máscara (MS axial). (b) Apertura morfológica. (c) Apertura por reconstrucción. (d) Ajuste de intensidad. Uno de los objetivos de la primera etapa fue resaltar las áreas de lesión en las muestras de MS (MRI axial) utilizando un SE en forma de disco de radio $r = 5$ . ...	93
8.8	Implementación de transformaciones morfológicas de apertura: (a) Imagen de máscara (MS saggital). (b) Apertura morfológica. (c) Apertura por reconstrucción. (d) Ajuste de intensidad. Uno de los objetivos de la primera etapa fue resaltar las áreas de lesión en las muestras de MS (MRI saggital) utilizando un SE en forma de disco de radio $r = 5$ . ..	94
8.9	Comparación de las mediciones de granulometría en imágenes cerebrales preprocesadas utilizando el algoritmo propuesto (Etapa 1): (a) Imagen axial de MS (SumVolGranu = 0,1011). (b) Imagen axial sana (SumVolGranu = 0,0786). (c) Imagen saggital de MS (SumVolGranu = 0,1389). (d) Imagen saggital sana (SumVolGranu = 0,1079). La suma de valores de intensidad de objetos (sumVolGranu) en los cerebros de MS es mayor que en los cerebros sanos, debido a la presencia de lesiones. ....	99
8.10	Comparación de la suma de granulometrías de objetos. La suma de los valores de granulometría en muestras de MS (axial y saggital) es mayor que en muestras sanas por la presencia de las lesiones. ....	100
8.11	Matriz de confusión de predicciones. Resultados de ANN (axial): $TP = 10$ (verdaderos positivos clasificados correctamente como MS), $TN = 10$ (verdaderos negativos clasificados correctamente como sanos), $FP = 0$ (falsos positivos clasificados correctamente como MS) y $FN = 0$ (falsos negativos clasificados correctamente como sanos). Resultados de ANN (sagital): $TP = 8$ (verdaderos positivos clasificados correctamente como MS), $TN = 10$ (verdaderos negativos clasificados correctamente como sanos), $FP = 2$ (falsos positivos clasificados incorrectamente como MS) y $FN = 0$ (falsos negativos clasificados correctamente como sanos). ....	101

8.12 Gráfico de la función de pérdida de entropía cruzada frente a la fuerza de regularización. Los hiperparámetros <i>lambda</i> correspondientes, $X = 0,005$ (axial) y $X = 0,02$ (saggital), se ajustaron para minimizar las funciones de pérdida. En estos puntos, las funciones de pérdida alcanzaron los valores más bajos: $Y = 0,0272$ (axial) e $Y = 0,0632$ (saggital), respectivamente. ....	102
8.13 Comportamiento de la función de pérdida de entropía cruzada de la ANN (axial). En la iteración 20 ( $X = 20$ ), la función de pérdida convergió aproximadamente en $Y = 0,0416$ . ....	102
8.14 Comportamiento de la función de pérdida de entropía cruzada de ANN (saggital). En la iteración 20 ( $X = 20$ ), la función de pérdida convergió aproximadamente en $Y = 0,100$ . Por lo tanto, la función de pérdida de la ANN axial alcanzó el valor más bajo en comparación con la función de pérdida de la ANN saggital. ....	103
8.15 Implementación de transformaciones morfológicas de cierre (MS axial): (a) Imagen máscara (MS axial). (b) Imagen máscara sin lesiones de MS. (c) Cierre morfológico. (d) Cierre por reconstrucción (imagen de referencia o aproximación de cerebro sano). ....	103
8.16 Implementación de transformaciones morfológicas de cierre (MS saggital): (a) Imagen máscara (MS saggital). (b) Imagen máscara sin lesiones de MS. (c) Cierre morfológico. (d) Cierre por reconstrucción (imagen de referencia o aproximación de cerebro sano). ....	104
8.17 Mediciones de granulometría (axial) de la imagen máscara y la imagen de referencia utilizando un SE en forma de disco con diferentes valores de radio ( $r = 1, 2, 3, \dots, 15$ ). El tamaño de las lesiones de MS se estima calculando la diferencia entre cada punto de datos. ....	105
8.18 Medidas de granulometría (saggital) de la imagen máscara y la imagen de referencia utilizando un SE en forma de disco con diferentes valores de radio ( $r = 1, 2, 3, \dots, 15$ ). ....	106

## ÍNDICE DE TABLAS

5-1	Características demográficas y clínicas de pacientes con RRMS tratados con IFN- $\beta - 1b$ durante dos años.....	52
5-2	Descripción de las variables lingüísticas (características demográficas, clínicas y la respuesta al IFN- $\beta 1b$ ). ....	53
5-3	Definición de reglas difusas (base de conocimientos primera parte). ....	54
5-4	Definición de reglas difusas (base de conocimientos segunda parte). ...	55
5-5	Definición de reglas difusas (base de conocimientos tercera parte). ....	56
5-6	Pseudocódigo para la optimización de parámetros de la ANN por el GA. ....	57
6-1	Resultados estadísticos de la prueba KS en las variables demográficas y clínicas. ....	58
6-2	Resultados estadísticos de la prueba KS en las variables genéticas (primera dosis). ....	60
6-3	Resultados de fuzzificación (Edad y EDSS al inicio). $\mu_B : \mu \text{ Bajo}$ , $\mu_M : \mu \text{ Medio}$ , $\mu_A : \mu \text{ Alto}$ . ....	66
6-4	Resultados de fuzzificación (EDSS después de 1 año y EDSS después de 2 años). $\mu_B : \mu \text{ Bajo}$ , $\mu_M : \mu \text{ Medio}$ , $\mu_A : \mu \text{ Alto}$ . ....	67
6-5	Resultados de inferencia de la evaluación de las reglas difusas 13 y 15 (muestra 7). ....	68
6-6	Clasificación de la respuesta al IFN- $\beta 1b$ . Los valores numéricos resultantes de la defuzzificación inferiores a 0.5 se consideran como respondedor bajo (LR), los iguales a 0.5 como respondedor medio (MR) y los superiores a 0.5 como respondedor alto (HR). Para fines de comparación, los datos de entrada de la Tabla 6-1 también se utilizaron para entrenar a un modelo de agrupamiento jerárquico (n_clusters = 3). ....	69
6-7	Optimización de hiperparámetros de la ANN por GA (parámetros óptimos con la precisión máxima). ....	70
6-8	Optimización de parámetros del MLP por <i>GridSearchCV</i> (parámetros seleccionados con la máxima precisión). ....	70

6-9	Resultados de rendimiento por modelo de aprendizaje. En el caso del modelo ANN, los resultados de las métricas se consideran en el intervalo de la época 0 hasta la 11, que es cuando alcanzan sus valores máximos. ....	70
7-1	Comparación de resultados de rendimiento de aplicaciones de ML en el estudio de MS. ....	72
8-1	Pseudocódigo implementado del algoritmo de preprocesamiento de MRI (Etapa 1). ....	97
8-2	Pseudocódigo implementado del algoritmo de preprocesamiento de MRI (Etapa 2). ....	98
8-3	Resultados de las métricas de rendimiento. La precisión de prueba es la proporción de predicciones correctas, el DSC mide la similitud entre los datos de prueba y las predicciones, el TPR es la proporción de positivos clasificados correctamente, el TNR es la proporción de negativos clasificados correctamente y la pérdida de entropía cruzada calcula la pérdida entre los datos de prueba y las predicciones. El rendimiento de la ANN entrenada con imágenes axiales fue superior al de la ANN entrenada con imágenes sagittales. ....	100
8-4	Comparación de resultados de rendimiento en diferentes estudios que analizan lesiones de MS. ....	107

## ABREVIATURAS Y SIGLAS

<b>MS</b> esclerosis múltiple . . . . .	XVII
<b>CNS</b> sistema nervioso central . . . . .	XVII
<b>IFN</b> interferón . . . . .	XVII
<b>ML</b> aprendizaje automático . . . . .	XVII
<b>AI</b> inteligencia artificial . . . . .	XVII
<b>RRMS</b> esclerosis múltiple recurrente-remitente . . . . .	XVII
<b>GA</b> algoritmo genético . . . . .	XVII
<b>DL</b> aprendizaje profundo . . . . .	24
<b>ANN</b> red neuronal artificial . . . . .	XVII
<b>SPMS</b> esclerosis múltiple progresiva secundaria . . . . .	19
<b>PPMS</b> esclerosis múltiple progresiva primaria . . . . .	19
<b>BBB</b> barrera hematoencefálica . . . . .	20

<b>Treg</b> reguladora T . . . . .	20
<b>Th</b> ayudante T . . . . .	20
<b>APC</b> células presentadoras de antígenos . . . . .	20
<b>MHC</b> complejo mayor de histocompatibilidad . . . . .	20
<b>AG</b> acetato de glatiramer . . . . .	20
<b>DEG</b> genes expresados diferencialmente . . . . .	21
<b>ADA</b> anticuerpos antidrogas . . . . .	22
<b>PCA</b> análisis de componentes principales . . . . .	22
<b>GBTM</b> modelado de trayectorias basado en grupos . . . . .	22
<b>DNA</b> ácido desoxirribonucleico . . . . .	22
<b>NAWM</b> sustancia blanca de apariencia normal . . . . .	22
<b>DMP</b> posiciones metiladas diferencialmente . . . . .	22
<b>DCE</b> genes correlacionados diferencialmente . . . . .	23
<b>SVM</b> máquina de vectores de soporte . . . . .	23



<b>SCM</b> métodos de computación blanda . . . . .	24
<b>TGF</b> factor de crecimiento transformante . . . . .	26
<b>CDSS</b> sistema de apoyo a la toma de decisiones clínicas . . . . .	26
<b>TNF</b> factor de necrosis tumoral . . . . .	27
<b>DT</b> árbol de decisión . . . . .	27
<b>RF</b> bosque aleatorio . . . . .	27
<b>AUC</b> área bajo la curva . . . . .	27
<b>KN</b> k vecinos . . . . .	28
<b>GNB</b> Gaussian Naive Bayes . . . . .	28
<b>CNN</b> red neuronal convolucional . . . . .	28
<b>EGB</b> extreme gradient boosting . . . . .	28
<b>UCSC</b> uncorrelated shrunken centroid . . . . .	29
<b>RNA</b> ácido ribonucleico . . . . .	29
<b>GEO</b> Gene Expression Omnibus . . . . .	32

<b>EDSS</b> escala ampliada del estado de discapacidad . . . . .	52
<b>KS</b> Kolmogorov-Smirnov . . . . .	35
<b>COG</b> centro de gravedad . . . . .	37
<b>MAFS</b> sistema difuso Mamdani-Assilan . . . . .	37
<b>LR</b> regresión logística . . . . .	42
<b>IG</b> ganancia de información . . . . .	44
<b>MLP</b> perceptrón multicapa . . . . .	46
<b>SSE</b> suma de errores al cuadrado . . . . .	47
<b>SGD</b> descenso de gradiente estocástico . . . . .	47
<b>Adam</b> estimación del momento adaptativo . . . . .	47
<b>CM</b> matriz de confusión . . . . .	49
<b>TP</b> positivos verdaderos . . . . .	49
<b>TN</b> negativos verdaderos . . . . .	49
<b>FP</b> positivos falsos . . . . .	49

<b>FN</b> negativos falsos . . . . .	49
<b>TPR</b> tasa de positivos verdaderos . . . . .	50
<b>TNR</b> tasa de negativos verdaderos . . . . .	50
<b>MRI</b> imagen por resonancia magnética . . . . .	25
<b>MM</b> morfología matemática . . . . .	84
<b>SE</b> elemento estructurante . . . . .	86
<b>DSC</b> coeficiente de similitud cubo . . . . .	94

## RESUMEN

La esclerosis múltiple (MS) es una enfermedad inflamatoria crónica del sistema nervioso central (CNS), caracterizada patológicamente por desmielinización, pérdida axonal y gliosis con la formación de múltiples placas en el cerebro y la médula espinal, y clínicamente por una variedad de signos y síntomas neurológicos diseminados en el tiempo y el espacio. Aunque la etiología (causas) y la patogenia (origen) de la MS no están claramente establecidas, es probable que sea el resultado de interacciones complejas entre factores genéticos y ambientales. La etiología autoinmune de la MS ha sido el objetivo del acercamiento terapéutico a los pacientes. El interferón (IFN)- $\beta$  es una de las terapias modificadoras de la enfermedad más prescritas en pacientes con esclerosis múltiple recurrente-remitente (RRMS). Sin embargo, este tratamiento es parcialmente efectivo y una proporción significativa de pacientes no responde a este fármaco. El aprendizaje automático (ML) es una rama de la inteligencia artificial (AI) que se enfoca en el desarrollo de modelos computacionales basados en datos para mejorar tareas específicas. En los últimos años, se han aplicado algunos métodos de aprendizaje no supervisado (agrupación jerárquica, *K - means*, etc.) en el estudio de la MS, para realizar tareas de clasificación. Aunque estos métodos son relativamente fáciles de implementar, rara vez proporcionan la mejor solución debido a la gran cantidad de decisiones arbitrarias y funcionan eficazmente sólo cuando el conjunto de datos contiene grupos de tamaño similar y no hay valores atípicos notables. Por lo tanto, en este trabajo de investigación se propone un algoritmo de ML que incluye un sistema experto difuso, basado en la opinión de un experto en neurología, para mejorar la eficiencia en la clasificación de la respuesta al IFN- $\beta$  en 25 pacientes con RRMS, y un algoritmo genético (GA) para optimizar la sintonización de hiperparámetros de un modelo red neuronal artificial (ANN) entrenado con biomarcadores genéticos, para estimar si los pacientes son candidatos potenciales a ser tratados con este fármaco. Los resultados experimentales muestran que el sistema experto difuso propuesto tuvo una mayor eficiencia de clasificación comparado con un método convencional de agrupamiento jerárquico. El modelo ANN logró una mayor precisión de rendimiento y una convergencia más rápida de la función de pérdida, comparado con otros métodos

convencionales de ML.

**Palabras clave:** esclerosis múltiple, aprendizaje automático, sistema experto difuso, biomarcadores.

## ABSTRACT

MS is a chronic inflammatory disease of the CNS, characterized pathologically by demyelination, axonal loss, and gliosis with the formation of multiple plaques in the brain and spinal cord, and clinically by a variety of neurological signs and symptoms disseminated in time and space. Although the etiology (causes) and pathogenesis (origin) of MS are not clearly established, it is likely the result of complex interactions between genetic and environmental factors. The autoimmune etiology of MS has been the target of therapeutic approach to patients. IFN- $\beta$  is one of the most prescribed disease-modifying therapies in patients with RRMS. However, this treatment is partially effective and a significant proportion of patients do not respond to this drug. ML is a branch of AI that focuses on the development of data-driven computational models to improve specific tasks. In recent years, some unsupervised learning methods (hierarchical clustering, K-means, etc.) have been applied in the study of MS to perform classification tasks. Although these methods are relatively easy to implement, they rarely provide the best solution due to the large number of arbitrary decisions and work effectively only when the dataset contains clusters of similar size and no notable outliers. Therefore, in this research work, we propose an ML algorithm that includes a fuzzy expert system, based on the opinion of a neurology expert, to improve the efficiency for classifying the response to IFN- $\beta$  in 25 patients with RRMS, and a GA to optimize the hyperparameter tuning of an ANN model trained with genetic biomarkers to estimate whether the patients are potential candidates to be treated with this drug. The experimental results show that the proposed fuzzy expert system had a higher classification efficiency compared to a conventional hierarchical clustering method. The ANN model achieved higher performance accuracy and faster convergence of the loss function compared to other conventional ML methods.

**Keywords:** multiple sclerosis, machine learning, fuzzy expert system, biomarkers.

## 1. INTRODUCCIÓN

La MS es una enfermedad inflamatoria crónica del CNS de naturaleza autoinmune, que se caracteriza patológicamente por diversos grados de desmielinización, pérdida axonal y gliosis con la formación de múltiples placas de desmielinización que se distribuyen a lo largo del CNS, y clínicamente por una variedad de signos y síntomas neurológicos (Milo, 2014). La desmielinización consiste en un daño en la capa de mielina, material graso que aísla los nervios. La mielina permite que un nervio transmita sus impulsos rápidamente. De forma general, se considera que los síntomas de inicio más frecuentes son alteraciones de la sensibilidad, alteraciones motoras y síntomas debidos a disfunción del tronco del encéfalo (Margarit et al., 2019). Aunque la MS puede adoptar varias formas diferentes, el tipo más común es la RRMS, caracterizada por períodos alternos de remisión y exacerbación de los síntomas. La esclerosis múltiple progresiva secundaria (SPMS) comienza como esclerosis múltiple progresiva primaria (PPMS) y los síntomas empeoran gradualmente sin períodos de remisión. La PPMS se caracteriza por un deterioro progresivo de los síntomas sin remisión y es una forma rara en la que la enfermedad tiene un desarrollo progresivo desde el inicio de la enfermedad con síntomas que empeoran continuamente (Martynova et al., 2023). Aunque la etiología y la patogenia de la MS no están claramente establecidas, es probable que sean el resultado de interacciones complejas entre factores genéticos y ambientales (Ghasemi et al., 2017). Los factores ambientales incluyen infecciones virales, deficiencia de vitamina D, fumar, etc. (Dominguez-Mozo et al., 2021; Tarlinton et al., 2020; Zarghami et al., 2021). Algunos de estos factores activan el sistema inmunológico humano, causando una desregulación inmune, lo que lleva a un ataque inmunológico contra las vainas de mielina del CNS. La desregulación inmune se manifiesta por la infiltración de células inmunes en el CNS, lo que desencadena en la reducción de la mielina, pérdida axonal, gliosis y degeneración neuronal (Rodríguez Murúa et al., 2022). Varios estudios han demostrado que la susceptibilidad a la MS es genéticamente dependiente, pero los factores genéticos específicos siguen siendo en gran parte desconocidos. Por lo tanto, la identificación de alelos de riesgo o genes candidatos que juegan un papel importante en la patogenia de la MS sigue siendo un desafío.

Se sabe que las células inmunes específicas-auto-antígenas periféricas se activan durante el proceso de presentación antígeno e ingresan al CNS a través de la barrera hematoencefálica (BBB) (Pinheiro et al., 2016). La ruta de entrada depende del fenotipo y del estado de activación de las células T. Las células T desempeñan funciones importantes en la inmunidad celular (Liu et al., 2022). Dependiendo de su función, las células T se dividen en células ayudante T (Th) y células reguladora T (Treg). Ma et al. (2023) resumieron el desarrollo de distintos subconjuntos de células T y sus citocinas en la patogénesis de la MS. Después de la transmigración al CNS, las células T específicas de mielina son reactivadas por células presentadoras de antígenos (APC) residentes en el CNS, que resulta en el desarrollo y/o exacerbación de reacciones inflamatorias y desmielinización. Szpakowski et al. (2021) presentaron el papel de las APC en la patogénesis de la MS. Las células T CD4+ reactivas a la mielina pueden ser activadas por células dendríticas, macrófagos infiltrantes, células de microglía o células B, las cuales expresan moléculas complejo mayor de histocompatibilidad (MHC) de clase II (Nishihara et al., 2020).

El tratamiento de la MS es un desafío e implica varios fármacos que actúan a través de diferentes mecanismos. La indicación depende fundamentalmente del curso clínico y la forma de la enfermedad. El tratamiento de la MS se puede dividir en: tratamiento de los síntomas, tratamiento de la recaída y tratamiento que modifica la progresión de la enfermedad. El objetivo principal del tratamiento de la MS es ralentizar la progresión de la enfermedad. El acetato de glatiramer (AG) y el IFN- $\beta$  se administran como terapias de primera línea en el tratamiento de la MS. En pacientes que no responden al tratamiento, hay disponibles varias terapias de segunda línea como el natalizumab y fingolimod (Lorscheider et al., 2018). El IFN- $\beta$  es una de las terapias modificadoras de la enfermedad más prescritas para pacientes con RRMS. El IFN- $\beta$  tiene múltiples vías de acción sobre el sistema inmunológico, puede inhibir la proliferación activada de células T, previene la migración de células inmunes activadas a través de la BBB, inhibe la producción de citocinas proinflamatorias (p. ej., IL-2, IL-12, IFN- $\gamma$ ), induce el aumento de citocinas antiinflamatorias (p. ej., IL-4, IL-5, IL-10 y TGF- $\beta$ ) y promueve la remielinización en el CNS (Kay et al., 2013; Mirandola et al., 2009). El IFN- $\beta$  también puede prevenir la diferenciación de las células inflamatorias Th1/Th17 y cambiar el fenotipo de las células Th de Th1 inflamatorias a Th2 anti-inflamatorias (Hočevár et al., 2019). Algunos estudios han analizado

los efectos del tratamiento IFN- $\beta$  sobre los patrones de citocinas en pacientes con RRMS (Ganji et al., 2020; Martínez-Aguilar et al., 2020). Otros estudios han demostrado que el IFN- $\beta$  tiene efectos positivos sobre la reducción de la actividad de la enfermedad (Bustamante et al., 2015; Cohan et al., 2021). Sin embargo, a pesar de su alta eficacia, un porcentaje de pacientes con MS no responde, o responde parcialmente a este tratamiento (Martínez-Aguilar et al., 2020; Río et al., 2002).

El ML es una rama de la AI que se enfoca en el desarrollo de modelos computacionales para mejorar tareas específicas mediante la realización de predicciones basadas en datos (M., 2020). El ML se clasifica en dos técnicas: aprendizaje supervisado, que entrena un modelo con datos de entrada y salida conocidos para predecir resultados futuros, y aprendizaje no supervisado, que identifica patrones ocultos en los datos de entrada sin salidas etiquetadas. Durante la última década, ha habido un aumento en la aplicación de algoritmos de ML en varios campos médicos, incluidos la radiología, la cardiología, la oftalmología, la oncología y la neurología (JM, 2023; O., 2020).

## 1.1 Descripción del problema

El agrupamiento jerárquico es un algoritmo de ML no supervisado que agrupa los datos en un árbol de conjuntos anidados. Algunos estudios han aplicado métodos de agrupación jerárquica y no jerárquica para clasificar la respuesta a los tratamientos en pacientes con MS (Giordani P., 2020; Vichi M., 2022). En específico, en este trabajo de Tesis se revisaron estudios que han analizado el genoma para identificar factores genéticos asociados con la respuesta al tratamiento con IFN- $\beta$ . Gurevich et al. (2015) clasificaron a 50 pacientes con SPMS (20 tratados con IFN- $\beta$  - 1a y 30 no tratados) utilizando agrupación jerárquica según el perfil de expresión génica inducible por IFN identificado en pacientes con RRMS clínicamente respondedores al tratamiento. El perfil de expresión génica se determinó mediante la búsqueda de genes expresados diferencialmente (DEG) entre pacientes con RRMS tratados con IFN ( $n = 10$ ) y no tratados ( $n = 25$ ). Como resultado se identificaron 104 DEG, enriquecidos por la vía de señalización de IFN ( $p = 7,4e-08$ ) en pacientes con RRMS tratados con IFN. Coelewij et al. (2024) identificaron bio-



marcadores tempranos de inmunogenicidad asociados con el desarrollo futuro de anticuerpos antidrogas (ADA) en pacientes con RRMS antes del primer tratamiento con IFN- $\beta$ . Se utilizaron DEG en pacientes con SPMS frente a RRMS ADAneg y ADApos frente a pacientes con RRMS ADAneg para clasificar a los pacientes mediante agrupamiento jerárquico y análisis de componentes principales (PCA). Los resultados mostraron que estos biomarcadores se podrían utilizar para ayudar a las decisiones de tratamiento futuro en pacientes con RRMS.

*K - means* es otro algoritmo de agrupamiento iterativo no supervisado que esta basado en el cálculo de centroides para dividir un conjunto de datos en grupos similares según la distancia entre sus centroides. Algunos estudios han implementado este método para realizar tareas de clasificación. Mésidor et al. (2021) propusieron un enfoque modelado de trayectorias basado en grupos (GBTM) para identificar grupos de pacientes con patrones de dosis homogéneos para los fármacos IFN- $\beta - 1a$  y amitriptilina. Las diferencias en las características entre los grupos se identificaron mediante  $\chi$ -cuadrada y análisis de varianza, ambos ponderados por la probabilidad posterior de pertenencia al grupo. Como resultado se identificaron siete patrones de dosis para el IFN- $\beta - 1a$  y cinco para la amitriptilina. Kular et al. (2022) analizaron la metilación del ácido desoxirribonucleico (DNA) en núcleos de células no neuronales aisladas de 38 muestras sustancia blanca de apariencia normal (NAWM) de pacientes con MS en comparación con individuos de control. Se identificaron 1226 posiciones metiladas diferencialmente (DMP) significativas con valor  $p$  ajustado en todo el genoma ( $p - adj < 0,05$ ) entre pacientes con MS y controles sanos. Los genes que intervienen en los procesos biológicos fueron clasificados mediante agrupamiento *K - means* establecido en 5 grupos. Los hallazgos sugieren que las células gliales de NAWM están altamente alteradas mostrando la existencia de cambios moleculares relevantes subyacentes a la neuropatología de la MS. Aunque las técnicas de ML no supervisado son relativamente fáciles de implementar se identifican algunas limitantes:

1. El método de agrupamiento jerárquico rara vez proporciona la mejor solución debido a la gran cantidad de decisiones arbitrarias.
2. El método *K - means* funciona eficazmente sólo cuando el conjunto de datos contiene grupos de tamaño similar y no hay valores atípicos notables ni

variaciones de densidad.

Tao Jin (2020) implementaron un método de selección de características para identificar pares de genes correlacionados diferencialmente (DCE) asociados con la respuesta al tratamiento con IFN- $\beta$  en pacientes con RRMS. Como resultado se consideraron 22 DCE estadísticamente significativos (valor  $p < 0,01$ ), que involucran 41 genes únicos. Utilizando estos 41 genes como predictores, se entrenó un modelo máquina de vectores de soporte (SVM) obteniéndose una precisión predictiva de 0.8095. Sin embargo, en este estudio se visualizan otras limitantes:

1. El conjunto de datos de entrenamiento (que es el mismo que se utiliza en el presente trabajo de Tesis) no contiene etiquetas de salida y Tao Jin (2020) no describen claramente el método utilizado para determinar las categorías de respondedores y no respondedores al IFN- $\beta$ .
2. El conjunto de datos de entrenamiento tiene solo 25 muestras en total y, para eliminar posibles resultados ambiguos causados por pacientes cuya primer recaída sucedió después de 5 años se excluyeron 7 pacientes del análisis, lo cual causa una pérdida de información que pudiera ser relevante.

Por otro lado, los modelos de aprendizaje basados en parámetros, como el modelo SVM, son entrenados por la sintonización de sus parámetros. Los hiperparámetros son configuraciones que por lo regular se sintonizan arbitrariamente antes de comenzar el proceso de entrenamiento para optimizar el rendimiento del modelo (Madhiarasan and Deepa, 2016). Sin embargo, en ocasiones la sintonización arbitraria de parámetros no siempre logra los mejores resultados de rendimiento. Por lo tanto, en este trabajo de Tesis se intenta resolver las limitaciones mencionadas en los trabajos anteriores.

## 1.2 Justificación

En las últimas décadas, el avance tecnológico ha permitido el estudio a gran escala de muchos genes, proteínas y metabolitos, provocando la creación de

las ciencias ómicas como la genómica, proteómica, metabolómica, entre otras. Cada una de estas áreas ha ayudado a un mejor entendimiento de la causa de ciertas enfermedades. En el caso de la genómica, se puede abordar el análisis de enormes cantidades de datos de una manera mas eficiente, algo que hasta hace unos años resultaba muy complicado realizar con las técnicas convencionales de análisis estadístico (Frigolet and Gutiérrez-Aguilar, 2017). En la actualidad es posible elegir qué regiones del genoma se desea secuenciar. Por ejemplo, a partir del DNA de un individuo se puede secuenciar el genoma completo, solo exones o regiones específicas del genoma, con la finalidad de encontrar la región del genoma que pudiera estar afectada o mutada y ser la posible causante de alguna enfermedad. El genoma humano contiene alrededor de 20,000 genes, si bien este volumen de datos es complejo y multidimensional, existe una gran cantidad de redundancia e información irrelevante que provoca un alto costo computacional en el procesamiento de los datos. Por lo tanto, existe la necesidad de utilizar nuevos métodos de computación blanda (SCM), como sistemas expertos, algoritmos de ML y aprendizaje profundo (DL), algoritmos evolutivos o metaheurísticos, entre otros, que puedan procesar grandes cantidades de datos de una manera más eficiente (Karlik, 2013).

Hasta el momento, el campo de los sistemas expertos ha sido uno de los más activos en diferentes áreas de investigación, como el diagnóstico de algunas enfermedades. Un sistema experto es un programa informático que emula el proceso de razonamiento de un experto humano y puede gestionar de forma eficiente información incierta e imprecisa (Saibene et al., 2021). Un sistema experto difuso incorpora conjuntos difusos y lógica difusa en su proceso de razonamiento y base de conocimientos. Algunos estudios han aplicado sistemas difusos para analizar enfermedades neurológicas (Sadoughi and Arani, 2022). Sin embargo, la mayoría de ellos se han enfocado sólo en el diagnóstico (Ayangbekun and Jimoh Ibrahim, 2015; Chen, 2024; Hosseini et al., 2020; Matinfar and Golpaygani, 2022). El estudio de la respuesta a tratamientos sigue siendo un desafío.

Los modelos de ML y DL se basan en algoritmos matemáticos que encuentran patrones naturales en los datos y están surgiendo como herramientas muy útiles en el campo de la neurología (Law et al., 2019). Estos modelos de aprendizaje, basados en las características genéticas de los pacientes, pueden ayudar a mejorar el

diagnóstico de algunas enfermedades, como la MS temprana (Casalino et al., 2021; Chen et al., 2021; Goyal et al., 2019), y a estimar la posible respuesta a algunos tratamientos para la MS, como el natalizumab, fingolimod e IFN- $\beta$  (Ebrahimkhani et al., 2020; Fagone et al., 2019; Tao Jin, 2020).

Los GA son algoritmos metaheurísticos que buscan la solución óptima en el espacio de búsqueda y pueden ser herramientas útiles en la optimización de parámetros de los modelos de aprendizaje para lograr un mayor rendimiento en los resultados (Vasuki, 2020). Por lo tanto, en este trabajo de Tesis se propone un algoritmo de ML que incluye: un método para verificar la distribución normal de los datos de entrada (variables demográficas, clínicas y genéticas), un sistema experto difuso basado en la opinión de un experto en neurología para mejorar la eficiencia en la clasificación de la respuesta al IFN- $\beta$  en 25 pacientes con RRMS, y un GA para optimizar la sintonización de hiperparámetros de un modelo ANN entrenado con biomarcadores genéticos. La implementación de este algoritmo puede ser un apoyo a la decisión de los especialistas en la estimación de la respuesta a tratamientos de la MS, evitando terapias ineficaces y reduciendo los costos de atención médica.

### 1.3 Estructura de la Tesis

El presente trabajo de Tesis se encuentra organizado de la siguiente manera. En el Capítulo 2, se mencionan los antecedentes sobre la investigación, se discuten las aplicaciones hasta ahora realizadas y se comparan los trabajos consultados. En el capítulo 3, se describen los supuestos de investigación. En el capítulo 4, se plantean los objetivos de investigación. En el capítulo 5, se describen los fundamentos teóricos y la metodología utilizada en la investigación. En el Capítulo 6, se presentan los resultados obtenidos y la discusión. En el Capítulo 7, se mencionan las conclusiones finales y las sugerencias de trabajos a futuros. Por último en el Capítulo 8 (Apéndice), se presenta la aplicación de un algoritmo de ML basado en imagen por resonancia magnética (MRI) en el estudio del progreso de la MS.

## 2. ANTECEDENTES

Entre los estudios que han analizado los efectos del tratamiento IFN- $\beta$  sobre los patrones de citocinas en pacientes con RRMS, Ganji et al. (2020) evaluaron los niveles en suero de algunas citocinas, incluidas la interleucina-5 (IL-5), IL-8, IL-9, IL-17A, el factor de crecimiento transformante (TGF)- $\beta$  y el IFN- $\gamma$ . Se analizaron las muestras de pacientes con RRMS tratados con dosis altas de IFN- $\beta$  – 1a, dosis bajas de IFN- $\beta$  – 1a, IFN- $\beta$  – 1b y AG, y se compararon con los resultados de pacientes sin tratamiento previo y controles sanos. Los hallazgos mostraron que los niveles séricos de IL-8, IL-9 e IFN- $\gamma$  en pacientes sin tratamiento previo fueron significativamente más altos que los controles sanos, mientras que no hubo diferencias significativas en términos de otras citocinas entre los grupos. IFN- $\beta$  – 1b resultó en una disminución significativa en los niveles de IL-9 y TGF- $\beta$ . En general, los resultados respaldaron los efectos inflamatorios y agravantes de IL-8, IL-9 e IFN- $\gamma$  en la MS. Martínez-Aguilar et al. (2020) realizaron una revisión sistemática de varios estudios que evaluaron el impacto de los polimorfismos genéticos sobre la respuesta al tratamiento con IFN- $\beta$ . Como resultado, aproximadamente el 50 % de los pacientes con RRMS no respondieron al tratamiento y la determinación de la respuesta requiere de un seguimiento clínico largo de hasta dos años. Específicamente, los polimorfismos genéticos en los genes CD46, CD58, FHIT, GAPVD1, GPC5, GRBRB3, IRF5, MxA, PELI3 y ZNF697 mostraron un vínculo con la respuesta al IFN- $\beta$  entre pacientes con RRMS, pero hay pocos estudios con pequeños grupos de pacientes, lo que complica la generalización de su influencia. Los resultados sugieren que los polimorfismos detectados pueden ser utilizados en el futuro como marcadores predictivos de respuesta al IFN- $\beta$  en pacientes diagnosticados con RRMS.

Por un lado, en los últimos años se han realizado algunos estudios que han propuesto el uso de sistemas expertos para el diagnóstico de algunas enfermedades neurológicas. Ayangbekun and Jimoh Ibrahim (2015) diseñaron un sistema de inferencia difusa para diagnosticar cinco enfermedades cerebrales: Alzheimer, Creutzfeldt-Jakob, Huntington, MS y Parkinson. Hosseini et al. (2020) desarrollaron un sistema de apoyo a la toma de decisiones clínicas (CDSS) para ayudar a los especialistas a diagnosticar la MS con un fenotipo remitente-recurrente. Hamedan

et al. (2020) desarrollaron un sistema experto basado en lógica difusa para diagnosticar y predecir la enfermedad renal crónica y evaluar su robustez frente a datos ruidosos. Matinfar and Golpaygani (2022) propusieron un sistema experto para el diagnóstico de MS basado en síntomas clínicos y características demográficas. Chen (2024) propusieron un sistema experto difuso para diagnosticar la enfermedad de Parkinson en base a los valores de entrada de cada síntoma mostrado. Los estudios anteriores sólo se enfocaron en el diagnóstico de enfermedades, la aplicación de sistemas expertos para la clasificación de la respuesta a tratamientos seguía siendo un desafío. Por esta razón, Ponce de Leon-Sanchez et al. (2023b) implementaron un sistema experto de lógica difusa basado en la opinión de un experto en neurología para clasificar la respuesta (alto, medio bajo) al tratamiento con IFN- $\beta$  en pacientes con RRMS. Los resultados mostraron que el sistema difuso tuvo una eficiencia de clasificación del 100 % en comparación con un método de agrupamiento jerárquico no supervisado (52 %). Por lo tanto, el uso de sistemas difusos para clasificar la respuesta a tratamientos puede ser un método confiable para asignar adecuadamente las etiquetas de clase no conocidas en problemas de aprendizaje no supervisado.

Por otro lado, en la última década se ha producido un incremento significativo de aplicaciones de ML y DL en el estudio de la MS, que incluyen: la clasificación de los subtipos dentro de la enfermedad, la predicción del diagnóstico de la MS, la predicción de la evolución o progreso del paciente (riesgo de convertirse en otra forma más evolucionada de la enfermedad), y la predicción de la posible respuesta del paciente a alguna terapia farmacológica para ayudar al especialista en la elección del tratamiento más adecuado Vázquez-Marrufo et al. (2021).

En lo que se refiere al diagnóstico de la MS, Goyal et al. (2019) analizaron el nivel de suero de 8 citocinas (IL-1 $\beta$ , IL-2, IL-4, IL-8, IL-10, IL-13, IFN- $\gamma$  y factor de necrosis tumoral (TNF)- $\alpha$  en pacientes con MS para identificar predictores de la enfermedad. Los conjuntos de datos se utilizaron como entrada en cuatro modelos de aprendizaje: SVM, árbol de decisión (DT), bosque aleatorio (RF) y ANN. Se identificó al modelo RF con el mayor rendimiento para el diagnóstico de la MS, ya que se desempeñó de manera notable en los criterios de precisión (0.909), área bajo la curva (AUC)=0.957, sensibilidad (0.756) y especificidad (0.857). Este estudio sugiere

que las concentraciones de citocinas en suero podrían utilizarse como biomarcadores para el diagnóstico temprano de la MS. Chen et al. (2021) analizaron perfiles de expresión génica de muestras de sangre periférica por bio-informática basada en múltiples conjuntos de datos de microarreglos. Como resultado se identificaron los genes CXCR4, ITGAM, ACTB, RHOA, RPS27A, UBA52 y RPL8 como biomarcadores potenciales para el diagnóstico de la MS. Se utilizó una SVM para establecer un modelo de diagnóstico de MS con un alto rendimiento de predicción en conjuntos de datos internos y externos y en diferentes chips de plataformas de conjuntos de datos (precisión=0.930, sensibilidad=0.969, especificidad=0.844 y AUC=0.907). Este estudio proporciona una nueva comprensión de la etiología/patogénesis de la MS facilitando un diagnóstico temprano. Ponce de Leon-Sanchez et al. (2023a) propusieron un modelo ANN entrenado con 35 perfiles de expresión génica relacionados con la etiología de la MS para predecir la susceptibilidad a la enfermedad. Las características genéticas fueron adquiridas de 144 individuos (99 con MS y 45 controles sanos). Se realizó una comparación del rendimiento de un modelo ANN con cuatro técnicas de ML convencionales: k vecinos (KN), Gaussian Naive Bayes (GNB), SVM y DT. Los resultados mostraron un mayor rendimiento de predicción del modelo ANN (precisión=0.896, sensibilidad=0.904, especificidad=0.875 y AUC=0.889). Por lo tanto, el modelo propuesto puede ser una alternativa de apoyo a la decisión para proporcionar predicciones a corto plazo de la susceptibilidad a la MS en función a la genética de cada individuo.

Entre los estudios que se han enfocado en el análisis del progreso de la MS, Viatkin et al. (2021) desarrollaron un sistema para medir los ángulos de las articulaciones de los dedos en base a la imagen de una cámara, su objetivo era rastrear el movimiento y los límites de la movilidad de la mano en pacientes con MS. La medición de los cambios en la movilidad de la mano permitió controlar el progreso de la enfermedad y su proceso de tratamiento. Para analizar la información recibida de la cámara se utilizaron red neuronal convolucional (CNN) basadas en diferentes arquitecturas y sus combinaciones (VGG16, MobileNet, MobileNetV2, InceptionV3, DenseNet, ResNet). Treaba et al. (2021) utilizaron imágenes de susceptibilidad 7-T de campo ultra-alto para caracterizar la presencia y evolución de las lesiones corticales y de borde, en pacientes con RRMS y SPMS. Se implementó un algoritmo de ML basado en técnicas de extreme gradient boosting (EGB) para evaluar el poder

acumulativo, así como la importancia individual de los tipos de lesiones corticales y de borde, para predecir la etapa de la enfermedad y la progresión de la discapacidad. Las características de imagen más influyentes que discriminaron entre las etapas de la MS ( $AUC \pm$  desviación estándar =  $0,82 \pm 0,08$ ) incluyeron la sustancia blanca normalizada, el volumen de la lesión de la sustancia blanca, y el volumen de la lesión leucocortical.

Por último, entre los estudios que sugieren que la genética puede proporcionar pistas relevantes para anticipar la posible respuesta de pacientes con MS a un tratamiento terapéutico, Fagone et al. (2019) propusieron el algoritmo uncorrelated shrunken centroid (UCSC) para identificar un subconjunto de genes predictores de la respuesta farmacológica al tratamiento con natalizumab en pacientes con RRMS. Los hallazgos encontrados sugieren que un perfil específico de expresión génica de células CD4+ T (OSBPL6, FABP3, ESR1, SNX10, SFN, OSBPL6, APLP1, PPA1, TPM3, PSMB7, RAB28, FABP3) puede caracterizar la capacidad de respuesta al natalizumab con una precisión importante (0.892). Ebrahimkhani et al. (2020) evaluaron la utilidad de los micro-ácido ribonucleico (RNA) de exosomas séricos como biomarcadores de la actividad de la enfermedad en pacientes con MS en respuesta a la terapia con fingolimod. Los datos se analizaron mediante modelos univariados/multivariados y ML para determinar firmas de miRNA con utilidad predictiva. Los resultados mostraron el poder de los miRNA individuales para predecir el estado de la MS después del fingolimod, varias combinaciones de 2 o 3 miRNA pudieron distinguir la enfermedad activa de la inactiva con una precisión superior a 0.9. Los resultados obtenidos en los estudios anteriores, incluyendo el de Tao Jin (2020) antes mencionado, muestran que el uso de biomarcadores como predictores de la respuesta a tratamientos de la MS puede conducir a notables beneficios para los pacientes y propicia el desarrollo de nuevas investigaciones en el estudio de tratamientos personalizados.



### 3. HIPÓTESIS

#### 3.1 Supuestos de investigación

Los sistemas de lógica difusa utilizan el razonamiento aproximado, basado en el conocimiento experto, para procesar información inexacta de una manera más eficiente que los métodos convencionales de agrupamiento no supervisado. Por lo tanto, los sistemas difusos son métodos alternativos convenientes para el soporte a la decisión en la clasificación de la respuesta al tratamiento con IFN- $\beta$  en pacientes con RRMS.

Los modelos de ML y DL entrenados con biomarcadores genéticos asociados con la respuesta al IFN- $\beta$  pueden ser herramientas útiles para determinar si los pacientes con RRMS son candidatos potenciales para ser tratados con este fármaco.

GA es un algoritmo que converge en un número finito de iteraciones y es capaz de evaluar diferentes soluciones candidatas que evolucionan hasta alcanzar una mayor aptitud o precisión en cada iteración (generación). La implementación de este algoritmo puede ser una alternativa para optimizar la sintonización de hiperparámetros de los modelos de aprendizaje y mejorar su rendimiento predictivo.

## 4. OBJETIVOS

### 4.1 Objetivo general

Diseñar un algoritmo de ML que pueda clasificar eficientemente la respuesta al tratamiento con IFN- $\beta$  en pacientes con RRMS y que optimice la sintonización de hiperparámetros de un modelo ANN entrenado con biomarcadores genéticos.

### 4.2 Objetivos específicos

1. Aplicar un método de validación de la distribución normal de un conjunto de datos que contiene variables demográficas, clínicas y genéticas de 25 pacientes con RRMS tratados con IFN- $\beta$  durante 2 años.
2. Diseñar e implementar un sistema experto difuso, con reglas lingüísticas basadas en el conocimiento de un experto en neurología, para clasificar de manera más eficiente las etiquetas no conocidas (respondedor alto, medio, y bajo al IFN- $\beta$ ) utilizando variables demográficas y clínicas como entradas al sistema.
3. Implementar un algoritmo genético que optimice la sintonización de los hiperparámetros de configuración de un modelo ANN, entrenada con biomarcadores genéticos asociados con la respuesta al IFN- $\beta$ , con el objetivo de estimar si nuevos pacientes son candidatos potenciales para ser tratados con este fármaco y obtener un mayor rendimiento de predicción que los trabajos previos.

## 5. MATERIAL Y MÉTODOS

### 5.1 Importación de datos

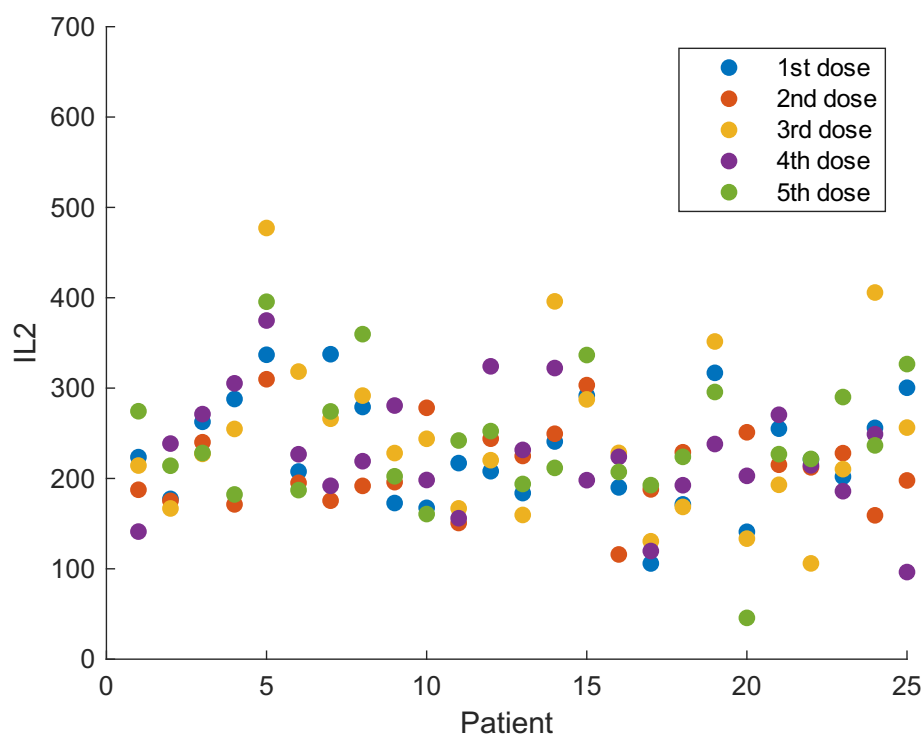
El conjunto de datos de entrenamiento y validación se recopiló del experimento de perfiles de expresión por matriz GSE24427, disponible en el repositorio Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>). Mediante el uso de la plataforma de micro matrices GPL96 [HG-U133A] (*Affymetrix Human Genome U133A Array*), se obtuvieron los perfiles de expresión genética en suero de 25 pacientes alemanes con RRMS que fueron tratados con IFN- $\beta$  – 1b subcutáneo (*Bayer Healthcare*), con dosis de 250  $\mu$ g por día durante 2 años. Las muestras de sangre de los pacientes se extrajeron antes de la primer inyección de IFN- $\beta$  – 1b, la segunda, y de los meses 1, 12 y 24. Los valores de expresión fueron analizados por GEO2R, una herramienta web interactiva que permite visualizar un valor de expresión genética específica a través de la función *Profile graph*. Por un lado, la plataforma GPL96 muestra las características demográficas y clínicas de los pacientes con RRMS, que se utilizan como variables de entrada para el sistema experto difuso propuesto, las cuales se presentan en la Tabla 5-1.

Por otro lado, se recopilaron los valores de expresión de 13 biomarcadores asociados con la respuesta al IFN- $\beta$  (IL-2, IFN- $\gamma$ , TNF- $\alpha$ , IL-4, IL-10, TGF- $\beta$ , CD46, CD58, FHIT, IRF5, GAPVD1, GRM3, GRIK2) y se integraron en una hoja de cálculo de Excel para entrenar al modelo ANN propuesto. Por ejemplo, los valores de expresión de las citocinas IL-2 e IL-4 se muestran en las Figuras 5.1 y 5.2.

### 5.2 Metodología

La estrategia de investigación implementada se describe en el diagrama de flujo de la Figura 5.3, que divide la propuesta en cinco etapas.

**Figura 5.1** Citocina IL-2. Valores de expresión de 25 pacientes con RRMS correspondientes a cinco dosis: antes de la primer inyección de IFN- $\beta$  – 1b, segunda, y los meses 1, 12, 24.



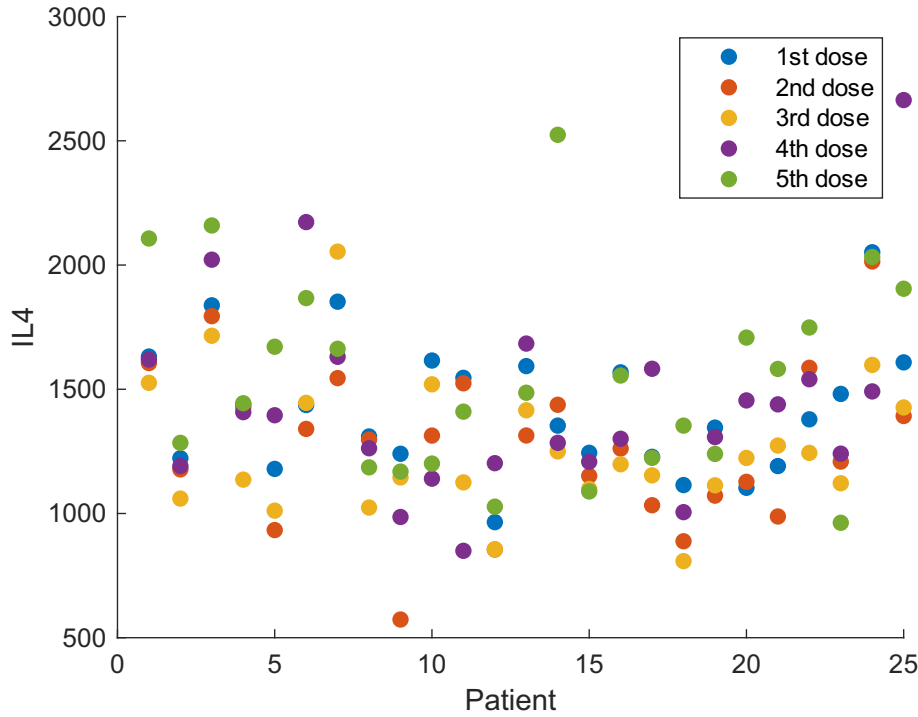
### 5.3 Asignación de etiquetas de clase

Como el conjunto de datos de entrenamiento no contiene información de las salidas, se utiliza un sistema de lógica difusa para asignar las etiquetas de clase no conocidas, es decir, para clasificar a los pacientes con RRMS como respondedor (alto, medio y bajo) al tratamiento con IFN- $\beta$  – 1b, en base a sus características demográficas y clínicas. También se utiliza un algoritmo de agrupamiento jerárquico para la clasificación de la respuesta al IFN- $\beta$  – 1b por motivos de comparación.

#### 5.3.1 Preprocesamiento de datos

Esta etapa de gestión de datos incluye la aplicación de alguna técnica de análisis de datos que permita mejorar la calidad del conjunto de datos de modo que las variables de interés se puedan comparar adecuadamente. La estandarización o normalización es una técnica de análisis de datos comúnmente utilizada por muchos

**Figura 5.2** Citocina IL-4. Los valores de expresión parecen más dispersos que los de la citocina IL-2.



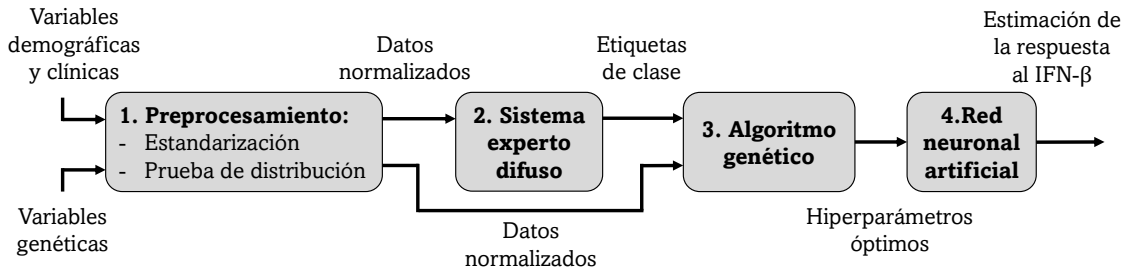
estimadores de ML, ya que es posible que no se comporten bien si las características individuales no se parecen a los datos estándar distribuidos normalmente.

### Estandarización

Para preprocesar los datos de entrada, primero se utiliza la técnica de estandarización para normalizar las variables demográficas, clínicas y genéticas de la base de datos de estudio. Este método normaliza las características eliminando la media  $\mu$ , y las escala dividiendo por su desviación estándar  $s$  (Casalino et al., 2021). La puntuación estándar de una muestra  $x$  se calcula como sigue:

$$z = \frac{x - \mu}{s} \quad (5.1)$$

**Figura 5.3 Metodología propuesta:** 1) Se recopilan las características demográficas, clínicas, y genéticas de pacientes con RRMS tratados con IFN- $\beta$  – 1b durante 2 años y se preprocesan los datos aplicando la técnica de estandarización y la prueba de KS. 2) Se ingresan las variables demográficas y clínicas a un sistema experto difuso para obtener las etiquetas de clase no conocidas. 3) Se ingresan las variables genéticas con las salidas etiquetadas a un GA que busque los hiperparámetros de configuración óptimos para entrenar a un modelo ANN que realice estimaciones de la respuesta al IFN- $\beta$  – 1b.



### Prueba de Kolmogorov-Smirnov (KS)

Antes de ingresar los datos estandarizados al sistema difuso y al GA, es conveniente evaluar que las variables de entrada se distribuyen normalmente. La prueba de KS es una prueba no paramétrica que se puede utilizar para comprobar que una variable se distribuye de forma normal (Robert, 2021). El procedimiento de KS para una muestra compara la función de distribución acumulada observada de una variable con una distribución teórica determinada, que puede ser la normal, la uniforme, la de Poisson o la exponencial. La salida de KS se calcula a partir de la diferencia mayor (en valor absoluto) entre las funciones de distribución acumuladas teórica y observada. Esta prueba de bondad de ajuste contrasta si las observaciones podrían razonablemente proceder de la distribución especificada.

La función de distribución empírica  $F_n$  para  $n$  observaciones ordenadas independientes e idénticamente distribuidas  $X_i$  se define como

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, x]} X_i \quad (5.2)$$

donde  $1_{(-\infty, x]} X_i$  es la función indicadora, que es igual a 1 si  $X_i \leq x$  o igual a cero de lo contrario.

La estadística de KS para una función de distribución acumulativa dada  $F(x)$  es

$$D_n = \sup_x |F_n(x) - F(x)| \quad (5.3)$$

donde  $\sup_x$  es el supremo del conjunto de distancias. Intuitivamente, la estadística toma la mayor diferencia absoluta entre las dos funciones de distribución en todos los valores de  $x$ . La prueba de KS también se puede usar para probar si dos distribuciones de probabilidad unidimensionales subyacentes difieren. En este caso, el estadístico de KS es

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)| \quad (5.4)$$

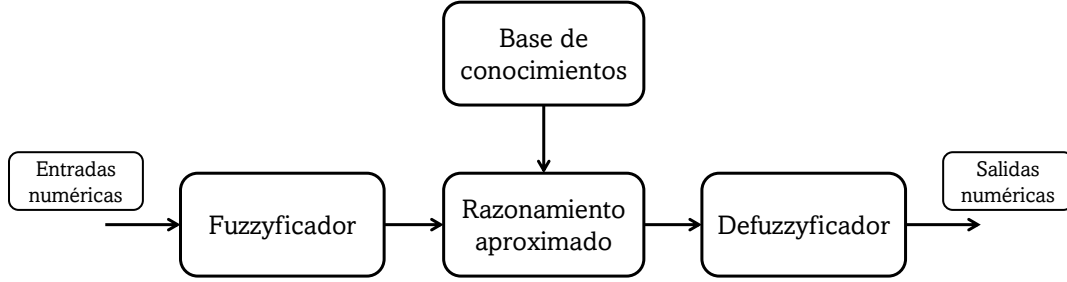
donde  $F_{1,n}(x)$  y  $F_{2,m}(x)$  son las funciones de distribución empírica de la primera y la segunda muestra respectivamente.

### 5.3.2 Sistema experto de lógica difusa

La principal inspiración detrás de la introducción de la teoría de conjuntos difusos fue la necesidad de modelar fenómenos del mundo real, que son inherentemente vagos y ambiguos. El conocimiento humano sobre problemas complejos se puede representar con éxito utilizando los términos imprecisos del lenguaje natural (Czabanski et al., 2017). Los sistemas difusos son estructuras basadas en las teorías de conjuntos difusos y de lógica difusa para el procesamiento eficiente de información inexacta (Rutkowska, 2001). Su propiedad principal incluye la representación simbólica del conocimiento en forma de reglas condicionales difusas (*if-then*). La estructura típica de un sistema difuso se describe en la Figura 5.4.

El fuzzyficador transforma los valores de las variables de entrada en un conjunto difuso N-dimensional  $A$  (valores lingüísticos de la variable de salida) definido

**Figura 5.4** Estructura de un sistema difuso, que consiste de tres etapas: 1) Fuzzyficador, 2) Razonamiento aproximado y 3) Defuzzyficador.



en el universo  $\mathbb{X}$ , mediante razonamiento aproximado (motor de inferencia) utilizando conocimiento experto, que se representa como un conjunto de reglas condicionales difusas (base de conocimientos). El fuzzyficador se puede definir como la función de membresía  $\mu_A(x)$  del conjunto difuso  $A$ . Las características demográficas y clínicas de los pacientes con RRMS se utilizan como variables de entrada para el fuzzyficador. El resultado del razonamiento aproximado es un conjunto difuso  $B(y)$ . El defuzzyficador calcula una salida numérica representativa  $y_0$  a partir del resultado del conjunto difuso  $B(y)$  definido en el universo  $\mathbb{Y}$ . La salida numérica  $y_0$  se calcula utilizando el método centro de gravedad (COG) (Van Leekwijck and Kerre, 1999), como en la Ecuación 5.5,

$$y_0 = \frac{\sum_{i=1}^n y_i \mu_B(y_i)}{\sum_{i=1}^n \mu_B(y_i)}, \quad (5.5)$$

donde  $\mu_B(y)$  representa la función de membresía del conjunto difuso  $B(y)$ . El sistema difuso propuesto es diseñado a través de la aplicación *Fuzzy Logic Designer* del software MATLAB R2023a. La estructura del sistema difuso propuesto está basado en el sistema difuso Mamdani-Assilan (MAFS) (Prokopowicz et al., 2017), que incluye un conjunto de reglas difusas condicionales de la forma de la Ecuación 5.6, que pueden ser determinadas por un experto humano,

$$\mathcal{R} = \{R^i\}_{i=1}^N = \{if \text{ and }_{n=1}^N (X_n \text{ is } L_{A_n}^{(i)}), \text{ then } Y \text{ is } L_B^{(i)}\}_{i=1}^I, \quad (5.6)$$



donde  $I$  es el número de reglas,  $X_n$  representa las variables lingüísticas de entrada,  $Y$  es la variable lingüística de salida y  $L_{A_n}$  y  $L_B$  son los valores lingüísticos, definidos por los conjuntos difusos  $A_N$  y  $B$  en los universos  $\mathbb{X}_N$  y  $\mathbb{Y}$ , respectivamente.

Para la implementación del sistema difuso, se definen las variables lingüísticas de entrada que describen las características demográficas y clínicas de la Tabla 5-1:  $\mathcal{N}_1 = "\mu$  edad";  $\mathcal{N}_2 = "\mu$  EDSS al inicio";  $\mathcal{N}_3 = "\mu$  EDSS después de un año";  $\mathcal{N}_4 = "\mu$  EDSS después de dos años". Los conjuntos de posibles valores lingüísticos son colecciones de diferentes etiquetas que describen las entradas edad, EDSS al inicio, EDSS después de un año y EDSS después de dos años:  $L_{A_1} = \{\text{"infantil", "adulto", "anciano"}\}$ ;  $L_{A_2} = \{\text{"bajo", "medio", "alto"}\}$ ;  $L_{A_3} = \{\text{"bajo", "medio", "alto"}\}$ ;  $L_{A_4} = \{\text{"bajo", "medio", "alto"}\}$ , y la salida que define la respuesta al IFN- $\beta 1b$ :  $L_B = \{\text{"bajo", "medio", "alto"}\}$ . A cada una de las etiquetas se le asignan los conjuntos difusos  $A_N^{(i)}$ , definidos en el universo  $\mathbb{X}_N$ , que representa el rango de valores posibles. La descripción completa de las variables lingüísticas definidas se presenta en la Tabla 5-2.

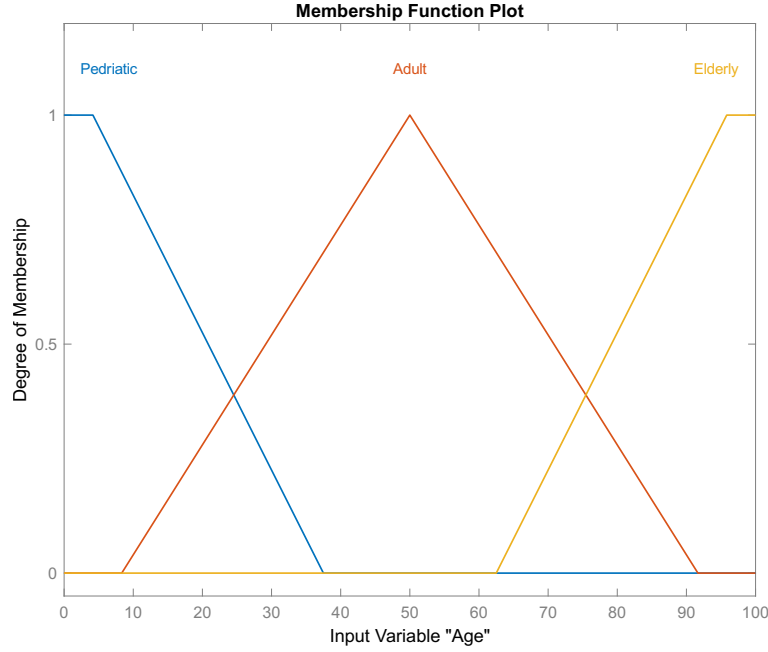
Por ejemplo, las gráficas de las funciones de membresía  $\mu_{A_1^{(1)}}(\text{edad})$ ,  $\mu_{A_4^{(1)}}(\text{EDSS después de dos años})$ , y  $\mu_B(\text{Respuesta al IFN-}\beta 1b)$  de los conjuntos difusos  $A_1^{(1)}$ ,  $A_4^{(1)}$ , y  $B(y)$  respectivamente se muestran en las Figuras 5.5–5.7.

Las reglas condicionales difusas (base de conocimientos) están destinadas a decidir la influencia de las variables de entrada en la respuesta al tratamiento con IFN- $\beta 1b$ . Las tablas 5-3, 5-4 y 5-5 muestran las 36 reglas definidas, en base a la opinión de un experto en neurología.

### 5.3.3 Agrupamiento jerárquico

La agrupación jerárquica es una familia general de algoritmos de ML no supervisado que construyen agrupaciones anidadas fusionándolas o dividiéndolas sucesivamente. Esta jerarquía de agrupamientos se representa como un árbol (o dendrograma). La raíz del árbol es el único grupo que reúne todas las muestras, siendo las hojas los grupos con una sola muestra. Este tipo de algoritmos realizan una agrupación jerárquica utilizando un enfoque de abajo hacia arriba: cada

**Figura 5.5** Conjunto de valores lingüísticos correspondientes a tres etiquetas que describen la variable de entrada "Edad" (conjunto difuso  $A_1^{(1)}$ ).



observación comienza en su propio grupo y los grupos se fusionan sucesivamente (Ackermann et al., 2014). Los criterios de vinculación determinan la métrica utilizada para la estrategia de fusión:

1. *Ward* minimiza la suma de diferencias al cuadrado dentro de todos los grupos. Es un enfoque que minimiza la varianza y en este sentido es similar a la función objetivo *K – Means* pero abordado con un enfoque jerárquico.

$$\frac{|\mathcal{A}| \cdot |\mathcal{B}|}{|\mathcal{A} \cup \mathcal{B}|} \|\mu_A - \mu_B\|^2 \quad (5.7)$$

2. El vínculo máximo o completo minimiza la distancia máxima entre elementos de cada grupo.

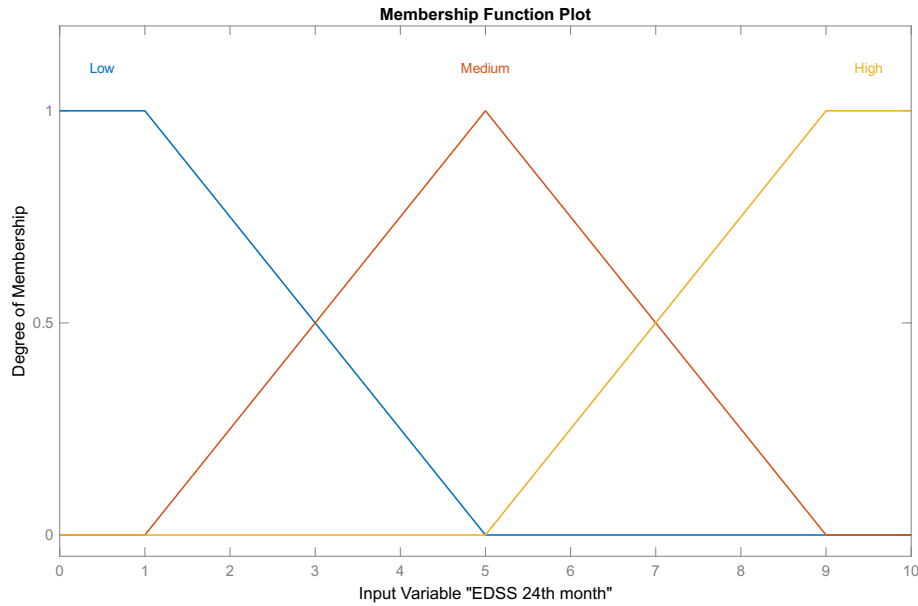
$$\max\{d(x, y) : x \in \mathcal{A}, y \in \mathcal{B}\} \quad (5.8)$$

3. El vínculo promedio minimiza la distancia promedio entre elementos de cada grupo.

$$\frac{1}{|\mathcal{A}| \cdot |\mathcal{B}|} \sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} d(x, y) \quad (5.9)$$

4. El vínculo único minimiza la distancia mínima entre los elementos de cada

**Figura 5.6** Conjunto de valores lingüísticos correspondientes a tres etiquetas que describen la variable de entrada "EDSS después de dos años" (conjunto difuso  $A_4^{(1)}$ ).



grupo.

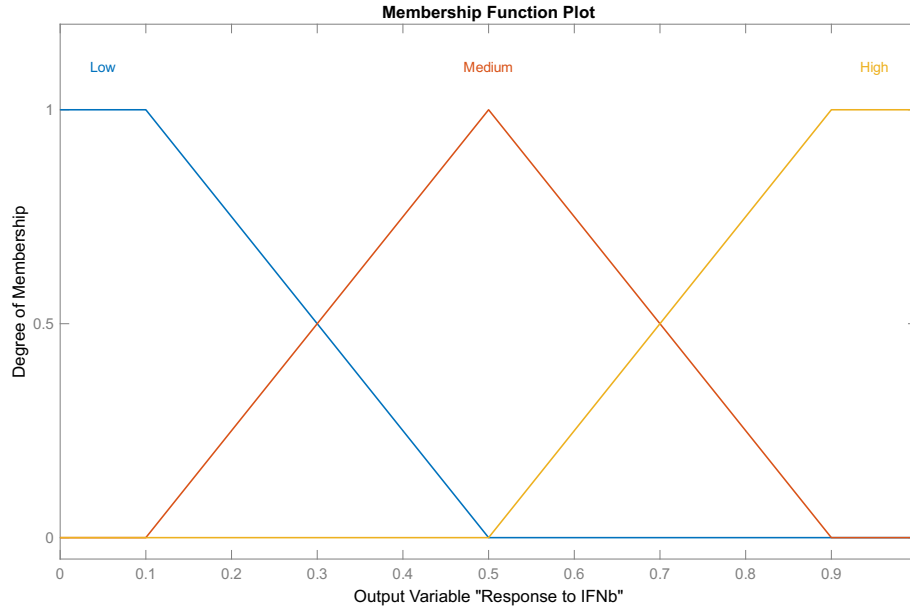
$$\min\{d(x, y) : x \in \mathcal{A}, y \in \mathcal{B}\} \quad (5.10)$$

El agrupamiento jerárquico también puede escalar a una gran cantidad de muestras cuando se usa junto con una matriz de conectividad, pero es computacionalmente costoso cuando no se agregan restricciones de conectividad entre muestras: considera en cada paso todas las fusiones posibles.

#### 5.4 Técnicas convencionales de ML

Una vez que se han asignado las etiquetas de clase no conocidas del conjunto de datos por medio del sistema difuso, se procede a entrenar algunos modelos de ML convencionales y un modelo ANN con las características genéticas previamente estandarizadas para realizar estimaciones de la respuesta al IFN- $\beta - 1b$  en pacientes con RRMS y comparar el rendimiento de los modelos. Para este propósito se utilizan el software Anaconda3 2021.05 (Python 3.8.8 de 64 bits) y la aplicación de internet de desarrollo de código abierto Jupyter Notebook para generar los pseudocódigos que se ejecutan en un procesador Intel core i5-1135G7 de 2,4 GHz.

**Figura 5.7** Conjunto de valores lingüísticos correspondientes a tres etiquetas que describen la variable de salida "Respuesta al IFN- $\beta$ " (conjunto difuso  $B(y)$ ).



Los hiperparámetros son configuraciones de un modelo de aprendizaje que se pueden sintonizar arbitrariamente antes de comenzar el proceso de entrenamiento para optimizar el rendimiento del modelo, por ejemplo, en los algoritmos basados en RF, el número de estimadores (número de árboles de decisión) y el criterio o medida de impureza. Por el contrario, los parámetros del modelo, como los pesos  $w$  en las redes neuronales, se aprenden durante el proceso de entrenamiento del modelo (Villegas-Mier et al., 2022).

#### 5.4.1 K-Vecinos

En el reconocimiento de patrones, el algoritmo KN es un método no paramétrico utilizado para clasificación y regresión. En ambos casos, la entrada consta de los  $k$  ejemplos de entrenamiento más cercanos en el espacio de características. Para clasificación, un objeto se clasifica por mayoría de votos de sus vecinos, con el objetivo de ser asignado a la clase más común entre sus  $k$ -vecinos ( $k$  es un número entero positivo, típicamente pequeño) (Mohammed et al., 2016). Para determinar el KN de un punto  $x_s$ , se utiliza una medida de distancia para determinar los puntos más cercanos  $k$  al punto  $x_s : \{x_{s1}, x_{s2}, \dots, x_{sk}\}$ . Asumiendo que  $d(x_s, x_j), j = 1, \dots, N$ , mide la distancia entre  $x_s$  y  $x_j$ , y  $x_{si} : i = 1, \dots, k$  es el

conjunto de  $k$ -vecinos de  $x_s$  de acuerdo con la distancia métrica  $d$ . Entonces, la aproximación de la Ecuación 5.11,

$$f(x_s) = \frac{1}{k} \sum_{j=1}^k f(x_{sj}) \quad (5.11)$$

asume que todos los puntos  $k$ -vecinos tienen la misma contribución a la clasificación del punto  $x_s$ . Se pueden utilizar varias métricas para determinar la distancia  $d$ , como la distancia Euclidiana, Minkowski y Manhattan.

#### 5.4.2 Modelos lineales

La regresión logística (LR) es uno de los modelos lineales más utilizados para clasificación. Para describir la LR como modelo probabilístico, se define la razón de probabilidades como las probabilidades de que ocurra un evento concreto. La razón de probabilidades se puede escribir como  $\frac{p}{1-p}$ , donde  $p$  es la probabilidad del evento a estimar (Mirjalili, 2020). Para estimar la probabilidad de que una determinada muestra pertenezca a una clase específica, se utiliza la función logística o *logit*, también llamada función *sigmoid*, la cual se define como:

$$\phi(z) = \frac{1}{1 + e^{-z}} \quad (5.12)$$

donde  $z = w^T x$  es la entrada de la red,  $w$  la combinación lineal de pesos y  $x$  las características de la muestra. Esta función toma valores de números reales como entrada y los transforma en valores del rango  $[0, 1]$ . La salida de la función *sigmoid* se interpreta como la probabilidad de que una muestra particular pertenezca a la clase  $i$ ,  $\phi(z) = p(y = i \mid x; w)$ , dadas sus características  $x$  parametrizadas por los pesos  $w$ . Las regularizaciones  $l1$  y  $l2$  se pueden utilizar para evitar el sobre ajuste. El parámetro  $C$  representa el inverso de la fuerza de la regularización, los valores más pequeños especifican una regularización más fuerte (Sarker, 2021).

### 5.4.3 Support Vector Machine

SVM es un clasificador binario que calcula el hiperplano óptimo que separa diferentes clases de datos. Este algoritmo utiliza funciones *kernel* no lineales para transformar los datos de entrada en un espacio de características de mayor dimensión, en el que los datos se vuelven más separables (Montolío et al., 2021). Para clases separables, el hiperplano óptimo maximiza el espacio que no contiene observaciones. Para clases no separables, el objetivo es el mismo, pero el algoritmo impone una penalización por cada observación que esté en el lado equivocado, lo que ayuda a evitar el sobre ajuste (Sarker, 2021). Intuitivamente, el hiperplano que tiene la mayor distancia a los puntos de datos de entrenamiento más cercanos de cualquier clase logra una buena separación (el llamado margen funcional), ya que en general, cuanto mayor es el margen, menor es el error de generalización del clasificador.

El objetivo de optimización de SVM es maximizar el margen funcional. El margen se define como la distancia entre el hiperplano de separación (límite de decisión) y las muestras de entrenamiento que están más cerca de ese hiperplano, que también se denominan vectores de soporte (Mirjalili, 2020). Una de las funciones *kernel* más ampliamente utilizadas es la función de base radial (RBF), también denominada *kernel* Gaussiana, que se define como:

$$K(x^{(i)}, x^{(j)}) = \exp(-\gamma \|x^{(i)} - x^{(j)}\|^2), \quad (5.13)$$

para dos puntos explícitamente  $x^{(i)T}, x^{(j)}$ . En este caso,  $\gamma$  conocido como el coeficiente *kernel*, es un parámetro libre que debe ser optimizado.

### 5.4.4 Decision Tree

Los DT son modelos secuenciales que combinan lógicamente una secuencia de pruebas simples; cada prueba compara un atributo numérico con un valor umbral o un atributo nominal con un conjunto de valores posibles. El algoritmo intenta generalizar o encontrar patrones en los datos, determinando qué pruebas dividen

mejor las instancias en clases separadas, formando un árbol. Este procedimiento puede concebirse como una búsqueda exhaustiva a través del espacio de todos los árboles de decisión posibles, escaneando las instancias en un nodo dado para determinar la ganancia de cada división y seleccionando la división única que proporcione la mayor ganancia de información (IG) (Kotsiantis, 2013). Con el fin de dividir los nodos en las características más informativas, se define una función objetivo que se pretende optimizar mediante el algoritmo de aprendizaje del árbol. El propósito de la función objetivo, definida en la Ecuación 5.14, es maximizar la ganancia de información en cada división,

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^m \frac{N_j}{N_p} I(D_j), \quad (5.14)$$

donde  $f$  es la característica para realizar la división,  $D_p$  y  $D_j$  el conjunto de datos de nodo padre y del nodo hijo,  $I$  la medida de impureza,  $N_p$  el número de muestras del nodo padre y  $N_j$  el número de muestras del nodo hijo. Luego, las instancias se particionan según la división y este procedimiento se aplica de forma recursiva hasta que todas las instancias de un nodo sean de la misma clase. En cada iteración, el algoritmo considera la partición del conjunto de entrenamiento utilizando el resultado de una función discreta de los atributos de entrada. La selección de la función más apropiada se realiza de acuerdo con algunas medidas de impurezas o criterios de división como la impureza de *Gini*, la entropía y el error de clasificación (Mirjalili, 2020).

## 5.5 Modelos de *Deep Learning*

Los modelos de DL han demostrado ser superiores a otros modelos de ML clásicos debido a su capacidad para aprender relaciones altamente complejas entre los datos de entrada y salida. Por lo general, en un modelo de ML clásico basado en aproximación, la ingeniería de características juega un papel crucial en el rendimiento general de un modelo entrenado. Sin embargo, un modelo de DL elimina la necesidad de crear características de forma manual. Con grandes cantidades de datos, los modelos de DL pueden funcionar muy bien sin requerir de características

diseñadas a mano y pueden superar a los modelos de ML tradicionales (Jha, 2021).

### 5.5.1 Redes Neuronales Artificiales

En el núcleo del DL se encuentran las ANN que son entidades matemáticas capaces de representar funciones complejas a través de una composición de funciones más simples. El término red neuronal sugiere un vínculo con la forma en que funciona nuestro cerebro. El componente básico de estas funciones complejas es la neurona, que es una transformación lineal de la entrada (por ejemplo, multiplicando la entrada por un número [el peso] y sumando una constante [el sesgo o *bias*]) seguido por la aplicación de una función no lineal fija, o función de activación (Stevens et al., 2020).

Matemáticamente, podemos escribir la salida como  $o = f(w * x + b)$ , con  $x$  como la entrada,  $w$  como el peso o factor de escalamiento, y  $b$  como el sesgo o compensación.  $f$  es la función de activación, establecida comúnmente en la tangente hiperbólica ( $\tanh$ ). En general,  $x$  y, por lo tanto,  $o$  pueden ser simples escalares, o vectores de escalares; y  $w$  puede ser un solo escalar o matriz, mientras que  $b$  es un escalar o vector (la dimensionalidad de las entradas y los pesos deben coincidir). En este último caso, la expresión anterior se conoce como una capa de neuronas, ya que representa muchas neuronas a través de los pesos y sesgos multidimensionales. Cuando el número de dichas capas es más de dos, el sistema neuronal se denomina red neuronal profunda (Jha, 2021).

Una red neuronal profunda o multicapa es una composición de funciones no lineales,

$$x_1 = f(w_0 * x + b_0) \quad (5.15)$$

$$x_2 = f(w_1 * x_1 + b_1) \quad (5.16)$$

$$\dots \quad (5.17)$$

$$y = f(w_n * x_n + b_n) \quad (5.18)$$

donde la salida de una capa de neuronas se utiliza como entrada para la



siguiente capa,  $w_0$  aquí es una matriz y  $x$  es un vector. El uso de un vector permite a  $w_0$  contener toda una capa de neuronas, no solo un peso único.

El *overfitting* es un problema común en ML y DL, donde un modelo funciona bien con los datos de entrenamiento pero no con los datos de validación, es decir, el modelo es demasiado complejo con una varianza alta (Mirjalili, 2020). Para evitar el *overfitting*, los datos de entrada se dividen en 80 % para entrenamiento ( $X_{train}$ ) y 20 % para validación ( $X_{test}$ ), basado en análisis de Pareto (Rocchetti et al., 2021). Además, las etiquetas de salida se dividen igualmente en 80 % para entrenamiento  $y_{train}$  y 20 % para validación  $y_{test}$ .

### 5.5.2 Perceptron Multi-capa

perceptrón multicapa (MLP) es un algoritmo de aprendizaje supervisado que utiliza la técnica de retropropagación para el aprendizaje. La estructura del MLP consta de una capa de entrada de neuronas que reciben las entradas de muestra  $X = x_1, x_2, \dots, x_m$ , una o más capas ocultas de neuronas que convierten los valores de la capa anterior en una suma lineal ponderada,  $w_1x_1 + w_2x_2 + \dots + w_mx_m$ , seguida de una función de activación no lineal que se utiliza para aprender los pesos, y luego la capa de salida que predice la etiqueta de clase de las muestras (Casalino et al., 2021). Durante la etapa de aprendizaje, MLP compara las etiquetas de clase verdaderas con los valores de salida continuos de la función de activación no lineal, para calcular el error de predicción y actualizar los pesos. Algunos hiperparámetros del modelo MLP son: *hidden\_layer\_sizes* = (*num\_neuronas\_ocultas*, *num\_capas*), *activation* = 'funcion\_activacion', *solver* = 'solucionador' para la optimización de los pesos, etc.

### 5.5.3 Esquema de optimización

Para entrenar a una red neuronal, es necesario adoptar un esquema de optimización. Como cualquier otro modelo de ML basado en parámetros, un modelo de DL es entrenado por la sintonización de sus parámetros. Los parámetros se sintonizan mediante el proceso de *backpropagation*, en el que la capa final o de

salida de la red neuronal produce una pérdida. Esta pérdida se calcula con la ayuda de una función de pérdida como la suma de errores al cuadrado (SSE) que toma en cuenta los valores objetivo  $y^{(i)}$  y las salidas  $\phi(z^{(i)})$  de la capa final de la red neuronal,

$$J(w) = \frac{1}{2} \sum_i (y^{(i)} - \phi(z^{(i)}))^2, \quad (5.19)$$

donde  $z = w^T x$  es la entrada de la red (combinación lineal de los parámetros  $w$  o pesos y los datos de entrada  $x$ ). Esta pérdida luego se propaga hacia atrás a las capas anteriores mediante el descenso de gradiente y la regla de la cadena de diferenciación. Los parámetros  $w$  en cada capa se modifican en consecuencia para minimizar la pérdida. El alcance de la modificación está determinado por la tasa de aprendizaje, que varía de 0 a 1 (Jha, 2021). Todo este procedimiento de actualización de los pesos de una red neuronal es llamado esquema de optimización y tiene un impacto significativo en qué tan bien se entrena a un modelo de aprendizaje. Los siguientes son algunos esquemas de optimización populares.

El descenso de gradiente estocástico (SGD) actualiza los parámetros  $w$  del modelo de la siguiente forma,

$$w = w - \eta * \frac{\partial J(x, y, w)}{\partial w} \quad (5.20)$$

donde  $x$  son los datos de entrenamiento,  $y$  las etiquetas correspondientes,  $J$  es la función de pérdida y  $\eta$  es la tasa de aprendizaje. SGD realiza esta actualización para cada par de ejemplos de entrenamiento  $(x, y)$ . Una variante de este descenso de gradiente por mini lotes realiza actualizaciones para cada  $k$  ejemplos, donde  $k$  es el tamaño del lote. Los gradientes se calculan en conjunto para todo el mini lote.

El estimación del momento adaptativo (Adam) es otro esquema de optimización que calcula tasas de aprendizaje personalizadas para cada parámetro. Adam utiliza el promedio decreciente de los gradientes cuadrados anteriores y también el promedio decreciente de los valores de gradiente anteriores,

$$SG_i^t = \gamma' * SG_i^{t-1} + (1 - \gamma') * \frac{\partial J(x, y, w)}{\partial w_i^t} \quad (5.21)$$

donde  $\gamma$  es el factor de decrecimiento que se desea elegir para la suma anterior de gradientes cuadrados. Con esta formulación, se asegura que la suma de gradientes al cuadrado no se acumule a un valor grande, gracias al promedio decreciente.

#### 5.5.4 Optimización de hiperparámetros con Algoritmo Genético (GA)

GA es un algoritmo metaheurístico basado en poblaciones que combina selección natural y genética. GA busca la solución óptima en el espacio de búsqueda incluyendo aleatoriedad. GA es un algoritmo iterativo que converge en un número finito de iteraciones (generaciones), y las soluciones candidatas evolucionan hacia una mayor aptitud (*fitness*) o calidad en cada iteración. El algoritmo termina cuando se alcanza la solución óptima (convergencia), cuando se completa el número máximo de iteraciones o se alcanza una condición de parada. Las características importantes de un GA incluyen la selección inicial de la población, la definición de un objetivo o función de aptitud, operadores genéticos y criterios de terminación (Vasuki, 2020).

El algoritmo comienza definiendo el problema y su función objetivo  $f(X)$  donde  $X$  es un vector multidimensional, con una dimensión típica  $d$ . La población inicial se elige aleatoriamente en el espacio de búsqueda y los miembros se codifican como un cromosoma en forma de una cadena de caracteres. El cromosoma se compone de genes. Los valores que toman los genes son los alelos, que pueden ser binarios sobre el alfabeto binario  $\{0,1\}$  o no binarios como el alfabeto decimal  $\{0,1,\dots,9\}$  o cualquier otro conjunto de alfabetos definidos. Las operaciones genéticas de selección, cruzamiento y mutación se aplican repetidamente sobre la población hasta que se alcanza el criterio de terminación. Al final de cada iteración, se calculan los valores de aptitud de la población. Los miembros con mayor aptitud se seleccionan para el apareamiento y la reproducción. Finalmente, el miembro con el valor de aptitud más alto es la solución óptima al problema.

En este trabajo de investigación se utilizan las variables genéticas asociadas con la respuesta al IFN- $\beta 1b$  como entradas de entrenamiento para configurar a un modelo ANN, cuya arquitectura esta basada en la configuración de hiperparáme-

tros. Para encontrar los hiperparámetros óptimos se propone el uso de un GA. Este algoritmo comienza definiendo el problema (hiperparámetros óptimos) y su función objetivo  $f(X)$  (máxima precisión de validación). En este caso,  $X$  representa las siguientes opciones de configuración de hiperparámetros:

1. 'num\_neuronas\_ocultas': [64, 128, 256, 512, 768, 1024],
2. 'num\_capas\_ocultas': [1, 2, 3, 4],
3. 'func\_activacion': ['relu', 'elu', 'tanh', 'sigmoid'],
4. 'optimizador': ['rmsprop', 'adam', 'sgd', 'adagrad', 'adadelat', 'adamax', 'nadam'].

La Tabla 5-6 presenta el pseudocódigo para la implementación del GA. El cromosoma (parámetros elegidos) con mayor aptitud representa la solución óptima global.

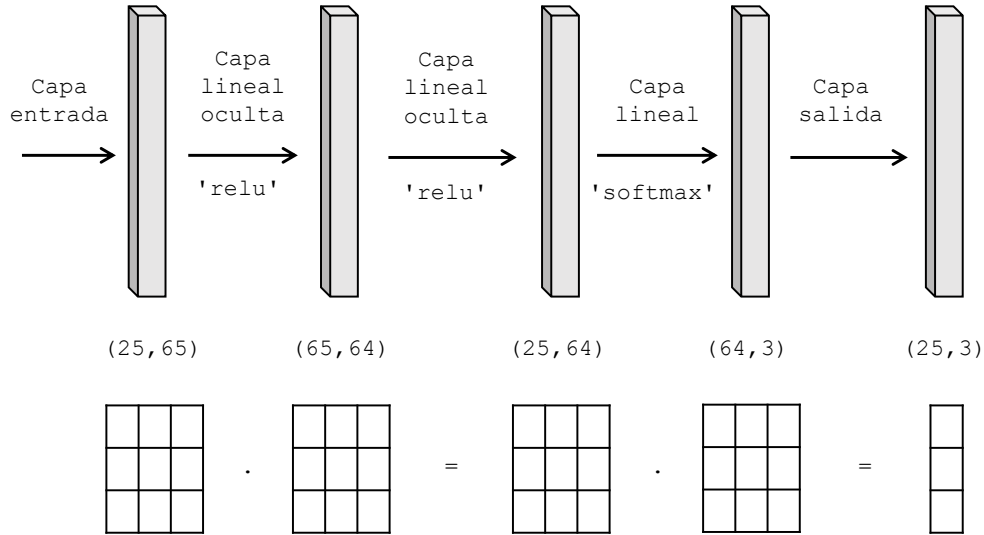
La configuración de la ANN propuesta está definida por los hiperparámetros optimizados por el GA. La arquitectura de la ANN se presenta en la Figura 5.8, que incluye una capa de entrada, donde 25 es el número de muestras (pacientes con RRMS) y 65 el total de neuronas de entrada (características genéticas); dos capas ocultas con función de activación *relu*, donde 64 es el número de neuronas ocultas; la tercer capa recibe los valores de la segunda capa oculta y los transforma en predicciones para la capa de salida (respondedor alto, medio, o bajo al IFN- $\beta 1b$ ), utilizando el optimizador de la función de pérdida *adamax*.

#### 5.5.5 Métricas de rendimiento

Es importante determinar qué tan efectivos son los modelos de aprendizaje por medio del cálculo de algunas métricas comúnmente utilizadas para evaluar su rendimiento predictivo (Goyal et al., 2019; Mirjalili, 2020; Salamai et al., 2021).

La matriz de confusión (CM) es una matriz que reporta el conteo de las predicciones: positivos verdaderos (TP), negativos verdaderos (TN), positivos falsos (FP) y negativos falsos (FN) de un modelo de predicción. Por ejemplo, la Figura

**Figura 5.8** Arquitectura de la ANN propuesta. La capa lineal aplica una transformación lineal a los datos de entrada:  $y = xA^T + b$ , donde  $A$  es la matriz de pesos y  $b$  es el vector de compensación.



5.9 muestra la CM con las tres etiquetas de clase asignadas por el sistema difuso propuesto.

La precisión (A) es el número de predicciones correctas como proporción de todas las predicciones realizadas y se puede calcular como,

$$A = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TH + TL + TM}{TH + TL + TM + FH + FL + FM}. \quad (5.22)$$

La sensibilidad o tasa de positivos verdaderos (TPR) mide la proporción de predicciones positivas que se identifican correctamente, se puede calcular mediante,

$$TPR = \frac{TP}{TP + FN} = \frac{TH}{TH + FH}. \quad (5.23)$$

La especificidad o tasa de negativos verdaderos (TNR) mide la proporción de predicciones negativas que se identifican correctamente, se puede calcular como,

**Figura 5.9** Diagrama de matriz de confusión para las etiquetas de clase: Responder alto (0), respondedor bajo (1) y respondedor medio (2) al IFN- $\beta$ 1b.

True Label	0	True High ( TH )	False Low ( FL )	False Medium ( FM )
	1	False High ( FH )	True Low ( TL )	False Medium ( FM )
	2	False High ( FH )	False Low ( FL )	True Medium ( TM )
		0	1	2
		Predicted Label		

$$TNR = \frac{TN}{TN + FP} = \frac{TL + TM}{(TL + TM) + (FL + FM)}. \quad (5.24)$$

El AUC indica el desempeño de la predicción entre la distribución de clases. El AUC se puede calcular directamente como el promedio de sensibilidad y especificidad. La precisión equilibrada o valor AUC se expresa matemáticamente como,

$$AUC = \frac{TPR + TNR}{2}. \quad (5.25)$$

**Tabla 5-1** *Características demográficas y clínicas de pacientes con RRMS tratados con IFN- $\beta$  – 1b durante dos años.*

Muestra	Edad	EDSS <sup>1</sup> al inicio	EDSS <sup>1</sup> después de 1 año	EDSS <sup>1</sup> después de 2 años
1	63	4	4.5	5.5
2	45	1	1	1
3	25	1	1	1
4	27	4	4	3.5
5	51	3	2.5	2.5
6	41	2	2	4.5
7	44	4	3	3
8	30	1.5	2	2
9	26	4	3.5	3.5
10	42	1	1	1
11	29	2	2	2.5
12	28	1.5	1	2.5
13	48	1	1	1
14	47	3.5	3.5	3
15	42	2	2.5	3
16	50	3.5	3.5	3.5
17	37	1.5	3.0	4.5
18	43	2	1.5	2
19	54	3	2	2
20	40	1	1	1
21	48	2	2	2
22	38	2	3.5	3
23	18	1.5	1.5	2
24	24	1	1	1
25	38	1	1	1

<sup>1</sup> escala ampliada del estado de discapacidad (EDSS).

**Tabla 5-2** Descripción de las variables lingüísticas (características demográficas, clínicas y la respuesta al IFN- $\beta 1b$ ).

Función de Membresía	Conjunto Difuso	Universo de Discurso	Parámetros y Tipo
$\mu_{A_1^{(1)}}(\text{Edad})$	$A_1^{(1)}$	$\mathbb{X}_1$ : [0 a 100] años	Infantil: [−37.5; −4.167; 4.167; 37.5] Trapezoidal Adulto: [8.333; 50; 91.666] Triangular Anciano: [62.5; 95.83; 104.2; 137.5] Trapezoidal
$\mu_{A_2^{(1)}}(\text{EDSS al inicio})$	$A_2^{(1)}$	$\mathbb{X}_2$ : [0 a 10] unidades	Bajo: [−3.75; −0.416; 1.0; 5.0] Trapezoidal Medio: [1.0; 5.0; 9.0] Triangular Alto: [5.0; 9.0; 10.42; 13.75] Trapezoidal
$\mu_{A_3^{(1)}}(\text{EDSS después de un año})$	$A_3^{(1)}$	$\mathbb{X}_3$ : [0 a 10] unidades	Bajo: [−3.75; −0.416; 1.0; 5.0] Trapezoidal Medio: [1.0; 5.0; 9.0] Triangular Alto: [5.0; 9.0; 10.42; 13.75] Trapezoidal
$\mu_{A_4^{(1)}}(\text{EDSS después de dos años})$	$A_4^{(1)}$	$\mathbb{X}_4$ : [0 a 10] unidades	Bajo: [−3.75; −0.416; 1.0; 5.0] Trapezoidal Medio: [1.0; 5.0; 9.0] Triangular Alto: [5.0; 9.0; 10.42; 13.75] Trapezoidal
$\mu_B(\text{Respuesta al IFN-}\beta 1b)$	$B(y)$	$\mathbb{Y}$ : [0 a 1] unidades	Bajo: [−0.375; −0.04167; 0.1, 0.5] Trapezoidal Medio: [0.1; 0.5; 0.9] Triangular Alto: [0.5; 0.9; 1.042; 1.375] Trapezoidal



**Tabla 5-3** Definición de reglas difusas (base de conocimientos primera parte).

#	Regla
1	Si la edad es Infantil y la EDSS al inicio es Baja y la EDSS después de un año es Baja y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta - 1b$ es Media
2	Si la edad es Infantil y la EDSS al inicio es Media y la EDSS después de un año es Media y la EDSS después de dos años es Media, entonces la respuesta al IFN- $\beta - 1b$ es Media
3	Si la edad es Infantil y la EDSS al inicio es Alta y la EDSS después de un año es Alta y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta - 1b$ es Media
4	Si la edad es Infantil y la EDSS al inicio es Media y la EDSS después de un año es Baja y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta - 1b$ es Alta
5	Si la edad es Infantil y la EDSS al inicio es Alta y la EDSS después de un año es Media y la EDSS después de dos años es Media, entonces la respuesta al IFN- $\beta - 1b$ es Alta
6	Si la edad es Infantil y la EDSS al inicio es Alta y la EDSS después de un año es Baja y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta - 1b$ es Alta
7	Si la edad es Infantil y la EDSS al inicio es Alta y la EDSS después de un año es Media y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta - 1b$ es Alta
8	Si la edad es Infantil y la EDSS al inicio es Baja y la EDSS después de un año es Media y la EDSS después de dos años es Media, entonces la respuesta al IFN- $\beta - 1b$ es Baja
9	Si la edad es Infantil y la EDSS al inicio es Baja y la EDSS después de un año es Alta y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta - 1b$ es Baja
10	Si la edad es Infantil y la EDSS al inicio es Baja y la EDSS después de un año es Media y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta - 1b$ es Baja
11	Si la edad es Infantil y la EDSS al inicio es Media y la EDSS después de un año es Alta y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta - 1b$ es Baja

**Tabla 5-4** Definición de reglas difusas (base de conocimientos segunda parte).

#	Regla
12	Si la edad es Adulto y la EDSS al inicio es Baja y la EDSS después de un año es Baja y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta - 1b$ es Media
13	Si la edad es Adulto y la EDSS al inicio es Media y la EDSS después de un año es Media y la EDSS después de dos años es Media, entonces la respuesta al IFN- $\beta - 1b$ es Media
14	Si la edad es Adulto y la EDSS al inicio es Alta y la EDSS después de un año es Alta y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta - 1b$ es Media
15	Si la edad es Adulto y la EDSS al inicio es Media y la EDSS después de un año es Baja y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta - 1b$ es Alta
16	Si la edad es Adulto y la EDSS al inicio es Alta y la EDSS después de un año es Media y la EDSS después de dos años es Media, entonces la respuesta al IFN- $\beta - 1b$ es Alta
17	Si la edad es Adulto y la EDSS al inicio es Alta y la EDSS después de un año es Baja y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta - 1b$ es Alta
18	Si la edad es Adulto y la EDSS al inicio es Alta y la EDSS después de un año es Media y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta - 1b$ es Alta
19	Si la edad es Adulto y la EDSS al inicio es Baja y la EDSS después de un año es Media y la EDSS después de dos años es Media, entonces la respuesta al IFN- $\beta - 1b$ es Baja
20	Si la edad es Adulto y la EDSS al inicio es Baja y la EDSS después de un año es Alta y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta - 1b$ es Baja
21	Si la edad es Adulto y la EDSS al inicio es Baja y la EDSS después de un año es Media y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta - 1b$ es Baja
22	Si la edad es Adulto y la EDSS al inicio es Media y la EDSS después de un año es Alta y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta - 1b$ es Baja

**Tabla 5-5** Definición de reglas difusas (base de conocimientos tercera parte).

#	Regla
23	Si la edad es Anciano y la EDSS al inicio es Baja y la EDSS después de un año es Baja y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta$ – 1b es Media
24	Si la edad es Anciano y la EDSS al inicio es Media y la EDSS después de un año es Media y la EDSS después de dos años es Media, entonces la respuesta al IFN- $\beta$ – 1b es Media
25	Si la edad es Anciano y la EDSS al inicio es Alta y la EDSS después de un año es Alta y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta$ – 1b es Media
26	Si la edad es Anciano y la EDSS al inicio es Media y la EDSS después de un año es Baja y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta$ – 1b es Alta
27	Si la edad es Anciano y la EDSS al inicio es Alta y la EDSS después de un año es Media y la EDSS después de dos años es Media, entonces la respuesta al IFN- $\beta$ – 1b es Alta
28	Si la edad es Anciano y la EDSS al inicio es Alta y la EDSS después de un año es Baja y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta$ – 1b es Alta
29	Si la edad es Anciano y la EDSS al inicio es Alta y la EDSS después de un año es Media y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta$ – 1b es Alta
30	Si la edad es Anciano y la EDSS al inicio es Baja y la EDSS después de un año es Media y la EDSS después de dos años es Media, entonces la respuesta al IFN- $\beta$ – 1b es Baja
31	Si la edad es Anciano y la EDSS al inicio es Baja y la EDSS después de un año es Alta y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta$ – 1b es Baja
32	Si la edad es Anciano y la EDSS al inicio es Baja y la EDSS después de un año es Media y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta$ – 1b es Baja
33	Si la edad es Anciano y la EDSS al inicio es Media y la EDSS después de un año es Alta y la EDSS después de dos años es Alta, entonces la respuesta al IFN- $\beta$ – 1b es Baja

**Tabla 5-6** Pseudocódigo para la optimización de parámetros de la ANN por el GA.

---

**Inicialización**

Seleccionar población inicial de tamaño  $N=100$   
Definir la función objetivo  $f(X)$  (precisión)  
Codificar la población como cromosomas (cadenas de bits)  
de longitud  $L_c$   
Calcular los valores de aptitud (precisión) de toda la población  
Definir la condición de terminación  
Elegir el número máximo de iteraciones ( $\text{Max\_Iter}=10$ )  
 $\text{iter} = 1$

**while**( $\text{iter} \leq \text{Max\_Iter}$ )

**Selección:** Seleccionar padres para la reproducción

**Cruzamiento:** Aplicar cruzamiento en padres para producir descendencia

**Mutación:** Aplicar mutación en cromosomas seleccionados

Calcular los valores de aptitud de la población

Seleccionar miembros para la próxima generación según los valores de aptitud

**If** se cumple la condición de terminación, salir; de lo contrario, continuar ( $\text{iter} = \text{iter} + 1$ )

**end while**

---

## 6. RESULTADOS Y DISCUSIÓN

En este trabajo de investigación, se implementó un sistema experto difuso utilizando variables demográficas y clínicas (previamente preprocesadas) como entradas para clasificar las etiquetas de clase no conocidas correspondientes a la respuesta al tratamiento con IFN- $\beta 1b$  (alto, medio o bajo) en pacientes con RRMS. Además, se implementó una técnica de agrupamiento jerárquico para clasificar a los mismos pacientes. Después se utilizaron 65 características genéticas (previamente preprocesadas) asociadas a la respuesta al IFN- $\beta 1b$  para entrenar a cinco modelos convencionales de ML y un modelo ANN, cuyos hiperparámetros de configuración fueron optimizados por un GA.

### 6.1 Distribución de los datos

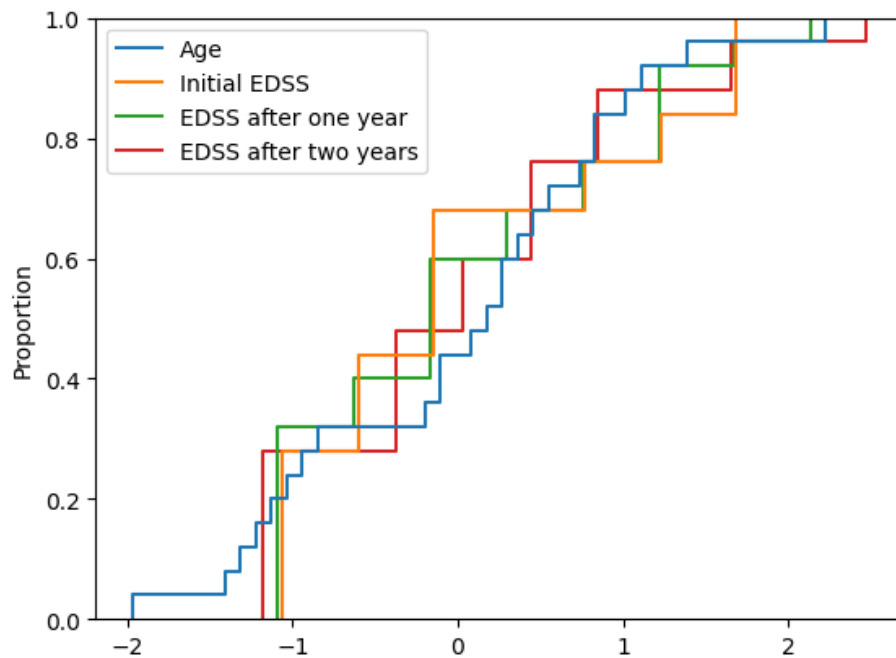
Después de estandarizar las variables demográficas y clínicas, y aplicar la prueba KS para verificar su distribución normal ( $valor - p > 0.05$ ), se obtuvieron los resultados de la Tabla 6-1. La Figura 6.1 muestra el gráfico de las funciones de distribución acumulativa empírica de las variables demográficas y clínicas.

**Tabla 6-1** Resultados estadísticos de la prueba KS en las variables demográficas y clínicas.

Pares de Variables	Estadística-D	Valor-p	Distribución
Edad / EDSS inicial	0.32	0.1557	Normal
EDSS inicial / EDSS después de 1 año	0.32	0.1557	Normal
EDSS después de 1 año / EDSS después de 2 años	0.28	0.2850	Normal

También se estandarizaron las variables genéticas y después de aplicar la prueba KS para validar su distribución normal ( $valor - p > 0.05$ ) se obtuvieron los resultados de la Tabla 6-2. La Figura 6.2 muestra el gráfico de las funciones de distribución acumulativa empírica de las variables genéticas.

**Figura 6.1** Gráfico de las funciones de distribución acumulativa empírica de las variables demográficas y clínicas.



## 6.2 Sistema Experto Difuso

En la etapa de fuzzyficación, se calcularon los valores de membresía para cada una de las variables de entrada. Las Tablas 6-3 y 6-4 muestran los valores obtenidos de cada función de membresía para todas las muestras.

En la etapa de razonamiento aproximado, cada una de las 33 reglas de inferencia de la base de conocimientos fueron evaluadas con los valores de membresía obtenidos de las Tablas 6-3 y 6-4. Por ejemplo, considerando los valores de entrada de la muestra 7 (Edad: “44”, EDSS al inicio: “4”, EDSS después de 1 año: “3” y EDSS después de 2 años: “3”), los resultados del motor de inferencia se muestran en la Tabla 6-5. En este caso, sólo dos reglas tuvieron un resultado de inferencia diferente de cero. La Figura 6.3 muestra el gráfico de evaluación de las reglas difusas 13 y 15.

En la etapa de defuzzificación se calcularon las salidas numéricas, en base a los gráficos de resultados de inferencia utilizando la Ecuación (5.5). Por ejemplo,

**Tabla 6-2** Resultados estadísticos de la prueba KS en las variables genéticas (primera dosis).

Pares de Variables	Estadística- $D$	Valor- $p$	Distribución
IL-2 / IFN- $\gamma$	0.12	0.9955	Normal
IFN- $\gamma$ / TNF- $\alpha$	0.16	0.9149	Normal
TNF- $\alpha$ / IL-4	0.12	0.9955	Normal
IL-4 / IL-10	0.16	0.9149	Normal
IL-10 / TGF- $\beta$	0.12	0.9955	Normal
TGF- $\beta$ / CD46	0.24	0.4754	Normal
CD46 / CD58	0.2	0.7102	Normal
CD58 / FHIT	0.16	0.9149	Normal
FHIT / IFR5	0.16	0.9149	Normal
IFR5 / GAPVD1	0.16	0.9149	Normal
GAPVD1 / GRM3	0.12	0.9955	Normal
GRM3 / GRIK2	0.16	0.9149	Normal

para obtener la salida numérica para la muestra 7 según el gráfico de inferencia de la Figura 6.4, se realizó el siguiente cálculo:

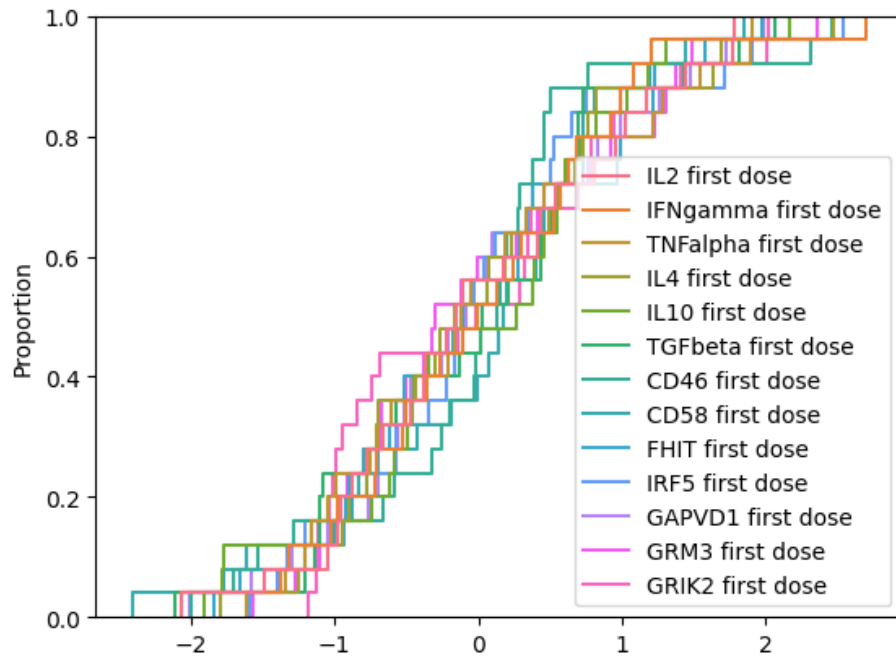
$$y_0 = \frac{0.0 * 0.0 + 0.1 * 0.0 + 0.2 * 0.25 + 0.3 * 0.5 + \dots + 0.8 * 0.25 + 0.9 * 0.25 + 1 * 0.25}{0.0 + 0.0 + 0.25 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.25 + 0.25 + 0.25} \quad (6.1)$$

$$y_0 = \frac{1.9025}{3.4} = 0.559 \approx 0.537 \quad (6.2)$$

Finalmente, se realizó una clasificación de respondedor alto, medio y bajo al tratamiento con IFN- $\beta$ , mediante tres métodos distintos: 1) Opinión de un experto en neurología, 2) Propuesta de sistema difuso y 3) Modelo de agrupamiento jerárquico. Los resultados obtenidos se muestran en la Tabla 6-6.

Como muestra la Tabla 6-6, el 80 % de los resultados fueron etiquetados correctamente mediante el sistema difuso propuesto, mientras que el 64 % fueron etiquetados correctamente mediante el agrupamiento jerárquico con respecto a la

**Figura 6.2** Gráfico de las funciones de distribución acumulativa empírica de las variables genéticas (primera dosis).



opinión de un experto.

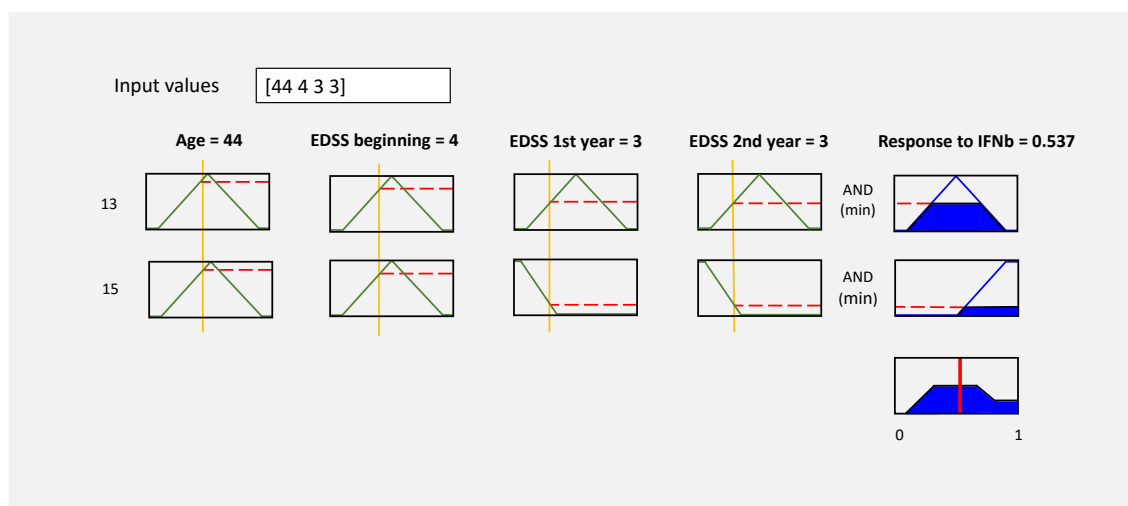
### 6.3 Optimización de hiperparámetros con GA

A continuación se presentan los resultados obtenidos después de implementar el algoritmo genético propuesto para la optimización de hiperparámetros del modelo ANN entrenado con las características genéticas. Para fines de comparación, también se realizó la optimización de parámetros de un modelo MLP (entrenado con las mismas características) por medio de la herramienta *GridSearchCV* considerando todas las combinaciones de hiperparámetros.

Por un lado, la Tabla 6-7 muestra los resultados obtenidos de la optimización de parámetros de la ANN por medio del GA. Por otro lado, la Tabla 6-8 presenta los resultados obtenidos de la optimización de hiperparámetros del MLP por medio de *GridSearchCV*.



**Figura 6.3** Gráfico de inferencia de reglas en la aplicación Fuzzy Logic Designer, que muestra los resultados de la combinación de los valores de inferencia de las reglas 13 y 15.



#### 6.4 Evaluación de rendimiento

Los modelos de aprendizaje: KN, LR, SVM, DT, MLP y ANN, fueron entrenados con 65 biomarcadores asociados con la respuesta al IFN- $\beta 1b$ . Luego, se calcularon las métricas: CM, precisión, sensibilidad, especificidad y AUC con las predicciones de salida para comparar el rendimiento de los modelos implementados. Los resultados de los indicadores de rendimiento se presentan en Tabla 6-9.

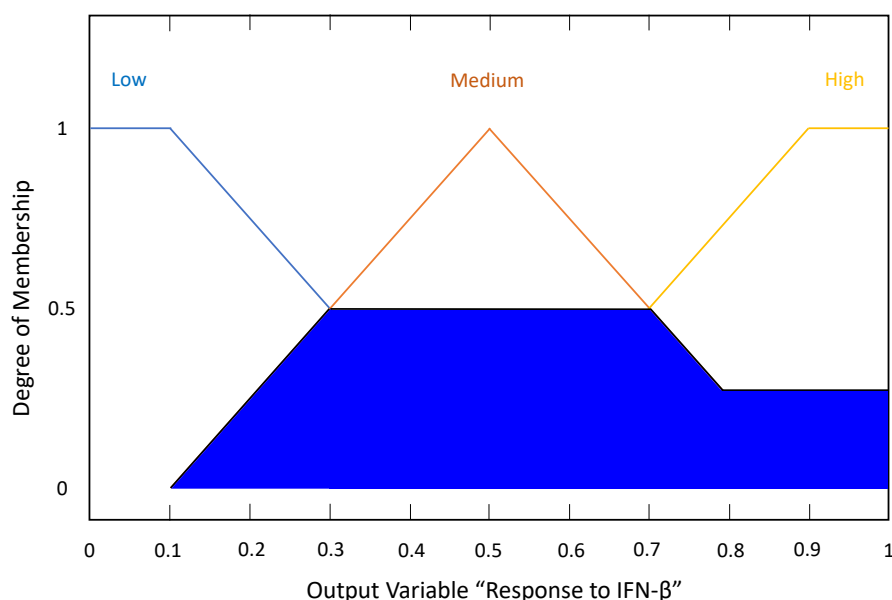
En el caso de la ANN propuesta, las Figuras 6.5 y 6.6 muestran los resultados de la función de pérdida y la precisión de entrenamiento respectivamente, durante 50 épocas. Como se puede ver la pérdida converge al valor mínimo desde la época 23 y la precisión es máxima a partir de la época 11, aproximadamente.

#### 6.5 Publicaciones

A partir de la presente investigación se han derivado los siguientes productos:

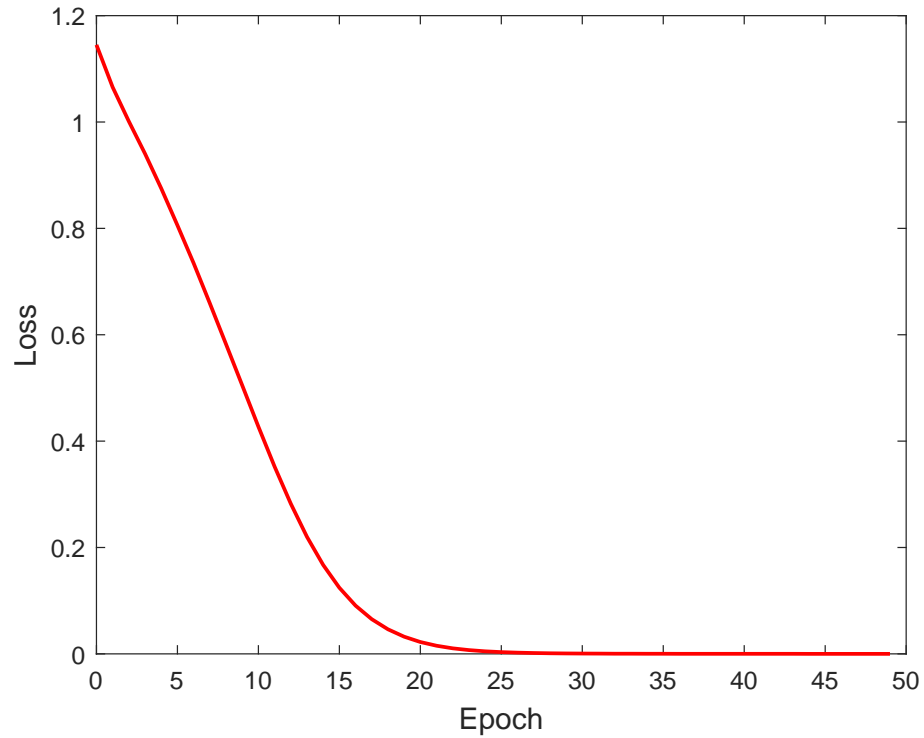
1. Ponce de Leon-Sanchez, E. R., Dominguez-Ramirez, O. A., Herrera-Navarro,

**Figura 6.4** *Gráfico de resultados de inferencia de la muestra 7, que incluye los valores de las etiquetas lingüísticas: bajo, medio y alto.*



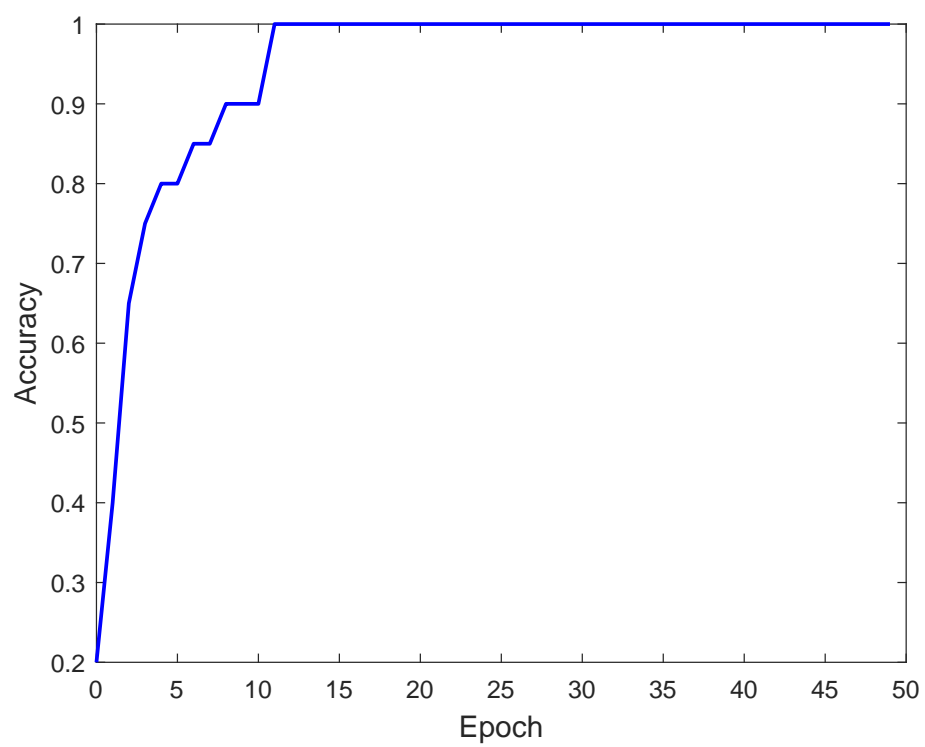
- A. M., Rodriguez-Resendiz, J., Paredes-Orta, C., Mendiola-Santibañez, J. D. (2023a). A deep learning approach for predicting multiple sclerosis. *Micromachines*, 14(4):749.
2. Ponce de Leon-Sanchez, E. R., Mendiola-Santibañez, J. D., Dominguez-Ramirez, O. A., Herrera-Navarro, A. M., Vazquez-Cervantes, A., Jimenez-Hernandez, H., Senties-Madrid, H. (2023b). Fuzzy logic system for classifying multiple sclerosis patients as high, medium, or low responders to interferon-beta. *Technologies*, 11(4):109.
  3. Ponce de Leon-Sanchez, E. R., Mendiola-Santibañez, J. D., Dominguez-Ramirez, O. A., Herrera-Navarro, A. M., Vazquez-Cervantes, A., Jimenez-Hernandez, H., Senties-Madrid, H. (2024). Training Artificial Neural Networks to Detect Multiple Sclerosis Lesions Using Granulometric Data from Preprocessed Magnetic Resonance Images with Morphological Transformations. *Technologies*, 12(9), 145.
  4. Ponce de León Sánchez, E. R., Santibañez, J. D. M., Ramírez, O. A. D., Duarte, M. G., Navarro, A. M. H., Madrid, H. S., Téllez, J. J. F. (2024). A based-rule

**Figura 6.5** Resultados de la función de pérdida del modelo ANN. Los parámetros de entrenamiento fueron configurados por el GA implementado: 'num\_neuronas\_ocultas' = 64, 'num\_capas\_ocultas' = 2, 'func\_activacion' = 'relu', 'optimizador' = 'adamax'.



fuzzy expert system to estimate the response to treatments in multiple sclerosis patients. Medical Research Archives, 12(9).

**Figura 6.6** Resultados de la métrica de precisión del modelo ANN.



**Tabla 6-3** Resultados de fuzzificación (Edad y EDSS al inicio).  $\mu_B$  :  $\mu$  Bajo,  $\mu_M$  :  $\mu$  Medio,  $\mu_A$  :  $\mu$  Alto.

Muestra	$\mu_{Infantil}(\text{Edad})$	$\mu_{Adulto}(\text{Edad})$	$\mu_{Anciano}(\text{Edad})$	$\mu_B(\text{EDSS}^1)$	$\mu_M(\text{EDSS}^1)$	$\mu_A(\text{EDSS}^1)$
1	0.0	0.687	0.015	0.25	0.75	0.0
2	0.0	0.88	0.0	1.0	0.0	0.0
3	0.375	0.4	0.0	1.0	0.0	0.0
4	0.315	0.448	0.0	0.25	0.75	0.0
5	0.0	0.975	0.0	0.5	0.5	0.0
6	0.0	0.784	0.0	0.75	0.25	0.0
7	0.0	0.856	0.0	0.25	0.75	0.0
8	0.225	0.479	0.0	0.875	0.125	0.0
9	0.345	0.424	0.0	0.25	0.75	0.0
10	0	0.808	0.0	1.0	0.0	0.0
11	0.255	0.496	0.0	0.75	0.25	0.0
12	0.258	0.472	0.0	0.875	0.125	0.0
13	0.0	0.952	0.0	1.0	0.0	0.0
14	0.0	0.928	0.0	0.375	0.625	0.0
15	0.0	0.808	0.0	0.75	0.25	0.0
16	0.0	1.0	0.0	0.375	0.625	0.0
17	0.015	0.688	0.0	0.875	0.125	0.0
18	0.0	0.832	0.0	0.75	0.25	0.0
19	0.0	0.903	0.0	0.5	0.5	0.0
20	0.0	0.76	0.0	1.0	0.0	0.0
21	0.0	0.952	0.0	0.75	0.25	0.0
22	0.0	0.712	0.0	0.75	0.25	0.0
23	0.585	0.232	0.0	0.875	0.125	0.0
24	0.405	0.376	0.0	1.0	0.0	0.0
25	0.0	0.712	0.0	1.0	0.0	0.0

<sup>1</sup> EDSS al inicio.

**Tabla 6-4** Resultados de fuzzificación (EDSS después de 1 año y EDSS después de 2 años).  $\mu_B$  :  $\mu$  Bajo,  $\mu_M$  :  $\mu$  Medio,  $\mu_A$  :  $\mu$  Alto.

Muestra	$\mu_B(\text{EDSS}^1)$	$\mu_M(\text{EDSS}^1)$	$\mu_A(\text{EDSS}^1)$	$\mu_B(\text{EDSS}^2)$	$\mu_M(\text{EDSS}^2)$	$\mu_A(\text{EDSS}^2)$
1	0.0	0.975	0.0	0.0	0.875	0.125
2	0.825	0.0	0.0	1.0	0.0	0.0
3	0.825	0.0	0.0	1.0	0.0	0.0
4	0.0	0.825	0.0	0.375	0.625	0.0
5	0.375	0.375	0.0	0.625	0.375	0.0
6	0.525	0.225	0.0	0.125	0.875	0.0
7	0.225	0.525	0.0	0.5	0.5	0.0
8	0.525	0.225	0.0	0.75	0.25	0.0
9	0.075	0.675	0.0	0.25	0.75	0.0
10	0.825	0.0	0.0	1.0	0.0	0.0
11	0.525	0.225	0.0	0.625	0.375	0.0
12	0.825	0.0	0.0	0.625	0.375	0.0
13	0.825	0.0	0.0	1.0	0.0	0.0
14	0.075	0.675	0.0	0.5	0.5	0.0
15	0.375	0.375	0.0	0.5	0.5	0.0
16	0.075	0.675	0.0	0.375	0.625	0.0
17	0.225	0.525	0.0	0.125	0.875	0.0
18	0.675	0.075	0.0	0.75	0.25	0.0
19	0.525	0.225	0.0	0.75	0.25	0.0
20	0.825	0.0	0.0	1.0	0.0	0.0
21	0.525	0.225	0.0	0.75	0.25	0.0
22	0.075	0.675	0.0	0.5	0.5	0.0
23	0.675	0.075	0	0.75	0.25	0.0
24	0.825	0.0	0.0	1.0	0.0	0.0
25	0.825	0.0	0.0	1.0	0.0	0.0

<sup>1</sup> EDSS después de 1 año. <sup>2</sup> EDSS después de 2 años.

**Tabla 6-5** *Resultados de inferencia de la evaluación de las reglas difusas 13 y 15 (muestra 7).*

#	Regla	Motor de Inferencia
13	Si la edad es Adulto y la EDSS al inicio es Media y la EDSS después de un año es Media y la EDSS después de dos años es Media, entonces la respuesta al IFN- $\beta$ es Media	$\min(0.856, 0.75, 0.525, 0.5) = 0.5$
15	Si la edad es Adulto y la EDSS al inicio es Media y la EDSS después de un año es Baja y la EDSS después de dos años es Baja, entonces la respuesta al IFN- $\beta$ es Alta	$\min(0.856, 0.75, 0.225, 0.5) = 0.225$

**Tabla 6-6** Clasificación de la respuesta al IFN- $\beta$ 1b. Los valores numéricos resultantes de la defuzzificación inferiores a 0.5 se consideran como respondedor bajo (LR), los iguales a 0.5 como respondedor medio (MR) y los superiores a 0.5 como respondedor alto (HR). Para fines de comparación, los datos de entrada de la Tabla 6-1 también se utilizaron para entrenar a un modelo de agrupamiento jerárquico ( $n\_clusters = 3$ ).

Muestra	Opinion de un experto	Sistema Difuso (Defuzzificación)	Agrupamiento Jerárquico
1	LR	0.5 $\Rightarrow$ MR	HR
2	MR	0.5 $\Rightarrow$ MR	MR
3	MR	0.5 $\Rightarrow$ MR	MR
4	MR	0.5 $\Rightarrow$ MR	HR
5	MR	0.537 $\Rightarrow$ HR	LR
6	LR	0.433 $\Rightarrow$ LR	LR
7	HR	0.537 $\Rightarrow$ HR	HR
8	MR	0.473 $\Rightarrow$ LR	MR
9	MR	0.513 $\Rightarrow$ HR	HR
10	MR	0.5 $\Rightarrow$ MR	MR
11	MR	0.5 $\Rightarrow$ MR	MR
12	LR	0.473 $\Rightarrow$ LR	MR
13	MR	0.5 $\Rightarrow$ MR	MR
14	MR	0.5 $\Rightarrow$ MR	HR
15	LR	0.443 $\Rightarrow$ LR	LR
16	MR	0.5 $\Rightarrow$ MR	HR
17	LR	0.247 $\Rightarrow$ LR	LR
18	MR	0.527 $\Rightarrow$ HR	LR
19	HR	0.595 $\Rightarrow$ HR	LR
20	MR	0.5 $\Rightarrow$ MR	MR
21	MR	0.5 $\Rightarrow$ MR	LR
22	LR	0.363 $\Rightarrow$ LR	LR
23	MR	0.5 $\Rightarrow$ MR	MR
24	MR	0.5 $\Rightarrow$ MR	MR
25	MR	0.5 $\Rightarrow$ MR	MR



**Tabla 6-7** Optimización de hiperparámetros de la ANN por GA (parámetros óptimos con la precisión máxima).

Neuronas ocultas	Capas ocultas	Funcion de activación	Optimizador	Precisión
256	3	'relu'	'adam' <sup>1</sup>	1.0
64	1	'sigmoid'	'nadam' <sup>2</sup>	1.0
64	2	'relu'	'adamax' <sup>3</sup>	1.0
512	1	'sigmoid'	'adamax'	1.0
128	4	'relu'	'nadam'	1.0

<sup>1</sup> Método de descenso de gradiente estocástico que se basa en la estimación adaptativa de momentos de primer y segundo orden. <sup>2</sup> Variante de Adam con momento Nesterov. <sup>3</sup> Variante de Adam basada en la norma infinita, es un método de optimización basado en gradiente de primer orden.

**Tabla 6-8** Optimización de parámetros del MLP por GridSearchCV (parámetros seleccionados con la máxima precisión).

Neuronas ocultas	Capas ocultas	Funcion de activación	Optimizador	Precisión
64	1	'logistic'	'sgd' <sup>1</sup>	0.8
128	1	'logistic'	'sgd'	0.8
256	4	'logistic'	'sgd'	0.8
512	3	'logistic'	'sgd'	0.8

<sup>1</sup> Optimizador descenso de gradiente estocástico.

**Tabla 6-9** Resultados de rendimiento por modelo de aprendizaje. En el caso del modelo ANN, los resultados de las métricas se consideran en el intervalo de la época 0 hasta la 11, que es cuando alcanzan sus valores máximos.

Modelo	Precisión	Sensibilidad	Especificidad	AUC
KN <sup>1</sup>	0.8	1.0	0.75	0.875
LR <sup>2</sup>	0.2	0.25	0.0	0.125
SVM <sup>3</sup>	0.4	0.4	0.0	0.2
DT <sup>4</sup>	0.6	0.5	0.667	0.583
MLP <sup>5</sup>	0.6	0.5	0.667	0.583
ANN	0.2-1.0	0.25-1.0	0.0-1.0	0.125-1.0

<sup>1</sup> K-Neighbors <sup>2</sup> Logistic Regression <sup>3</sup> Support Vector Machine <sup>4</sup> Decision Tree <sup>5</sup> Multi-layer Perceptron.

## 7. CONCLUSIONES

Mientras que la lógica binaria genera sólo dos tipos de salida  $[0, 1]$ , los motores de inferencia difusa utilizan un razonamiento aproximado basado en reglas de inferencia generalizadas. Por lo tanto, los sistemas difusos son métodos convenientes para resolver problemas de clasificación, debido a su capacidad para procesar información inexacta. En esta investigación, se implementó un sistema experto difuso alternativo basado en conocimiento experto como apoyo a la decisión en la clasificación de la respuesta al tratamiento con IFN- $\beta - 1b$  en pacientes con RRMS. Se utilizaron características demográficas y clínicas previamente estandarizadas como variables de entrada al sistema difuso. Como se muestra en la Tabla 6-6, la clasificación del sistema difuso propuesto logró mejores resultados que el agrupamiento jerárquico, debido a que este método no siempre proporciona la mejor solución debido a la gran cantidad de decisiones arbitrarias. Una limitante de software en el diseño del sistema difuso fue establecer pocas variables de entrada ya que durante la experimentación cuanto mayor era la cantidad de variables, mayor era el tiempo de procesamiento de datos.

Es importante mencionar que al inicio del diseño del sistema difuso, se realizó una propuesta de definición de reglas difusas que consideraba sólo dos etiquetas lingüísticas de salida: respondedor “alto” y “bajo” al IFN- $\beta - 1b$ , y en una primera etapa de experimentación se obtuvo una eficiencia de 63 % en los resultados. Por tal motivo, las reglas difusas fueron evaluadas por un neurólogo experto. Después de validar los resultados, el experto recomendó agregar una etiqueta adicional (respondedor “medio”) para clasificar a los pacientes con RRMS que tenían niveles similares de EDSS al inicio y al final del tratamiento. Tras redefinir las reglas difusas, se logró el 80 % de eficiencia.

Una vez que el sistema difuso clasificó las etiquetas de salida del conjunto de datos, se estandarizaron las características genéticas y se aplicó la prueba KS para verificar su distribución normal. Después, se implementó un GA para optimizar los parámetros de configuración de un modelo ANN entrenado con los biomarcadores asociados con la respuesta al IFN- $\beta - 1b$  para estimar si los pacientes con RRMS

son candidatos potenciales para ser tratados con este fármaco. Los resultados de rendimiento de las tablas 6-7 y 6-8 muestran que el GA funciona mejor como optimizador de parámetros de la ANN que la herramienta *GridSearchCV*. Una desventaja que se presentó al evaluar el rendimiento del algoritmo propuesto fue que el tamaño de las muestras de prueba era demasiado pequeño.

Los algoritmos de ML y DL pueden encontrar patrones naturales en los datos y son una alternativa útil en el campo de la bioinformática. Estos algoritmos se han implementado para mejorar el diagnóstico de MS Chen et al. (2021); Goyal et al. (2019) y ayudar a los especialistas a predecir la respuesta a los tratamientos farmacológicos en pacientes con MS Fagone et al. (2019); Tao Jin (2020). La tabla 7-1 presenta una comparación de los resultados de rendimiento de algunas aplicaciones de ML en el estudio de la MS con los resultados obtenidos en esta investigación.

**Tabla 7-1** *Comparación de resultados de rendimiento de aplicaciones de ML en el estudio de MS.*

Autor	Predicción	Técnica	Precisión
Fagone et al. (2019)	Respuesta al Natalizumab	UCRC <sup>1</sup>	0.892
Goyal et al. (2019)	Diagnóstico MS	RF <sup>2</sup>	0.909
Tao Jin (2020)	Respuesta al IFN- $\beta$	SVM <sup>3</sup>	0.809
Chen et al. (2021)	Diagnóstico MS	SVM <sup>3</sup>	0.930
Trabajo actual	Respuesta al IFN- $\beta$	ANN	1.0

<sup>1</sup> Uncorrelated Reduced Centroid <sup>2</sup> Random Forest <sup>3</sup> Support Vector Machine

## REFERENCIAS

- Acar, Z. Y., Başçiftçi, F., and Ekmekci, A. H. (2022). A convolutional neural network model for identifying multiple sclerosis on brain flair mri. *Sustainable Computing: Informatics and Systems*, 35(100706). (Citado en página 107.)
- Ackermann, M. R., Blömer, J., Kuntze, D., and Sohler, C. (2014). Analysis of agglomerative clustering. *Algorithmica*, 69:184–215. (Citado en página 39.)
- Ayangbekun, O. and Jimoh Ibrahim, A. (2015). Fuzzy logic application to brain diseases diagnosis. *Journal of Emerging Trends in Computing and Information Sciences*, 6(3):144–148. (Citado en páginas 24 y 26.)
- Bandyopadhyay, R., Kundu, R., Oliva, D., and Sarkar, R. (2021). Segmentation of brain mri using an altruistic harris hawks' optimization algorithm. *Knowledge-Based Systems*, 232:107468. (Citado en página 96.)
- Bustamante, M. F., Morcillo-Suárez, C., Malhotra, S., Rio, J., Leyva, L., Fernández, O., Zettl, U. K., Killestein, J., Brassat, D., García-Merino, J. A., et al. (2015). Pharmacogenomic study in patients with multiple sclerosis: responders and nonresponders to ifn- $\beta$ . *Neurology-Neuroimmunology Neuroinflammation*, 2(5). (Citado en página 21.)
- Casalino, G., Castellano, G., Consiglio, A., Nuzziello, N., and Vessio, G. (2021). Mi-crona expression classification for pediatric multiple sclerosis identification. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–10. (Citado en páginas 25, 34 y 46.)
- Chen, J. & Gustientiedina, G. (2024). Implementation of fuzzy expert system to detect parkinson's disease based on mobile. *Journal of Applied Business and Technology*, 5(2):72–81. (Citado en páginas 24 y 27.)
- Chen, X., Hou, H., Qiao, H., Fan, H., Zhao, T., and Dong, M. (2021). Identification of blood-derived candidate gene markers and a new 7-gene diagnostic model for multiple sclerosis. *Biological Research*, 54(1):1–12. (Citado en páginas 25, 28 y 72.)

- Coelewij, L., Adriani, M., Dönnies, P., Waddington, K. E., Ciurtin, C., Havrdova, E. K., Farrell, R., Nytrova, P., Pineda-Torra, I., Jury, E. C., et al. (2024). Patients with multiple sclerosis who develop immunogenicity to interferon-beta have distinct transcriptomic and proteomic signatures prior to treatment which are associated with disease severity. *Clinical Immunology*, 267:110339. (Citado en página 21.)
- Cohan, S. L., Hendin, B. A., Reder, A. T., Smoot, K., Avila, R., Mendoza, J. P., and Weinstock-Guttman, B. (2021). Interferons and multiple sclerosis: lessons from 25 years of clinical and real-world experience with intramuscular interferon beta-1a (avonex). *CNS drugs*, 35:743–767. (Citado en página 21.)
- Czabanski, R., Jezewski, M., and Leski, J. (2017). Introduction to fuzzy systems. *Theory and applications of ordered fuzzy numbers: A tribute to professor Witold Kosiński*, pages 23–43. (Citado en página 36.)
- de Arruda, A. L. C., Vital, D. A., Kitamura, F. C., Abdala, N., and Moraes, M. C. (2020). Multiple sclerosis segmentation method in magnetic resonance imaging using fuzzy connectedness, binarization, mathematical morphology, and 3d reconstruction. *Research on Biomedical Engineering*, 36:291–301. (Citado en páginas 84 y 96.)
- de Oliveira, M., Piacenti-Silva, M., da Rocha, F. C. G., Santos, J. M., Cardoso, J. d. S., and Lisboa-Filho, P. N. (2022). Lesion volume quantification using two convolutional neural networks in mris of multiple sclerosis patients. *Diagnostics*, 12(2):230. (Citado en página 107.)
- Dominguez-Mozo, M. I., Perez-Perez, S., Villarrubia, N., Costa-Frossard, L., Fernandez-Velasco, J. I., Ortega-Madueño, I., Garcia-Martinez, M. A., Garcia-Calvo, E., Estevez, H., Luque Garcia, J. L., et al. (2021). Herpesvirus antibodies, vitamin d and short-chain fatty acids: their correlation with cell subsets in multiple sclerosis patients and healthy controls. *Cells*, 10(1):119. (Citado en página 19.)
- Ebrahimkhani, S., Beadnall, H. N., Wang, C., Suter, C. M., Barnett, M. H., Buckland, M. E., and Vafaei, F. (2020). Serum exosome micrornas predict multiple sclerosis disease activity after fingolimod treatment. *Molecular Neurobiology*, 57:1245–1258. (Citado en páginas 25 y 29.)

- Fagone, P., Mazzon, E., Mammana, S., Di Marco, R., Spinasant, F., Basile, M. S., Petralia, M. C., Bramanti, P., Nicoletti, F., and Mangano, K. (2019). Identification of cd4+ t cell biomarkers for predicting the response of patients with relapsing-remitting multiple sclerosis to natalizumab treatment. *Molecular medicine reports*, 20(1):678–684. (Citado en páginas 25, 29 y 72.)
- Frigolet, M. and Gutiérrez-Aguilar, R. (2017). Ciencias “ómicas”, ¿cómo ayudan a las ciencias de la salud. *Revista Digital Universitaria*, 18(7):1–7. (Citado en página 24.)
- Ganji, A., Monfared, M. E., Shapoori, S., Nourbakhsh, P., Ghazavi, A., Ghasami, K., and Mosayebi, G. (2020). Effects of interferon and glatiramer acetate on cytokine patterns in multiple sclerosis patients. *Cytokine*, 126:154911. (Citado en páginas 21 y 26.)
- Ghasemi, N., Razavi, S., and Nikzad, E. (2017). Multiple sclerosis: pathogenesis, symptoms, diagnoses and cell-based therapy. *Cell Journal (Yakhteh)*, 19(1):1. (Citado en página 19.)
- Ghosh, S., Huo, M., Shawkat, M. S. A., and McCalla, S. (2021). Using convolutional encoder networks to determine the optimal magnetic resonance image for the automatic segmentation of multiple sclerosis. *Applied Sciences*, 11(18):8335. (Citado en páginas 96 y 107.)
- Giordani P., Ferraro M. B., M. F. (2020). Non-hierarchical clustering. *An Introduction to Clustering with R*, pages 75–109. (Citado en página 21.)
- Goyal, M., Khanna, D., Rana, P. S., Khaibullin, T., Martynova, E., Rizvanov, A. A., Khaiboullina, S. F., and Baranwal, M. (2019). Computational intelligence technique for prediction of multiple sclerosis based on serum cytokines. *Frontiers in Neurology*, 10. (Citado en páginas 25, 27, 49 y 72.)
- Gurevich, M., Miron, G., Falb, R. Z., Magalashvili, D., Dolev, M., Stern, Y., and Achiron, A. (2015). Transcriptional response to interferon beta-1a treatment in patients with secondary progressive multiple sclerosis. *BMC neurology*, 15(1):1–8. (Citado en página 21.)

- Hamedan, F., Orooji, A., Sanadgol, H., and Sheikhtaheri, A. (2020). Clinical decision support system to predict chronic kidney disease: A fuzzy expert system approach. *International Journal of Medical Informatics*, 138:104134. (Citado en página 26.)
- Hashemi, M., Akhbari, M., and Jutten, C. (2022). Delve into multiple sclerosis (ms) lesion exploration: A modified attention u-net for ms lesion segmentation in brain mri. *Computers in Biology and Medicine*, 145:105402. (Citado en página 107.)
- Hočevár, K., Ristić, S., and Peterlin, B. (2019). Pharmacogenomics of multiple sclerosis: a systematic review. *Frontiers in neurology*, 10:134. (Citado en página 20.)
- Hosseini, A., Asadi, F., and Arani, L. A. (2020). Development of a knowledge-based clinical decision support system for multiple sclerosis diagnosis. *Journal of Medicine and Life*, 13(4):612. (Citado en páginas 24 y 26.)
- Jha, A. R. (2021). *Mastering PyTorch: build powerful neural network architectures using advanced PyTorch 1. x features*. Packt Publishing Ltd. (Citado en páginas 45 y 47.)
- JM, H. C. . D. (2023). Artificial intelligence and machine learning in clinical medicine. *N Engl J Med*, 388(1201–8). (Citado en página 21.)
- Karlik, B. (2013). Soft computing methods in bioinformatics: a comprehensive review. *Mathematical and Computational Applications*, 18(3):176–197. (Citado en página 24.)
- Kay, M., Hojati, Z., and Dehghanian, F. (2013). The molecular study of ifn $\beta$  pleiotropic roles in ms treatment. *Iranian journal of neurology*, 12(4):149. (Citado en página 20.)
- Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39:261–283. (Citado en página 44.)
- Kular, L., Ewing, E., Needhamsen, M., Pahlevan Kakhki, M., Covacu, R., Gomez-Cabrero, D., Brundin, L., and Jagodic, M. (2022). Dna methylation changes in glial cells of the normal-appearing white matter in multiple sclerosis patients. *Epigenetics*, 17(11):1311–1330. (Citado en página 22.)

- La Rosa, F., Yu, T., Barquero, G., Thiran, J.-P., Granziera, C., and Cuadra, M. B. (2021). Mprage to mp2rage uni translation via generative adversarial network improves the automatic tissue and lesion segmentation in multiple sclerosis patients. *Computers in biology and medicine*, 132:104297. (Citado en página 96.)
- Law, M. T., Traboulsee, A. L., Li, D. K., Carruthers, R. L., Freedman, M. S., Kollind, S. H., and Tam, R. (2019). Machine learning in secondary progressive multiple sclerosis: an improved predictive model for short-term disability progression. *Multiple Sclerosis Journal—Experimental, Translational and Clinical*, 5(4):2055217319885983. (Citado en página 24.)
- Liu, X., Song, L., Liu, S., and Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, 13(3):1224. (Citado en página 95.)
- Liu, Y., Zheng, M., Ma, Z., Zhou, Y., Huo, J., Zhang, W., Liu, Y., Guo, Y., Zhou, X., Li, H., et al. (2022). Design, synthesis, and evaluation of pd-l1 degraders to enhance t cell killing activity against melanoma. *Chinese Chemical Letters*, page 107762. (Citado en página 20.)
- Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J. C., Quiles, A., Valls, L., Ramió-Torrentà, L., and Rovira, À. (2012). Segmentation of multiple sclerosis lesions in brain mri: a review of automated approaches. *Information Sciences*, 186(1):164–185. (Citado en página 84.)
- Lorscheider, J., Benkert, P., Lienert, C., Hänni, P., Derfuss, T., Kuhle, J., Kappos, L., and Yaldizli, Ö. (2018). Comparative analysis of natalizumab versus fingolimod as second-line treatment in relapsing–remitting multiple sclerosis. *Multiple Sclerosis Journal*, 24(6):777–785. (Citado en página 20.)
- M., M. B. . G. (2020). A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med*, 3(126). (Citado en página 21.)
- Ma, X., Ma, R., Zhang, M., Qian, B., Wang, B., and Yang, W. (2023). Recent progress in multiple sclerosis treatment using immune cells as targets. *Pharmaceutics*, 15(3):728. (Citado en página 20.)



- Macin, G., Tasci, B., Tasci, I., Faust, O., Barua, P. D., Dogan, S., Tuncer, T., Tan, R.-S., and Acharya, U. R. (2022). An accurate multiple sclerosis detection model based on exemplar multiple parameters local phase quantization: Exmplpq. *Applied Sciences*, 12(10):4920. (Citado en páginas 85 y 107.)
- Madhiarasan, M. and Deepa, S. (2016). A novel criterion to select hidden neuron numbers in improved back propagation networks for wind speed forecasting. *Applied intelligence*, 44:878–893. (Citado en página 23.)
- Margarit, B. P., Monteiro, G. C., Herán, I. S., Delgado, F. R., and Izquierdo, A. Y. (2019). Esclerosis múltiple. *Medicine-Programa de Formación Médica Continuada Acreditado*, 12(78):4587–4597. (Citado en página 19.)
- Marrodan M., Gaitán M.I., C. J. (2020). Spinal cord involvement in ms and other demyelinating diseases. *Biomedicines*, 8(5):130. (Citado en página 84.)
- Martínez-Aguilar, L., Pérez-Ramírez, C., del Mar Maldonado-Montoro, M., Carrasco-Campos, M. I., Membrive-Jiménez, C., Martínez-Martínez, F., García-Collado, C., Calleja-Hernández, M. Á., Ramírez-Tortosa, M. C., and Jiménez-Morales, A. (2020). Effect of genetic polymorphisms on therapeutic response in multiple sclerosis relapsing-remitting patients treated with interferon-beta. *Mutation Research/-Reviews in Mutation Research*, 785:108322. (Citado en páginas 21 y 26.)
- Martynova, E., Khaibullin, T., Salafutdinov, I., Markelova, M., Laikov, A., Lopukhov, L., Liu, R., Sahay, K., Goyal, M., Baranwal, M., et al. (2023). Seasonal changes in serum metabolites in multiple sclerosis relapse. *International Journal of Molecular Sciences*, 24(4):3542. (Citado en página 19.)
- Matinfar, F. and Golpaygani, A. T. (2022). A fuzzy expert system for early diagnosis of multiple sclerosis. *Journal of Biomedical Physics & Engineering*, 12(2):181. (Citado en páginas 24 y 27.)
- Mendiola J., Gallegos M., O. J. (2007). Segmentación y análisis granulométrico de sustancia blanca y gris en irm para el estudio del estrabismo usando transformaciones morfológicas. *Revista Mexicana De Ingenieria Biomedica*, 28(2):13. (Citado en página 92.)

- Mésidor, M., Rousseau, M.-C., Duquette, P., and Sylvestre, M.-P. (2021). Classification and visualization of longitudinal patterns of medication dose: An application to interferon-beta-1a and amitriptyline in patients with multiple sclerosis. *Pharmacoepidemiology and Drug Safety*, 30(9):1214–1223. (Citado en página 22.)
- Milo, Ron y Miller, A. (2014). Revised diagnostic criteria of multiple sclerosis. *Autoimmunity reviews*, 13(4-5):518–524. (Citado en página 19.)
- Mirandola, S. R., Hallal, D. E., Farias, A. S., Oliveira, E. C., Brandão, C. O., Ruocco, H. H., Damasceno, B. P., and Santos, L. M. (2009). Interferon-beta modifies the peripheral blood cell cytokine secretion in patients with multiple sclerosis. *International immunopharmacology*, 9(7-8):824–830. (Citado en página 20.)
- Mirjalili, Vahid y Raschka, S. (2020). *Python machine learning*. Marcombo. (Citado en páginas 42, 43, 44, 46 y 49.)
- Mohammed, M., Khan, M. B., and Bashier, E. B. M. (2016). *Machine learning: algorithms and applications*. Crc Press. (Citado en página 41.)
- Montolío, A., Martín-Gallego, A., Cegoñino, J., Orduna, E., Vilades, E., Garcia-Martin, E., and Del Palomar, A. P. (2021). Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography. *Computers in Biology and Medicine*, 133:104416. (Citado en página 43.)
- Nishihara, H., Soldati, S., Mossu, A., Rosito, M., Rudolph, H., Muller, W. A., Latorre, D., Sallusto, F., Sospedra, M., Martin, R., et al. (2020). Human cd4+ t cell subsets differ in their abilities to cross endothelial and epithelial brain barriers in vitro. *Fluids and Barriers of the CNS*, 17(1):1–18. (Citado en página 20.)
- O., B. G. . L. M. (2020). Artificial intelligence in medicine: today and tomorrow. *Front Med (Lausanne)*, 7(27). (Citado en página 21.)
- Pinheiro, M. A. L., Kooij, G., Mizze, M. R., Kamermans, A., Enzmann, G., Lyck, R., Schwaninger, M., Engelhardt, B., and de Vries, H. E. (2016). Immune cell trafficking across the barriers of the central nervous system in multiple sclerosis and stroke. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1862(3):461–471. (Citado en página 20.)

- Ponce de Leon-Sanchez, E. R., Dominguez-Ramirez, O. A., Herrera-Navarro, A. M., Rodriguez-Resendiz, J., Paredes-Orta, C., and Mendiola-Santibañez, J. D. (2023a). A deep learning approach for predicting multiple sclerosis. *Micromachines*, 14(4):749. (Citado en página 28.)
- Ponce de Leon-Sanchez, E. R., Mendiola-Santibañez, J. D., Dominguez-Ramirez, O. A., Herrera-Navarro, A. M., Vazquez-Cervantes, A., Jimenez-Hernandez, H., Cordova-Esparza, D. M., Cuán Hernández, M. d. I. A., and Senties-Madrid, H. (2024). Training artificial neural networks to detect multiple sclerosis lesions using granulometric data from preprocessed magnetic resonance images with morphological transformations. *Technologies*, 12(9):145. (Citado en página 84.)
- Ponce de Leon-Sanchez, E. R., Mendiola-Santibañez, J. D., Dominguez-Ramirez, O. A., Herrera-Navarro, A. M., Vazquez-Cervantes, A., Jimenez-Hernandez, H., and Senties-Madrid, H. (2023b). Fuzzy logic system for classifying multiple sclerosis patients as high, medium, or low responders to interferon-beta. *Technologies*, 11(4):109. (Citado en página 27.)
- Prokopowicz, P., Czerniak, J., Mikołajewski, D., Apiecione, Ł., and Ślęzak, D. (2017). *Theory and applications of ordered fuzzy numbers: a tribute to Professor Witold Kosiński*. Springer Nature. (Citado en página 37.)
- Río, J., Nos, C., Tintoré, M., Borrás, C., Galán, I., Comabella, M., and Montalban, X. (2002). Assessment of different treatment failure criteria in a cohort of relapsing–remitting multiple sclerosis patients treated with interferon  $\beta$ : Implications for clinical trials. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 52(4):400–406. (Citado en página 21.)
- Robert, C. (2021). Statistics and analysis of scientific data. (Citado en página 35.)
- Rocchetti, M., Delnevo, G., Casini, L., and Mirri, S. (2021). An alternative approach to dimension reduction for pareto distributed data: a case study. *Journal of big Data*, 8(1):1–23. (Citado en página 46.)
- Rodríguez Murúa, S., Farez, M. F., and Quintana, F. J. (2022). The immune response in multiple sclerosis. *Annual Review of Pathology: Mechanisms of Disease*, 17:121–139. (Citado en página 19.)

- Rondinella, A., Crispino, E., Guarnera, F., Giudice, O., Ortis, A., Russo, G., Di Lorenzo, C., Maimone, D., Pappalardo, F., and Battiato, S. (2023). Boosting multiple sclerosis lesion segmentation through attention mechanism. *Computers in Biology and Medicine*, 161:107021. (Citado en página 107.)
- Rutkowska, D. (2001). *Neuro-fuzzy architectures and hybrid learning*, volume 85. Springer Science & Business Media. (Citado en página 36.)
- Sadoughi, F. and Arani, L. A. (2022). Intelligent computer systems for multiple sclerosis diagnosis. In *Augmenting Neurological Disorder Prediction and Rehabilitation Using Artificial Intelligence*, pages 75–105. Elsevier. (Citado en página 24.)
- Saibene, A., Assale, M., and Giltri, M. (2021). Expert systems: Definitions, advantages and issues in medical field applications. *Expert Systems with Applications*, 177:114900. (Citado en página 24.)
- Salamai, A. A., El-kenawy, E.-S. M., and Abdelhameed, I. (2021). Dynamic voting classifier for risk identification in supply chain 4.0. *CMC-COMPUTERS MATERIALS & CONTINUA*, 69(3):3749–3766. (Citado en página 49.)
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3):1–21. (Citado en páginas 42 y 43.)
- Schlaeger, S., Shit, S., Eichinger, P., Hamann, M., Opfer, R., Krüger, J., Dieckmeyer, M., Schön, S., Mühlau, M., Zimmer, C., et al. (2023). Ai-based detection of contrast-enhancing mri lesions in patients with multiple sclerosis. *Insights into imaging*, 14(1):123. (Citado en página 84.)
- Shen, D., Wu, G., and Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19(1):221–248. (Citado en página 84.)
- Soille, P. (1999). *Morphological Image Analysis: Principles and Applications*, volume 2. Springer. (Citado en páginas 87 y 92.)
- Soille, P. (2004). *Morphological Image Analysis: Principles and Applications*. Springer. (Citado en páginas 88 y 90.)
- Spagnolo, F., Depeursinge, A., Schädelin, S., Akbulut, A., Müller, H., Barakovic, M., Melie-Garcia, L., Cuadra, M. B., and Granziera, C. (2023). Automated ms lesion

- detection and segmentation in clinical workflow: a systematic review. *NeuroImage: Clinical*, page 103491. (Citado en página 84.)
- Stevens, E., Antiga, L., and Viehmann, T. (2020). *Deep learning with PyTorch*. Manning Publications. (Citado en página 45.)
- Szpakowski, P., Ksiazek-Winiarek, D., and Glabinski, A. (2021). Targeting antigen-presenting cells in multiple sclerosis treatment. *Applied Sciences*, 11(18):8557. (Citado en página 20.)
- Tao Jin, Chi Wang, S. T. (2020). Feature selection based on differentially correlated gene pairs reveals the mechanism of ifn-therapy for multiple sclerosis. *Bioinformatics and Genomics*, 8(e8812). (Citado en páginas 23, 25, 29 y 72.)
- Tarlinton, R. E., Martynova, E., Rizvanov, A. A., Khaiboullina, S., and Verma, S. (2020). Role of viruses in the pathogenesis of multiple sclerosis. *Viruses*, 12(6):643. (Citado en página 19.)
- Treaba, C. A., Conti, A., Klawiter, E. C., Barletta, V. T., Herranz, E., Mehndiratta, A., Russo, A. W., Sloane, J. A., Kinkel, R. P., Toschi, N., et al. (2021). Cortical and phase rim lesions on 7 t mri as markers of multiple sclerosis disease progression. *Brain communications*, 3(3):fcab134. (Citado en página 28.)
- Van Leekwijck, W. and Kerre, E. E. (1999). Defuzzification: criteria and classification. *Fuzzy sets and systems*, 108(2):159–178. (Citado en página 37.)
- Vasuki, A. (2020). *Nature Inspired Optimization Algorithms*. CRC Press. (Citado en páginas 25 y 48.)
- Vázquez-Marrufo, M., Sarrias-Arrabal, E., García-Torres, M., Martín-Clemente, R., and Izquierdo, G. (2021). Revisión sistemática de la aplicación de algoritmos de «machine learning» en la esclerosis múltiple. *Neurología*. (Citado en página 27.)
- Viatkin, D., Garcia-Zapirain, B., and Méndez Zorrilla, A. (2021). Deep learning techniques applied to predict and measure finger movement in patients with multiple sclerosis. *Applied Sciences*, 11(7):3137. (Citado en página 28.)
- Vichi M., Cavicchia C., G. P. J. (2022). Hierarchical means clustering. *Journal of Classification*, 39(3):553–577. (Citado en página 21.)

- Villegas-Mier, C. G., Rodriguez-Resendiz, J., Álvarez-Alvarado, J. M., Jiménez-Hernández, H., and Odry, Á. (2022). Optimized random forest for solar radiation prediction using sunshine hours. *Micromachines*, 13(9):1406. (Citado en página 41.)
- Wang, Z., Xiao, P., and Tan, H. (2023). Spinal magnetic resonance image segmentation based on u-net. *Journal of Radiation Research and Applied Sciences*, 16(3):100627. (Citado en página 107.)
- Wildner, P., Stasiołek, M., and Matysiak, M. (2020). Differential diagnosis of multiple sclerosis and other inflammatory cns diseases. *Multiple sclerosis and related disorders*, 37:101452. (Citado en página 84.)
- Zarghami, A., Li, Y., Claflin, S. B., van der Mei, I., and Taylor, B. V. (2021). Role of environmental factors in multiple sclerosis. *Expert Review of Neurotherapeutics*, 21(12):1389–1408. (Citado en página 19.)

## **8. APÉNDICE (APLICACIÓN DE UN ALGORITMO DE ML EN EL ESTUDIO DEL PROGRESO DE LA MS)**

### **8.1 Introducción**

En este trabajo de Tesis también se desarrolló otro algoritmo de ML basado en MRI para estimar el tamaño de las lesiones en pacientes con MS utilizando transformaciones morfológicas, lo cual puede ser un apoyo en el análisis del progreso de la enfermedad (Ponce de Leon-Sanchez et al., 2024). Inicialmente, la patología de la MS se definió como un proceso inflamatorio asociado con placas focales de desmielinización primaria en la sustancia blanca del cerebro y la médula espinal (Marrodan M., 2020). Un diagnóstico confiable y preciso de la MS es necesario para introducir tratamientos tempranos para la enfermedad. Actualmente, la MRI es la herramienta más común para diagnosticar la MS, detectar lesiones, comprender el progreso de la enfermedad y analizar los efectos de los tratamientos (Lladó et al., 2012; Wildner et al., 2020). La MRI es un método no invasivo para obtener imágenes detalladas de la estructura interna del cuerpo, como órganos y tejidos.

Aunque en los últimos años se han propuesto varios métodos de DL para la detección/segmentación automatizada de lesiones de MS utilizando MRI, como algoritmos computacionales basados en CNN y U-Nets (CNN desarrolladas para segmentación de imágenes) (Schlaeger et al., 2023; Shen et al., 2017; Spagnolo et al., 2023), en algunos casos, solo se ha realizado un preprocesamiento simple de datos previo al entrenamiento de los algoritmos, como el cambio de tamaño de la imagen para estandarizar las dimensiones de entrada. La morfología matemática (MM) es una técnica clásica de análisis de imágenes que ayuda a filtrar la imagen y extraer estructuras relevantes. La granulometría es una herramienta de medición de imágenes para medir la MM, que determina la distribución del tamaño de los objetos en una imagen sin segmentar explícitamente cada objeto (de Arruda et al., 2020). Por lo tanto, en este trabajo de Tesis se propone un método alternativo de preprocesamiento de MRI que incluye dos etapas:

1. La primer etapa consiste en realizar transformaciones de apertura morfológica

en MRI cerebrales (de pacientes con MS diagnosticados por expertos médicos y de individuos sanos) para eliminar el ruido y otros componentes indeseables, y luego calcular la granulometría de objetos para caracterizar las lesiones de desmielinización en la materia blanca causadas por la enfermedad. Los datos resultantes se utilizan para entrenar dos modelos de ANN para estimar el diagnóstico de la MS.

2. La segunda etapa consiste en realizar transformaciones de cierre morfológico en las MRI cerebrales (de pacientes con MS) para crear una imagen de referencia (sin lesiones) y calcular la granulometría de objetos en la imagen que contiene lesiones y en la imagen de referencia para compararlas. Luego, se estima el tamaño de las lesiones de MS calculando las diferencias en las mediciones de granulometría. Estas mediciones podrían ser apoyo a la decisión de los especialistas para determinar la progresión de la enfermedad.

## 8.2 Materiales y métodos

### 8.2.1 Base de datos

La base de datos analizada es la misma que Macin et al. (2022) y se puede descargar en <https://www.kaggle.com/datasets/buraktaci/multiple-sclerosis> (consultado el 20 de octubre de 2023). El conjunto de datos adquirido incluye 200 MRI FLAIR cerebrales (100 axiales y 100 sagitales) de pacientes con MS y personas sanas que asistieron a la Facultad de Medicina de la Universidad de Oza en 2021. Los expertos médicos analizaron las secciones de imágenes axiales y sagitales de los pacientes con MS con lesiones identificables de sustancia blanca y se asignaron a la clase de MS, y las secciones de imágenes de personas sanas con una apariencia normal se asignaron a la clase de personas sanas.

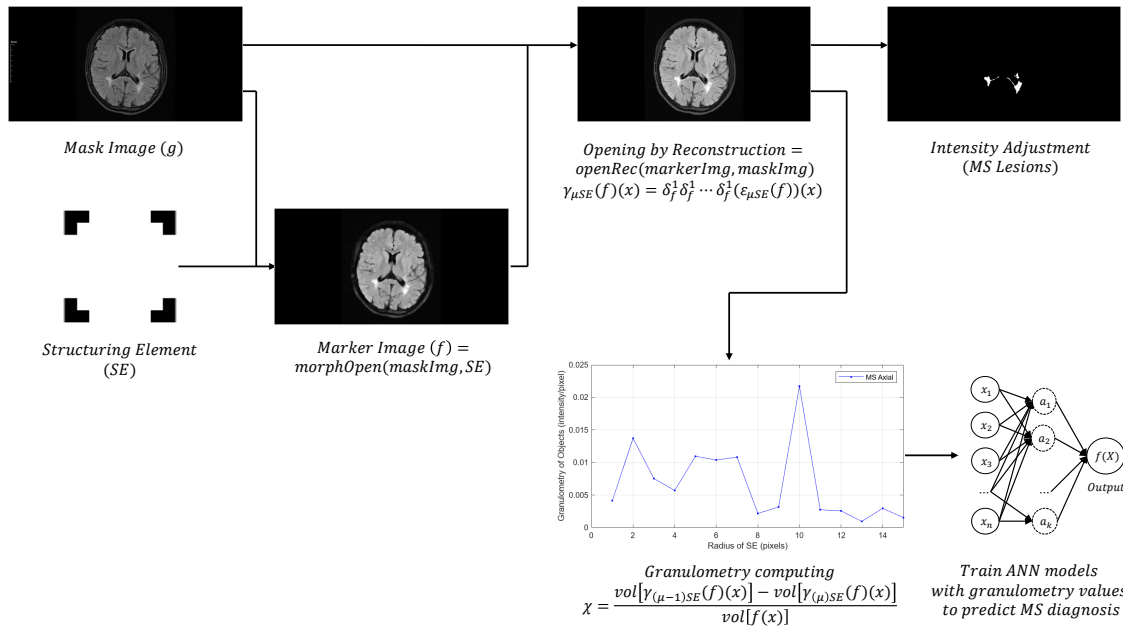
### 8.2.2 Algoritmo propuesto

El algoritmo de preprocesamiento propuesto incluye transformaciones morfológicas elementales en MRI cerebrales (axiales y sagitales) de pacientes con MS



e individuos sanos, y mediciones de granulometría para caracterizar las lesiones de MS sin afectar las dimensiones originales de las imágenes. Las figuras 8.1 y 8.2 describen el procedimiento de dos etapas del algoritmo propuesto. Las tablas 8-1 y 8-2 presentan el pseudocódigo implementado del algoritmo de preprocesamiento.

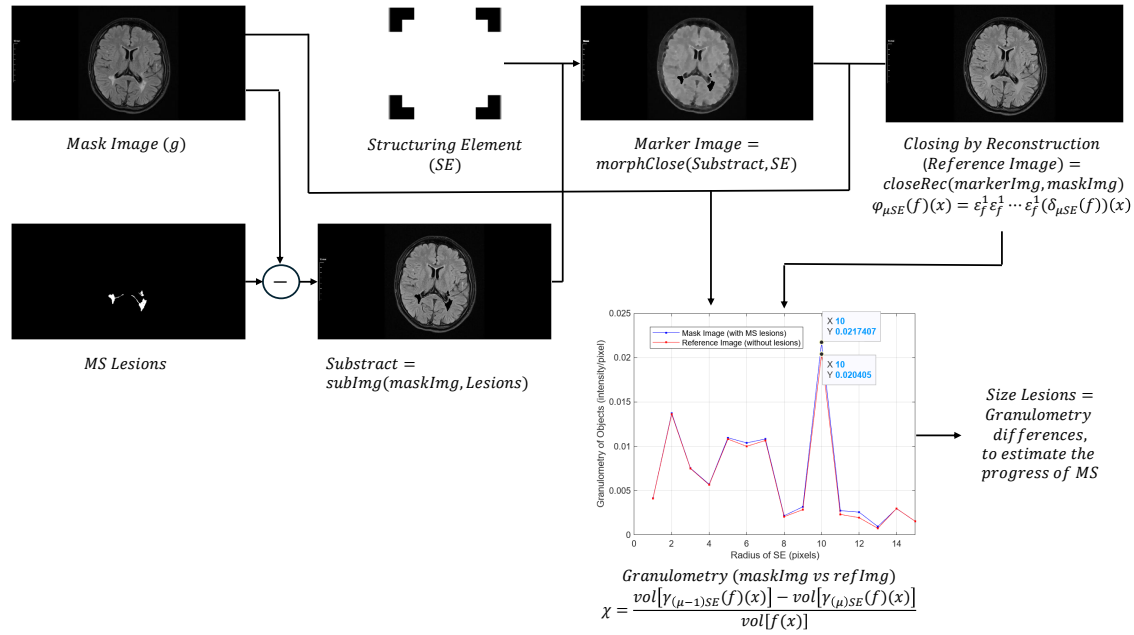
**Figura 8.1** Procedimiento de la primer etapa del algoritmo propuesto: Transformaciones morfológicas apertura en MRI de cerebros sanos y con MS, y mediciones de granulometría para estimar el diagnóstico de MS.



### 8.2.3 Morfología matemática

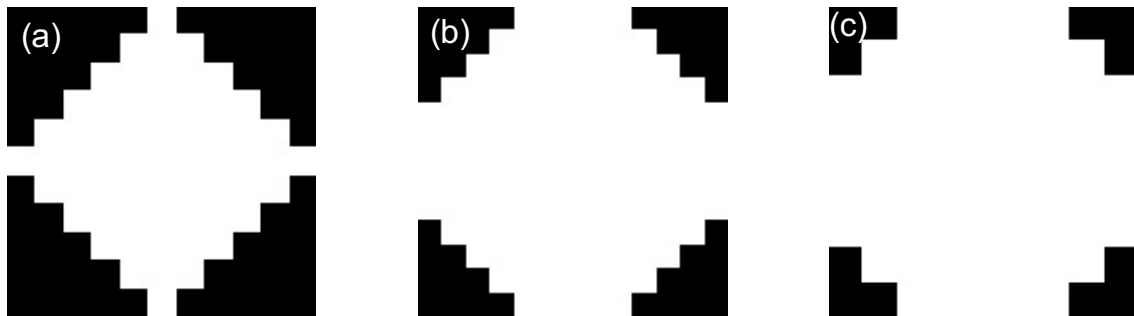
MM es una técnica de análisis de imágenes basada en la teoría de conjuntos, la geometría integral y el álgebra de redes. El objetivo de los operadores morfológicos es extraer estructuras relevantes de una imagen. Esto se logra evaluando la imagen con otro conjunto de formas conocidas llamadas elemento estructurante (SE). Los operadores morfológicos fundamentales necesitan la definición de un origen para cada SE. Este origen permite el posicionamiento del SE en un punto o píxel determinado: un SE en el punto  $x$  significa que su origen coincide con  $x$ . Los SE isótropos planos elementales se presentan en la Figura 8.3 para gráficos octagonales y de disco. Los píxeles que coinciden con cada SE plano elemental centrado en un píxel dado  $p$  corresponden a los vecinos de  $p$  en el grafo  $G$  más el propio píxel

**Figura 8.2** Procedimiento de la segunda etapa del algoritmo propuesto: Transformaciones morfológicas de cierre en MRI de cerebros de pacientes con MS y mediciones de granulometría para estimar el tamaño de las lesiones de la enfermedad.



$p$ , como  $\bar{N}_G(p)$ . La forma y el tamaño de cada SE se eligen de acuerdo a las propiedades geométricas de las estructuras relevantes e irrelevantes en la imagen. Las estructuras irrelevantes hacen referencia a ruidos u objetos que deben eliminarse (Soille, 1999).

**Figura 8.3** SE isótopos planos elementales para grafos octagonales y de disco: (a) Diamante. (b) Octágono. (c) Disco. El origen de cada SE está en su centro.



#### 8.2.4 Transformaciones geodésicas

Algunas transformaciones morfológicas incluyen combinaciones de una imagen de entrada con SE determinados. El enfoque de transformación geodésica considera dos imágenes de entrada, se aplica una transformación morfológica a la primera imagen y luego se mantiene por encima o por debajo de la segunda imagen. En este caso, las transformaciones morfológicas se limitan a erosiones y dilataciones elementales. En la práctica, las transformaciones geodésicas se iteran hasta la estabilidad (Soille, 2004).

La dilatación geodésica también involucra dos imágenes: una imagen marcador y una imagen máscara. Por definición, ambas imágenes deben tener el mismo dominio y la imagen de máscara debe ser mayor o igual que la imagen de marcador. Primero, la imagen marcador se dilata por el SE isotrópico elemental. Luego, la imagen dilatada resultante permanece por debajo de la imagen máscara. Por lo tanto, la imagen máscara actúa como un límite a la propagación de la dilatación de la imagen marcador. Sea  $f$  la imagen marcador y  $g$  la imagen de máscara ( $D_f = D_g$  y  $f \leq g$ ). La dilatación geodésica de tamaño 1 de la imagen marcador  $f$  respecto a la imagen máscara  $g$  se denota por la Ecuación (8.1) y se define como el punto mínimo entre la imagen de máscara y la dilatación elemental  $\delta^{(1)}$  de la imagen marcador. La dilatación geodésica de una imagen binaria se muestra en la Figura 8.4.

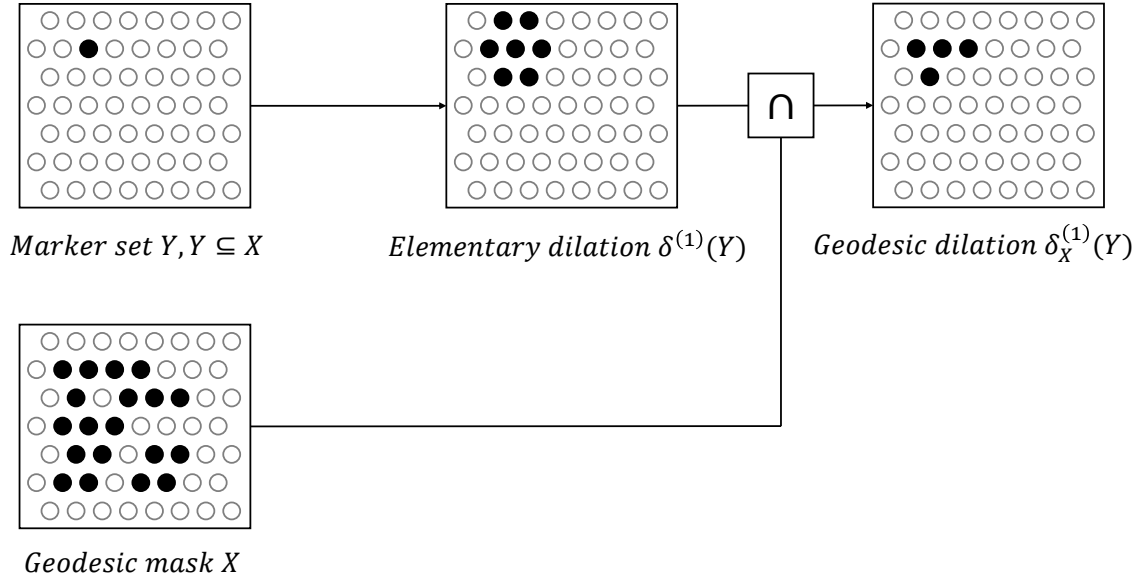
$$\delta_g^{(1)}(f) = \delta^{(1)}(f) \wedge g. \quad (8.1)$$

La erosión geodésica es la doble transformación de la dilatación geodésica respecto a la complementación de conjuntos:

$$\varepsilon_g^{(1)}(f) = \frac{[\delta^{(1)}(f^c) \wedge g^c]^c}{\varepsilon^{(1)}(f) \vee g}, \quad (8.2)$$

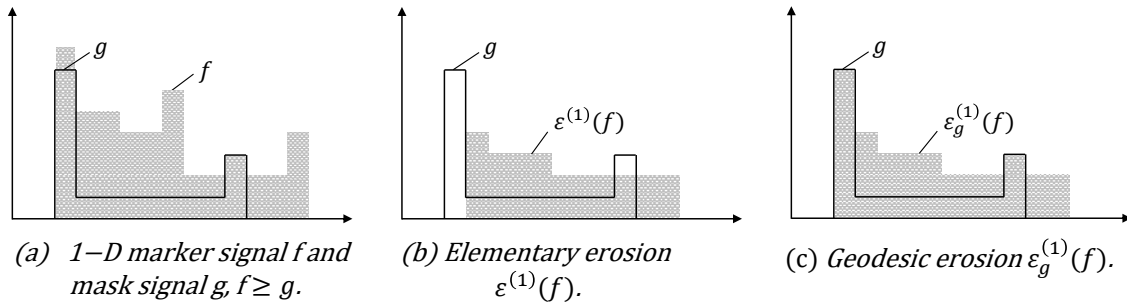
donde  $f \geq g$  y  $\varepsilon^{(1)}$  es la erosión elemental. Primero, se erosiona la imagen marcador y luego se calcula el punto máximo con la imagen máscara. La figura 8.5

**Figura 8.4** Dilatación geodésica de una imagen binaria o un conjunto  $Y$  dentro de una máscara geodésica  $X$ . El conjunto marcador se dilata primero mediante el SE isótropo elemental y luego se intersecta con la máscara geodésica:  $\delta_X^{(1)}(Y) = \delta^{(1)}(Y) \cap X$ .



muestra un ejemplo de erosión geodésica. Aquí, la imagen de máscara actúa como un límite para la reducción de la imagen marcador.

**Figura 8.5** Erosión geodésica de una señal marcador  $f$  1D con respecto a una señal máscara  $g$ . Debido al operador máximo puntual, todos los píxeles de la erosión elemental  $f$  con valores inferiores a  $g$  se establecen en el valor de  $g$ .



## 8.2.5 Reconstrucción morfológica

En la práctica, cuando se iteran transformaciones geodésicas hasta la estabilidad, se pueden crear algoritmos de reconstrucción morfológica muy eficientes.

Las transformaciones geodésicas de imágenes acotadas siempre convergen después de un período finito de iteraciones (es decir, hasta que la imagen máscara obstruye por completo la propagación de la imagen marcador y evita que se reduzca). La reconstrucción morfológica de una imagen máscara  $g$  a partir de una imagen marcador  $f$  se basa en este principio. Las transformaciones geodésicas por reconstrucción permiten la eliminación de algunos componentes no deseados sin afectar considerablemente las estructuras restantes (Soille, 2004).

Al igual que las dilataciones y erosiones geodésicas, la iteración de la transformación geodésica autodual  $v$  siempre logra estabilidad para un dominio de definición de una imagen acotada. La transformación correspondiente se conoce como reconstrucción autodual de una imagen máscara  $g$  a partir de una imagen marcador  $f$ , y se define como:

$$R_g^v(f) = v_g^{(i)}(f), \quad (8.3)$$

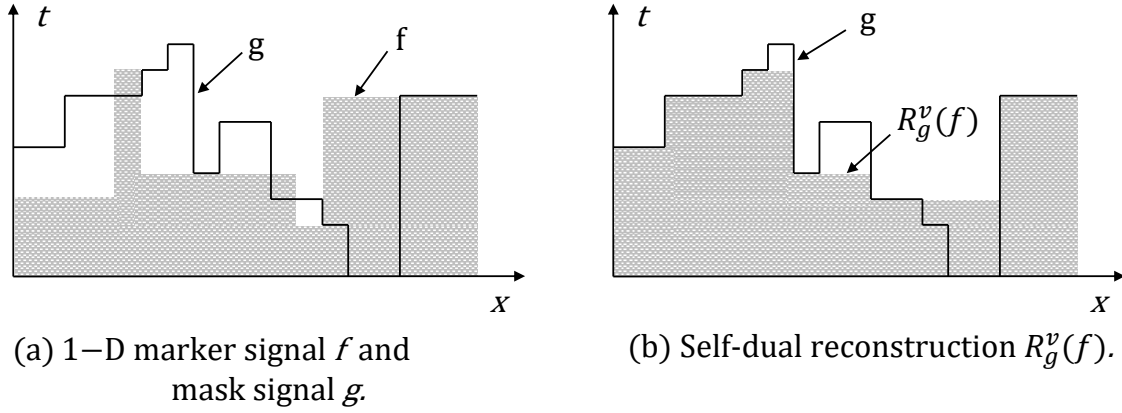
donde  $i$  existe de manera que  $v_g^{(i+1)}(f) = v_g^{(i)}(f)$ . La Figura 8.6 muestra una reconstrucción autodual. En la práctica, la reconstrucción autodual se puede obtener calculando la reconstrucción por erosión y dilatación en paralelo, determinando la salida en un píxel dado  $x$  alternando entre la reconstrucción por erosión y dilatación dependiendo de si  $f(x)$  está por encima o por debajo de  $g(x)$ :

$$[R_g^v(f)](x) = \begin{cases} [R_g^\delta(f)](x), & \text{if } f(x) \leq g(x) \\ [R_g^\varepsilon(f)](x), & \text{otherwise.} \end{cases} \quad (8.4)$$

### 8.2.6 Apertura y cierre por reconstrucción

Para la reconstrucción morfológica, la erosión y la dilatación de tamaño 1 se iteran hasta que se logra la estabilidad. La dilatación geodésica  $\delta_f^1(g)$  y la erosión geodésica  $\varepsilon_f^1(g)$  de tamaño 1 se dan por  $\delta_f^1(g) = f \wedge \delta(g)$  con  $g \leq f$ , y  $\varepsilon_f^1(g) = f \vee \varepsilon(g)$  con  $g \geq f$ , respectivamente. Cuando la función  $g$  es igual a la erosión o dilatación morfológica de la función original, la apertura por reconstrucción  $\tilde{\gamma}_{\mu B}(f)(x)$  se da por,

**Figura 8.6** Reconstrucción morfológica autodual  $R^v$  de una imagen máscara  $g$  a partir de una imagen marcador  $f$ .



$$\tilde{\gamma}_{\mu B}(f)(x) = \underbrace{\delta_f^1 \delta_f^1 \dots \delta_f^1 (\varepsilon_{\mu B}(f))}_{\text{Until stability}}(x), \quad (8.5)$$

o el cierre por reconstrucción  $\tilde{\varphi}_{\mu B}(f)(x)$  se da por,

$$\tilde{\varphi}_{\mu B}(f)(x) = \underbrace{\varepsilon_f^1 \varepsilon_f^1 \dots \varepsilon_f^1 (\delta_{\mu B}(f))}_{\text{Until stability}}(x), \quad (8.6)$$

donde  $B$  representa el SE y  $\mu$  es un parámetro de tamaño. Para esta investigación, se eligió un SE en forma de disco con radio  $r = \mu$  ( $rB = \mu B$ ). Por ejemplo, un disco de radio  $r = 5$  representa un disco de 9x9 píxeles; por lo tanto, se analizan 9 vecinos. Por lo tanto, un SE de tamaño  $n$  consiste en un disco de  $(2n-1) \times (2n-1)$  píxeles.

### 8.2.7 Medidas de imagen

Las mediciones de imagen tienen como objetivo caracterizar objetos en una imagen mediante el cálculo de algunos valores numéricos. Para un criterio dado, la medición es discriminante si los valores obtenidos para los objetos que satisfacen este criterio difieren de los demás. La MM proporciona varias herramientas de medición de imágenes, como la espectrografía de patrones o granulometría, análisis de dirección, análisis de textura, descripción de formas, etc. Estas medidas definen

un vector de características que se puede utilizar como entrada para un análisis estadístico o un método de clasificación (Soille, 1999).

La granulometría determina la distribución del tamaño de los objetos en una imagen sin segmentar (detectar) explícitamente cada objeto, y se utiliza en algunas áreas para describir las características del tamaño y la forma de los gránulos individuales dentro de un objeto (Mendiola J., 2007). La definición formal de este concepto se presenta a continuación.

Sea  $(\psi_\lambda)_{\lambda \in \mathbb{Z}}$  una familia de transformación que depende de un único parámetro positivo  $\lambda$ . Esta familia comprende una granulometría si, y solo si, se cumplen las siguientes propiedades: (i)  $\forall \lambda$  es positivo y  $\Psi_\lambda$  es creciente; (ii)  $\forall \lambda$  es positivo y  $\Psi_\lambda$  es antiextensional; (iii)  $\forall \lambda$  y  $\mu$  son positivos y  $\Psi_\lambda \Psi_\mu = \Psi_{\max(\lambda, \mu)}$ . En particular, la familia de aperturas y cierres morfológicos para la instancia numérica  $\gamma_{\mu B}$ ,  $\varphi_{\mu B}$  con  $\mu = \{1, 2, \dots, n\}$  satisface la definición anterior.

La medida granulométrica asociada con las regiones claras se denota como  $\chi$ , y se calcula por,

$$\chi = \frac{\text{vol}[\gamma_{(\mu-1)B}(f)(x)] - \text{vol}[\gamma_{\mu B}(f)(x)]}{\text{vol}[f(x)]}, \quad (8.7)$$

donde *vol* se refiere al número total de píxeles de una imagen. La medida granulométrica asociada a las regiones oscuras se denota como  $\zeta$  y se calcula por,

$$\zeta = \frac{\text{vol}[\varphi_{\mu B}(f)(x)] - \text{vol}[\varphi_{(\mu-1)B}(f)(x)]}{\text{vol}[f(x)]}. \quad (8.8)$$

### 8.3 Resultados

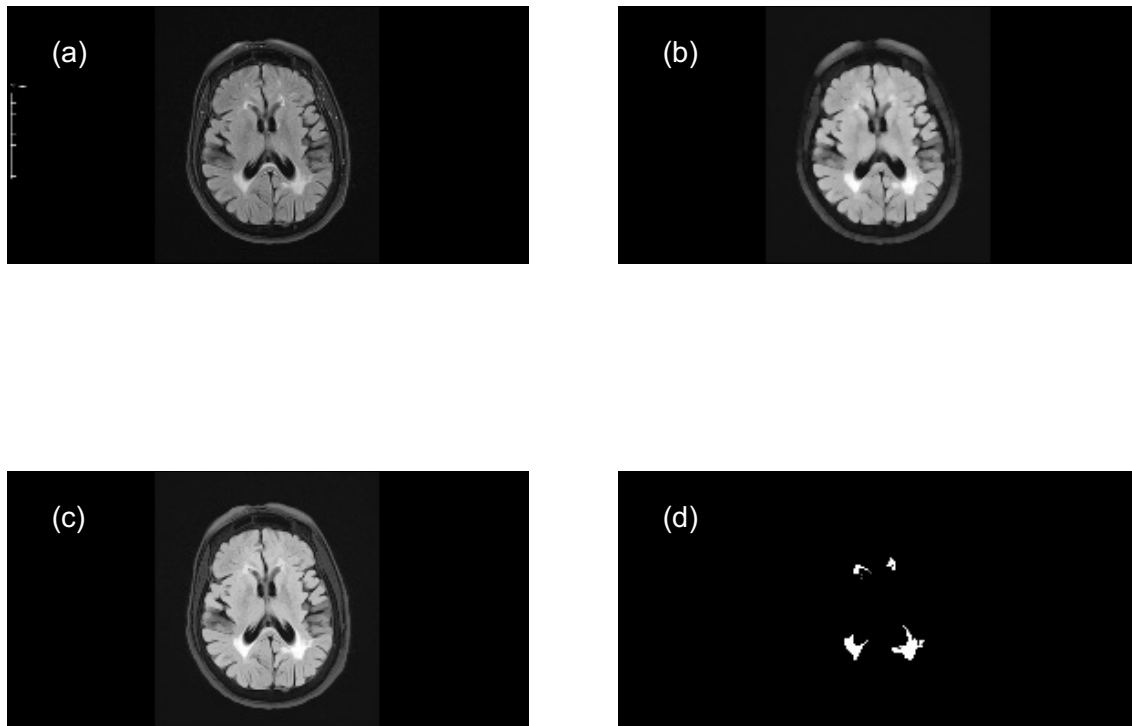
El algoritmo de preprocesamiento de MRI propuesto se implementó en 100 imágenes de muestra (axiales) de pacientes con MS y 100 imágenes de muestra (sagitales) de individuos sanos. Se calculó la granulometría de objetos en las imágenes cerebrales para adquirir información relevante sobre algunas estructuras. En particular, se utilizó la transformación de apertura morfológica como un filtro que

satisface la definición de granulometría y que afecta principalmente a las regiones claras directamente asociadas con lesiones de MS.

### 8.3.1 Algoritmo (Etapa 1)

En la primera etapa, se implementó el algoritmo de preprocesamiento para todas las imágenes de pacientes con MS y e individuos sanos (MRI axial y saggital). Las figuras 8.7 y 8.8 describen los resultados obtenidos con las transformaciones morfológicas de apertura aplicadas.

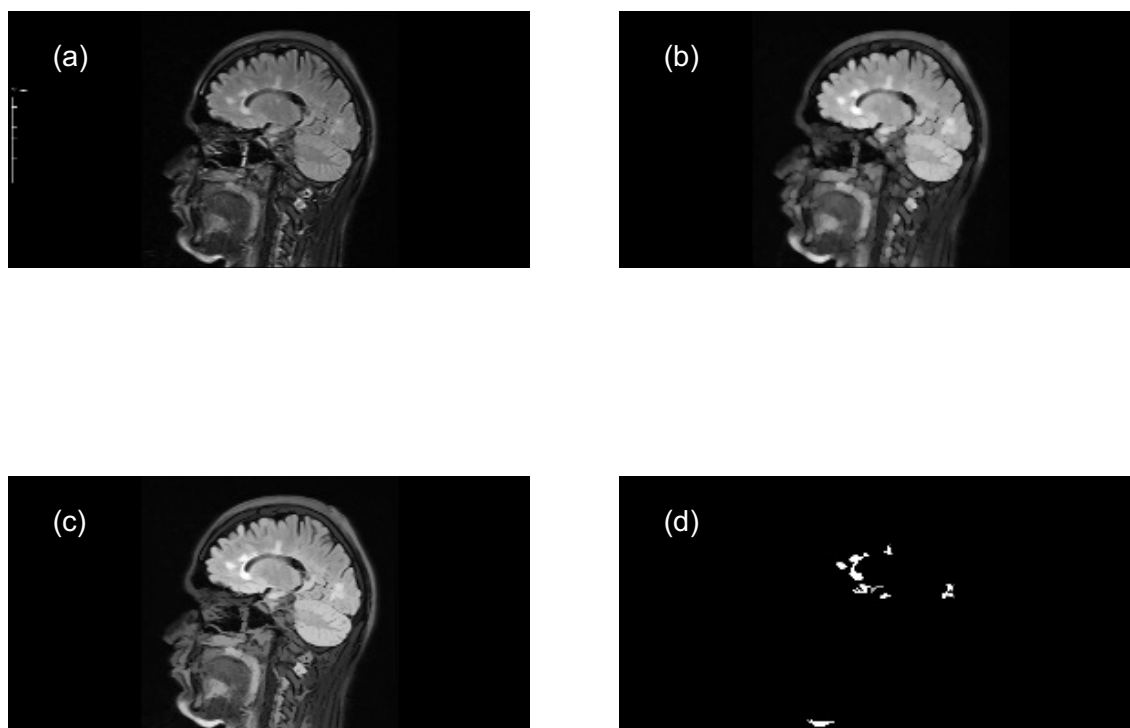
**Figura 8.7** Implementación de transformaciones morfológicas de apertura: (a) Imagen de máscara (MS axial). (b) Apertura morfológica. (c) Apertura por reconstrucción. (d) Ajuste de intensidad. Uno de los objetivos de la primera etapa fue resaltar las áreas de lesión en las muestras de MS (MRI axial) utilizando un SE en forma de disco de radio  $r = 5$ .



Luego, se calculó la granulometría de los objetos para todas las muestras de MRI para SE de diferentes valores de radio ( $r = 1, 2, 3, \dots, 15$ ). La figura 8.9 compara los resultados de la medición de granulometría entre las imágenes cerebrales de dos pacientes con MS (axial y saggital) y las de dos individuos sanos (axial y saggital)



**Figura 8.8** Implementación de transformaciones morfológicas de apertura: (a) Imagen de máscara (MS sagittal). (b) Apertura morfológica. (c) Apertura por reconstrucción. (d) Ajuste de intensidad. Uno de los objetivos de la primera etapa fue resaltar las áreas de lesión en las muestras de MS (MRI sagittal) utilizando un SE en forma de disco de radio  $r = 5$ .



con dimensiones similares ( $569 \times 1158$  píxeles, respectivamente).

Después de calcular la granulometría, los datos resultantes se ingresaron en dos matrices ( $100 \times 15$  muestras) para entrenar a dos modelos ANN. La figura 8.10 muestra los valores de la suma de granulometría para 80 muestras de MRI (axial y sagittal).

Luego, se calcularon las métricas de rendimiento coeficiente de similitud cubo (DSC), sensibilidad y especificidad utilizando los resultados de la matriz de confusión para evaluar los modelos ANN. La Figura 8.11 muestra las clasificaciones de las predicciones de las ANN.

Además, se calculó la pérdida de entropía cruzada con diferentes valores de fuerza de regularización; los resultados correspondientes se muestran en la Figura 8.12.

Se entrenaron dos modelos ANN utilizando las mejores fuerzas de regularización  $\lambda$ . Las Figuras 8.13 y 8.14 muestran el comportamiento de la función de pérdida a lo largo de diferentes iteraciones.

La Tabla 8-3 compara los resultados de las métricas de rendimiento entre los modelos ANN implementados.

### 8.3.2 Algoritmo (Etapa 2)

En la segunda etapa del algoritmo implementado, las áreas de lesiones de MS (adquiridas en el paso de ajuste de intensidad) se restaron de la imagen máscara (imagen original) y se calculó la transformación de cierre morfológico de la imagen resultante (con huecos de lesión). Luego, la imagen fue cerrada por reconstrucción para rellenar los huecos y se creó una imagen de referencia (sin lesiones) para compararla con la imagen máscara (con lesiones). El procedimiento anterior se aplicó a dos imágenes muestra de MS (axial y sagittal), y los resultados se muestran en las Figuras 8.15 y 8.16.

Finalmente, se calculó la granulometría de los objetos en la imagen máscara y en la imagen de referencia. Las figuras 8.17 y 8.18 muestran los resultados de la granulometría para dos imágenes muestra de MS (axial y sagittal). Las diferencias observadas demuestran que es posible estimar el tamaño de las lesiones de MS calculando la granulometría de objetos de una imagen máscara (con lesiones) y una imagen de referencia (sin lesiones).

## 8.4 Discusión

La segmentación es una técnica útil para analizar e identificar tejidos dañados u otras anomalías en MRI y para diferentes tareas como el diagnóstico de enfermedades (Liu et al., 2021). Aunque se han propuesto varios modelos computacionales para la segmentación automática de lesiones de MS en MRI, en algunos casos, solo se ha realizado un preprocesamiento simple de datos (cambio de tamaño de la imagen para estandarizar las dimensiones de entrada) antes del entrenamiento del

algoritmo (Bandyopadhyay et al., 2021; de Arruda et al., 2020; Ghosh et al., 2021; La Rosa et al., 2021). Por lo tanto, en este trabajo de investigación se propone un algoritmo de preprocesamiento de MRI basado en transformaciones morfológicas elementales y cálculo de granulometría. El algoritmo ayuda a filtrar la imagen eliminando componentes no deseados sin afectar sus dimensiones originales. Además, mejora la caracterización de las lesiones de MS al extraer solo las estructuras relevantes.

Para la primer etapa del algoritmo propuesto, se realizaron transformaciones de apertura morfológica en las imágenes cerebrales para identificar objetos de alta intensidad (objetos claros que pueden asociarse con lesiones de MS), seguido de un ajuste de intensidad para aumentar el contraste y extraer las estructuras relevantes. Además, se calculó la granulometría de objetos en las imágenes preprocesadas. Los resultados de granulometría de la Figura 8.9 muestran que los valores de intensidad para cerebros con MS son más altos que los de cerebros sanos, lo que significa que estas mediciones están directamente asociadas con las lesiones. Luego, estos valores de granulometría se utilizaron para entrenar a dos modelos ANN para predecir el diagnóstico de MS. Los resultados de las métricas de rendimiento de la Tabla 8-3 prueban que las mediciones de granulometría funcionan como predictores eficientes para entrenar a modelos ANN. De esta manera, este algoritmo puede respaldar las decisiones de los especialistas para diagnosticar MS en base al cálculo de granulometría. Además, el comportamiento de la función de pérdida de las Figuras 8.13 y 8.14 indica que no es necesario dedicar tanto tiempo a ajustar los hiperparámetros de un modelo de aprendizaje; es posible minimizar la pérdida ajustando únicamente el hiperparámetro de fuerza de regularización ( $\lambda$ ). La Tabla 8-4 presenta los resultados de rendimiento obtenidos en otros estudios relacionados con el análisis de lesiones de MS en MRI. Los resultados obtenidos en este trabajo de investigación muestran una alta precisión (0.9753) y el valor DSC más alto (1.0) en comparación con otros estudios. Una limitación de este artículo es que algunas MRI disponibles en la base de datos analizada tenían diferentes dimensiones, por lo que solo se eligieron 100 imágenes de pacientes con MS (axial y sagittal) y 100 imágenes de individuos sanos (axial y sagittal), con dimensiones de  $569 \times 1158 \times 3$  píxeles.

**Tabla 8-1** *Pseudocódigo implementado del algoritmo de preprocesamiento de MRI (Etapa 1).*

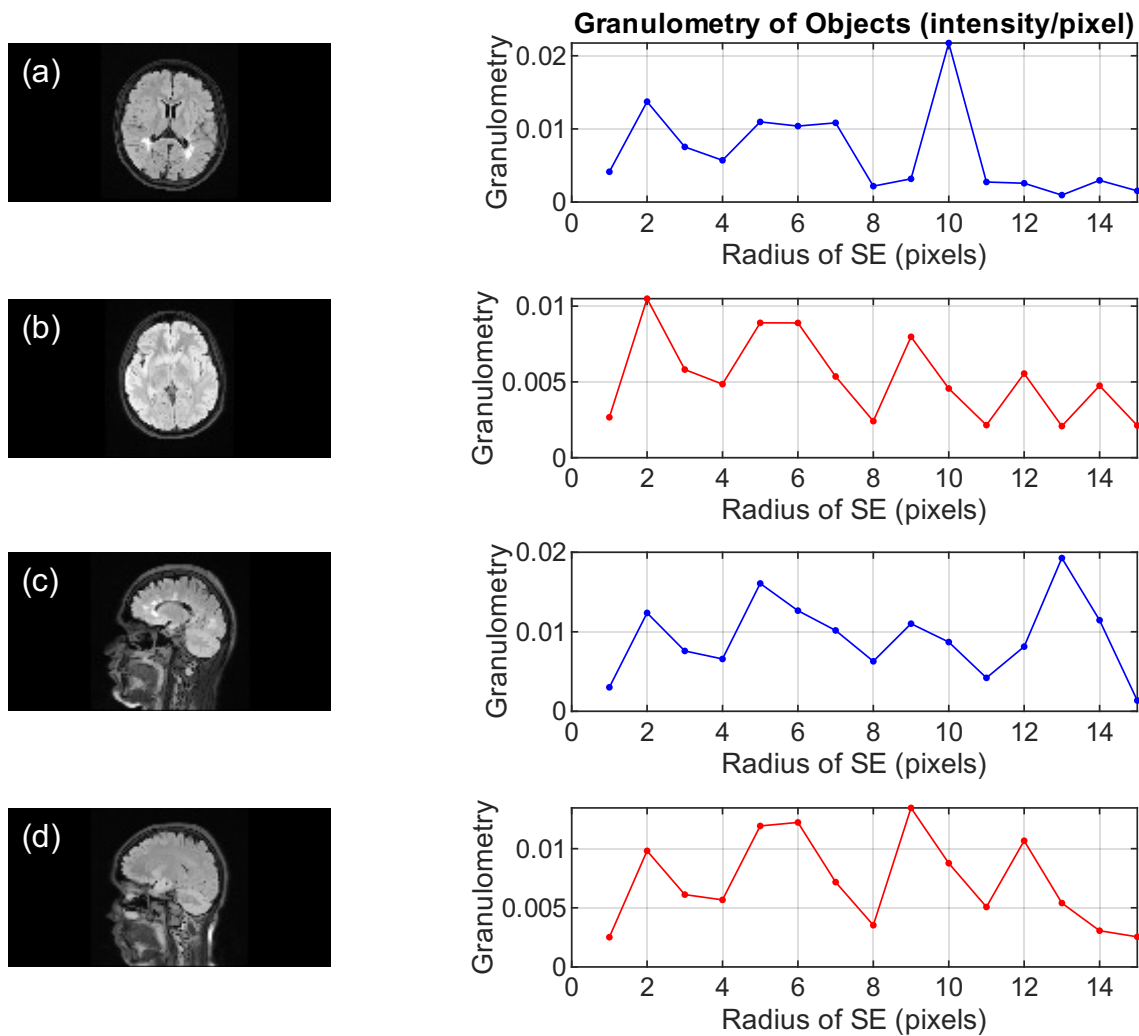
<b>Pseudocódigo</b>	
	Start
	rgb = readImg('sample'.png)
	grayScale = rgbTogray(rgb)
	maskImg = grayScale
	markerImg1 = morphOpen(maskImg, SE('disk', 5))
	openRec1 = openRec(markerImg1, maskImg)
	MSlesions = intensityAdjust(openRec1)
	volMaskImg1 = 1.0*sum(maskImg)
For	radius = 1:15
	markerImg2 = morphOpen(maskImg, SE('disk', radius - 1))
	openRec2 = openRec(markerImg2, maskImg)
	volOpenRec1 = sum(openRec2)
	markerImg3 = morphOpen(maskImg, SE('disk', radius))
	openRec3 = openRec(markerImg3, maskImg)
	volOpenRec2 = sum(openRec3)
	volGranu(radius) = (volOpenRec1 - volOpenRec2)/volMaskImg1
	End

**Tabla 8-2** *Pseudocódigo implementado del algoritmo de preprocesamiento de MRI (Etapa 2).*

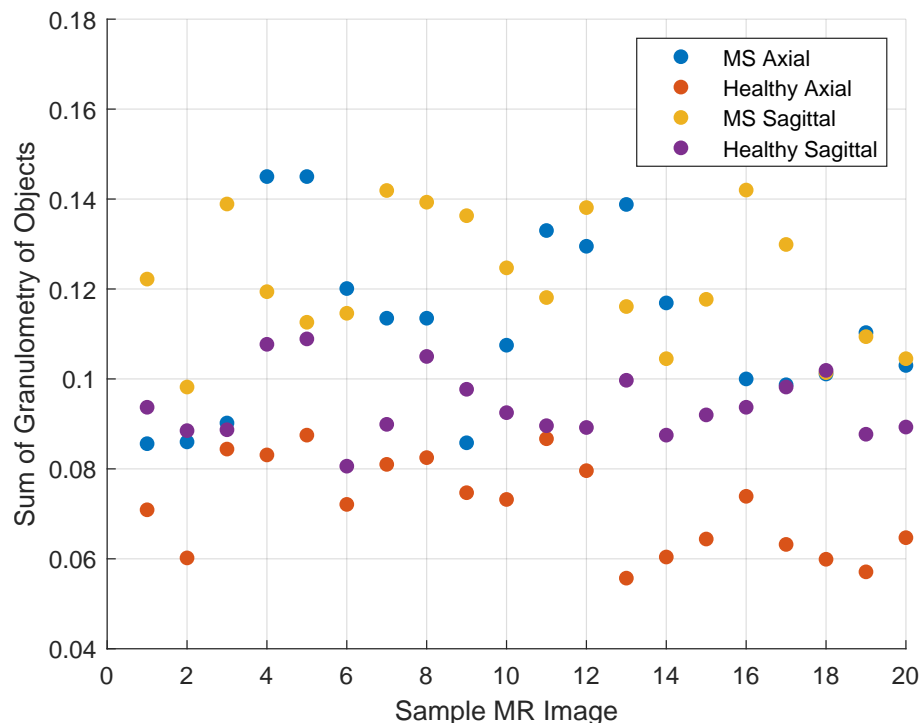
<b>Pseudocódigo</b>	
	Start
	subtract = imgSub(maskImg, MSLesions)
	markerImg4 = morphClose(subtract, SE('disk', radius = 5))
	closeRec1 <sup>1</sup> = closeRec(markerImg4, maskImg)
	volMaskImg2 = 1.0*sum(closeRec1)
For	radius = 1:15
	markerImg5 = morphOpen(closeRec1, SE('disk', radius – 1))
	openRec5 = openRec(markerImg5, closeRec1)
	volOpenRec3 = sum(openRec5)
	markerImg6 = morfOpen(closeRec1, SE('disk', radius))
	openRec6 = openRec(markerImg6, closeRec1)
	volOpenRec4 = sum(openRec6)
	volGranu(radius) = (volOpenRec3 – volOpenRec4)/volMaskImg2
	End

<sup>1</sup>Imagen de referencia (sin lesiones de MS).

**Figura 8.9** Comparación de las mediciones de granulometría en imágenes cerebrales preprocesadas utilizando el algoritmo propuesto (Etapa 1): (a) Imagen axial de MS (SumVolGranu = 0,1011). (b) Imagen axial sana (SumVolGranu = 0,0786). (c) Imagen saggital de MS (SumVolGranu = 0,1389). (d) Imagen saggital sana (SumVolGranu = 0,1079). La suma de valores de intensidad de objetos (sumVolGranu) en los cerebros de MS es mayor que en los cerebros sanos, debido a la presencia de lesiones.



**Figura 8.10** Comparación de la suma de granulometrías de objetos. La suma de los valores de granulometría en muestras de MS (axial y sagittal) es mayor que en muestras sanas por la presencia de las lesiones.



**Tabla 8-3** Resultados de las métricas de rendimiento. La precisión de prueba es la proporción de predicciones correctas, el DSC mide la similitud entre los datos de prueba y las predicciones, el TPR es la proporción de positivos clasificados correctamente, el TNR es la proporción de negativos clasificados correctamente y la pérdida de entropía cruzada calcula la pérdida entre los datos de prueba y las predicciones. El rendimiento de la ANN entrenada con imágenes axiales fue superior al de la ANN entrenada con imágenes sagittales.

Modelo	Precisión	DSC <sup>1</sup>	TPR <sup>2</sup>	TNR <sup>3</sup>	Pérdida de entropía cruzada
ANN (axial)	0.9753	1.0	1.0	1.0	0.0247
ANN (sagittal)	0.9197	0.888	1.0	0.833	0.0803

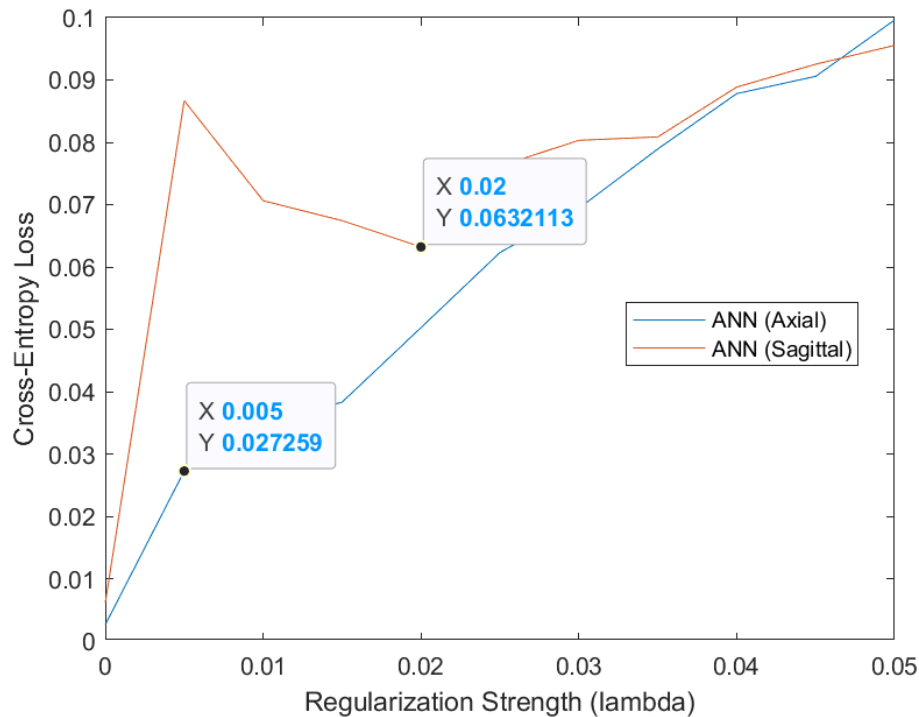
<sup>1</sup> Dice similarity coefficient. <sup>2</sup> True positive rate (sensitivity). <sup>3</sup> True negative rate (specificity).

**Figura 8.11** Matriz de confusión de predicciones. Resultados de ANN (axial):  $TP = 10$  (verdaderos positivos clasificados correctamente como MS),  $TN = 10$  (verdaderos negativos clasificados correctamente como sanos),  $FP = 0$  (falsos positivos clasificados correctamente como MS) y  $FN = 0$  (falsos negativos clasificados correctamente como sanos). Resultados de ANN (sagittal):  $TP = 8$  (verdaderos positivos clasificados correctamente como MS),  $TN = 10$  (verdaderos negativos clasificados correctamente como sanos),  $FP = 2$  (falsos positivos clasificados incorrectamente como MS) y  $FN = 0$  (falsos negativos clasificados correctamente como sanos).

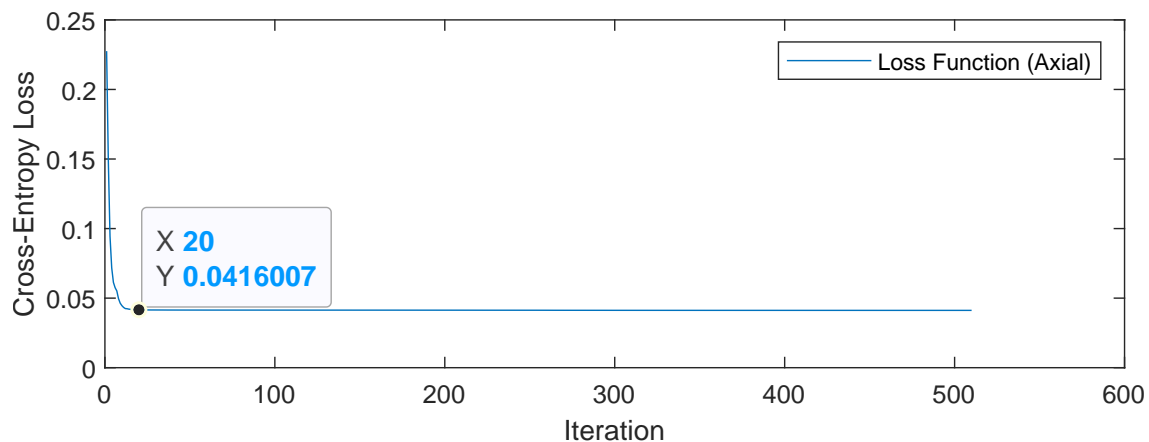




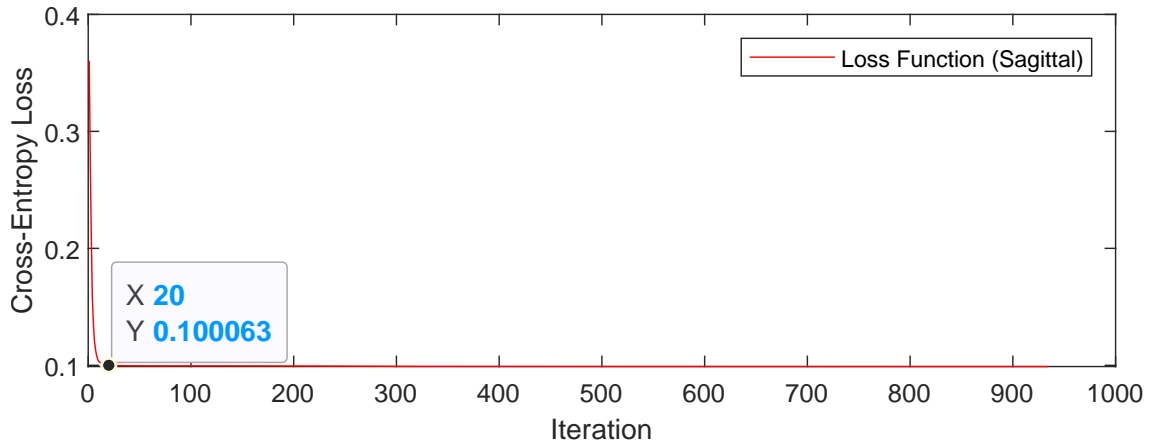
**Figura 8.12** Gráfico de la función de pérdida de entropía cruzada frente a la fuerza de regularización. Los hiperparámetros  $\lambda$  correspondientes,  $X = 0,005$  (axial) y  $X = 0,02$  (sagittal), se ajustaron para minimizar las funciones de pérdida. En estos puntos, las funciones de pérdida alcanzaron los valores más bajos:  $Y = 0,0272$  (axial) e  $Y = 0,0632$  (sagittal), respectivamente.



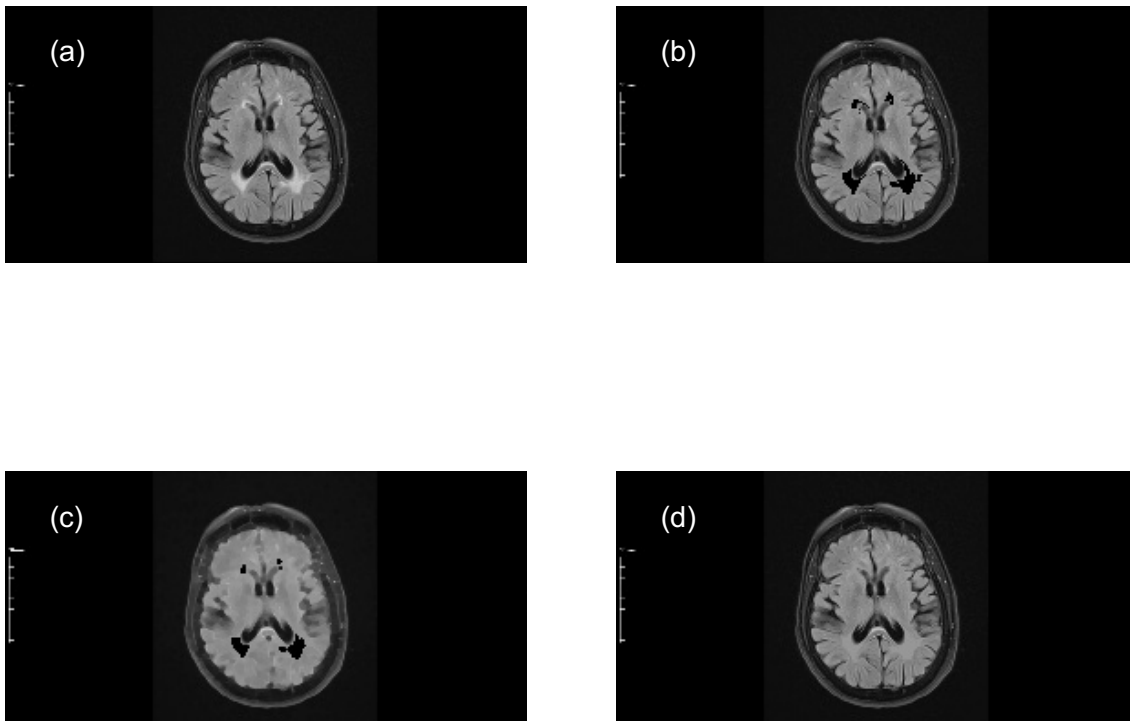
**Figura 8.13** Comportamiento de la función de pérdida de entropía cruzada de la ANN (axial). En la iteración 20 ( $X = 20$ ), la función de pérdida convergió aproximadamente en  $Y = 0,0416$ .



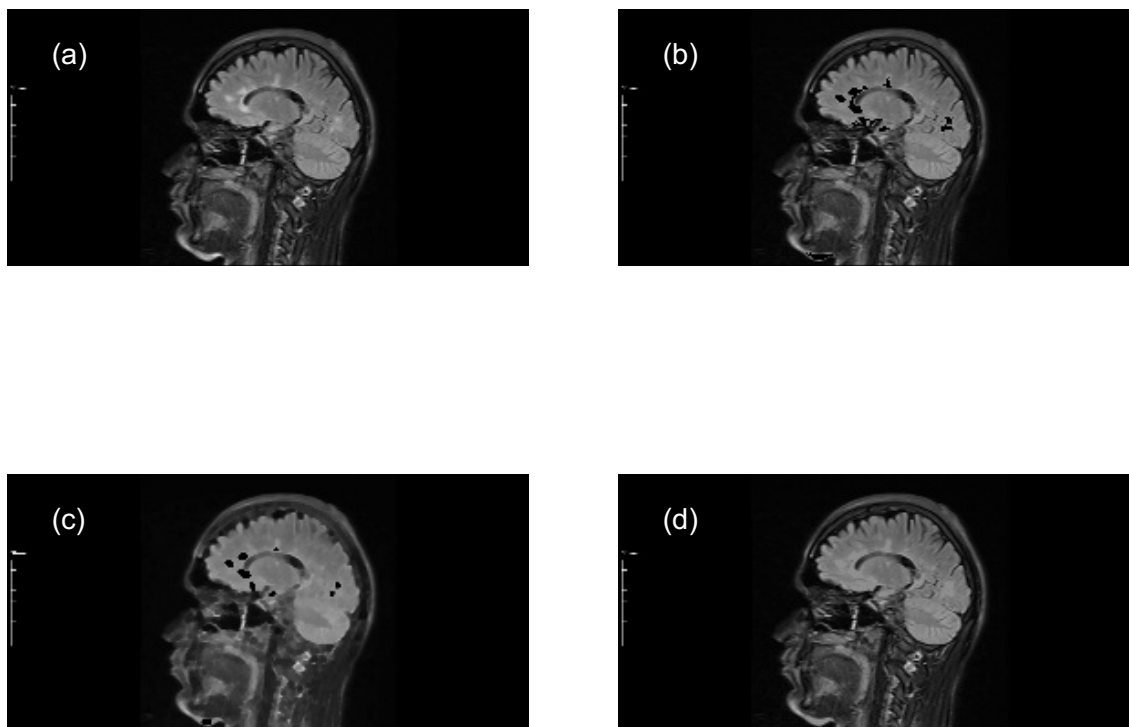
**Figura 8.14** Comportamiento de la función de pérdida de entropía cruzada de ANN (saggital). En la iteración 20 ( $X = 20$ ), la función de pérdida convergió aproximadamente en  $Y = 0,100$ . Por lo tanto, la función de pérdida de la ANN axial alcanzó el valor más bajo en comparación con la función de pérdida de la ANN saggital.



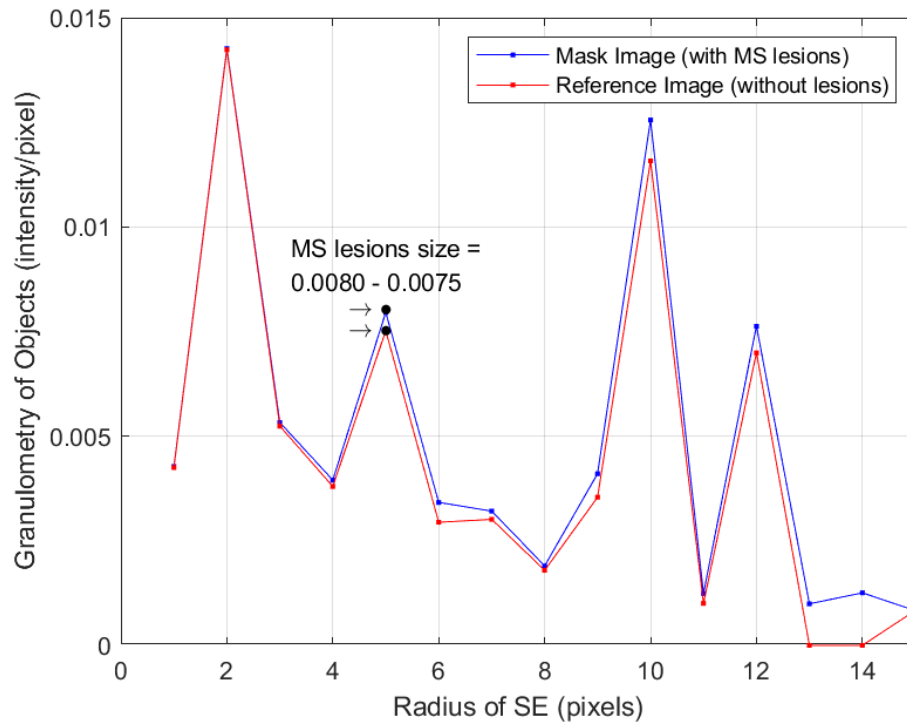
**Figura 8.15** Implementación de transformaciones morfológicas de cierre (MS axial): (a) Imagen máscara (MS axial). (b) Imagen máscara sin lesiones de MS. (c) Cierre morfológico. (d) Cierre por reconstrucción (imagen de referencia o aproximación de cerebro sano).



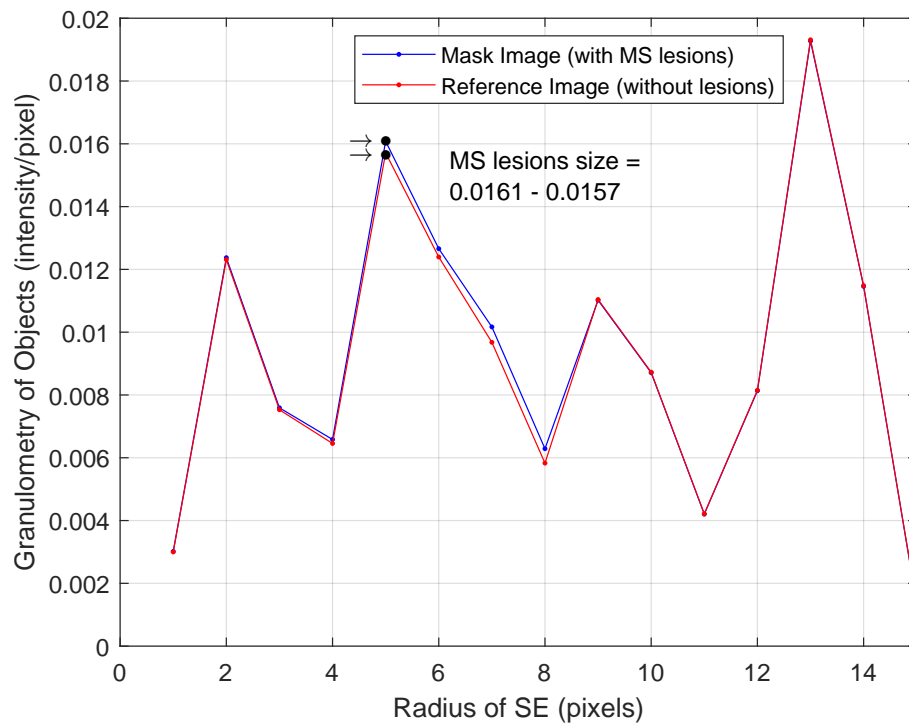
**Figura 8.16** Implementación de transformaciones morfológicas de cierre (MS sagittal): (a) Imagen máscara (MS sagittal). (b) Imagen máscara sin lesiones de MS. (c) Cierre morfológico. (d) Cierre por reconstrucción (imagen de referencia o aproximación de cerebro sano).



**Figura 8.17** Mediciones de granulometría (axial) de la imagen máscara y la imagen de referencia utilizando un SE en forma de disco con diferentes valores de radio ( $r = 1, 2, 3, \dots, 15$ ). El tamaño de las lesiones de MS se estima calculando la diferencia entre cada punto de datos.



**Figura 8.18** Medidas de granulometría (saggital) de la imagen máscara y la imagen de referencia utilizando un SE en forma de disco con diferentes valores de radio ( $r = 1, 2, 3, \dots, 15$ ).



**Tabla 8-4** Comparación de resultados de rendimiento en diferentes estudios que analizan lesiones de MS.

Autor	Técnica		Precisión	DSC <sup>1</sup>
	Preprocesamiento Imagen	Clasificador		
Ghosh et al. (2021)	CEN	U-Net, U-Net++, Linknet	-	0.7159
Macin et al. (2022)	ExMPLPQ	kNN	0.9837 <sup>2</sup> 0.9775 <sup>3</sup>	- -
de Oliveira et al. (2022)	Lesion volume quantification	CNN	0.9969	0.9786
Acar et al. (2022)	CNN	CNN	0.98 <sup>4</sup> 0.903 <sup>5</sup>	- -
Hashemi et al. (2022)	Attention U-Net	Modified U-Net	-	0.823
Wang et al. (2023)	U-Net	U-Net++	-	0.88
Rondinella et al. (2023)	Augmented U-Net	LSTM	-	0.89
Algoritmo propuesto	Morphology and granulometry	ANN	0.9753 <sup>2</sup> 0.9197 <sup>3</sup>	1.0

<sup>1</sup> Dice similarity coefficient. <sup>2</sup> Axial. <sup>3</sup> Sagittal. <sup>4</sup> Cutting-level. <sup>5</sup> Patient-level.