



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

SISTEMA BASADO EN INTELIGENCIA ARTIFICIAL PARA EL ANÁLISIS DE TEXTURAS MUSCULARES EN PACIENTES CON ENFERMEDADES NEURODEGENERATIVAS (ELA) A PARTIR DE BIOMARCADORES ECOGRÁFICOS

Autor: Candela Paradinas Solera

Director: Constantino Malagón Luque

Madrid

Declaración de originalidad

Declaro bajo mi responsabilidad que el Proyecto presentado con el título **[Sistema basado en Inteligencia Artificial para el análisis de texturas musculares en pacientes con enfermedades neurodegenerativas (ELA) a partir de biomarcadores ecográficos]** e la ETS de Ingeniería – ICAI de la Universidad Pontificia Comillas en el curso académico 2025/2026 es de mi autoría y no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido tomada de otros documentos está debidamente referenciada.

Uso de Inteligencia Artificial¹

Declaro bajo mi responsabilidad que (indicar la opción correcta):

- ☐ No he utilizado Inteligencia Artificial en la elaboración del presente documento.
- ☒ He utilizado Inteligencia Artificial en la elaboración del presente documento y/o del Anexo B siempre en las condiciones permitidas por la Universidad Pontificia Comillas, es decir, aplicando el Nivel 2 de la [Escala de Evaluación de Perkins et al. \(2024\)](#): *“La IA puede utilizarse para actividades previas a la tarea, como la lluvia de ideas, la descripción y la investigación inicial. Este nivel se centra en el uso de la IA para la planificación, las síntesis y la generación de ideas, pero las evaluaciones deben hacer hincapié en la capacidad de desarrollar y refinar estas ideas de forma independiente”*. En concreto, las Inteligencia Artificial ha sido empleada para:

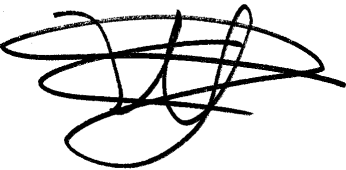
La Inteligencia Artificial se ha utilizado como herramienta de apoyo durante la elaboración de este Trabajo de Fin de Grado, principalmente para ayudar en la organización y mejora de algunos apartados del documento. Además, ha servido para aclarar conceptos técnicos.

Firmado (alumno): Candela Paradinas Solera

Fecha: 28/05/2026

¹ Esta declaración se refiere al uso de la Inteligencia Artificial generativa para realizar los documentos del Proyecto (Anexo B y Memoria). No aplica a Proyectos donde, por su naturaleza, deban emplear Inteligencia Artificial como parte de los mismos (aplicación de técnicas de aprendizaje automático, redes neuronales, análisis de datos...)

Autorización para la entrega del Proyecto

El Director del Proyecto	El co-Director del Proyecto (si aplica)
	
Fdo: Constantino Malagón Luque	Fdo:
Fecha: 03/06/2026	Fecha:



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

SISTEMA BASADO EN INTELIGENCIA ARTIFICIAL PARA EL ANÁLISIS DE TEXTURAS MUSCULARES EN PACIENTES CON ENFERMEDADES NEURODEGENERATIVAS (ELA) A PARTIR DE BIOMARCADORES ECOGRÁFICOS

Autor: Candela Paradinas Solera

Director: Constantino Malagón Luque

Madrid

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas que me han acompañado y apoyado durante la realización de este Trabajo Fin de Grado y a lo largo de toda mi etapa universitaria.

En primer lugar, a mis padres, Montse y Álvaro, por vuestro apoyo incondicional, vuestro esfuerzo constante y por haber estado siempre a mi lado en cada paso de este camino. Gracias por confiar en mí incluso cuando yo misma dudaba y por darme siempre la oportunidad de perseguir mis objetivos. Gracias por animarme en los momentos difíciles, por celebrar conmigo cada pequeño logro y por apoyarme en todas las decisiones que me han traído hasta aquí. Todo lo que he conseguido durante estos años ha sido posible gracias a vuestro cariño, vuestra confianza y vuestro ejemplo.

A mi hermana Lola, porque la vida decidió que empezáramos esta aventura el mismo día y no podría haberme tocado mejor compañera para recorrerla. Gracias por estar siempre ahí, por compartir conmigo tantos momentos, por apoyarme en los días difíciles y por celebrar conmigo los buenos. Tener una hermana es una suerte; tenerte a ti, después de haber compartido incluso el día de nacimiento, lo es todavía más.

A mis abuelos María, Juan y Luis, por todo el cariño, los valores y las enseñanzas que me habéis transmitido a lo largo de mi vida. Gracias por vuestro apoyo, por vuestra confianza y por haber estado siempre presentes en los momentos importantes. Vuestro esfuerzo, vuestra forma de afrontar la vida y todo lo que me habéis enseñado han contribuido a convertirme en la persona que soy hoy. Vuestro ejemplo ha sido y seguirá siendo una inspiración constante para mí.

Y, de manera muy especial, a mi abuela Lola. Aunque ya no estés físicamente con nosotros, tu recuerdo me ha acompañado siempre. Gracias por todo el amor que me diste y por seguir siendo una referencia en mi vida. Me habría gustado poder compartir contigo este momento y enseñarte todo lo que he conseguido, pero sé que, desde el cielo, has estado presente durante todo este camino. Tu cariño, tus enseñanzas y los recuerdos que compartimos seguirán acompañándome siempre.

Por último, quiero agradecer a mi director de TFG, Constantino Malagón, por tu ayuda, dedicación y orientación durante el desarrollo de este trabajo. Gracias por tu disponibilidad, por tus consejos y por todo el apoyo recibido a lo largo de este proyecto. Ha sido un placer trabajar contigo durante estos meses.

A todos vosotros, gracias.

SISTEMA BASADO EN INTELIGENCIA ARTIFICIAL PARA EL ANÁLISIS DE TEXTURAS MUSCULARES EN PACIENTES CON ENFERMEDADES NEURODEGENERATIVAS (ELA) A PARTIR DE BIOMARCADORES ECOGRÁFICOS

Autor: Paradinas Solera, Candela.

Director: Malagón Luque, Constantino.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

Este Trabajo Fin de Grado desarrolla un sistema basado en Inteligencia Artificial para clasificar pacientes con Esclerosis Lateral Amiotrófica (ELA) utilizando biomarcadores obtenidos mediante ecografía muscular. Los resultados muestran una alta capacidad para diferenciar pacientes con ELA de personas sanas, alcanzando valores de *AUC* superiores a 0.97.

Palabras clave: ELA, ecografía muscular, *Machine Learning*, biomarcadores, *GLCM*.

1. Introducción

La Esclerosis Lateral Amiotrófica (ELA) es una enfermedad neurodegenerativa que afecta a las motoneuronas [1,2]. La pérdida de estas neuronas provoca debilidad muscular, atrofia, dificultades funcionales y, en fases avanzadas, problemas respiratorios [1,6]. Actualmente no existe una prueba única que permita diagnosticar la enfermedad de forma temprana por lo que el diagnóstico sigue basándose en la evaluación clínica y en pruebas complementarias [7].

En la actualidad, uno de los principales retos es encontrar biomarcadores objetivos que permitan detectar la enfermedad de forma rápida y precisa [7]. La ecografía muscular ha despertado un gran interés porque es una técnica no invasiva, económica, disponible en muchos hospitales y capaz de detectar cambios estructurales en el músculo [8,9].

Además de los biomarcadores ecográficos tradicionales, en los últimos años ha aumentado el uso del análisis cuantitativo de imágenes [8,9]. Esta técnica permite extraer información que no puede apreciarse a simple vista y ofrece una descripción detallada de la textura muscular [10].

2. Estado del arte

El uso de la Inteligencia Artificial en enfermedades neurodegenerativas ha crecido durante los últimos años [11]. Varios estudios han demostrado que los algoritmos de aprendizaje automático pueden identificar patrones complejos en imágenes médicas, datos clínicos y biomarcadores, mejorando tanto el diagnóstico como la predicción de la evolución de la enfermedad [11,12].

Los trabajos de Martínez-Payá y colaboradores demostraron que hay biomarcadores ecográficos que permiten diferenciar pacientes con ELA de personas sanas mediante métodos estadísticos tradicionales [8,9]. Sin embargo, el uso de algoritmos de *Machine Learning* sobre estos biomarcadores ha sido menos estudiado.

Además, investigaciones recientes han mostrado que los descriptores de textura obtenidos mediante *GLCM* permiten detectar cambios en la estructura interna del músculo [10]. Esto sugiere que el análisis de textura puede aportar información adicional a la proporcionada por los biomarcadores ecográficos clásicos [8,10].

3. Metodología

El estudio se realizó utilizando una base de datos formada por pacientes con ELA y personas sanas [8,9]. De cada participante se disponía de información clínica, biomarcadores obtenidos mediante ecografía muscular y parámetros relacionados con la textura del músculo.

Los biomarcadores ecográficos incluían medidas como el grosor muscular, la ecogenicidad y la textura de distintos músculos, entre ellos el bíceps braquial, los flexores del antebrazo, el cuádriceps y el tibial anterior [8]. Además, se calcularon diferentes descriptores *GLCM*, como el contraste, la correlación y la entropía [10].

Tras una fase inicial de preparación de los datos, se realizaron distintos experimentos. Primero se entrenaron modelos utilizando únicamente los biomarcadores clásicos. Después se analizaron solo los biomarcadores de textura y, por último, se combinaron todas las variables disponibles [12]. También se estudió el efecto de añadir variables clínicas como la edad, el sexo o el índice de masa corporal (*BMI*) [8].

Los modelos evaluados fueron Regresión Logística, Máquinas de Vectores de Soporte (*SVM*), *Random Forest*, *K-Nearest Neighbors (KNN)*, Árboles de Decisión, *Gradient Boosting* y *Naive Bayes* [12]. Su rendimiento se evaluó mediante validación cruzada utilizando métricas como la *accuracy*, la sensibilidad, el *F1-score* y el área bajo la curva *ROC (AUC)* [13].

La Figura 1 resume las principales etapas seguidas durante el desarrollo del TFG, desde la extracción de variables hasta la clasificación final de los pacientes.

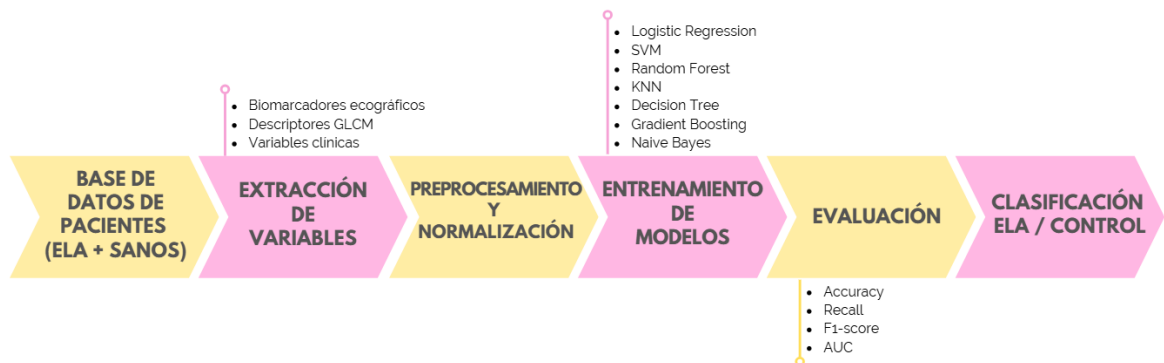


Figura 1. Esquema de la metodología empleada para la clasificación de pacientes con ELA

4. Resultados

Los resultados mostraron que la combinación de biomarcadores ecográficos clásicos y parámetros de textura mejoró el rendimiento de la mayoría de los modelos analizados [8,10], tal y como se observa en la Figura 2, donde los mejores resultados se obtuvieron al combinar todos los conjuntos de variables. En general, las *accuracies* obtenidas se situaron entre el 80% y el 90%, y aumentaron todavía más cuando se añadieron variables clínicas.

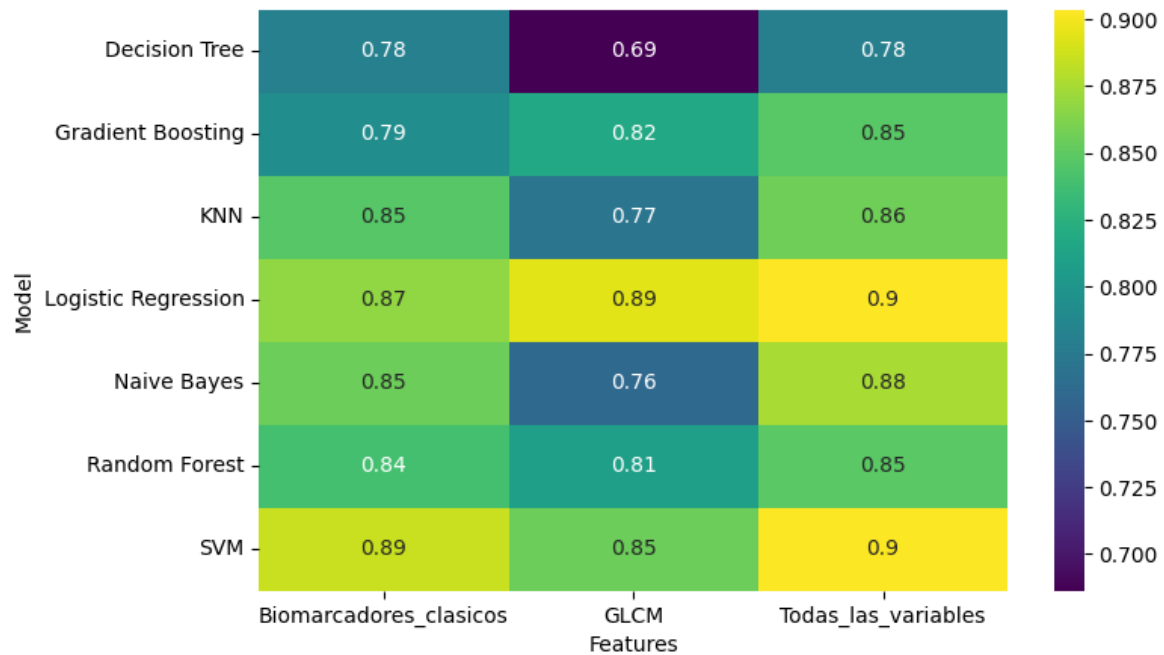


Figura 2. Comparación del rendimiento de los modelos según el conjunto de variables utilizado

La Regresión Logística fue el modelo que ofreció los mejores resultados, alcanzando *accuracies* cercanas al 90% y valores de *AUC* superiores a 0.97 [12]. Estos resultados indican una alta capacidad para distinguir entre pacientes con ELA y personas sanas [8,9].

El análisis de importancia de variables mostró que los biomarcadores obtenidos del músculo cuádriceps fueron los que más información aportaban [8]. Entre ellos destacó COR.Cdr, que corresponde al descriptor de correlación *GLCM* del cuádriceps y fue la variable que más influyó en el modelo final [10]. Estos resultados demuestran que los descriptores texturales contienen información relevante sobre las alteraciones musculares asociadas a la enfermedad.

5. Conclusiones

Este Trabajo Fin de Grado ha demostrado que las técnicas de *Machine Learning* pueden utilizarse para analizar biomarcadores obtenidos mediante ecografía muscular y ayudar a identificar pacientes con ELA [11,12]. Los resultados muestran que combinar biomarcadores ecográficos tradicionales, parámetros de textura y variables clínicas permite obtener mejores resultados que utilizar cada tipo de información por separado [8,10].

Entre todos los modelos estudiados, la Regresión Logística fue la que obtuvo los mejores resultados [12]. Además, el análisis de importancia de variables permitió identificar los biomarcadores más útiles, destacando los obtenidos del músculo cuádriceps [8,9].

En conclusión, estos resultados muestran el potencial de la ecografía muscular y la Inteligencia Artificial como herramientas de apoyo al diagnóstico de la ELA [7,11]. Además, indican que el análisis de textura puede proporcionar información útil, objetiva y no invasiva que complemente las técnicas de diagnóstico utilizadas actualmente [8,10].

6. Referencias

- [1] Rowland LP, Shneider NA. Amyotrophic lateral sclerosis. *New England Journal of Medicine*. 2001;344(22):1688-1700. <https://doi.org/10.1056/NEJM200105313442207>
- [2] Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W, Shaw PJ, Simmons Z, van den Berg LH. Amyotrophic lateral sclerosis. *Nature Reviews Disease Primers*. 2017;3:17071. <https://doi.org/10.1038/nrdp.2017.85>
- [3] Riancho J, Gonzalo I, Ruiz-Soto M, Berciano J. ¿Por qué degeneran las motoneuronas? Actualización en la patogenia de la esclerosis lateral amiotrófica. *Neurología*. 2019; 34(1):27-37. <https://doi.org/10.1016/j.nrl.2015.12.001>
- [4] Brooks BR, Miller RG, Swah M, Munsat TL; World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*. 2000;1(5):293-299. <https://doi.org/10.1080/146608200300079536>
- [5] Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, Nakanishi A. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*. 1999;169(1-2):13-21. [https://doi.org/10.1016/S0022-510X\(99\)00210-5](https://doi.org/10.1016/S0022-510X(99)00210-5)
- [6] Wijesekera LC, Leigh PN. Amyotrophic lateral sclerosis. *Orphanet Journal of Rare Diseases*. 2009;4:3. <https://doi.org/10.1186/1750-1172-4-3>
- [7] Turner MR, Hardiman O, Benatar M, Brooks BR, Chio A, de Carvalho M, Couratier P, Shaw PJ, Veldink JH, van den Berg LH. Controversies and priorities in amyotrophic lateral sclerosis. *The Lancet Neurology*. 2013;12(3):310-322. [https://doi.org/10.1016/S1474-4422\(13\)70036-X](https://doi.org/10.1016/S1474-4422(13)70036-X)
- [8] Martínez-Payá JJ, Ríos-Díaz J, del Baño-Aledo ME, Tembl-Ferrairó JI, Vázquez-Costa JF, Medina-Mirapeix F. Quantitative muscle ultrasonography using textural analysis in amyotrophic lateral sclerosis. *Ultrasonic Imaging*. 2017;39(6):357-368. <https://doi.org/10.1177/0161734617711370>
- [9] Martínez-Payá JJ, Ríos-Díaz J, Medina-Mirapeix F, Vázquez-Costa JF, del Baño-Aledo ME. Monitoring progression of amyotrophic lateral sclerosis using ultrasound morpho-textural muscle biomarkers: a pilot study. *Ultrasound in Medicine and Biology*. 2018;44(1):102-109. <https://doi.org/10.1016/j.ultrasmedbio.2017.09.013>
- [10] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*. 1973;3(6):610-621. <https://doi.org/10.1109/TSMC.1973.4309314>
- [11] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak J, van Ginneken B, Sánchez CI. A survey on Deep Learning in medical image analysis. *Medical Image Analysis*. 2017;42:60-88. <https://doi.org/10.1016/j.media.2017.07.005>

- [12] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-2830.
- [13] Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006;27(8):861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>

ARTIFICIAL INTELLIGENCE-BASED SYSTEM FOR THE ANALYSIS OF MUSCLE TEXTURES IN PATIENTS WITH NEURODEGENERATIVE DISEASES (ALS) USING ULTRASOUND BIOMARKERS

Author: Paradinas Solera, Candela.

Supervisor: Malagón Luque, Constantino.

Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

This Final Degree Project develops an Artificial Intelligence-based system for the classification of patients with Amyotrophic Lateral Sclerosis (ALS) using biomarkers obtained from muscle ultrasound. The results demonstrate a high capability to distinguish ALS patients from healthy controls, achieving AUC values above 0.97.

Keywords: ALS, muscle ultrasound, Machine Learning, biomarkers, GLCM.

1. Introduction

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease that affects motor neurons [1,2]. The loss of these neurons leads to muscle weakness, atrophy, functional impairment, and, in advanced stages, respiratory complications [1,6]. Currently, there is no single test capable of diagnosing the disease at an early stage; therefore, diagnosis still relies on clinical evaluation and complementary tests [7].

One of the main challenges today is the identification of objective biomarkers that enable rapid and accurate disease detection [7]. Muscle ultrasound has attracted considerable interest because it is a non-invasive, low-cost technique, widely available in hospitals, and capable of detecting structural changes in muscle tissue [8,9].

In addition to traditional ultrasound biomarkers, the use of quantitative image analysis has increased in recent years [8,9]. This technique makes it possible to extract information that cannot be detected by visual inspection alone and provides a detailed characterization of muscle texture [10].

2. State of the Art

The use of Artificial Intelligence in neurodegenerative diseases has grown significantly in recent years [11]. Several studies have shown that Machine Learning algorithms can identify complex patterns in medical images, clinical data, and biomarkers, improving both diagnosis and disease progression prediction [11,12].

The studies conducted by Martínez-Payá and colleagues demonstrated that ultrasound biomarkers can differentiate ALS patients from healthy controls using traditional statistical methods [8,9]. However, the application of Machine Learning techniques to these biomarkers has been less extensively explored.

Furthermore, recent research has shown that texture descriptors obtained through Gray-Level Co-occurrence Matrices (GLCM) can detect changes in the internal structure of

muscle tissue [10]. This suggests that texture analysis may provide additional information beyond that offered by conventional ultrasound biomarkers [8,10].

3. Methodology

This study was carried out using a database composed of ALS patients and healthy controls [8,9]. Clinical information, ultrasound-derived biomarkers, and muscle texture-related parameters were available for each participant.

The ultrasound biomarkers included measurements such as Muscle Thickness, echogenicity, and texture from different muscles, including the biceps brachii, forearm flexors, quadriceps, and tibialis anterior [8]. In addition, several GLCM descriptors were calculated, including contrast, correlation, and entropy [10].

After an initial data preparation stage, different experiments were performed. First, models were trained using only conventional ultrasound biomarkers. Then, texture-based biomarkers were analyzed independently, and finally, all available variables were combined [12]. The effect of adding clinical variables such as age, sex, and Body Mass Index (BMI) was also evaluated [8].

The models assessed included Logistic Regression, Support Vector Machines (SVM), Random Forest, K-Nearest Neighbors (KNN), Decision Trees, Gradient Boosting, and Naive Bayes [12]. Their performance was evaluated through cross-validation using metrics such as accuracy, recall, F1-score, and the Area Under the ROC Curve (AUC) [13].

Figure 3 summarizes the main stages followed during the development of this project, from feature extraction to the final classification of patients.

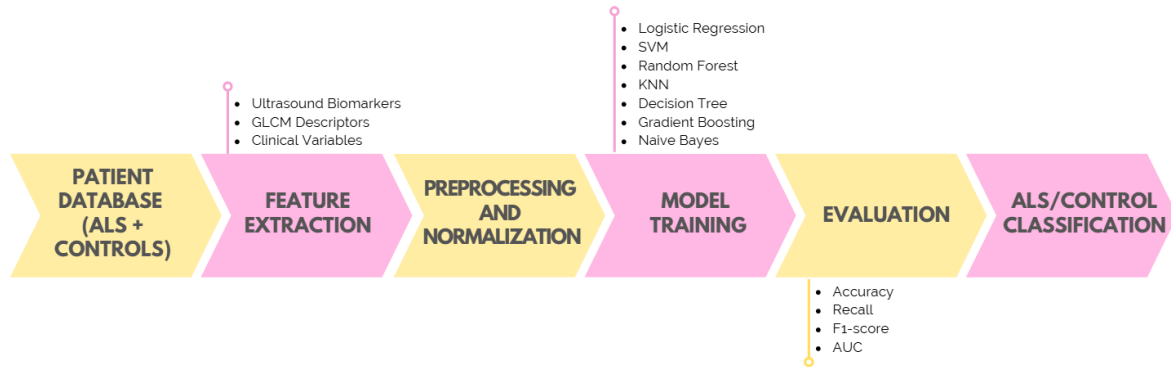


Figure 3. Methodology used for ALS patient classification

4. Results

The results showed that combining conventional ultrasound biomarkers with texture parameters improved the performance of most of the analyzed models [8,10], as shown in Figure 4, where the best results were obtained when all feature sets were combined. In general, the accuracies obtained ranged between 80% and 90%, with additional improvements observed when clinical variables were incorporated.

Logistic Regression was the best-performing model, achieving accuracies close to 90% and AUC values above 0.97 [12]. These results indicate a high capability to distinguish ALS patients from healthy controls [8,9].

The variable importance analysis revealed that biomarkers derived from the quadriceps muscle provided the most relevant information [8]. Among them, COR.Cdr, corresponding to the GLCM correlation descriptor of the quadriceps, was identified as the most influential variable in the final model [10]. These findings demonstrate that texture descriptors contain valuable information regarding the muscle alterations associated with the disease.

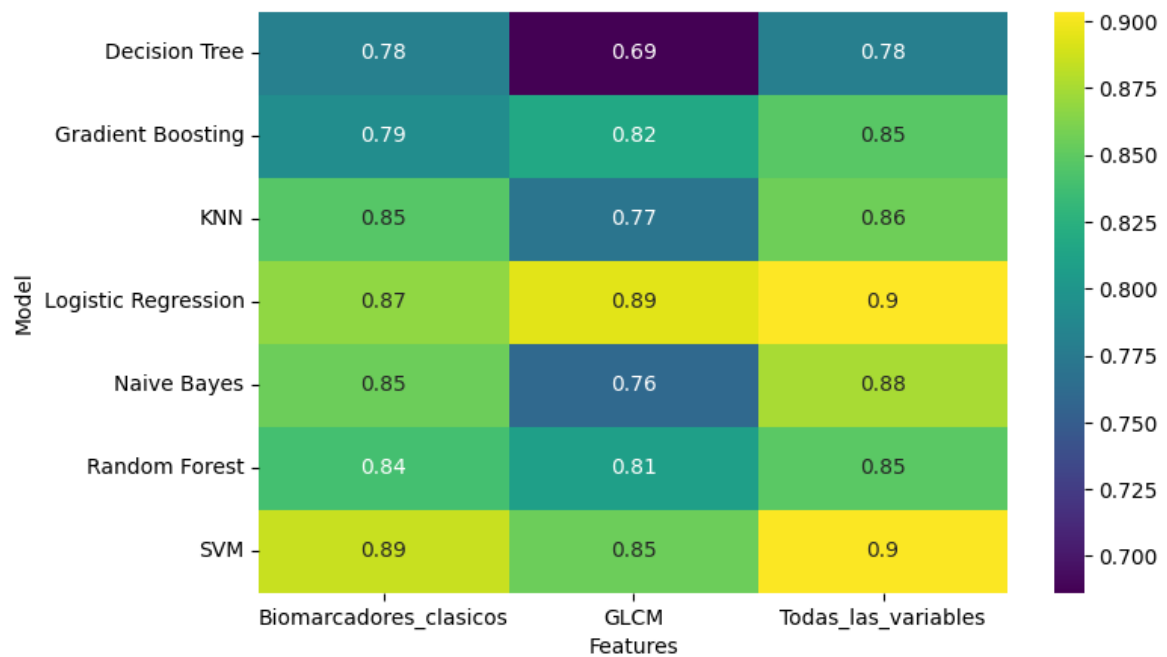


Figura 4. Comparison of model performance according to the feature set used

5. Conclusions

This Final Degree Project has demonstrated that Machine Learning techniques can be successfully applied to the analysis of muscle ultrasound biomarkers for the identification of ALS patients [11,12]. The results show that combining conventional ultrasound biomarkers, texture parameters, and clinical variables provides better performance than using each type of information independently [8,10].

Among all the evaluated models, Logistic Regression achieved the best results [12]. In addition, the variable importance analysis made it possible to identify the most informative biomarkers, particularly those obtained from the quadriceps muscle [8,9].

In conclusion, these findings highlight the potential of muscle ultrasound and Artificial Intelligence as supporting tools for ALS diagnosis [7,11]. Furthermore, they suggest that texture analysis can provide useful, objective, and non-invasive information that complements the diagnostic techniques currently used in clinical practice [8,10].

6. References

- [1] Rowland LP, Shneider NA. Amyotrophic lateral sclerosis. *New England Journal of Medicine*. 2001;344(22):1688-1700. <https://doi.org/10.1056/NEJM200105313442207>
- [2] Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W, Shaw PJ, Simmons Z, van den Berg LH. Amyotrophic lateral sclerosis. *Nature Reviews Disease Primers*. 2017;3:17071. <https://doi.org/10.1038/nrdp.2017.85>
- [3] Riancho J, Gonzalo I, Ruiz-Soto M, Berciano J. ¿Por qué degeneran las motoneuronas? Actualización en la patogenia de la esclerosis lateral amiotrófica. *Neurología*. 2019; 34(1):27-37. <https://doi.org/10.1016/j.nrl.2015.12.001>
- [4] Brooks BR, Miller RG, Swah M, Munsat TL; World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*. 2000;1(5):293-299. <https://doi.org/10.1080/146608200300079536>
- [5] Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, Nakanishi A. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*. 1999;169(1-2):13-21. [https://doi.org/10.1016/S0022-510X\(99\)00210-5](https://doi.org/10.1016/S0022-510X(99)00210-5)
- [6] Wijesekera LC, Leigh PN. Amyotrophic lateral sclerosis. *Orphanet Journal of Rare Diseases*. 2009;4:3. <https://doi.org/10.1186/1750-1172-4-3>
- [7] Turner MR, Hardiman O, Benatar M, Brooks BR, Chio A, de Carvalho M, Couratier P, Shaw PJ, Veldink JH, van den Berg LH. Controversies and priorities in amyotrophic lateral sclerosis. *The Lancet Neurology*. 2013;12(3):310-322. [https://doi.org/10.1016/S1474-4422\(13\)70036-X](https://doi.org/10.1016/S1474-4422(13)70036-X)
- [8] Martínez-Payá JJ, Ríos-Díaz J, del Baño-Aledo ME, Tembl-Ferrairó JI, Vázquez-Costa JF, Medina-Mirapeix F. Quantitative muscle ultrasonography using textural analysis in amyotrophic lateral sclerosis. *Ultrasonic Imaging*. 2017;39(6):357-368. <https://doi.org/10.1177/0161734617711370>
- [9] Martínez-Payá JJ, Ríos-Díaz J, Medina-Mirapeix F, Vázquez-Costa JF, del Baño-Aledo ME. Monitoring progression of amyotrophic lateral sclerosis using ultrasound morpho-textural muscle biomarkers: a pilot study. *Ultrasound in Medicine and Biology*. 2018;44(1):102-109. <https://doi.org/10.1016/j.ultrasmedbio.2017.09.013>
- [10] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*. 1973;3(6):610-621. <https://doi.org/10.1109/TSMC.1973.4309314>
- [11] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak J, van Ginneken B, Sánchez CI. A survey on Deep Learning in medical image analysis. *Medical Image Analysis*. 2017;42:60-88. <https://doi.org/10.1016/j.media.2017.07.005>

- [12] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-2830.
- [13] Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006;27(8):861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>

Índice de la memoria

Capítulo 1. Introducción	7
1.1 Contexto clínico general.....	7
1.2 Impacto personal, familiar y sanitario	8
1.3 Historia y fisiopatología	9
1.4 Epidemiología y factores de riesgo	11
1.5 Presentación clínica y variantes	12
1.6 Diagnóstico actual y sus limitaciones.....	13
1.7 Neuroimagen avanzada en ELA.....	14
1.8 La ecografía muscular como herramienta alternativa.....	15
1.9 Inteligencia artificial aplicada al diagnóstico de la ELA.....	19
1.10 Objetivo del presente trabajo.....	21
1.11 Estructura del presente trabajo	22
Capítulo 2. Estado del arte	23
2.1 Introducción.....	23
2.2 Biomarcadores en enfermedades neurodegenerativas.....	25
2.3 Evolución de la ecografía muscular en neurología.....	27
2.4 Biomarcadores ecográficos clásicos.....	29
2.4.1 Grosor muscular.....	29
2.4.2 Ecogenicidad muscular	30
2.5 Análisis textural en imagen médica.....	31
2.6 Matriz de co-ocurrencia de niveles de gris (GLCM).....	32
2.7 Otras técnicas de análisis textural.....	34
2.8 Inteligencia artificial aplicada a imagen médica	35
2.9 Algoritmos de Machine Learning utilizados en diagnóstico médico	36
2.10 Limitaciones actuales del estado del arte	38
2.11 Motivación del presente trabajo	39
Capítulo 3. Metodología	41
3.1 Introducción.....	41
3.2 Diseño del estudio	42
3.3 Descripción del dataset.....	44

3.3.1 Variables clínicas	45
3.3.2 Biomarcadores ecográficos clásicos	45
3.3.3 Descriptores texturales GLCM.....	46
3.4 Músculos analizados.....	47
3.5 Variable objetivo	48
3.6 Selección de variables	49
3.7 Preprocesamiento de los datos	50
3.8 Modelos de Machine Learning utilizados	51
3.8.1 Regresión logística	52
3.8.2 Máquinas de soporte vectorial (SVM).....	53
3.8.3 Random Forest	53
3.8.4 K-Nearest Neighbors (KNN)	54
3.8.5 Árboles de decisión.....	55
3.8.6 Gradient Boosting	55
3.8.7 Naive Bayes	56
3.9 Pipeline de entrenamiento	56
3.10 Validación cruzada	57
3.11 Métricas de evaluación.....	58
3.11.1 Accuracy.....	58
3.11.2 Precisión.....	59
3.11.3 Recall o sensibilidad.....	59
3.11.4 F1-score.....	60
3.11.5 Área bajo la curva ROC (AUC)	60
3.12 Selección de variables y reducción de dimensionalidad	61
3.13 Prevención del sobreajuste	62
3.14 Interpretabilidad clínica de los modelos.....	63
3.15 Consideraciones éticas y limitaciones metodológicas.....	64
Capítulo 4. Resultados.....	65
4.1 Introducción.....	65
4.2 Comparación de modelos de Machine Learning.....	66
4.2.1 Configuración experimental.....	66
4.2.2 Detección y corrección del sobreajuste.....	67
4.2.3 Resultados globales por modelo.....	67

4.3	Impacto del conjunto de variables.....	69
4.3.1	<i>Análisis comparativo por tipo de biomarcador.....</i>	69
4.3.2	<i>Resultados por conjunto de variables</i>	70
4.4	Evaluación del mejor modelo.....	73
4.4.1	<i>Selección y justificación del modelo.....</i>	73
4.4.2	<i>Métricas de clasificación detalladas</i>	74
4.4.3	<i>Análisis de la matriz de confusión.....</i>	74
4.4.4	<i>Curva ROC y AUC</i>	76
4.5	Comparación con el estado del arte.....	77
4.5.1	<i>Contexto de la comparación.....</i>	77
4.5.2	<i>Comparación de rendimientos</i>	78
4.5.3	<i>Ventajas del enfoque propuesto frente a estudios previos.....</i>	79
4.6	Análisis por músculo: el caso del cuádriceps femoral.....	79
4.6.1	<i>Motivación del análisis específico.....</i>	79
4.6.2	<i>Resultados del análisis del cuádriceps.....</i>	80
4.6.3	<i>Explicación fisiopatológica de la relevancia del cuádriceps.....</i>	82
4.6.4	<i>Limitaciones del análisis por músculo único</i>	83
4.7	Análisis de importancia de variables.....	83
4.7.1	<i>Metodología del análisis</i>	83
4.7.2	<i>Variables más relevantes identificadas</i>	85
4.7.3	<i>Interpretación clínica de los resultados</i>	86
4.8	Incorporación de variables	87
4.8.1	<i>Justificación de la incorporación.....</i>	87
4.8.2	<i>Resultados de la incorporación de variables clínicas.....</i>	88
4.8.3	<i>Importancia de las variables clínicas.....</i>	91
4.8.4	<i>Limitaciones y perspectivas futuras</i>	92
4.9	Análisis preliminar de progresión de la enfermedad.....	93
4.9.1	<i>Planteamiento del problema.....</i>	93
4.9.2	<i>Resultados preliminares</i>	93
4.9.3	<i>Correlaciones entre biomarcadores y escala ALSFRS</i>	95
4.9.4	<i>Limitaciones y perspectivas del análisis de progresión</i>	95
4.10	Discusión de resultados	96
4.10.1	<i>Síntesis de los hallazgos principales</i>	96

4.10.2 Relevancia clínica de los resultados	97
4.10.3 Limitaciones identificadas	97
4.10.4 Perspectivas de trabajo futuro	98
Capítulo 5. Conclusiones.....	99
5.1 Síntesis del trabajo realizado	99
5.2 Principales resultados obtenidos	101
5.3 Interpretación clínica de los resultados	104
5.4 Limitaciones del estudio.....	106
5.5 Líneas futuras de trabajo	108
5.6 Conclusión final	111
Capítulo 6. Bibliografía.....	113
ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS	119

Índice de figuras

Figura 1. Esquema de la metodología empleada para la clasificación de pacientes con ELA	10
Figura 2. Comparación del rendimiento de los modelos según el conjunto de variables utilizado	11
Figura 3. Methodology used for ALS patient classification.....	15
Figura 4. Comparison of model performance according to the feature set used	16
Figura 5. Comparación de la accuracy de los diferentes modelos de Machine Learning para los tres conjuntos de variables analizados	71
Figura 6. Distribución de la accuracy media obtenida por los modelos según el conjunto de variables empleado	72
Figura 7. Matriz de confusión del modelo de Regresión Logística utilizando el conjunto completo de variables	75
Figura 8. Curva ROC del modelo de Regresión Logística utilizando el conjunto completo de variables.....	76
Figura 9. Curva ROC del modelo basado en biomarcadores ecográficos del músculo cuádriceps	81
Figura 10. Matriz de confusión del modelo basado en biomarcadores ecográficos del músculo cuádriceps.....	82
Figura 11. Ranking de los biomarcadores más importantes para clasificar pacientes con ELA según el modelo de Regresión Logística	85
Figura 12. Matriz de confusión obtenida mediante Regresión Logística tras incorporación de variables clínicas y fasciculaciones	89
Figura 13. Curva ROC del modelo de Regresión Logística tras incorporar variables clínicas y fasciculaciones.....	90
Figura 14. Variables clínicas con mayor importancia en el modelo de Regresión Logística tras incorporar biomarcadores ecográficos y variables clínicas	91
Figura 15. Distribución de las puntuaciones ALSFRS en los pacientes con ELA.....	94

Índice de tablas

Tabla 1. Accuracy media (10-Fold CV estratificado) por modelo y conjunto de variables	69
Tabla 2. Informe de clasificación del mejor modelo (Regresión Logística con todas las variables)	74
Tabla 3. Resultados del modelo específico del músculo cuádriceps femoral	80
Tabla 4. Comparación entre el modelo basado únicamente en biomarcadores ecográficos y el modelo que incorpora variables clínicas.....	90

Capítulo 1. INTRODUCCIÓN

1.1 CONTEXTO CLÍNICO GENERAL

La Esclerosis Lateral Amiotrófica (ELA) es una enfermedad neurodegenerativa progresiva que afecta a las motoneuronas del sistema nervioso, tanto las superiores como las inferiores [1,2]. Estas neuronas se encargan de enviar señales nerviosas desde el cerebro y la médula espinal hasta los músculos, permitiendo realizar movimientos voluntarios como caminar, escribir, hablar, respirar o tragar. Cuando estas neuronas empiezan a deteriorarse y a morir, esa comunicación entre el sistema nervioso y los músculos se rompe de manera irreversible, lo que provoca una pérdida progresiva de fuerza muscular, movilidad y autonomía.

La ELA se considera una de las enfermedades neurodegenerativas más graves que existen, tanto por lo que implica en el contexto clínico como por el enorme impacto que tiene en la vida de quienes la padecen [8]. A medida que la enfermedad avanza, los pacientes sufren un deterioro físico que termina afectando a la mayoría de las funciones motoras del cuerpo. Al principio pueden aparecer síntomas leves o poco específicos, como torpeza en una mano, pérdida de fuerza en una pierna, dificultad para realizar movimientos precisos, cansancio muscular, calambres o pequeños problemas para hablar [4]. Sin embargo, a medida que aumenta el deterioro de las neuronas, el número de músculos afectados comienza a aumentar y termina dificultando funciones esenciales del cuerpo humano como la respiración y la deglución.

La principal causa de muerte de estos pacientes es la insuficiencia respiratoria. Aunque la supervivencia varía bastante de una persona a otra, el tiempo medio desde los primeros síntomas hasta el fallecimiento suele situarse entre tres y cinco años [6,15]. Aun así, la ELA no sigue siempre el mismo patrón. Algunos pacientes experimentan una progresión muy rápida, mientras que otros consiguen mantener parcialmente ciertas funciones motoras permitiéndoles vivir durante más de diez años [7]. Esta variabilidad es uno de los principales

desafíos tanto para la investigación biomédica como para el tratamiento clínico de la enfermedad.

Estas diferencias en la evolución de la ELA complican, no solo la predicción del pronóstico individual de cada paciente, sino también el diseño de ensayos clínicos, la comparación de resultados entre distintos hospitales y el desarrollo de herramientas para diagnósticos que sean comunes y estandarizados [7].

1.2 IMPACTO PERSONAL, FAMILIAR Y SANITARIO

Además del deterioro físico que provoca la esclerosis lateral amiotrófica, esta enfermedad cambia por completo la vida del paciente y de las personas de su entorno [2]. La pérdida progresiva de movilidad hace que, en muchos casos, los pacientes necesiten ayuda de familiares o cuidadores para realizar las actividades básicas del día a día. A medida que avanza la enfermedad, tareas habituales como vestirse, ducharse, subir escaleras, escribir o incluso comer se vuelven cada vez más difíciles. Esta pérdida de independencia tiene un gran impacto en el bienestar emocional y psicológico del paciente.

La ELA también produce importantes consecuencias emocionales. Muchos pacientes sufren ansiedad, miedo, frustración o síntomas depresivos debido al avance continuo de la enfermedad y a la incertidumbre sobre el futuro [2]. Además, a esto se suma el gran desgaste psicológico que sufren las familias y cuidadores, que deben adaptarse constantemente a las nuevas necesidades del paciente y asumir una carga física y emocional cada vez mayor.

Desde el punto de vista sanitario, la ELA representa también un importante desafío para los sistemas de salud. El aumento progresivo de la dependencia hace necesarios cuidados continuos, adaptaciones en el hogar, dispositivos de ayuda a la movilidad y atención especializada [5]. Todo ello supone un coste, tanto económico como social, muy elevado. En muchos casos, los familiares deben reorganizar su vida personal y laboral para poder cuidar al paciente, lo que también puede tener consecuencias emocionales y económicas dentro de la familia.

El impacto económico de la enfermedad no afecta solo al entorno familiar, sino también al sistema sanitario. Los costes médicos asociados a la ELA son muy elevados e incluyen hospitalizaciones, consultas especializadas, dispositivos de respiración, fisioterapia, logopedia y tratamientos farmacológicos [5]. Conocer toda esta carga es fundamental para impulsar la investigación en diagnóstico y tratamiento de esta enfermedad.

1.3 HISTORIA Y FISIOPATOLOGÍA

El término “esclerosis lateral amiotrófica” fue utilizado por primera vez por Jean Martin Charcot en 1874, cuando relacionó el daño en zonas de la médula espinal con la pérdida progresiva de masa muscular [3,6]. Desde entonces, la ciencia ha avanzado mucho gracias al desarrollo de la neurología, la genética y las técnicas de imagen médica. Aun así, más de cien años después, la ELA sigue siendo una enfermedad sin cura definitiva y todavía existen muchas dudas sobre sus causas, evolución y tratamiento [9].

Charcot observó que la enfermedad afectaba tanto a las neuronas encargadas de enviar las órdenes de movimiento desde el cerebro como a las que conectan con los músculos. Este descubrimiento fue muy importante porque permitió definir las principales características clínicas de la enfermedad y mejorar su diagnóstico. Con el paso de los años, la investigación ha permitido comprender mucho mejor los procesos biológicos que provocan la degeneración de estas neuronas.

Actualmente, se considera que la ELA es una enfermedad muy compleja y que no tiene una única causa. Su aparición parece estar relacionada con la combinación de diferentes factores genéticos, celulares y ambientales que terminan dañando progresivamente las motoneuronas [3]. Entre los mecanismos más estudiados se encuentran el estrés oxidativo, la acumulación de sustancias tóxicas, el mal funcionamiento de las mitocondrias, los problemas en el transporte dentro de las neuronas, la acumulación anormal de proteínas y la inflamación del sistema nervioso [3]. Todos estos procesos contribuyen al daño y muerte de las neuronas motoras.

Uno de los mecanismos más importantes es el estrés oxidativo. Las motoneuronas necesitan mucha energía para funcionar y, por ello, producen una gran cantidad de sustancias que pueden resultar dañinas si se acumulan en exceso [3]. Cuando las células no son capaces de eliminar correctamente estas sustancias, se produce un daño progresivo en componentes importantes de la neurona, conduciendo finalmente a su muerte. Además, este problema empeora por el deterioro de las mitocondrias, que son las estructuras encargadas de producir energía dentro de las células.

Otro proceso importante es la toxicidad causada por el glutamato, una sustancia que utiliza el sistema nervioso para transmitir señales entre neuronas. En pacientes con ELA, el glutamato puede acumularse en exceso y provocar una estimulación demasiado intensa de las neuronas, lo que termina dañándolas [3]. Este mecanismo está relacionado con el funcionamiento del riluzol, uno de los principales medicamentos utilizados para tratar la enfermedad, ya que ayuda a reducir la liberación del glutamato [9].

También se ha observado que las mitocondrias funcionan de forma incorrecta en muchas personas con ELA [3]. Esto provoca problemas en la producción de energía y dificulta el correcto funcionamiento de las neuronas. Además, aparecen alteraciones en el transporte interno de sustancias dentro de la célula, haciendo que proteínas y otros componentes se acumulen en zonas donde no deberían estar. Todo ello afecta al funcionamiento normal de las neuronas y favorece su degeneración.

Otro aspecto importante es la acumulación de proteínas dentro de las neuronas. Algunas proteínas como TDP-43 o FUS aparecen agrupadas de forma incorrecta en pacientes con ELA [3,14]. Esto altera procesos esenciales de la célula, como la regulación genética y el procesamiento de la información celular, haciendo que las neuronas pierdan poco a poco su capacidad de funcionar correctamente.

Por último, la inflamación del sistema nervioso también juega un papel importante en la enfermedad. Algunas células del sistema nervioso liberan sustancias inflamatorias que, aunque inicialmente pueden intentar proteger las neuronas, con el tiempo terminan aumentando el daño y acelerando la muerte de las motoneuronas [3].

1.4 EPIDEMIOLOGÍA Y FACTORES DE RIESGO

Desde el punto de vista epidemiológico, la esclerosis lateral amiotrófica se considera una enfermedad rara, aunque es la enfermedad de las motoneuronas más común en adultos [4,12]. Cada año se diagnostican entre uno y tres casos por cada 100.000 habitantes, y se calcula que entre cuatro y diez personas por cada 100.000 habitantes viven con esta enfermedad [12]. La ELA aparece con mayor frecuencia en hombres y suele diagnosticarse entre los 55 y los 70 años, aunque también puede presentarse en personas jóvenes o de edad avanzada [5,8].

En España, actualmente viven entre 3.500 y 4.000 personas con ELA y cada año se diagnostican alrededor de 900 nuevos casos [5]. Aunque estas cifras son menores comparadas con otras enfermedades neurológicas, la ELA tiene un gran impacto debido a la gravedad de la enfermedad y a que todavía no existe una cura definitiva.

Algunos estudios han encontrado diferencias en la frecuencia de la enfermedad entre distintos países, regiones y grupos de población, lo que sugiere que pueden influir tanto factores genéticos como ambientales [12]. Entre los posibles factores de riesgo más estudiados se encuentran el tabaco, la exposición a pesticidas, determinados traumatismos y algunos factores ambientales. Sin embargo, en muchos casos todavía no se ha demostrado que estos factores sean una causa directa de la enfermedad [7].

También se han investigado posibles relaciones entre la ELA y determinadas profesiones, estilos de vida o exposición a sustancias tóxicas. Por ejemplo, algunos estudios han observado más casos de ELA en militares, deportistas profesionales o personas expuestas a metales pesados y productos químicos [7]. Aun así, los resultados no son totalmente claros y todavía no existe un acuerdo definitivo sobre la influencia real de estos factores. Por ello, actualmente se piensa que la ELA podría aparecer por la combinación de una predisposición genética y diferentes factores ambientales acumulados a lo largo de la vida [3,13].

1.5 PRESENTACIÓN CLÍNICA Y VARIANTES

Desde el punto de vista clínico, la esclerosis lateral amiotrófica puede presentarse de formas muy diferentes según las neuronas y los músculos que se vean afectados al inicio de la enfermedad [8]. En algunos pacientes predominan síntomas relacionados con la motoneurona superior, como rigidez muscular, reflejos exagerados o alteraciones en algunos reflejos neurológicos. En otros casos, destacan síntomas relacionados con la motoneurona inferior, como la pérdida de masa muscular, debilidad, pequeños movimientos involuntarios en los músculos o disminución de los reflejos [4].

La forma más común es la ELA clásica o de Charcot, en la que desde etapas tempranas se ven afectadas tanto las motoneuronas superiores como las inferiores [1]. Sin embargo, también existen otras variantes clínicas. Una de ellas es la esclerosis lateral primaria, que afecta sobre todo a las motoneuronas superiores y suele avanzar más lentamente. Otra es la atrofia muscular progresiva, en la que predominan los síntomas relacionados con la motoneurona inferior. También existe la parálisis bulbar progresiva, en la que los primeros síntomas aparecen en los músculos relacionados con hablar, masticar y tragar. Esta última variante suele tener peor pronóstico porque afecta rápidamente a funciones importantes como la alimentación y la respiración [2].

Estas variantes pueden diferenciarse en los síntomas, la velocidad de progresión y la supervivencia de los pacientes [2]. Aunque con el tiempo muchas terminan evolucionando hacia un cuadro clínico parecido, las diferencias iniciales son importantes para realizar un buen diagnóstico, planificar la atención médica y explicar la evolución de la enfermedad al paciente.

Toda esta variabilidad hace que el diagnóstico de la ELA sea especialmente complicado. No existe ningún síntoma característico que permita identificar la enfermedad con seguridad desde el principio. Por ello, los médicos deben descartar otras enfermedades que pueden producir síntomas parecidos, como algunas neuropatías, enfermedades musculares, lesiones en la médula espinal, esclerosis múltiple y otras enfermedades neurodegenerativas [4]. Esta

dificultad explica que, en muchos casos, el diagnóstico de la ELA tarde bastante tiempo en confirmarse.

1.6 DIAGNÓSTICO ACTUAL Y SUS LIMITACIONES

Uno de los principales problemas de esta enfermedad sigue siendo la dificultad para diagnosticarla de forma temprana. Actualmente, no existe un biomarcador definitivo, objetivo y no invasivo capaz de detectar la enfermedad en sus primeras fases [7,16]. Por ello, el diagnóstico se basa principalmente en la evaluación clínica de los síntomas y en diferentes pruebas que ayuden a descartar otras enfermedades neurológicas.

Los criterios de El Escorial, establecidos por la World Federation of Neurology en 1994 y revisados en 1998, son los más utilizados a nivel internacional para diagnosticar la ELA [1]. Estos criterios clasifican la enfermedad según el grado de certeza del diagnóstico, dependiendo de las zonas del cuerpo afectadas y de la presencia de signos relacionados con daño en las motoneuronas superiores e inferiores. Más adelante, en 2008, los criterios de Awaji mejoraron la sensibilidad del diagnóstico al incluir las fasciculaciones detectadas mediante electromiografía como un signo importante de afectación neuronal temprana [11].

La electromiografía es una de las pruebas más importantes para el diagnóstico de la ELA. Esta técnica permite detectar alteraciones en los músculos causadas por el daño de las motoneuronas, incluso antes de que aparezcan síntomas visibles [16]. Sin embargo, en las primeras fases de la enfermedad los resultados pueden no ser claros, especialmente en formas de evolución lenta o poco comunes. Además, la electromiografía es una prueba invasiva, puede resultar incómoda para el paciente y necesita ser realizada por profesionales especializados, lo que dificulta su uso frecuente [5].

Por otro lado, la resonancia magnética convencional se utiliza principalmente para descartar otras enfermedades que puedan producir síntomas parecidos a los de la ELA [4]. En muchas ocasiones, las imágenes son normales o muestran cambios poco específicos en las fases iniciales, por lo que esta técnica no es suficiente para confirmar el diagnóstico por sí sola.

En los últimos años, también se han estudiado biomarcadores presentes en el líquido cefalorraquídeo y en sangre. Entre ellos destacan los neurofilamentos, unas proteínas que aumentan cuando existe daño neuronal. Los niveles elevados de estas proteínas se relacionan con la velocidad de progresión y el pronóstico de la enfermedad [7]. Sin embargo, estos biomarcadores todavía no son completamente específicos para la ELA, ya que también pueden aparecer elevados en otras enfermedades neurológicas.

Debido a todas estas dificultades, el tiempo entre la aparición de los primeros síntomas y el diagnóstico suele ser de entre 10 y 14 meses [2,5]. Este retraso puede tener consecuencias importantes, ya que dificulta comenzar el tratamiento de forma temprana, retrasa la planificación de los cuidados y limita el seguimiento inicial de la enfermedad. Por ello, uno de los principales objetivos actuales de la investigación en ELA es conseguir diagnósticos más rápidos y precisos, para poder iniciar antes el tratamiento y mejorar la calidad de vida de los pacientes [7].

1.7 NEUROIMAGEN AVANZADA EN ELA

Debido a las dificultades para diagnosticar la esclerosis lateral amiotrófica en fases tempranas, en los últimos años ha aumentado mucho el interés por encontrar biomarcadores que permitan detectar la enfermedad de manera rápida y fiable [7]. Por ello, las técnicas de imagen médica se han convertido en herramientas muy importantes para estudiar las enfermedades neurodegenerativas.

La neuroimagen avanzada ha mejorado mucho el estudio de la ELA durante los últimos años. Técnicas como la morfometría basada en vóxeles (VBM), la imagen por tensor de difusión (DTI), la curtosis de difusión (DKI), el marcaje de spin arterial (ASL) o la resonancia magnética funcional en reposo (rs-fMRI) permiten detectar cambios en el cerebro relacionados con la enfermedad [7]. Gracias a estas técnicas, se ha descubierto que la ELA no afecta solo a las zonas encargadas del movimiento, sino también a otras regiones del cerebro relacionadas con funciones cognitivas, emocionales y de coordinación.

La imagen por tensor de difusión (DTI) ha demostrado ser especialmente útil para detectar daños en la sustancia blanca del cerebro, sobre todo en las vías nerviosas relacionadas con el movimiento [7]. Algunas medidas obtenidas con esta técnica, como la fracción de anisotropía (FA) y la difusividad media (MD), muestran diferencias claras entre personas con ELA y personas sanas. Además, algunos estudios han relacionado estos cambios con la evolución de la enfermedad.

Sin embargo, aunque estas técnicas ofrecen resultados muy prometedores, también tienen algunas limitaciones. Son pruebas costosas, no están disponibles en todos los hospitales y requieren equipos y personal especializado para analizar los resultados [7]. Además, algunas necesitan bastante tiempo para realizarse y procesarse, lo que dificulta su uso habitual en la práctica clínica. Por eso, actualmente se busca desarrollar herramientas más rápidas, sencillas y accesibles para poder utilizarlas de forma más frecuente en hospitales y centros médicos.

1.8 LA ECOGRAFÍA MUSCULAR COMO HERRAMIENTA ALTERNATIVA

La ecografía muscular se ha convertido en una técnica muy prometedora para estudiar enfermedades neuromusculares y neurodegenerativas, como la esclerosis lateral amiotrófica [16,17]. La ecografía tiene varias ventajas frente a otras técnicas de imagen médica: es una prueba no invasiva, accesible, más económica, rápida y puede repetirse fácilmente [18]. Además, permite obtener imágenes del músculo en tiempo real sin necesidad de usar radiación ni contraste intravenoso. Por todo ello, la ecografía muscular resulta muy útil para evaluar y seguir la evolución de pacientes con enfermedades neurodegenerativas [22].

En los pacientes con ELA, la pérdida progresiva de las motoneuronas provoca cambios importantes en los músculos, como pérdida de masa muscular, acumulación de grasa y cambios en la estructura del tejido muscular [16,17]. Muchos de estos cambios pueden detectarse mediante ecografía incluso antes de que los síntomas sean claramente visibles, lo que hace que esta técnica sea de gran interés para el diagnóstico temprano de la enfermedad [23].

Cuando las motoneuronas dejan de funcionar correctamente, el músculo se debilita, pierde volumen y su estructura interna comienza a cambiar poco a poco. Además, parte del tejido muscular puede ser sustituido por grasa y tejido conectivo [27]. Todos estos cambios modifican la apariencia del músculo en las imágenes ecográficas.

Tradicionalmente, la ecografía muscular se utilizaba sobre todo para medir el grosor del músculo o detectar fasciculaciones, que son pequeños movimientos involuntarios de las fibras musculares [18,25]. Sin embargo, gracias al desarrollo de nuevas técnicas informáticas, actualmente es posible obtener mucha más información a partir de las imágenes. Hoy en día pueden analizarse características como el grosor muscular, la ecogenicidad y la ecotextura, que permiten describir de forma más objetiva el estado del músculo [16].

Durante muchos años, el análisis de las ecografías dependía principalmente de la experiencia del especialista, que interpretaba visualmente las imágenes para detectar posibles alteraciones musculares. El problema de este método es que los resultados podían variar según la persona que realizara la evaluación y era difícil medir pequeños cambios de forma precisa [18].

Con el avance del procesamiento digital de imágenes, comenzó a ser posible analizar las ecografías de forma más objetiva y cuantitativa. En lugar de basarse solo en la observación visual, las imágenes podían convertirse en datos numéricos que describieran matemáticamente la estructura del músculo [24]. Esto supuso un avance muy importante, ya que permitió estudiar los cambios musculares de manera más precisa, medible y reproducible.

Uno de los parámetros más utilizados es el grosor muscular, también llamado *Muscle Thickness*. La pérdida progresiva de masa muscular es una de las características más visibles de la ELA, especialmente en fases avanzadas [20]. Mediante ecografía se puede medir fácilmente el grosor de distintos músculos y observar cómo cambia con el tiempo. Diferentes estudios han demostrado que los pacientes con ELA presentan una reducción importante del grosor muscular en comparación con personas sanas [18,20].

Sin embargo, el grosor muscular no siempre es suficiente para detectar cambios tempranos. En las primeras fases de la enfermedad, algunos músculos pueden parecer normales en tamaño, aunque ya estén sufriendo alteraciones internas importantes [16]. Por esta razón, comenzó a estudiarse también la ecogenicidad muscular.

La ecogenicidad se refiere al brillo o intensidad de la imagen ecográfica dentro del músculo. Un músculo sano suele tener una apariencia homogénea y relativamente oscura en la ecografía [27]. Cuando aparece acumulación de grasa y fibrosis debido a la degeneración muscular, el músculo se vuelve más brillante en la imagen. Diferentes estudios han demostrado que los pacientes con ELA presentan un aumento de la ecogenicidad muscular, especialmente en fases más avanzadas de la enfermedad [19,23].

Sin embargo, la ecogenicidad por sí sola también tiene algunas limitaciones. Aunque permite detectar cambios generales en el músculo, no siempre muestra correctamente cambios más complejos relacionados con la organización interna del tejido muscular [16]. Por esta razón, en los últimos años ha aumentado mucho el interés por el análisis de textura en imágenes ecográficas.

El análisis de textura consiste en estudiar matemáticamente cómo se distribuyen los tonos grises dentro de una imagen ecográfica. En lugar de analizar solo el brillo medio de la imagen, esta técnica estudia patrones más complejos que pueden reflejar cambios en la estructura interna del músculo [28,32]. Esto es especialmente útil en enfermedades neurodegenerativas, donde pueden aparecer alteraciones internas difíciles de detectar visualmente, pero que sí quedan reflejadas en la distribución de los píxeles de la imagen.

Una de las técnicas más utilizadas para este análisis es la matriz de co-ocurrencia de niveles de gris, conocida como *GLCM (Gray-Level Co-occurrence Matrix)*. Esta técnica estudia cómo se relacionan entre sí los diferentes tonos grises de la imagen según su posición y distancia [28]. A partir de esta información se pueden obtener distintos parámetros matemáticos que describen con bastante precisión la textura del músculo.

Entre los parámetros más importantes obtenidos con *GLCM* destacan el contraste, la correlación, la homogeneidad, la energía y la entropía [28,35]. El contraste mide las diferencias de intensidad entre píxeles cercanos; la homogeneidad indica si la imagen es uniforme; la energía mide el grado de orden de la imagen; y la entropía refleja el nivel de desorden o complejidad de la textura. Todos estos parámetros permiten describir de manera mucho más detallada la estructura muscular.

Además de la *GLCM*, también existen otras técnicas de análisis textural. Una de ellas es la matriz de longitud de corrida de niveles de gris (*Gray-Level Run-Length Matrix, GLRLM*), que analiza grupos de píxeles consecutivos con la misma intensidad y aporta información sobre la organización del tejido [29]. Otra técnica son los patrones binarios locales (*Local Binary Patterns, LBP*), que estudian la relación entre cada píxel y sus vecinos cercanos para obtener descriptores de textura más precisos [30].

En pacientes con esclerosis lateral amiotrófica, la degeneración muscular modifica progresivamente la estructura interna del músculo. Como consecuencia, los parámetros texturales también cambian de forma importante respecto a personas sanas. Diferentes estudios han demostrado que algunos de estos parámetros permiten diferenciar pacientes con ELA de personas sanas incluso en fases tempranas de la enfermedad, cuando otros indicadores todavía no muestran cambios claros [16,17].

Además, algunos músculos parecen ser especialmente sensibles a estos cambios, como el tibial anterior, el cuádriceps, el bíceps braquial o los músculos flexores del antebrazo. Estos músculos muestran alteraciones importantes en los parámetros ecográficos y texturales, posiblemente porque son más vulnerables al proceso neurodegenerativo asociado a la ELA [16,17].

En muchos casos, los cambios observados mediante ecografía aparecen incluso antes de que existan síntomas claramente visibles [17]. Esto convierte a la ecografía muscular en una herramienta muy útil para el diagnóstico temprano y el seguimiento de la enfermedad. Además, al tratarse de una técnica no invasiva, puede repetirse fácilmente y realizar controles periódicos sin causar molestias ni riesgos para el paciente.

1.9 INTELIGENCIA ARTIFICIAL APLICADA AL DIAGNÓSTICO DE LA ELA

Al mismo tiempo que avanzaba el análisis de imágenes médicas, la Inteligencia Artificial también ha crecido enormemente en el ámbito médico y científico. Actualmente, técnicas como el *Machine Learning* y el *Deep Learning* están transformando muchas áreas de la medicina, especialmente las relacionadas con la imagen médica y el análisis de datos complejos [39].

El aprendizaje automático permite crear modelos capaces de encontrar patrones complejos dentro de grandes cantidades de datos. A diferencia de los métodos estadísticos tradicionales, estos algoritmos pueden detectar relaciones difíciles de identificar manualmente [36]. Esto es especialmente útil en enfermedades como la esclerosis lateral amiotrófica, donde intervienen muchos biomarcadores y factores diferentes.

En medicina, estos algoritmos se utilizan cada vez más para ayudar en el diagnóstico, predecir la evolución de enfermedades, analizar imágenes médicas y apoyar la toma de decisiones clínicas [39,51]. En las enfermedades neurodegenerativas, donde existe una gran complejidad biológica, la Inteligencia Artificial tiene un potencial muy importante.

En enfermedades como el Alzheimer y el Parkinson, la Inteligencia Artificial ya ha mostrado resultados muy prometedores utilizando resonancia magnética, PET, electroencefalografía y otras técnicas de imagen médica [39]. Algunos modelos alcanzan niveles de precisión superiores al 90%, especialmente cuando se utilizan redes neuronales profundas capaces de analizar automáticamente las imágenes médicas [50].

En el caso de la ELA, la investigación con Inteligencia Artificial todavía está menos desarrollada. Esto se debe principalmente a que existen menos bases de datos disponibles y menos pacientes, ya que se trata de una enfermedad poco frecuente [46]. Aun así, en los últimos años han aparecido estudios prometedores que utilizan imágenes médicas, electromiografía, datos clínicos y datos genéticos para clasificar pacientes de forma

automática. Estos trabajos muestran que la Inteligencia Artificial podría convertirse en una herramienta muy útil para ayudar en el diagnóstico y seguimiento de la ELA [46].

Uno de los aspectos más interesantes del *Machine Learning* aplicado a ecografías musculares es que puede analizar muchas variables al mismo tiempo. Mientras que los métodos tradicionales suelen estudiar cada parámetro por separado, los algoritmos de aprendizaje automático pueden combinar diferentes características y encontrar patrones mucho más complejos [36]. Por ejemplo, un modelo puede analizar simultáneamente el grosor muscular, la ecogenicidad y diferentes parámetros de textura de varios músculos para distinguir automáticamente entre pacientes con ELA y personas sanas. Gracias a esto, el rendimiento diagnóstico puede mejorar considerablemente [16,17].

Entre los algoritmos más utilizados en medicina destacan la regresión logística, las máquinas de soporte vectorial (*SVM*), los árboles de decisión, *Random Forest*, *KNN*, *Gradient Boosting* y las redes neuronales artificiales [36,37,38,40]. Cada uno tiene ventajas y limitaciones dependiendo del tipo de datos y del tamaño de la muestra disponible. Algunos modelos son más fáciles de interpretar, mientras que otros pueden detectar relaciones más complejas entre las variables.

Uno de los principales problemas en el desarrollo de modelos de Inteligencia Artificial es el sobreajuste u *overfitting*. Esto ocurre cuando el modelo aprende demasiado bien los datos utilizados para entrenarlo, pero después funciona peor con nuevos pacientes [44]. Este problema es especialmente importante en enfermedades raras como la ELA, donde el número de pacientes disponibles suele ser pequeño. Por ello, es necesario utilizar métodos de validación adecuados y evaluar el rendimiento de los modelos mediante diferentes métricas de precisión y fiabilidad [44,45].

Además, en medicina no es suficiente con que un algoritmo sea preciso, también es importante entender cómo toma sus decisiones. Por esta razón, en los últimos años ha aumentado el interés por desarrollar modelos de Inteligencia Artificial más interpretables. Técnicas como *SHAP* o *LIME* permiten identificar qué variables influyen más en la

clasificación y ayudan a que los médicos comprendan mejor el funcionamiento del modelo [48,49].

1.10 OBJETIVO DEL PRESENTE TRABAJO

Este Trabajo Fin de Grado se centra en el desarrollo y evaluación de un sistema basado en Inteligencia Artificial para analizar imágenes musculares de pacientes con esclerosis lateral amiotrófica, utilizando datos obtenidos mediante ecografía muscular [16,17].

El objetivo principal es estudiar si diferentes modelos de *Machine Learning* son capaces de diferenciar entre pacientes con ELA y personas sanas a partir de información obtenida de las imágenes ecográficas.

Para lograr este objetivo principal, se plantean los siguientes objetivos secundarios:

1. Comparar varios algoritmos de *Machine Learning* utilizando los mismos datos y las mismas condiciones de evaluación, con el fin de identificar cuál ofrece mejores resultados diagnósticos con este tipo de datos médicos [36,43,44].
2. Estudiar distintos tipos de biomarcadores ecográficos, analizando tanto los parámetros clásicos como los descriptores texturales obtenidos mediante *GLCM*, así como su combinación, para determinar qué información es más útil para diferenciar pacientes con ELA de personas sanas [16,28].
3. Analizar la importancia de las variables en los modelos desarrollados, identificando qué biomarcadores ecográficos y qué músculos contribuyen más a la clasificación, con el fin de obtener conclusiones clínicamente relevantes [16,48,49].
4. Realizar un análisis específico del músculo cuádriceps femoral y comparar los resultados obtenidos mediante *Machine Learning* con los de investigaciones previas basadas en métodos estadísticos [16].
5. Llevar a cabo un análisis exploratorio sobre la progresión de la enfermedad a lo largo del tiempo, como base para futuras investigaciones con seguimiento longitudinal [17].

1.11 ESTRUCTURA DEL PRESENTE TRABAJO

El documento se organiza en cinco capítulos principales, además de la bibliografía y los anexos.

El Capítulo 2 recoge el estado del arte, donde se revisan los principales biomarcadores utilizados en el diagnóstico de la ELA, la evolución de la ecografía muscular como herramienta diagnóstica, las técnicas de análisis textural aplicadas a imagen médica y el uso de la Inteligencia Artificial en enfermedades neurodegenerativas. También se presentan los algoritmos de *Machine Learning* más utilizados en diagnóstico biomédico y se discuten las principales limitaciones de los estudios actuales.

El Capítulo 3 describe la metodología empleada en el trabajo. Se explica el diseño del estudio, las características del *dataset* utilizado, las variables analizadas y el proceso de preprocesamiento de los datos. También se detallan los modelos de *Machine Learning* implementados, la estructura de los *pipelines* de entrenamiento, la técnica de validación cruzada utilizada y las métricas de evaluación seleccionadas.

El Capítulo 4 presenta los resultados obtenidos en los distintos experimentos realizados. Se compara el rendimiento de los diferentes modelos según el conjunto de variables utilizado, se analiza el comportamiento específico del músculo cuádriceps femoral, se estudia la importancia de cada biomarcador en la clasificación y se evalúa el efecto de incorporar variables clínicas. El capítulo termina con una discusión de los resultados y su comparación con estudios previos.

El Capítulo 5 recoge las conclusiones del trabajo, resumiendo los principales hallazgos obtenidos e interpretándolos desde el punto de vista clínico. También se identifican las limitaciones del estudio y se proponen líneas de investigación futura que podrían ampliar y mejorar los resultados alcanzados.

Finalmente, se incluyen la bibliografía con todas las referencias citadas a lo largo del documento y los anexos con información complementaria al trabajo.

Capítulo 2. ESTADO DEL ARTE

2.1 INTRODUCCIÓN

Durante los últimos años, el estudio de las enfermedades neurodegenerativas ha avanzado mucho gracias al desarrollo de nuevas técnicas de imagen médica, análisis computacional e Inteligencia Artificial [39]. La esclerosis lateral amiotrófica sigue siendo uno de los mayores retos tanto para la neurología como para la ingeniería biomédica [7,8]. Aunque la investigación ha progresado considerablemente, el diagnóstico de esta enfermedad continúa siendo complicado, especialmente en las primeras fases, debido a que todavía no existen biomarcadores rápidos, objetivos y de fácil acceso [7,16].

Tradicionalmente, el diagnóstico de la ELA se ha basado en la evaluación clínica, la exploración neurológica y pruebas como la electromiografía [1]. Sin embargo, estos métodos presentan varias limitaciones. Muchas veces los primeros síntomas pueden confundirse con otras enfermedades neurológicas o musculares. Además, algunas pruebas son invasivas, dependen mucho de la experiencia del especialista y no siempre permiten detectar la enfermedad en etapas tempranas [5].

El retraso en el diagnóstico es un problema importante en la ELA. Cuanto más tiempo pasa sin un diagnóstico definitivo, más avanza la enfermedad sin que el paciente pueda recibir los tratamientos y cuidados adecuados [2,5]. Por esta razón, uno de los principales objetivos actuales es conseguir diagnósticos más rápidos y precisos mediante biomarcadores fiables [7].

Debido a estas dificultades, en los últimos años ha aumentado mucho el interés por desarrollar biomarcadores cuantitativos capaces de detectar la enfermedad de forma más temprana y objetiva [7]. En este campo, las técnicas de imagen médica han adquirido una gran importancia, ya que permiten obtener información sobre la estructura y el

funcionamiento de músculos y nervios. Esto ha abierto nuevas líneas de investigación centradas en el diagnóstico precoz y el seguimiento de la evolución de la enfermedad [16,17].

Entre las diferentes técnicas disponibles, la ecografía muscular ha despertado un gran interés por sus ventajas clínicas. Es una técnica rápida, económica, no invasiva y muy accesible, que además permite obtener imágenes en tiempo real sin utilizar radiación ni procedimientos agresivos para el paciente [18,22]. Gracias a estas características, la ecografía se ha convertido en una herramienta muy útil para estudiar enfermedades neuromusculares y neurodegenerativas.

Al principio, la ecografía muscular se utilizaba principalmente para detectar fasciculaciones o medir el grosor muscular [18,25]. Sin embargo, el desarrollo de nuevas técnicas de procesamiento digital de imágenes ha permitido obtener mucha información a partir de las ecografías. Actualmente, además de estudiar el grosor muscular o la ecogenicidad, también es posible analizar características de textura relacionadas con la estructura interna del músculo [16,32].

El análisis textural es una de las áreas más innovadoras dentro del análisis de imagen médica. Estas técnicas permiten estudiar cómo se distribuyen los tonos de grises dentro de la imagen para obtener información más detallada sobre la estructura del tejido muscular [28,34]. En enfermedades como la ELA, donde aparecen cambios como pérdida de inervación, fibrosis y acumulación de grasa muscular, la textura de la imagen ecográfica también cambia y puede analizarse mediante métodos computacionales [16,17].

Además, la Inteligencia Artificial y el *Machine Learning* han crecido enormemente dentro del ámbito biomédico. Actualmente, muchos estudios utilizan algoritmos capaces de analizar grandes cantidades de datos médicos y encontrar patrones complejos que serían difíciles de detectar con métodos tradicionales [36,39]. Estos modelos son especialmente útiles en enfermedades complejas como la ELA, donde intervienen muchos factores diferentes.

La combinación entre biomarcadores obtenidos mediante ecografía e Inteligencia Artificial es actualmente una de las líneas de investigación más prometedoras para mejorar el

diagnóstico de la ELA [16,17,46]. Gracias a estos sistemas, es posible analizar muchas variables al mismo tiempo y desarrollar herramientas automáticas capaces de diferenciar pacientes con ELA de personas sanas con alta precisión, manteniendo además el bajo coste y la accesibilidad de la ecografía muscular.

Este capítulo revisa el estado actual de la investigación relacionada con biomarcadores ecográficos, análisis textural e Inteligencia Artificial aplicados al estudio de la ELA. También se presentan las principales técnicas utilizadas, los resultados obtenidos en investigaciones anteriores y las ventajas y limitaciones de cada metodología.

2.2 BIOMARCADORES EN ENFERMEDADES NEURODEGENERATIVAS

Uno de los principales objetivos de la investigación médica actual es encontrar biomarcadores que permitan detectar enfermedades neurodegenerativas de forma temprana, objetiva y fiable [7]. Un biomarcador es cualquier característica del organismo que pueda medirse y que aporta información sobre una enfermedad o sobre la respuesta a un tratamiento. Para que un biomarcador sea útil, debería ser fácil de obtener, dar resultados similares en distintos hospitales y especialistas, detectar cambios importantes de la enfermedad y diferenciar la enfermedad correctamente de otras patologías [34].

En enfermedades como el alzheimer, el parkinson, la esclerosis múltiple o la esclerosis lateral amiotrófica, los biomarcadores tienen una gran importancia. Ayudan a detectar la enfermedad antes, permiten seguir su evolución con el tiempo y también sirven para comprobar si los tratamientos funcionan correctamente durante los ensayos clínicos [7,8].

Actualmente, existen diferentes tipos de biomarcadores en neurología, dependiendo de la información que proporcionan y de la técnica utilizada para obtenerlos.

Los biomarcadores moleculares son sustancias que pueden medirse en fluidos del cuerpo, como la sangre o el líquido cefalorraquídeo. En la ELA, algunos de los más estudiados son los neurofilamentos, unas proteínas que aumentan cuando existe daño en las neuronas [7].

Diferentes estudios han demostrado que los pacientes con ELA suelen tener niveles más altos de estas proteínas que las personas sanas. Además, estos niveles pueden estar relacionados con la velocidad de progresión de la enfermedad [7]. Sin embargo, estos biomarcadores no son totalmente específicos, ya que también pueden aumentar en otras enfermedades neurológicas.

También existen biomarcadores genéticos, que permiten detectar mutaciones relacionadas con formas hereditarias de la ELA. Algunas mutaciones importantes se encuentran en genes como SOD1, C9ORF72, TDP-43 o FUS [3,14]. Estos biomarcadores pueden ayudar tanto en el diagnóstico como en el pronóstico de algunos pacientes. Aun así, la mayoría de los casos de ELA no son hereditarios, por lo que los biomarcadores genéticos tienen una aplicación limitada como herramienta diagnóstica general.

Otro grupo importante son los biomarcadores funcionales, como los obtenidos mediante electromiografía (EMG) y estudios de conducción nerviosa. La electromiografía permite detectar daños en los músculos y en las motoneuronas incluso antes de que algunos síntomas sean claramente visibles [11]. Sin embargo, es una prueba invasiva y, en fases tempranas, los resultados pueden no ser completamente claros [5].

Por último, los biomarcadores obtenidos mediante imagen médica han ganado mucha importancia en los últimos años porque permiten detectar cambios en músculos y nervios de forma no invasiva [7,39]. Entre las técnicas más utilizadas destacan la resonancia magnética, el PET, la tomografía computarizada y, más recientemente, la ecografía muscular cuantitativa [16,18].

En la ELA, encontrar biomarcadores útiles es especialmente difícil debido a la gran variabilidad de la enfermedad [8]. Cada paciente puede presentar síntomas diferentes, velocidades de progresión distintas y grados variables de afectación muscular y cognitiva. Esta diversidad hace más complicado desarrollar herramientas diagnósticas válidas para todos los pacientes [7].

Por esta razón, actualmente existe un gran interés por desarrollar biomarcadores más objetivos y fiables que ayuden a reducir el tiempo de diagnóstico, mejorar el seguimiento de la enfermedad y facilitar los ensayos clínicos [7,16]. La combinación de distintos tipos de biomarcadores, integrando información molecular, funcional y de imagen médica, parece ser una de las estrategias más prometedoras para avanzar en este campo.

2.3 EVOLUCIÓN DE LA ECOGRAFÍA MUSCULAR EN NEUROLOGÍA

La ecografía muscular ha mejorado muchísimo durante las últimas décadas. Al principio era una técnica con pocas aplicaciones, pero hoy en día se ha convertido en una herramienta muy importante para estudiar enfermedades neuromusculares [22]. Esta evolución ha sido posible gracias a las mejoras tecnológicas de los equipos de ultrasonidos, que ahora ofrecen imágenes con mayor calidad, mejor resolución y sistemas digitales más avanzados.

En los años 80, cuando comenzó a utilizarse en neurología, la calidad de las imágenes era bastante limitada y resultaba difícil observar con detalle la estructura interna de los músculos. Las exploraciones se utilizaban principalmente para detectar lesiones grandes o analizar el músculo de manera muy general. Sin embargo, con el desarrollo de sondas de alta frecuencia y la mejora de los sistemas de procesamiento de imagen, las ecografías empezaron a mostrar estructuras mucho más pequeñas y cambios más difíciles de detectar [22].

Actualmente, la ecografía muscular se utiliza de forma habitual para estudiar enfermedades neuromusculares, lesiones nerviosas, neuropatías, enfermedades musculares inflamatorias y enfermedades de motoneurona como la ELA [22,26]. Sus principales ventajas son que es una técnica rápida, económica, segura y fácil de utilizar tanto en investigación como en hospitales y consultas médicas.

Una de las mayores ventajas de la ecografía es su accesibilidad. A diferencia de otras técnicas más complejas, puede realizarse rápidamente en consulta, tiene un coste bajo y no requiere radiación ni procedimientos invasivos [18]. Esto es especialmente importante en

enfermedades progresivas como la ELA, donde es necesario repetir las pruebas periódicamente para controlar la evolución del paciente [17].

Además, la ecografía permite obtener imágenes en tiempo real, algo que no ocurre con otras técnicas como la resonancia magnética o el PET. Gracias a ello, es posible observar fasciculaciones musculares y otros movimientos involuntarios mientras ocurren, lo que resulta muy útil en la ELA, ya que las fasciculaciones son uno de los signos más importantes de la enfermedad [25].

En los pacientes con ELA, la ecografía muscular comenzó utilizándose principalmente para detectar estas fasciculaciones. Algunos estudios demostraron incluso que la ecografía podía detectar mejor las fasciculaciones que la electromiografía en ciertos músculos, sobre todo en músculos profundos o difíciles de explorar con electrodos [25]. Este descubrimiento hizo que la ecografía empezara a utilizarse como una herramienta complementaria para el diagnóstico de la ELA.

Más adelante, los investigadores comenzaron a estudiar también los cambios estructurales del músculo. Se observó que la degeneración de las motoneuronas provoca pérdida de masa muscular, fibrosis y acumulación de grasa en el músculo [27]. Todos estos cambios modifican la apariencia del músculo en la ecografía haciendo que se vea más brillante, menos uniforme y más delgado. Para clasificar estos cambios se desarrolló la escala de Heckmatt, que divide la ecogenicidad muscular en cuatro niveles según el grado de afectación. Aunque esta escala se creó inicialmente para enfermedades musculares, después también empezó a utilizarse en enfermedades de motoneuronas [22].

Gracias a estos avances, comenzaron a desarrollarse métodos capaces de analizar las imágenes musculares de forma objetiva y obtener datos numéricos relacionados con la degeneración muscular [16,24]. Así surgió la denominada ecografía muscular cuantitativa (*Quantitative Muscle Ultrasound, QMUS*), que es la base metodológica utilizada en este trabajo.

2.4 BIOMARCADORES ECOGRÁFICOS CLÁSICOS

Los biomarcadores ecográficos más utilizados en enfermedades neuromusculares pueden dividirse en diferentes grupos según la información que aportan y el tipo de análisis necesario para obtenerlos [16,18].

2.4.1 GROSOR MUSCULAR

El grosor muscular, también conocido como *Muscle Thickness*, es uno de los biomarcadores más sencillos y utilizados en ecografía muscular [18]. Este parámetro mide el tamaño o espesor del músculo en la imagen ecográfica y permite estudiar la pérdida progresiva de masa muscular causada por la degeneración de las motoneuronas.

En pacientes con esclerosis lateral amiotrófica, muchos estudios han demostrado que el grosor muscular suele ser menor que en personas sanas [16,18,20]. Esta reducción suele observarse especialmente en músculos más afectados por la enfermedad. En general, la disminución del grosor muscular refleja la atrofia que aparece cuando el músculo deja de recibir correctamente las señales nerviosas y parte de sus fibras se sustituyen por grasa y tejido conectivo [27].

Además, el grosor muscular también puede servir para seguir la evolución de la enfermedad. Algunos estudios han observado que el grosor muscular disminuye progresivamente con el tiempo y que esta reducción se relaciona con el empeoramiento funcional del paciente medido mediante la escala ALSFRS-R [17,20]. Esto sugiere que el grosor muscular puede ser útil para controlar la progresión de la ELA.

Sin embargo, este parámetro también tiene limitaciones importantes. En las primeras fases de la enfermedad, algunos músculos pueden mantener un tamaño aparentemente normal, aunque ya existan cambios internos importantes [16]. Además, factores como la edad, el sexo, el peso o el nivel de actividad física pueden influir en el grosor muscular y dificultar la comparación entre pacientes. Por esta razón, es necesario utilizar valores de referencia adecuados [18].

2.4.2 ECOGENICIDAD MUSCULAR

La ecogenicidad mide el brillo medio de la imagen ecográfica dentro de la zona muscular analizada. Este parámetro está relacionado con la capacidad del tejido muscular para reflejar los ultrasonidos y depende de la composición interna del músculo [16,24].

En músculos sanos, el tejido muscular suele verse relativamente uniforme y oscuro en la ecografía [27]. Sin embargo, cuando aparecen procesos como fibrosis o acumulación de grasa debido a la degeneración muscular, el músculo refleja más los ultrasonidos y se vuelve más brillante en la imagen.

Diferentes estudios han demostrado que los pacientes con ELA presentan un aumento de ecogenicidad en comparación con personas sanas [16,18,19,23]. Estos cambios están relacionados con el daño muscular causado por la pérdida progresiva de motoneuronas y pueden aparecer tanto en músculos cercanos como alejados del tronco.

La ecogenicidad tiene la ventaja de ser relativamente fácil de medir utilizando una región concreta de la imagen ecográfica. Sin embargo, también presenta algunas limitaciones. Factores técnicos como el tipo de ecógrafo, la sonda utilizada, la profundidad de la exploración o la presión aplicada sobre la piel pueden modificar los resultados y dificultar la comparación entre hospitales o estudios diferentes [16]. Por esta razón algunos investigadores han desarrollado métodos para corregir y normalizar estas medidas.

Además, la ecogenicidad solo informa del brillo medio de la imagen, pero no describe cómo están organizados internamente los tejidos musculares. Es decir, no tiene en cuenta la distribución espacial de los píxeles ni las relaciones entre ellos [16,28]. Esta limitación hizo que comenzaran a desarrollarse técnicas más avanzadas de análisis de textura, que permiten estudiar con mayor detalle la estructura interna del músculo.

2.5 ANÁLISIS TEXTURAL EN IMAGEN MÉDICA

Para superar las limitaciones de los biomarcadores clásicos, especialmente el hecho de que la ecogenicidad solo mide el brillo medio de la imagen, comenzaron a desarrollarse técnicas más avanzadas de análisis de textura aplicadas a la imagen médica [28,32].

El análisis textural consiste en estudiar matemáticamente cómo se distribuyen los tonos grises dentro de una imagen. Mientras que parámetros simples como la ecogenicidad solo analizan el brillo general, las técnicas texturales permiten estudiar relaciones más complejas entre píxeles cercanos y describir con más detalle la estructura interna del tejido [28,35]. Gracias a ello, es posible obtener información sobre la uniformidad, el contraste, el orden o la complejidad de la imagen, características que no pueden analizarse con los biomarcadores clásicos.

La textura de una imagen médica refleja la estructura interna del tejido que se está estudiando. En la ecografía muscular, esta textura depende de diferentes componentes del músculo, como las fibras musculares, el tejido conectivo, los vasos sanguíneos y la grasa [32]. Cuando estas estructuras cambian debido a una enfermedad, también cambia la textura de la imagen ecográfica, y esos cambios pueden medirse utilizando métodos matemáticos.

Estas técnicas se utilizan actualmente en muchas áreas de la medicina. En oncología, por ejemplo, el análisis textural de imágenes de resonancia magnética o PET se utiliza para estudiar tumores y predecir la respuesta al tratamiento [34]. En cardiología, ayuda a detectar fibrosis en el corazón. En neurología, se ha utilizado para estudiar enfermedades como la esclerosis múltiple, el alzheimer y los traumatismos craneales. Además, en enfermedades musculares y neuromusculares, el análisis textural de ecografías musculares está mostrando resultados muy prometedores [33].

En el caso de la esclerosis lateral amiotrófica, el análisis textural es especialmente útil porque la enfermedad modifica progresivamente la estructura interna del músculo. La pérdida de fibras musculares, la aparición de fibrosis y la acumulación de grasa producen cambios complejos en la imagen ecográfica [16,17]. Estos cambios son difíciles de detectar utilizando

solo parámetros simples como el grosor muscular o la ecogenicidad, pero pueden analizarse con mucho más detalle mediante técnicas texturales.

Entre las técnicas de análisis textural más utilizadas destacan la matriz de co-ocurrencia de niveles de gris (*GLCM*) [28], la matriz de longitud de corrida de niveles de gris (*GLRLM*) [29], los patrones binarios locales (*LBP*) [30], la transformada wavelet y el análisis fractal [33]. Cada una de estas técnicas proporciona información diferente sobre la organización interna del músculo, y la combinación de varias de ellas puede mejorar la capacidad diagnóstica del sistema.

2.6 MATRIZ DE CO-OCURRENCIA DE NIVELES DE GRIS (*GLCM*)

La matriz de co-ocurrencia de niveles de gris, conocida como *GLCM* (*Gray-Level Co-occurrence Matrix*), es una de las técnicas más utilizadas para analizar la textura de imágenes médicas. Fue desarrollada en 1973 por Haralick y sus colaboradores, y desde entonces se ha convertido en una herramienta muy importante para estudiar imágenes biomédicas [28].

La *GLCM* funciona analizando cómo se relacionan entre sí los diferentes tonos de gris de una imagen. Concretamente, estudia cuántas veces aparecen juntos ciertos niveles de gris en píxeles cercanos y en diferentes direcciones de la imagen [28,35]. A partir de esta información se crea una matriz matemática que permite describir la textura del tejido de forma objetiva.

Normalmente, la *GLCM* se calcula en varias direcciones diferentes (0°, 45°, 90° y 135°) para obtener resultados más completos y evitar que la orientación del músculo influya demasiado en el análisis [16,28]. Esto es especialmente importante en ecografía muscular, ya que las fibras musculares pueden verse en distintas posiciones dependiendo del músculo y de cómo se coloque la sonda.

A partir de esta matriz pueden calcularse diferentes parámetros matemáticos que describen características de la textura muscular. Los más importantes son: [16,28,35]

- **Contraste (CON):** mide las diferencias de intensidad entre píxeles cercanos. Valores altos indican que existen muchos cambios bruscos en la imagen, algo frecuente en músculos degenerados.
- **Correlación (COR):** analiza si los tonos grises de los píxeles cercanos cambian de forma parecida. Valores altos suelen indicar una estructura más organizada.
- **Homogeneidad (IDM, *Inverse Difference Moment*):** mide si la imagen es uniforme. Los músculos sanos suelen tener valores más altos porque su estructura interna es más regular.
- **Energía o ASM (*Angular Second Moment*):** evalúa el grado de orden de la imagen. Una textura más uniforme y organizada tendrá valores más altos.
- **Entropía (ENT):** mide el nivel de desorden o complejidad de la textura. Los músculos degenerados suelen presentar valores más altos de entropía debido a su estructura más irregular.

En pacientes con esclerosis lateral amiotrófica, estos parámetros suelen ser diferentes a los de personas sanas. Diversos estudios han encontrado aumentos del contraste y la entropía, junto con disminuciones de la energía, la homogeneidad y la correlación [16,17]. Estos cambios reflejan que el músculo afectado por la enfermedad tiene una estructura más irregular y desorganizada.

Uno de los estudios más importantes sobre este tema fue realizado por Martínez-Payá y sus colaboradores en 2017 [16]. Este trabajo analizó por primera vez los descriptores *GLCM* en ecografías musculares de pacientes con ELA y encontró diferencias claras entre pacientes y personas sanas en todos los músculos estudiados.

Posteriormente, otro estudio realizado en 2018 por estos mismos investigadores analizó cómo evolucionaban estos parámetros con el tiempo en pacientes con ELA [17]. Los resultados mostraron que medidas como el contraste y la entropía eran especialmente útiles para detectar cambios progresivos en la estructura muscular. Esto sugiere que estos parámetros podrían servir no solo para ayudar al diagnóstico, sino también para controlar la evolución de la enfermedad a lo largo del tiempo.

2.7 OTRAS TÉCNICAS DE ANÁLISIS TEXTURAL

Además de la *GLCM*, existen otras técnicas de análisis textural que también se utilizan en imágenes médicas y que aportan información complementaria sobre la estructura interna de los tejidos [32,34].

Una de ellas es la matriz de longitud de corrida de niveles de gris, conocida como *GLRLM* (*Gray-Level Run-Length Matrix*). Esta técnica estudia grupos de píxeles consecutivos que tienen el mismo nivel de gris dentro de la imagen [29]. Mientras que la *GLCM* analiza la relación entre pares de píxeles, la *GLRLM* se centra en detectar regiones uniformes y estructuras lineales dentro del tejido.

Entre los parámetros más importantes de la *GLRLM* destacan: [29]

- *Short Run Emphasis* (SRE): da más importancia a grupos cortos de píxeles y suele relacionarse con texturas finas.
- *Long Run Emphasis* (LRE): detectan estructuras largas y más uniformes.
- *Run Length Non-Uniformity* (RLN): mide si las longitudes de los grupos de píxeles son muy diferentes entre sí.

Otra técnica importante son los patrones binarios locales, conocidos como *LBP* (*Local Binary Patterns*). Esta técnica analiza cada píxel comparándolo con los píxeles que tiene alrededor. A partir de esa comparación se genera un código que describe la textura local de la imagen [30].

Una de las ventajas principales de los *LBP* es que son menos sensibles a cambios de brillo o contraste en la ecografía [30]. Esto es importante porque las imágenes ecográficas pueden variar dependiendo del equipo utilizado o de la configuración del ecógrafo. Gracias a ello, los *LBP* ayudan a obtener resultados más estables y comparables entre distintos hospitales y médicos.

Otra técnica utilizada es la transformada wavelet. Esta herramienta permite analizar la imagen en diferentes niveles de detalle y estudiar la textura a distintas escalas [32]. Esto es especialmente útil en el músculo, ya que su estructura tiene distintos niveles de organización, desde las fibras musculares hasta el músculo completo. Los parámetros obtenidos mediante

wavelets han mostrado buenos resultados en algunos estudios de ecografía muscular, aunque su análisis es más complejo desde el punto de vista computacional.

También existe el análisis fractal, que se utiliza para medir la complejidad geométrica de las estructuras presentes en la imagen [33]. En imagen médica, la dimensión fractal puede reflejar cambios en la organización interna del tejido causados por enfermedades. Aunque esta técnica se utiliza menos en ecografía muscular que la *GLCM*, algunos estudios han encontrado relación entre la dimensión fractal y el grado de afectación muscular en enfermedades neuromusculares, incluida la esclerosis lateral amiotrófica [33].

2.8 INTELIGENCIA ARTIFICIAL APLICADA A IMAGEN MÉDICA

El desarrollo de la Inteligencia Artificial ha cambiado mucho el análisis de imágenes médicas en los últimos años. Gracias a estas tecnologías, ahora es posible analizar grandes cantidades de datos y detectar patrones muy complejos que serían difíciles de encontrar con métodos estadísticos tradicionales [36,39].

Actualmente, la Inteligencia Artificial se utiliza en imagen médica para muchas tareas diferentes, como ayudar en el diagnóstico automático de enfermedades, detectar lesiones, analizar estructuras del cuerpo, predecir la evolución de los pacientes y combinar información obtenida de distintas pruebas médicas [39,51]. En algunos casos, estos sistemas han conseguido resultados similares o incluso mejores que los de médicos especialistas expertos [50].

En las enfermedades neurodegenerativas, donde intervienen muchos factores diferentes y la enfermedad es muy compleja, la Inteligencia Artificial tiene un gran potencial [51]. Estos algoritmos pueden analizar al mismo tiempo muchos biomarcadores distintos y encontrar relaciones entre ellos que serían muy difíciles de detectar manualmente [36].

En el alzheimer, por ejemplo, algunos modelos basados en resonancia magnética han conseguido detectar cambios cerebrales varios años antes de que aparezcan los primeros

síntomas [39]. También se han obtenido resultados muy precisos utilizando imágenes PET y biomarcadores relacionados con proteínas como el amiloide y la tau.

En el parkinson, la Inteligencia Artificial también ha mostrado resultados prometedores mediante el análisis de imágenes cerebrales, la forma de caminar de los pacientes y diferentes biomarcadores obtenidos en pruebas médicas. Además, el uso de técnicas de *Deep Learning* y redes neuronales convolucionales (*CNN*) ha supuesto un gran avance, ya que estos modelos pueden aprender automáticamente las características importantes de las imágenes sin necesidad de seleccionarlás manualmente [50].

En el caso de la esclerosis lateral amiotrófica, la investigación en Inteligencia Artificial todavía es más limitada porque existen menos bases de datos disponibles y menos pacientes para estudiar [46]. Aun así, los resultados publicados en los últimos años son muy prometedores. Diferentes investigadores han utilizado algoritmos de *Machine Learning* para analizar imágenes médicas, electromiografía, datos clínicos y datos genéticos, consiguiendo en muchos casos mejores resultados diagnósticos que los métodos estadísticos tradicionales [16,17,46].

2.9 ALGORITMOS DE MACHINE LEARNING UTILIZADOS EN DIAGNÓSTICO MÉDICO

Existen muchos algoritmos de *Machine Learning* utilizados en medicina para clasificar pacientes y ayudar en el diagnóstico de enfermedades [36]. Cada modelo tiene características diferentes y puede funcionar mejor o peor dependiendo del tipo de datos utilizados.

La regresión logística es uno de los modelos más sencillos y utilizados en medicina [41,42]. Sirve para calcular la probabilidad de que un paciente pertenezca a un grupo concreto, por ejemplo, tener o no una enfermedad. Su principal ventaja es que es fácil de interpretar y suele funcionar bien incluso cuando se dispone de pocos datos [42]. Por esta razón, se utiliza frecuentemente como modelo de referencia en muchos estudios biomédicos.

Las máquinas de soporte vectorial (*SVM*) son especialmente útiles para problemas donde hay pocas muestras pero muchas variables diferentes [38]. Estos modelos intentan encontrar la mejor separación posible entre dos grupos de datos. Además, pueden detectar relaciones complejas y no lineales entre las variables, lo que los hace muy útiles en problemas médicos complicados.

Random Forest es un modelo basado en múltiples árboles de decisión [37]. En lugar de utilizar un único árbol, combina muchos árboles diferentes y obtiene la respuesta final mediante votación. Esto hace que el modelo sea más estable y preciso. Además, una de sus ventajas más importantes es que permite identificar qué variables son las más relevantes para el diagnóstico, algo muy útil en medicina [37,40].

Gradient Boosting también utiliza varios árboles de decisión, pero en este caso cada nuevo árbol intenta corregir los errores del anterior [40,43]. Gracias a esto, puede detectar relaciones complejas entre las variables y suele ofrecer resultados muy buenos. Sin embargo, es un modelo más complejo y puede ajustarse demasiado a los datos si no se configura correctamente, especialmente cuando hay pocas muestras disponibles [44].

K-Nearest Neighbors (KNN) es un algoritmo sencillo que clasifica una muestra según las muestras más parecidas que tiene cerca [43]. Por ejemplo, un paciente puede clasificarse según los pacientes más similares en el conjunto de datos. Sin embargo, su rendimiento puede empeorar cuando existen muchas variables diferentes.

Los árboles de decisión funcionan creando reglas basadas en preguntas sobre los datos [43]. Por ejemplo, el modelo puede ir tomando decisiones según valores concretos de las variables hasta llegar a una clasificación final. Son modelos fáciles de interpretar y visualizar, pero pueden volverse demasiado específicos si no se controlan adecuadamente.

Naive Bayes gaussiano es un modelo probabilístico basado en el teorema de Bayes [36,43]. Aunque hace algunas simplificaciones matemáticas que no siempre se cumplen en la realidad, suele funcionar bastante bien en muchos problemas médicos. Además, tiene la ventaja de ser rápido y eficiente.

En muchos estudios médicos se comparan varios algoritmos diferentes para elegir cuál ofrece mejores resultados diagnósticos [36,43]. Esto es importante porque el rendimiento de cada modelo depende mucho del tipo de datos disponibles, del número de pacientes y de la relación entre las distintas variables analizadas.

2.10 LIMITACIONES ACTUALES DEL ESTADO DEL ARTE

A pesar de todos los avances realizados, todavía existen varias limitaciones importantes que dificultan el desarrollo y la aplicación clínica de sistemas basados en biomarcadores ecográficos e Inteligencia Artificial para el diagnóstico de la esclerosis lateral amiotrófica [16,17].

La principal limitación es el pequeño número de pacientes incluido en muchos estudios. Como la ELA es una enfermedad poco frecuente, resulta difícil conseguir bases de datos amplias y variadas que representen correctamente todos los tipos de pacientes [12,46]. La mayoría de los estudios cuentan con pocos participantes, lo que puede reducir la fiabilidad de los resultados y hacer que los modelos de *Machine Learning* aprendan demasiado los datos concretos del estudio en lugar de generalizar correctamente. Además, cuando hay pocos datos, es más complicado dividir la información en grupos independientes para entrenar y evaluar los modelos [44].

Otro problema importante es que muchos estudios utilizan metodologías diferentes. Por ejemplo, pueden analizar músculos distintos o utilizar configuraciones diferentes en los ecógrafos [16]. También pueden variar aspectos como la posición del paciente, la zona analizada o el programa utilizado para procesar las imágenes. Todo esto dificulta comparar resultados entre investigadores y puede generar diferencias importantes entre estudios.

Además, muchos modelos todavía no han sido probados en hospitales o grupos de pacientes diferentes [16,17]. En muchos casos, los algoritmos se entrenan y se evalúan utilizando la misma base de datos. Esto puede hacer que el modelo funcione muy bien en ese conjunto

concreto de datos, pero peor cuando se utiliza con pacientes reales de otros centros médicos o con otros equipos de ecografía.

También existe el problema del sobreajuste u *overfitting*. Esto ocurre cuando el modelo aprende demasiado bien los datos de entrenamiento, incluyendo errores o características poco importantes, y después no es capaz de funcionar correctamente con nuevos pacientes [44]. Este problema es especialmente frecuente cuando se utilizan modelos complejos y pocas muestras.

Por último, aunque muchos modelos de Inteligencia Artificial consiguen resultados muy buenos, todavía es necesario desarrollar sistemas más fáciles de interpretar por los médicos [48,49]. En medicina, no es suficiente con que el algoritmo dé una respuesta correcta, también es importante entender por qué toma esa decisión y qué biomarcadores están influyendo más en el resultado. Por esta razón, técnicas como *SHAP* o *LIME*, que ayudan a explicar el funcionamiento de los modelos, están despertando un interés cada vez mayor en el estudio de la ELA [48,49].

2.11 MOTIVACIÓN DEL PRESENTE TRABAJO

Este Trabajo Fin de Grado surge con el objetivo de ayudar al desarrollo de herramientas basadas en ecografía muscular e Inteligencia Artificial para el estudio de la esclerosis lateral amiotrófica, intentando mejorar algunas limitaciones existentes en investigaciones anteriores [16,17].

En primer lugar, el trabajo compara varios algoritmos de *Machine Learning* utilizando los mismos datos y las mismas condiciones de evaluación [36,43]. Para ello, se utiliza una técnica de validación cruzada que permite obtener resultados más fiables sobre el rendimiento real de cada modelo [44]. Gracias a esta comparación, es posible identificar qué algoritmos funcionan mejor con este tipo de datos médicos.

En segundo lugar, el trabajo estudia distintos tipos de biomarcadores ecográficos, analizando tanto los parámetros clásicos como los descriptores texturales obtenidos mediante técnicas

como *GLCM* [16,28]. Además, también se estudia la combinación de todas las variables disponibles. Esto permite conocer qué biomarcadores son más útiles para diferenciar pacientes con ELA de personas sanas.

En tercer lugar, se realiza un análisis específico del músculo cuádriceps femoral, ya que en estudios anteriores ha mostrado que este músculo puede ser especialmente útil para el diagnóstico ecográfico de la ELA [16]. Este análisis permite comparar los resultados obtenidos mediante *Machine Learning* con los resultados de investigaciones previas realizadas con métodos estadísticos tradicionales.

Por último, el trabajo incluye un análisis inicial orientado a estudiar la progresión de la enfermedad a lo largo del tiempo [17]. Aunque esta parte todavía es exploratoria y los resultados son preliminares, puede servir como base para futuras investigaciones con un seguimiento más largo de los pacientes.

La combinación de ingeniería biomédica, procesamiento de imágenes e Inteligencia Artificial representa actualmente una de las áreas más innovadoras en la investigación de enfermedades neurodegenerativas [39,46]. Estas herramientas podrían ayudar en el futuro a desarrollar sistemas de diagnóstico más rápidos, precisos, accesibles y personalizados para la ELA [7,16].

Capítulo 3. METODOLOGÍA

3.1 INTRODUCCIÓN

En este capítulo se explica de forma detallada la metodología utilizada en este Trabajo Fin de Grado, centrado en el estudio de biomarcadores ecográficos y en el uso de técnicas de *Machine Learning* para clasificar pacientes con esclerosis lateral amiotrófica.

La principal motivación de este trabajo es la necesidad de desarrollar herramientas objetivas, fiables y no invasivas que ayuden tanto en el diagnóstico temprano como en el seguimiento de la enfermedad [7,16]. Como se ha explicado anteriormente, la ELA es una enfermedad muy compleja y diferente en cada paciente, lo que hace que su diagnóstico en las primeras fases sea especialmente difícil [8]. Además, actualmente no existe una única prueba capaz de confirmar la enfermedad de forma definitiva, por lo que investigar nuevas herramientas diagnósticas es una prioridad importante en neurología [7].

Esta necesidad no es solo un problema de investigación, sino también un problema clínico real. Cada mes de retraso en el diagnóstico puede afectar al bienestar del paciente, retrasar el inicio de tratamientos y dificultar la planificación de los cuidados necesarios [2,5]. Por ello, cualquier herramienta que ayude a reducir ese retraso puede tener un gran impacto en la atención médica de los pacientes. En este trabajo, se propone la combinación de ecografía muscular e Inteligencia Artificial como una posible solución para este problema [16,17].

La ecografía muscular ha despertado mucho interés en los últimos años porque es una técnica no invasiva, accesible, relativamente económica y fácil de repetir durante el seguimiento clínico [18,22]. Además, el análisis computacional de las imágenes ecográficas permite obtener información cuantitativa que muchas veces no puede detectarse únicamente observando la imagen de forma visual [16,24]. Esta combinación entre una técnica sencilla y herramientas avanzadas de análisis hace que este enfoque sea una alternativa interesante frente a otras técnicas más complejas o costosas.

El principal objetivo de esta metodología es estudiar si determinados biomarcadores obtenidos mediante ecografía muscular, junto con técnicas de aprendizaje automático, permiten diferenciar pacientes con ELA de personas sanas [16,36]. Para ello, se ha desarrollado un proceso de trabajo dividido en varias fases: obtención y organización de los datos, preprocesamiento de las imágenes, selección de variables, entrenamiento de modelos de *Machine Learning*, validación estadística y análisis de los resultados obtenidos.

La metodología utilizada busca que los experimentos puedan repetirse fácilmente y que otros investigadores puedan obtener los mismos resultados utilizando los mismos datos [36]. Además, se ha prestado especial atención a la fiabilidad estadística de los modelos, utilizando técnicas de validación adecuadas al tamaño de la base de datos disponible [44]. También se ha considerado muy importante que los resultados sean fáciles de interpretar desde el punto de vista clínico, ya que un sistema de ayuda al diagnóstico solo es útil si los médicos pueden entender cómo funciona y confiar en sus resultados [48,49]. Por último, se ha intentado evitar el sobreajuste de los modelos para que puedan funcionar correctamente con nuevos pacientes y no solo con los datos utilizados durante el entrenamiento [44].

Además, uno de los aspectos más importantes de este trabajo ha sido comparar distintos algoritmos de clasificación para identificar cuál ofrece mejores resultados diagnósticos probando distintos conjuntos de variables [36,43]. Esta comparación permite obtener conclusiones más fiables que estudiar un único modelo y es especialmente importante en investigaciones biomédicas donde el número de pacientes disponibles suele ser reducido [46].

3.2 DISEÑO DEL ESTUDIO

Este trabajo es un estudio basado en datos clínicos y biomarcadores ecográficos obtenidos de pacientes con esclerosis lateral amiotrófica y personas sanas utilizadas como grupo de control. Al ser un estudio retrospectivo, los datos ya habían sido recogidos anteriormente en investigaciones realizadas en el Centro Universitario San Rafael [16,17], y después se utilizaron para aplicar técnicas de *Machine Learning*.

La metodología utilizada en este trabajo se basa en aplicar técnicas de Inteligencia Artificial y *Machine Learning* a biomarcadores obtenidos mediante ecografía muscular de pacientes con esclerosis lateral amiotrófica y personas sanas [36,39]. Este enfoque supone una evolución respecto a estudios anteriores, que se centraban principalmente en analizar los parámetros ecográficos utilizando métodos estadísticos tradicionales [16,17].

Mientras que muchos trabajos anteriores estudiaban cada variable por separado, en este trabajo se utilizan algoritmos de *Machine Learning* capaces de analizar muchas variables al mismo tiempo y encontrar patrones complejos dentro de los datos [36,43].

El principal objetivo es desarrollar modelos que sean capaces de diferenciar automáticamente entre pacientes con ELA y personas sanas utilizando información obtenida de las ecografías musculares. Para ello, durante el entrenamiento, los algoritmos reciben tanto los datos de entrada como el diagnóstico real de cada persona, aprendiendo así las características asociadas a cada grupo [36].

Se eligió este tipo de enfoque porque todos los datos ya tienen una clasificación conocida y fiable. Esto permite comprobar directamente si el modelo es capaz de distinguir correctamente entre pacientes y controles sanos [44,45].

El estudio se desarrolló siguiendo varias fases. Primero se recopilaron y organizaron los datos para asegurarse de que estuvieran correctamente preparados. Después se realizó un análisis inicial de las variables mediante estadísticas y gráficos para comprender mejor la información y detectar posibles errores o valores extraños.

A continuación, se aplicaron técnicas de preprocesamiento y normalización para preparar los datos antes de entrenar los modelos. También se crearon diferentes grupos de variables para estudiar cómo influían distintos tipos de biomarcadores en los resultados. Después, se entrenaron varios modelos de *Machine Learning* y se evaluó su rendimiento utilizando validación cruzada [44]. Finalmente, se compararon los resultados obtenidos y se analizaron las variables más importantes [48].

Todo este proceso permite seguir una metodología ordenada y reproducible, algo muy importante en estudios biomédicos con Inteligencia Artificial [36]. Además, ayuda a obtener resultados más fiables y a reducir el riesgo de que los modelos funcionen bien solo con los datos del estudio, pero no con nuevos pacientes [44].

3.3 DESCRIPCIÓN DEL DATASET

El *dataset* utilizado en este trabajo está formado por un total de 104 muestras, obtenidas a partir de 52 sujetos: 26 pacientes diagnosticados con esclerosis lateral amiotrófica y 26 personas sanas utilizadas como grupo de control. Dado que las mediciones ecográficas se realizaron de forma bilateral, cada sujeto contribuye con dos registros al *dataset*, correspondientes al lado derecho e izquierdo de cada músculo, lo que explica que el total de muestras sea de 104. Tener el mismo número de pacientes y controles es algo positivo, ya que evita problemas relacionados con el desbalanceo de clases, algo frecuente en estudios médicos de enfermedades raras [12,46].

De los 26 pacientes con ELA, 18 eran hombres y 8 mujeres, con una edad media de $58,6 \pm 11,2$ años. El grupo control estaba formado por 9 hombres y 17 mujeres, con una edad media de $59,6 \pm 6,5$ años. Esta distribución por sexo es coherente con lo descrito anteriormente, donde la ELA es ligeramente más frecuente en hombres que en mujeres [12].

Los datos proceden de ecografías musculares realizadas en el Centro Universitario San Rafael de la Universidad Pontificia Comillas, dentro de estudios previos realizados por Martínez-Payá y sus colaboradores [16,17]. Antes de utilizar los datos en este trabajo, toda la información fue anonimizada para proteger la identidad de los participantes y cumplir la normativa de protección de datos.

Uno de los aspectos más importantes del *dataset* es que combina información clínica con biomarcadores obtenidos mediante análisis computacional de imágenes ecográficas [16]. Esto permite estudiar distintos tipos de información al mismo tiempo y analizar cuáles son más útiles para diferenciar pacientes con ELA de personas sanas.

Las variables del *dataset* pueden dividirse en tres grupos principales.

3.3.1 VARIABLES CLÍNICAS

Las variables clínicas incluyen datos relacionados con el estado general del paciente. Entre las más importantes se encuentran la edad, el sexo, el índice de masa corporal (*BMI*), la escala funcional ALSFRS, medidas de fuerza mediante la escala MRC y la presencia de fasciculaciones musculares observadas durante la exploración [16,17].

Estas variables aportan información sobre el estado clínico y el grado de evolución de la enfermedad en el momento de la prueba. Además, permiten estudiar si combinar datos clínicos con biomarcadores ecográficos mejora la capacidad diagnóstica del sistema [46].

La escala ALSFRS (*Amyotrophic Lateral Sclerosis Functional Rating Scale*) es una de las herramientas más utilizadas para evaluar a pacientes con ELA [10]. Esta escala analiza aspectos relacionados con el movimiento, el habla, la respiración, la deglución y las actividades diarias. La puntuación máxima es 48 puntos, que corresponde a una funcionalidad normal, mientras que la mínima es 0. Esta escala es útil tanto para describir el estado del paciente como para estudiar la progresión de la enfermedad con el tiempo [10,17].

La escala MRC se utiliza para medir la fuerza muscular [16,17]. Sus valores van de 0 a 5, donde 0 significa ausencia total de fuerza y 5 corresponde a fuerza normal. Esta escala aporta información complementaria sobre el estado funcional de los músculos.

3.3.2 BIOMARCADORES ECOGRÁFICOS CLÁSICOS

El segundo grupo de variables corresponde a biomarcadores obtenidos directamente de las ecografías musculares. Estos parámetros describen características del músculo relacionadas con su estructura y con el grado de afectación de la enfermedad [16,18].

Entre las variables más importantes se encuentran el grosor muscular (*Muscle Thickness*), la ecogenicidad (*Echointensity*) y la ecotextura (*Echotexture*).

- El grosor muscular mide el tamaño del músculo en la ecografía y suele disminuir en pacientes con ELA debido a la pérdida de masa muscular [18,20].
- La ecogenicidad mide el brillo medio del músculo en la imagen y suele aumentar cuando aparece grasa o fibrosis en el tejido muscular [27].
- La ecotextura analiza cómo varía la intensidad dentro del músculo y aporta información sobre la uniformidad o irregularidad del tejido [24].

En pacientes con ELA, la degeneración de las motoneuronas produce cambios en el músculo como pérdida de fibras musculares, acumulación de grasa y fibrosis [17,27]. Todos estos cambios modifican la apariencia del músculo en la ecografía y se reflejan en estos biomarcadores. Además, estos parámetros son relativamente fáciles de obtener utilizando equipos modernos de ecografía y programas de análisis de imagen [16].

3.3.3 DESCRIPTORES TEXTURALES GLCM

El tercer grupo de variables corresponde a descriptores texturales obtenidos mediante análisis avanzado de las imágenes ecográficas. En este trabajo se han utilizado parámetros derivados de la matriz de co-ocurrencia de niveles de gris (*GLCM*), una de las técnicas más utilizadas en análisis de textura en imagen médica [28].

La *GLCM* permite estudiar cómo se relacionan los distintos tonos grises dentro de la imagen y analizar la organización interna del músculo [28,35]. A diferencia de los biomarcadores clásicos, estos descriptores permiten obtener información más detallada sobre la estructura del tejido muscular [16,32].

Entre los parámetros utilizados destacan: [16,28,35]

- ASM o Energía: mide el grado de uniformidad de la imagen.
- CON o Contraste: analiza las diferencias de intensidad entre píxeles cercanos.
- COR o Correlación: estudia la relación entre los niveles de gris de distintos píxeles.
- IDM u Homogeneidad: mide si la imagen es uniforme y organizada.
- ENT o Entropía: refleja el nivel de desorden o complejidad de la textura.

En músculos sanos suele existir una estructura más organizada, con valores altos de homogeneidad y energía y valores bajos de contraste y entropía. En cambio, en músculos afectados por ELA, la estructura es más irregular debido a la degeneración muscular, la grasa y la fibrosis, produciendo el efecto contrario [16,17].

El uso de estos biomarcadores texturales es una de las partes más innovadoras de este trabajo, ya que permiten convertir imágenes ecográficas en datos numéricos que pueden analizarse mediante Inteligencia Artificial para detectar patrones difíciles de observar visualmente [16,28].

3.4 MÚSCULOS ANALIZADOS

Las medidas ecográficas de este trabajo se realizaron en cuatro grupos musculares diferentes: el bíceps braquial, los flexores del antebrazo, el cuádriceps femoral y el tibial anterior [16,17]. Estos músculos fueron seleccionados porque tienen una gran importancia en la esclerosis lateral amiotrófica y además son adecuados para realizar ecografías de buena calidad.

La elección de estos músculos no fue aleatoria. Para seleccionarlos se tuvieron en cuenta varios aspectos importantes. Primero, se eligieron músculos fáciles de explorar mediante ecografía y que permitieran obtener imágenes claras y reproducibles [16]. También se seleccionaron músculos que suelen verse afectados con frecuencia en pacientes con ELA. Además, se buscó que las medidas pudieran compararse fácilmente entre distintos pacientes y diferentes exploraciones, algo muy importante para estudiar la evolución de la enfermedad con el tiempo [17].

El bíceps braquial es uno de los músculos más estudiados en ecografía muscular porque es grande, superficial y fácil de explorar [16,18]. En la ELA puede verse afectado cuando la enfermedad comienza en los brazos. Su estudio permite analizar el grado de atrofia y degeneración muscular en el miembro superior.

Los flexores del antebrazo son músculos muy importantes para realizar actividades diarias como agarrar objetos o mover las manos. También suelen verse afectados en pacientes cuya enfermedad comienza en los miembros superiores [16]. Su análisis complementa la información obtenida con el bíceps braquial y permite estudiar músculos más alejados del hombro.

El cuádriceps femoral es el músculo más grande del cuerpo y uno de los más utilizados en estudios de ecografía muscular [16,17]. Su gran tamaño facilita obtener imágenes de alta calidad. En pacientes con ELA puede verse afectado tanto en las formas que comienzan en las piernas como en fases más avanzadas de la enfermedad. Estudios anteriores obtuvieron especialmente buenos resultados utilizando este músculo [16], por lo que en este trabajo se realizó un análisis específico del cuádriceps.

El tibial anterior es un músculo situado en la parte frontal de la pierna y suele verse afectado en pacientes con debilidad en el pie o dificultad para levantarlo al caminar [16,17]. Su posición superficial hace que sea fácil de estudiar mediante ecografía y permite obtener imágenes de buena calidad.

Analizar varios músculos diferentes es una de las ventajas más importantes de este trabajo, ya que permite estudiar si algunos músculos son más útiles que otros para detectar la enfermedad [16]. Además, combinar información de distintos músculos puede mejorar la capacidad diagnóstica del sistema y ofrecer una visión más completa de cómo afecta la ELA al cuerpo.

3.5 VARIABLE OBJETIVO

La variable objetivo utilizada en todos los modelos de clasificación de este trabajo es la variable llamada “Group”. Esta variable indica si cada muestra pertenece a un paciente con esclerosis lateral amiotrófica o a una persona sana [16].

Para ello, se asignó el valor 1 a los pacientes con ELA y el valor 0 a los sujetos de control sanos. Este tipo de problema se conoce como clasificación binaria supervisada, ya que el

modelo debe aprender a diferenciar entre dos grupos utilizando ejemplos previamente clasificados [36,41].

La utilización de los valores 0 y 1 sigue la forma habitual de trabajar en estudios médicos. El valor 1 representa la presencia de la enfermedad y el valor 0 indica que la persona está sana. Esto es importante para interpretar métricas como la sensibilidad, que mide la capacidad del modelo para detectar pacientes con ELA, y la especificidad, que mide su capacidad para identificar correctamente a las personas sanas [45].

Todos los modelos desarrollados en este trabajo intentan aprender patrones presentes en los biomarcadores ecográficos para poder predecir correctamente el valor de esta variable en nuevos pacientes que no hayan sido utilizados durante el entrenamiento del modelo [44]. La calidad de estas predicciones se evalúa mediante métricas que se describen más adelante en la metodología.

3.6 SELECCIÓN DE VARIABLES

Uno de los aspectos más importantes al desarrollar modelos predictivos es elegir correctamente las variables que se van a utilizar [36,43]. En este trabajo se disponía de muchas variables diferentes, incluyendo biomarcadores ecográficos de cuatro músculos, variables clínicas y descriptores texturales. Como el número de variables era bastante grande en comparación con el número total de muestras, fue necesario seleccionar cuidadosamente qué información utilizar en cada modelo [44].

Con el objetivo de estudiar qué biomarcadores eran más útiles para el diagnóstico, se crearon distintos grupos de variables y se evaluó el rendimiento de los modelos utilizando cada grupo por separado [16,36]. Esto permitió analizar qué tipo de información aportaba mejores resultados y comprobar si combinar varias fuentes de datos mejoraba la capacidad diagnóstica del sistema.

Los grupos de variables analizados fueron los siguientes:

- Primero, se utilizaron únicamente los biomarcadores ecográficos clásicos, como el grosor muscular, la ecogenicidad y la ecotextura [16,18].
- Después, se analizaron solo los descriptores texturales *GLCM* de los cuatro músculos estudiados [28].
- Más adelante, se combinaron todas las variables ecográficas y texturales disponibles.
- Finalmente, en el último experimento, también se añadieron variables clínicas como la edad, el sexo, el índice de masa corporal (*BMI*) y las puntuaciones de las escalas funcionales [10,46].

Además, al inicio del estudio se realizó un análisis de correlación entre cada variable y la variable objetivo para identificar cuáles estaban más relacionadas con el diagnóstico de esclerosis lateral amiotrófica [36]. Este análisis ayudó a detectar algunas variables con correlaciones extremadamente altas, lo que en algunos casos indicaba posible redundancia de información o incluso riesgo de fuga de información (*data leakage*), un problema que se explica más adelante en la sección 3.13 dedicada a la prevención del sobreajuste.

3.7 PREPROCESAMIENTO DE LOS DATOS

Antes de entrenar los modelos de *Machine Learning*, fue necesario preparar los datos para asegurarse de que estaban en las condiciones adecuadas para el análisis [36,43]. Esta fase de preprocesamiento es muy importante en cualquier estudio de Inteligencia Artificial, ya que influye directamente en la calidad y fiabilidad de los resultados.

En los datos biomédicos, el preprocesamiento es especialmente importante porque las variables pueden tener escalas muy diferentes [36]. Por ejemplo, el grosor muscular se mide en milímetros, la ecogenicidad utiliza valores de intensidad de gris y los descriptores *GLCM* tiene rangos numéricos distintos. Estas diferencias pueden afectar negativamente al rendimiento de algunos modelos de *Machine Learning*, como las *SVM*, el *KNN* o la regresión logística [38,41,42].

La principal técnica utilizada fue la normalización mediante la herramienta *StandardScaler* de la librería *scikit-learn* [36]. Esta técnica transforma todas las variables para que tengan

una escala similar, con media cero y desviación estándar igual a uno. De esta forma, ninguna variable tiene más importancia que otra solo por tener valores numéricos altos, y todos los biomarcadores contribuyen de manera equilibrada al entrenamiento de los modelos.

Un aspecto muy importante es que esta normalización se realizó dentro del proceso de validación cruzada [44]. Esto significa que los parámetros necesarios para normalizar los datos se calculaban únicamente utilizando los datos de entrenamiento de cada partición y después se aplicaban a los datos de validación. Este procedimiento evita que el modelo utilice información que no debería conocer durante el entrenamiento y ayuda a obtener resultados más fiables.

Además de la normalización, también se utilizó el parámetro *class_weight="balanced"* en algunos algoritmos, como la regresión logística, las *SVM* y el *Random Forest* [36,37,38]. Este parámetro ajusta automáticamente la importancia de cada grupo durante el entrenamiento. Aunque en este estudio había el mismo número de pacientes y controles, esta estrategia ayuda a reducir posibles pequeños sesgos y mejora la estabilidad de los modelos.

Por último, se realizó una revisión inicial de los datos para detectar posibles errores, valores ausentes o valores extremadamente diferentes al resto. Se comprobó que las variables principales utilizadas en el análisis no tenían valores vacíos. En algunos análisis secundarios donde sí aparecieron datos faltantes, como en el análisis de progresión de la enfermedad, se aplicaron métodos específicos para tratarlos, que se explican más adelante en el capítulo de resultados [43].

3.8 **MODELOS DE MACHINE LEARNING UTILIZADOS**

Uno de los principales objetivos de este trabajo fue comparar diferentes algoritmos de *Machine Learning* para estudiar cuál ofrecía mejores resultados al clasificar pacientes con esclerosis lateral amiotrófica utilizando biomarcadores ecográficos [36,43]. Para ello, se seleccionaron modelos de distintos tipos, desde modelos sencillos y lineales hasta algoritmos más complejos capaces de detectar relaciones no lineales entre las variables.

Comparar varios modelos es importante porque permite entender mejor cómo se relacionan los biomarcadores con el diagnóstico [36]. Si los modelos lineales consiguen buenos resultados, significa que las diferencias entre pacientes y controles pueden separarse de manera relativamente simple. En cambio, si los modelos no lineales funcionan claramente mejor, eso indica que existen relaciones más complejas entre las variables y que se necesitan métodos más avanzados para detectarlas [43].

3.8.1 REGRESIÓN LOGÍSTICA

La regresión logística es uno de los modelos más clásicos y utilizados en medicina y análisis biomédico [41,42]. Se trata de un modelo lineal que calcula la probabilidad de que una muestra pertenezca a una determinada clase, en este caso ser paciente con ELA o estar sano.

El modelo funciona combinando todas las variables de entrada mediante una ecuación lineal y aplicando después una función llamada sigmoide. Esta función transforma cualquier valor numérico en un número entre 0 y 1, que puede interpretarse como una probabilidad [41]. Por ejemplo, un resultado cercano a 1 indicaría alta probabilidad de que la muestra pertenezca al grupo ELA.

En este trabajo, la regresión logística se implementó utilizando el método “liblinear”, adecuado para *datasets* pequeños o medianos [36,42]. Además, se utilizó regularización L2 para reducir el riesgo de sobreajuste y se permitió un máximo de 1000 iteraciones para asegurar que el algoritmo encontrara una solución estable.

Aunque es un modelo relativamente sencillo, tiene varias ventajas importantes en medicina. Una de las principales es su interpretabilidad, ya que permite analizar fácilmente cuánto influye cada variable en la clasificación mediante los coeficientes del modelo [41,42]. Esto ayuda a identificar qué biomarcadores tienen mayor importancia clínica. Además, suele funcionar de forma estable con *datasets* pequeños y presenta bajo riesgo de sobreajuste [44].

3.8.2 MÁQUINAS DE SOPORTE VECTORIAL (SVM)

Las máquinas de soporte vectorial o *SVM* (*Support Vector Machines*) son modelos muy utilizados en bioinformática y estudios biomédicos, especialmente cuando el número de pacientes es pequeño pero existe un gran número de variables [38].

El objetivo principal del *SVM* es encontrar la mejor separación posible entre las dos clases del problema, es decir, entre pacientes con ELA y sujetos sanos [38]. Para ello, el algoritmo busca un hiperplano de separación que deje la máxima distancia posible entre ambos grupos. Las muestras más cercanas a esa frontera se denominan vectores soporte y son las más importantes para construir el modelo.

En este trabajo se utilizó un kernel RBF (*Radial Basis Function*), que permite que el modelo aprenda relaciones no lineales entre las variables [38]. Este tipo de kernel transforma internamente los datos a un espacio de mayor dimensión donde las clases pueden separarse más fácilmente. Los parámetros utilizados fueron los valores por defecto de *scikit-learn* ($C=1.0$ y $\text{gamma}=\text{"scale"}$) [36], ya que suelen ofrecer un buen equilibrio entre capacidad de ajuste y capacidad de generalización.

Una de las grandes ventajas del *SVM* es que suele funcionar muy bien con *datasets* relativamente pequeños, algo especialmente importante en enfermedades poco frecuentes como la ELA [38]. Sin embargo, una de sus limitaciones es que resulta más difícil de interpretar que modelos lineales sencillos, ya que la frontera de decisión puede ser bastante compleja.

3.8.3 RANDOM FOREST

El *Random Forest* es un algoritmo basado en la combinación de múltiples árboles de decisión [37]. En lugar de construir un único árbol, el modelo crea muchos árboles independientes utilizando diferentes partes de los datos y diferentes variables. Después, todos los árboles votan y la clasificación final se obtiene mediante mayoría de votos.

Cada árbol se entrena utilizando una selección aleatoria de muestras del *dataset* y un subconjunto aleatorio de variables en cada división [37]. Esta combinación de aleatoriedad hace que los árboles sean diferentes entre sí y evita que todos aprendan exactamente los mismos patrones. Gracias a ello, el modelo completo es más robusto y tiene menor riesgo de sobreajuste que un único árbol de decisión [40].

En este trabajo se utilizaron 100 árboles ($n_estimators=100$) y el parámetro *class_weight*="balanced" para equilibrar el peso de ambas clases durante el entrenamiento [36,37].

Además de ofrecer buenos resultados predictivos, el *Random Forest* tiene una ventaja muy importante: permite calcular la importancia relativa de cada variable [37]. Esto hace posible identificar qué biomarcadores ecográficos son más relevantes para diferenciar pacientes con ELA de sujetos sanos.

3.8.4 K-NEAREST NEIGHBORS (KNN)

El algoritmo *KNN* es uno de los modelos más sencillos desde el punto de vista conceptual [43]. Su funcionamiento se basa en comparar cada nueva muestra con las muestras más parecidas del *dataset* de entrenamiento.

Cuando el modelo recibe una nueva muestra, calcula cuáles son sus vecinos más cercanos utilizando normalmente la distancia euclidiana. Después, asigna la clase más frecuente entre esos vecinos [43]. En este trabajo se utilizó un valor de $k=5$, es decir, el modelo analizaba los cinco vecinos más próximos antes de decidir la clasificación final.

A diferencia de otros modelos, el *KNN* no construye realmente un modelo matemático durante el entrenamiento. Simplemente almacena los datos y realiza las comparaciones directamente en el momento de la clasificación [43].

Por este motivo, la normalización de las variables es especialmente importante en este algoritmo. Si algunas variables tienen valores mucho mayores que otras, podrían dominar el cálculo de distancias y afectar negativamente al rendimiento del modelo [36].

3.8.5 ÁRBOLES DE DECISIÓN

Los árboles de decisión son modelos que clasifican las muestras mediante una serie de reglas organizadas de forma jerárquica [43]. El algoritmo va realizando preguntas sobre las variables y dividiendo los datos en diferentes grupos hasta llegar a una decisión final.

En cada paso, el modelo selecciona la variable y el valor que mejor separan a pacientes y controles según el criterio de impureza de Gini [43]. Este proceso continúa hasta que se alcanza un límite de profundidad o un número mínimo de muestras en cada nodo.

La principal ventaja de los árboles de decisión es su gran interpretabilidad. El resultado puede representarse gráficamente y permite entender fácilmente qué variables utiliza el modelo y qué reglas sigue para clasificar cada muestra [43].

Sin embargo, tienen una limitación importante: si el árbol crece demasiado, puede aprender demasiado bien los datos de entrenamiento y perder capacidad para generalizar correctamente a nuevos pacientes [44]. Esto se conoce como sobreajuste.

3.8.6 GRADIENT BOOSTING

El *Gradient Boosting* es otro algoritmo basado en árboles de decisión, pero funciona de forma diferente al *Random Forest* [40]. En lugar de crear todos los árboles de manera independiente, los construye de forma secuencial.

Cada nuevo árbol intenta corregir los errores cometidos por los árboles anteriores [40]. De esta manera, el modelo va mejorando progresivamente sus predicciones en cada iteración. Este proceso permite detectar relaciones complejas y no lineales entre las variables con gran eficacia.

En este trabajo se utilizó la implementación de *scikit-learn* con 100 árboles y una tasa de aprendizaje de 0.1 [36].

El *Gradient Boosting* suele ofrecer buenos resultados en problemas de clasificación, aunque también puede presentar mayor riesgo de sobreajuste si sus parámetros no se ajustan correctamente, especialmente cuando el *dataset* es pequeño [40,44].

3.8.7 NAIVE BAYES

El clasificador *Naive Bayes* gaussiano es un modelo probabilístico basado en el teorema de Bayes [36,43]. El algoritmo calcula la probabilidad de que una muestra pertenezca a cada clase utilizando las variables disponibles.

El modelo asume que todas las variables son independientes entre sí y que siguen una distribución normal o gaussiana dentro de cada clase [43]. Aunque esta suposición no suele cumplirse completamente en datos reales, el modelo suele funcionar bien en muchos problemas prácticos.

La clasificación final se realiza asignando cada muestra a la clase con mayor probabilidad calculada por el modelo.

Las principales ventajas de *Naive Bayes* son su sencillez, la rapidez de entrenamiento y su buen comportamiento con *datasets* pequeños [36,43]. Además, al ser un modelo simple, tiene menos riesgo de sobreajuste que otros algoritmos más complejos [44].

3.9 PIPELINE DE ENTRENAMIENTO

Para garantizar que los experimentos fueran reproducibles y evitar problemas como la fuga de información (*data leakage*), todos los modelos se entrenaron utilizando *pipelines* implementados con la librería *scikit-learn* [36].

Un *pipeline* permite organizar de forma automática y ordenada todas las etapas necesarias para entrenar un modelo [36,43]. De esta manera, cada paso se realiza siempre en el orden correcto y utilizando únicamente la información adecuada en cada fase del proyecto.

En este trabajo, cada *pipeline* estaba formado por dos etapas principales. Primero, se realizaba la normalización de las variables mediante *StandardScaler*. Después, se entrenaba el modelo de clasificación correspondiente [36].

Esta estructura es especialmente importante durante la validación cruzada, ya que asegura que los parámetros utilizados para normalizar los datos se calculan solo con las muestras de entrenamiento de cada partición y posteriormente se aplican a los datos de validación [44]. Así se evita que el modelo tenga acceso a información que no debería conocer durante el entrenamiento, lo que ayuda a obtener resultados más fiables y realistas.

Además, el uso de *pipelines* facilita mucho la organización y reproducibilidad del código, ya que todo el proceso de preprocesamiento y entrenamiento queda integrado en un único objeto [36]. Esto permite reutilizar fácilmente los modelos, compartir código y repetir los experimentos en las mismas condiciones.

En estudios biomédicos, esta reproducibilidad es especialmente importante, ya que permite que otros investigadores puedan comprobar y validar los resultados obtenidos.

3.10 VALIDACIÓN CRUZADA

Para evaluar de forma fiable el rendimiento de los modelos, se utilizó una técnica llamada validación cruzada estratificada de 10 particiones (*Stratified 10-Fold Cross Validation*) [44]. Esta técnica se implementó utilizando la clase *StratifiedKFold* de la librería *scikit-learn* [36]. Además, se fijó una semilla aleatoria (*random_state=42*) para asegurar que las divisiones de los datos fueran siempre las mismas y los resultados pudieran reproducirse.

La validación cruzada es especialmente útil cuando el número de muestras disponibles es pequeño, como ocurre en este trabajo [44]. En lugar de dividir los datos una sola vez en entrenamiento y prueba, esta técnica divide el *dataset* en 10 partes diferentes o *folds*.

En cada iteración, el modelo se entrena utilizando 9 partes de los datos y se evalúa con la parte restante. Este proceso se repite 10 veces, cambiando cada vez el grupo utilizado para

validación, hasta que todas las muestras han sido utilizadas tanto para entrenamiento como para prueba [44]. Finalmente, el rendimiento total del modelo se calcula haciendo la media de los resultados obtenidos en las 10 particiones.

Además de la media, también se calcula la desviación estándar entre particiones, lo que permite analizar si el modelo es estable o si sus resultados cambian mucho dependiendo de los datos utilizados en cada iteración [44].

La validación cruzada utilizada en este trabajo es estratificada, lo que significa que mantiene aproximadamente la misma proporción de pacientes y controles en todas las particiones [44]. Esto es importante para que cada grupo de validación sea representativo del *dataset* completo.

En este caso, como el *dataset* contiene 52 muestras con esclerosis lateral amiotrófica y 52 muestras sanas, cada partición contiene aproximadamente 5 muestras con ELA y 5 controles en el conjunto de validación.

3.11 MÉTRICAS DE EVALUACIÓN

El rendimiento de los modelos se evaluó utilizando varias métricas diferentes, ya que una sola medida no es suficiente para describir completamente cómo funciona un clasificador en un problema biomédico [45]. Cada métrica analiza un aspecto distinto del comportamiento del modelo, por lo que es importante interpretarlas de manera conjunta.

3.11.1 ACCURACY

La *accuracy* representa el porcentaje total de muestras clasificadas correctamente sobre el total de muestras analizadas [44,45]. Se calcula mediante la fórmula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Donde:

- TP (True Positives): pacientes con esclerosis lateral amiotrófica correctamente identificados.
- TN (True Negatives): sujetos sanos correctamente identificados.
- FP (False Positives): sujetos sanos clasificados incorrectamente como ELA.
- FN (False Negatives): pacientes con ELA clasificados incorrectamente como sanos.

La *accuracy* es la métrica más sencilla e intuitiva, ya que indica el porcentaje total de aciertos del modelo. Sin embargo, puede resultar engañosa cuando las clases están desbalanceadas, por lo que siempre debe complementarse con otras métricas.

3.11.2 PRECISIÓN

La precisión mide cuántas de las muestras clasificadas por el modelo como ELA realmente corresponden a pacientes con la enfermedad [45]. Se calcula como:

$$\text{Precisión} = \frac{TP}{TP + FP}$$

Una precisión alta significa que cuando el modelo predice que un paciente tiene ELA, normalmente está acertando. Es decir, el modelo genera pocos falsos positivos.

Esta métrica es especialmente importante cuando un diagnóstico incorrecto podría tener consecuencias importantes, como el impacto emocional y psicológico que puede producir un falso diagnóstico de ELA.

3.11.3 RECALL O SENSIBILIDAD

El *recall* o sensibilidad mide la capacidad del modelo para detectar correctamente a los pacientes que realmente tienen la enfermedad [45]. Se calcula mediante la fórmula:

$$\text{Sensibilidad} = \frac{TP}{TP + FN}$$

En diagnóstico médico, esta suele ser una de las métricas más importantes. Un falso negativo significa que un paciente enfermo no es detectado por el modelo, lo que puede retrasar el diagnóstico y el inicio del tratamiento.

Por esta razón, en muchos sistemas de ayuda al diagnóstico se intenta priorizar una sensibilidad alta, incluso aunque eso implique aceptar algunos falsos positivos adicionales.

3.11.4 F1-SCORE

El *F1-score* combina la precisión y el *recall* en una sola métrica. Se calcula como la media armónica entre ambas: [45]

$$F1 - Score = 2 \times \frac{Precisión \times Recall}{Precisión + Recall}$$

Esta métrica es útil porque permite evaluar el equilibrio general del modelo entre detectar correctamente los casos positivos y evitar errores en esas detecciones.

El *F1-score* suele ser más informativo que la *accuracy* cuando existe desbalanceo entre clases o cuando tanto los falsos positivos como los falsos negativos tienen consecuencias importantes.

3.11.5 ÁREA BAJO LA CURVA ROC (AUC)

El área bajo la curva *ROC*, conocida como *AUC* (*Area Under the Curve*), es una de las métricas más utilizadas en clasificación biomédica [45].

La curva *ROC* representa gráficamente la relación entre la sensibilidad y la tasa de falsos positivos para todos los posibles umbrales de decisión del modelo. El *AUC* mide el área bajo esa curva y refleja la capacidad general del modelo para diferenciar correctamente entre pacientes y controles.

Un valor de:

- *AUC* = **1.0** indica un clasificador perfecto.
- *AUC* = **0.5** indica un rendimiento equivalente al azar, es decir, sin capacidad real de clasificación.

Cuanto más alto sea el *AUC*, mejor será la capacidad del modelo para distinguir entre pacientes con ELA y sujetos sanos.

3.12 SELECCIÓN DE VARIABLES Y REDUCCIÓN DE DIMENSIONALIDAD

Uno de los principales problemas en estudios biomédicos con pocos pacientes es que muchas veces hay un número muy alto de variables en comparación con el tamaño del *dataset* [44]. En este trabajo se disponía de una gran cantidad de biomarcadores ecográficos obtenidos de cuatro músculos diferentes, incluyendo variables clínicas, biomarcadores clásicos y numerosos descriptores texturales. Todo esto generaba un número elevado de variables para un total de 104 muestras.

Cuando ocurre esta situación, existe un mayor riesgo de sobreajuste [44]. Esto significa que el modelo puede aprender demasiado bien los datos de entrenamiento, pero sin ser capaz de generalizar correctamente a nuevos pacientes. Con suficientes variables, incluso es posible que el modelo memorice los datos en lugar de aprender patrones reales de la enfermedad.

Por esta razón, se realizó un proceso de análisis y selección de variables para identificar cuáles eran realmente las más útiles para diferenciar pacientes con esclerosis lateral amiotrófica y sujetos sanos [36,43].

Durante el análisis inicial de los datos se detectaron algunas variables con correlaciones extremadamente altas con la variable objetivo, muy cercanas a 1.0. Esto podría indicar que existía información repetida o incluso riesgo de *data leakage* [44].

Estas variables se revisaron cuidadosamente y, cuando su comportamiento no parecía biológicamente lógico o podía afectar a la fiabilidad del modelo, fueron eliminadas del análisis. De esta manera, se intentó garantizar que los resultados obtenidos fueran más fiables y metodológicamente correctos [36].

Después de este proceso de limpieza y selección, se trabajó con los distintos grupos de variables descritos anteriormente, utilizando conjuntos de biomarcadores que fueran tanto útiles desde el punto de vista diagnóstico como coherentes desde el punto de vista clínico [16,28].

3.13 PREVENCIÓN DEL SOBREAJUSTE

El sobreajuste (*overfitting*) es uno de los principales problemas en *Machine Learning* biomédico, especialmente cuando se trabaja con *datasets* pequeños [44]. Ocurre cuando el modelo aprende demasiado bien los datos de entrenamiento, incluyendo incluso el ruido o pequeñas variaciones propias de la muestra, y después no es capaz de funcionar correctamente con nuevos pacientes.

En lugar de aprender patrones generales de la enfermedad, un modelo sobreajustado “memoriza” los datos utilizados durante el entrenamiento [44]. Como consecuencia, puede obtener resultados muy buenos en esos datos, pero rendir peor cuando se aplica a muestras nuevas.

Durante los primeros experimentos de este trabajo se obtuvieron valores de *accuracy* extremadamente altos, cercanos al 100% en algunos modelos y conjuntos de variables. Aunque inicialmente estos resultados parecían muy positivos, un análisis más detallado mostró que algunas variables tenían correlaciones artificialmente muy elevadas con la variable objetivo. Esto sugería la existencia de problemas como fuga de información (*data leakage*) o variables redundantes [44].

Después de detectar estas situaciones, se eliminaron las variables problemáticas y se aplicaron medidas más estrictas durante el entrenamiento, como el uso correcto de *pipelines* y validación cruzada [36,44].

Tras estas correcciones, los resultados descendieron a valores más realistas, con *accuracies* entre el 70% y el 90% dependiendo del modelo y de las variables utilizadas. Estos resultados son similares a los publicados en otros estudios sobre clasificación de esclerosis lateral amiotrófica con tamaños de muestra parecidos [16,46], lo que indica que los modelos estaban aprendiendo patrones reales y no simplemente memorizando los datos.

3.14 INTERPRETABILIDAD CLÍNICA DE LOS MODELOS

En aplicaciones médicas de Inteligencia Artificial, no solo es importante que un modelo tenga buenos resultados, sino también que sea capaz de explicar por qué toma una determinada decisión [48,49]. Un sistema que no pueda justificar sus predicciones tiene poca utilidad clínica, ya que los profesionales sanitarios necesitan entender cómo funciona el modelo antes de confiar en él y utilizarlo en la práctica médica.

Por esta razón, además de comparar la precisión de los modelos, también se estudió qué variables tenían más importancia en las predicciones [48]. Este análisis se realizó en aquellos algoritmos que permiten medir directamente la relevancia de cada variable.

En modelos como *Random Forest* y *Gradient Boosting*, la importancia de cada variable se calculó analizando cuánto ayudaba esa variable a mejorar las decisiones de los árboles del modelo [37,40].

En el caso de regresión logística, la importancia de las variables se estudió utilizando los coeficientes del modelo después de normalizar los datos, lo que permitía comparar correctamente el peso de cada biomarcador [41,42].

Este análisis permitió identificar qué biomarcadores ecográficos eran más importantes para diferenciar pacientes con esclerosis lateral amiotrófica y sujetos sanos [16,28]. Además de ser útil desde el punto de vista técnico, esta información también tiene interés clínico, ya que ayuda a comprender mejor qué músculos y qué características ecográficas están más relacionadas con la enfermedad.

Los resultados obtenidos mostraron bastante relación con los hallazgos descritos en estudios anteriores sobre afectación muscular en la ELA [16,17], lo que aporta mayor confianza y solidez al enfoque utilizado en este trabajo.

3.15 CONSIDERACIONES ÉTICAS Y LIMITACIONES METODOLÓGICAS

El uso de Inteligencia Artificial en medicina implica tener en cuenta varios aspectos éticos importantes, tanto durante el desarrollo del estudio como al interpretar los resultados [39,51].

En este trabajo, todos los datos utilizados fueron anonimizados previamente por los investigadores del Centro Universitario San Rafael. Esto significa que la información personal de los participantes fue eliminada para proteger su privacidad y garantizar el cumplimiento de las normas de protección de datos y datos de salud.

Desde el punto de vista metodológico, la principal limitación del estudio es el tamaño reducido del *dataset*. Este trabajo incluye un total de 104 registros, por lo que los resultados deben interpretarse con cierta precaución [12,46]. Aunque la validación cruzada estratificada de 10 particiones ayuda a obtener resultados más fiables que una única división entre entrenamiento y prueba [44], no elimina completamente las limitaciones asociadas a trabajar con pocas muestras.

Otra limitación importante es que los modelos fueron entrenados y evaluados utilizando datos de un único centro hospitalario. Esto significa que todavía no se ha comprobado si funcionaría igual de bien en otros hospitales, utilizando diferentes equipos de ecografía o distintos protocolos de adquisición de imágenes [16,17]. Por ello, sería necesario realizar estudios multicéntricos antes de aplicar estos modelos en la práctica clínica real.

Por último, es importante destacar que los modelos de *Machine Learning* desarrollados en este trabajo no pretenden sustituir al médico especialista. Su objetivo es servir como herramientas de apoyo que ayuden a complementar la evaluación clínica tradicional [39,48], facilitando el diagnóstico precoz, la identificación de biomarcadores importantes y el seguimiento de la progresión de la esclerosis lateral amiotrófica.

Capítulo 4. RESULTADOS

4.1 INTRODUCCIÓN

En este capítulo se presentan los resultados obtenidos tras aplicar diferentes modelos de *Machine Learning* al análisis de biomarcadores ecográficos relacionados con la esclerosis lateral amiotrófica [16,36]. El principal objetivo fue estudiar si estos modelos eran capaces de diferenciar correctamente entre pacientes con ELA y sujetos sanos utilizando variables ecográficas, datos clínicos y descriptores texturales obtenidos de imágenes musculares.

Para ello, se realizaron distintos experimentos diseñados de forma sistemática y comparativa con el fin de analizar el rendimiento de los diferentes algoritmos estudiados. Todos los experimentos se prepararon de forma que se garantice que los resultados sean reproducibles y para evitar problemas frecuentes en estudios biomédicos con pocos datos, como el sobreajuste o el *data leakage* [44].

Además, en todos los experimentos se utilizó validación cruzada estratificada de 10 particiones [44]. Esto permite que las métricas obtenidas reflejen mejor la verdadera capacidad de los modelos para funcionar con nuevos pacientes y no solo su capacidad para ajustarse a los datos de entrenamiento.

Además de comparar distintos modelos de clasificación, también se estudió cómo influían diferentes factores, como el conjunto de variables utilizado, la aportación de cada músculo analizado, la importancia de los biomarcadores ecográficos y el efecto de incluir variables clínicas adicionales [16,46]. Este enfoque más amplio permite obtener conclusiones más completas que simplemente comparar la *accuracy* de los modelos.

Los resultados obtenidos muestran que las técnicas de *Machine Learning* son capaces de detectar patrones relevantes relacionados con la ELA con niveles altos de precisión, especialmente cuando se combinan diferentes tipos de biomarcadores [36,43]. Además, los

resultados coinciden con los descritos en estudios previos de la literatura científica [16,17], lo que aporta mayor confianza al enfoque utilizado y sugiere que los biomarcadores ecográficos analizados contienen información importante sobre los cambios musculares producidos por la enfermedad.

4.2 COMPARACIÓN DE MODELOS DE MACHINE LEARNING

4.2.1 CONFIGURACIÓN EXPERIMENTAL

Con el objetivo de identificar qué algoritmo era más adecuado para clasificar pacientes con esclerosis lateral amiotrófica, se compararon de forma sistemática siete modelos de *Machine Learning* utilizando validación cruzada estratificada de 10 particiones [44]. Los modelos analizados fueron la Regresión Logística, las Máquinas de Vectores de Soporte (*SVM*), *Random Forest*, *K-Nearest Neighbors (KNN)*, los Árboles de Decisión, *Gradient Boosting* y *Naive Bayes* [36,41,37,38,40,43].

La comparación se realizó utilizando exactamente los mismos datos y las mismas condiciones experimentales para todos los modelos. Esto es importante porque garantiza que las diferencias observadas en el rendimiento se deben realmente a las características de cada algoritmo y no a cambios en el proceso de evaluación [43]. Todos los modelos se implementaron mediante *pipelines* que incluían tanto la normalización de las variables como el entrenamiento del clasificador, evitando problemas de fuga de información entre entrenamiento y validación [36].

La evaluación de los modelos se realizó utilizando varias métricas habituales en clasificación biomédica: *accuracy*, *recall* o sensibilidad, *F1-score* y área bajo la curva *ROC (AUC)* [45]. Cada una de estas métricas analiza un aspecto diferente del comportamiento del modelo. La *accuracy* indica el porcentaje total de clasificaciones correctas; el *recall* mide la capacidad de detectar correctamente a los pacientes enfermos; el *F1-score* combina precisión y sensibilidad; y el *AUC* evalúa la capacidad general del modelo para diferenciar entre pacientes y controles independientes del umbral de decisión utilizado.

4.2.2 DETECCIÓN Y CORRECCIÓN DEL SOBREAJUSTE

En los primeros experimentos se obtuvieron valores de *accuracy* extremadamente altos, cercanos al 100% en algunos modelos. Aunque inicialmente estos resultados parecían excelentes, un análisis más detallado mostró que algunas variables tenían correlaciones demasiado altas con la variable objetivo [44]. Esto podía deberse a información repetida entre variables, fuga de información entre entrenamiento y validación o variables que reflejaban indirectamente el diagnóstico sin representar biomarcadores reales desde el punto de vista biológico.

Después de identificar y eliminar estas variables problemáticas, y tras comprobar que los *pipelines* estaban correctamente implementados [36], los resultados pasaron a valores más realistas y clínicamente creíbles. Este proceso fue uno de los aspectos metodológicos más importantes del trabajo, ya que permitió asegurar que los modelos estaban aprendiendo patrones reales y no simplemente aprovechando errores metodológicos.

Esta experiencia demuestra uno de los mayores riesgos del *Machine Learning* aplicado a datos biométricos: obtener resultados aparentemente muy buenos debido a pequeños problemas metodológicos que solo pueden detectarse mediante una revisión cuidadosa de las variables y del proceso de evaluación [44,46].

4.2.3 RESULTADOS GLOBALES POR MODELO

Tras aplicar las correcciones necesarias, los resultados mostraron que la mayoría de los modelos eran capaces de clasificar pacientes con ELA con un rendimiento alto y clínicamente relevante. Las *accuracies* obtenidas se situaron entre el 78% y el 90%, dependiendo del modelo y del conjunto de variables utilizado.

Los modelos que mostraron un comportamiento más estable y consistente fueron la Regresión Logística, el *SVM*, el *KNN* y el *Random Forest* [36,41,38]. En particular, la Regresión Logística destacó por ofrecer el mejor equilibrio entre precisión, estabilidad, facilidad de interpretación y capacidad de generalización [41,42]. Este modelo alcanzó una *accuracy* media del 90% utilizando el conjunto completo de variables y un *AUC* de 0.973.

Este resultado es especialmente interesante porque indica que gran parte de las diferencias entre pacientes con ELA y controles pueden explicarse mediante relaciones relativamente lineales entre los biomarcadores musculares. El hecho de que un modelo lineal consiga resultados tan altos indica que los biomarcadores ecográficos contienen una señal biológica clara y consistente [16,17].

El modelo *SVM* obtuvo resultados prácticamente iguales a los de la Regresión Logística, con una *accuracy* del 90% y un *AUC* de 0.973. También mostró valores elevados de sensibilidad y gran capacidad para separar ambos grupos, incluso cuando las relaciones entre variables eran más complejas [38]. Estos resultados coinciden con otros estudios biomédicos donde el *SVM* suele funcionar muy bien con *datasets* pequeños o medianos [38].

El *Random Forest* alcanzó una *accuracy* del 85% y mostró buena capacidad de generalización gracias al uso combinado de múltiples árboles de decisión [37]. Además, este modelo permite analizar qué variables tienen más importancia en la clasificación, algo que se estudia más adelante en la sección 4.7 [47].

El *KNN* también obtuvo buenos resultados, alcanzando una *accuracy* del 86% con el conjunto completo de variables [43]. Esto indica que los biomarcadores ecográficos permiten separar de forma bastante clara a pacientes y controles dentro del espacio de características analizado.

El *Gradient Boosting* alcanzó una *accuracy* del 85% [40], mientras que el *Naive Bayes* obtuvo un 88% [43]. Este último resultado es especialmente interesante porque el modelo es relativamente sencillo, lo que sugiere que algunas variables individuales tienen una gran capacidad para diferenciar pacientes con ELA y sujetos sanos.

Por otro lado, los Árboles de Decisión individuales mostraron peor capacidad de generalización y mayor tendencia al sobreajuste, alcanzando una *accuracy* del 78% [43,44]. Además, sus resultados variaban más entre las distintas particiones de validación cruzada. Este comportamiento era esperado, ya que este tipo de modelos tiende a memorizar los datos cuando no se controla adecuadamente su complejidad.

La Tabla 1 resume las *accuracies* medias obtenidas mediante validación cruzada estratificada de 10 particiones para cada modelo y conjunto de variables.

	Biomarcadores clásicos	GLCM	Todas las variables
Decision Tree	0.780000	0.686364	0.779091
Gradient Boosting	0.791818	0.816364	0.847273
KNN	0.846364	0.770909	0.855455
Logistic Regression	0.867273	0.893636	0.903636
Naive Bayes	0.854545	0.760000	0.875455
Random Forest	0.839091	0.808182	0.848182
SVM	0.885455	0.854545	0.902727

Tabla 1. *Accuracy media (10-Fold CV estratificado) por modelo y conjunto de variables*

4.3 IMPACTO DEL CONJUNTO DE VARIABLES

4.3.1 ANÁLISIS COMPARATIVO POR TIPO DE BIOMARCADOR

Uno de los aspectos más importantes estudiados en este trabajo fue analizar cómo influye el tipo de variables utilizadas en el rendimiento de los modelos de clasificación [16,36]. Para ello, se compararon tres configuraciones experimentales diferentes: usar únicamente biomarcadores ecográficos clásicos, utilizar solo descriptores texturales *GLCM* y combinar todos los biomarcadores disponibles junto con variables clínicas.

Esta comparación es muy importante desde el punto de vista clínico porque permite conocer qué tipo de información aporta más utilidad al sistema diagnóstico. Si los biomarcadores clásicos, que son más fáciles de obtener en la práctica clínica diaria, fueran suficientes para conseguir buenos resultados, sería más sencillo aplicar este sistema en hospitales sin necesidad de herramientas avanzadas de análisis de imagen [18,22]. En cambio, si los descriptores texturales mejoran claramente el rendimiento, esto justificaría el uso de técnicas más avanzadas de procesamiento de imágenes [28,32].

Los biomarcadores ecográficos clásicos incluían medidas de grosor muscular, ecogenicidad y ecotextura en los cuatro músculos analizados [16,18]. Estos parámetros reflejan cambios visibles en la imagen ecográfica, como la pérdida de masa muscular producida por la denervación o el aumento de ecogenicidad causado por infiltración grasa [27]. Además, pueden obtenerse utilizando herramientas presentes en la mayoría de los equipos modernos de ecografía.

Por otro lado, los descriptores texturales *GLCM* incluían cinco parámetros principales: ASM, CON, COR, IDM y ENT para cada uno de los cuatro músculos [28,35]. Estas variables permiten estudiar la organización interna del tejido muscular y describir cambios microestructurales que no siempre son visibles a simple vista en la imagen ecográfica [16,17].

4.3.2 RESULTADOS POR CONJUNTO DE VARIABLES

Los resultados mostraron de forma consistente que combinar todos los tipos de biomarcadores disponibles proporcionaba el mejor rendimiento en la mayoría de los modelos analizados [16,36].

Para visualizar de forma más clara las diferencias de rendimiento entre los distintos algoritmos, se compararon las *accuracies* medias obtenidas por cada modelo utilizando los tres conjuntos de variables analizados: descriptores texturales *GLCM*, biomarcadores clásicos y la combinación de todas las variables disponibles. Los resultados obtenidos se muestran en la Figura 5.

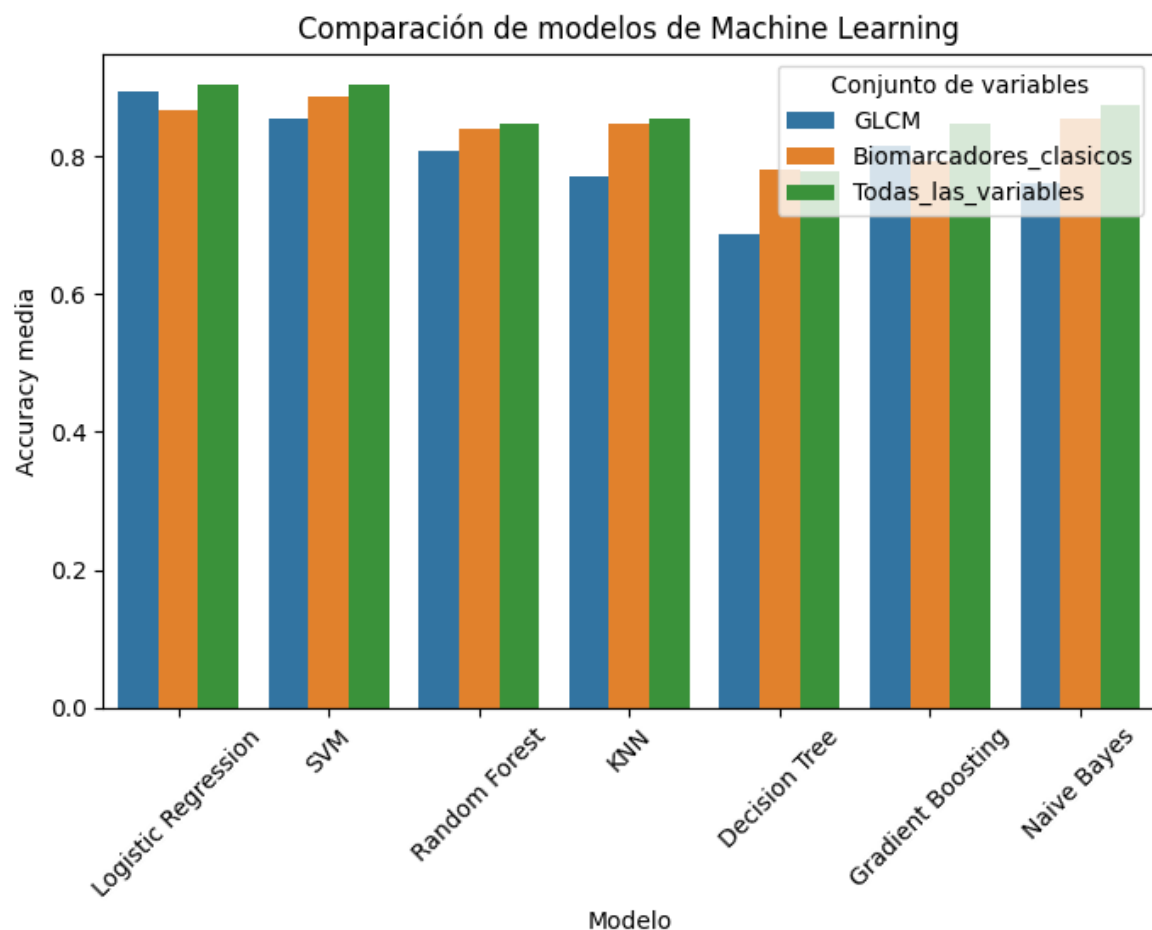


Figura 5. Comparación de la accuracy de los diferentes modelos de Machine Learning para los tres conjuntos de variables analizados

La Regresión Logística obtuvo una *accuracy* del 87% utilizando solo biomarcadores clásicos y del 89% utilizando únicamente descriptores *GLCM*. Sin embargo, cuando se combinaron todas las variables, la *accuracy* aumentó hasta el 90% [41,42]. El modelo *SVM* mostró un comportamiento parecido [38]. Además, el *Naive Bayes* fue el modelo que experimentó la mejora más clara al utilizar todas las variables, pasando del 76% con solo *GLCM* al 88% con el conjunto completo.

Estos resultados tienen una implicación importante: los biomarcadores clásicos y los descriptores texturales aportan información complementaria sobre el estado del músculo

[16,17]. Por ello, combinarlos permite obtener una descripción más completa y precisa del proceso neurodegenerativo asociado a la esclerosis lateral amiotrófica.

Los biomarcadores clásicos reflejan cambios globales visibles en el músculo, como la atrofia o el aumento de ecogenicidad [18,27]. En cambio, los descriptores texturales permiten detectar cambios más sutiles relacionados con la organización interna del tejido muscular, que pueden aparecer en fases más tempranas de la enfermedad, cuando los cambios todavía no son muy evidentes [16,28].

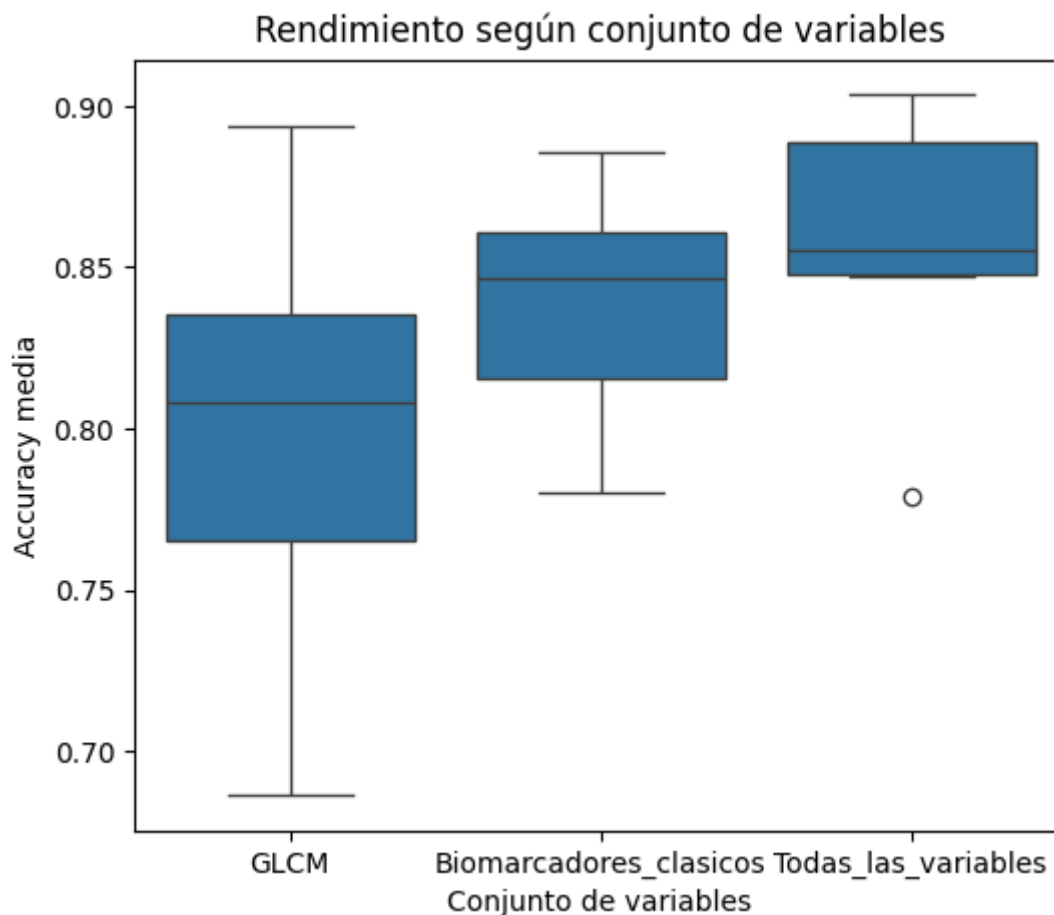


Figura 6. Distribución de la accuracy media obtenida por los modelos según el conjunto de variables empleado

La distribución de *accuracies* de la Figura 6 muestra que el conjunto formado por todas las variables no solo obtiene mayores valores de rendimiento, sino también una menor

dispersión entre modelos. Esto sugiere que la combinación de biomarcadores clásicos y descriptores texturales proporciona resultados más robustos y consistentes [16,36].

También es interesante destacar que algunos modelos respondieron mejor a determinados tipos de variables. Por ejemplo, el *Gradient Boosting* obtuvo mejores resultados utilizando solo descriptores *GLCM* (*accuracy* de 0.84) que utilizando únicamente biomarcadores clásicos (*accuracy* de 0.78) [40]. Esto sugiere que, para algunos algoritmos, la información textural puede ser especialmente útil [28].

Por el contrario, el *SVM* consiguió su mejor resultado individual utilizando únicamente biomarcadores clásicos (*accuracy* de 0.89), lo que indica que este modelo aprovecha muy bien este tipo de información.

4.4 EVALUACIÓN DEL MEJOR MODELO

4.4.1 SELECCIÓN Y JUSTIFICACIÓN DEL MODELO

A partir de los resultados obtenidos en los diferentes experimentos, se seleccionó la Regresión Logística entrenada con el conjunto completo de variables como el modelo con mejor rendimiento global [41,42]. Aunque otros algoritmos consiguieron resultados similares en algunos experimentos, la Regresión Logística destacó por ser más estable entre las distintas particiones de validación cruzada, por su facilidad de interpretación, por tener menor tendencia al sobreajuste con un *dataset* de tamaño moderado y por ofrecer un buen equilibrio entre sensibilidad y especificidad [41,44].

La elección de la Regresión Logística frente al *SVM*, que obtuvo una *accuracy* prácticamente igual (0.9036 frente a 0.9027), también se debe a que la Regresión Logística es más fácil de interpretar [41,42]. Los coeficientes del modelo permiten conocer directamente cuánto influye cada variable en la clasificación, algo muy útil tanto para validar clínicamente el modelo como para identificar qué biomarcadores son más importantes [48,49]. Por esta razón, la Regresión Logística se considera especialmente adecuada como base para un posible sistema de apoyo al diagnóstico clínico en el futuro.

4.4.2 MÉTRICAS DE CLASIFICACIÓN DETALLADAS

La Tabla 2 muestra los resultados detallados del mejor modelo obtenidos mediante validación cruzada estratificada de 10 particiones.

	Precisión	Recall	F1-score	Soporte
Control	0.89	0.92	0.91	52
ELA	0.92	0.88	0.90	52
Accuracy global			0.90	104

Tabla 2. Informe de clasificación del mejor modelo (Regresión Logística con todas las variables)

El modelo alcanzó una *accuracy* global del 90%, con valores bastante equilibrados entre ambas clases. El *recall* para la clase ELA fue del 88%, lo que significa que el modelo detectó correctamente el 88% de los pacientes con Esclerosis lateral amiotrófica presentes en el *dataset* [45]. Además, la precisión del 92% indica que cuando el modelo predice que un sujeto tiene ELA, acierta en el 92% de los casos, lo que refleja una tasa baja de falsos positivos.

Desde el punto de vista clínico, el equilibrio entre precisión y *recall* es muy importante [45]. Un *recall* del 88% implica que aproximadamente 1 de cada 8 pacientes con ELA no sería detectado por el modelo, es decir, se producirían falsos negativos. Sin embargo, es importante recordar que este sistema no pretende sustituir al médico, sino actuar como una herramienta de apoyo adicional junto con la evaluación clínica, la electromiografía y otras pruebas diagnósticas [11,16].

La clase Control también obtuvo resultados sólidos: precisión del 89%, *recall* del 92% y un *F1-score* de 0.91. Esto significa que el modelo también identifica correctamente a la mayoría de los sujetos sanos y produce pocos falsos positivos [45].

4.4.3 ANÁLISIS DE LA MATRIZ DE CONFUSIÓN

La matriz de confusión del mejor modelo en la Figura 7 mostró que, de las 104 muestras analizadas mediante validación cruzada, el modelo clasificó correctamente 48 registros

sanos (verdaderos negativos) y 46 con ELA (verdaderos positivos). Además, cometió 4 falsos positivos, es decir, controles clasificados incorrectamente como ELA, y 6 falsos negativos, es decir, pacientes con ELA clasificados como controles sanos [45].

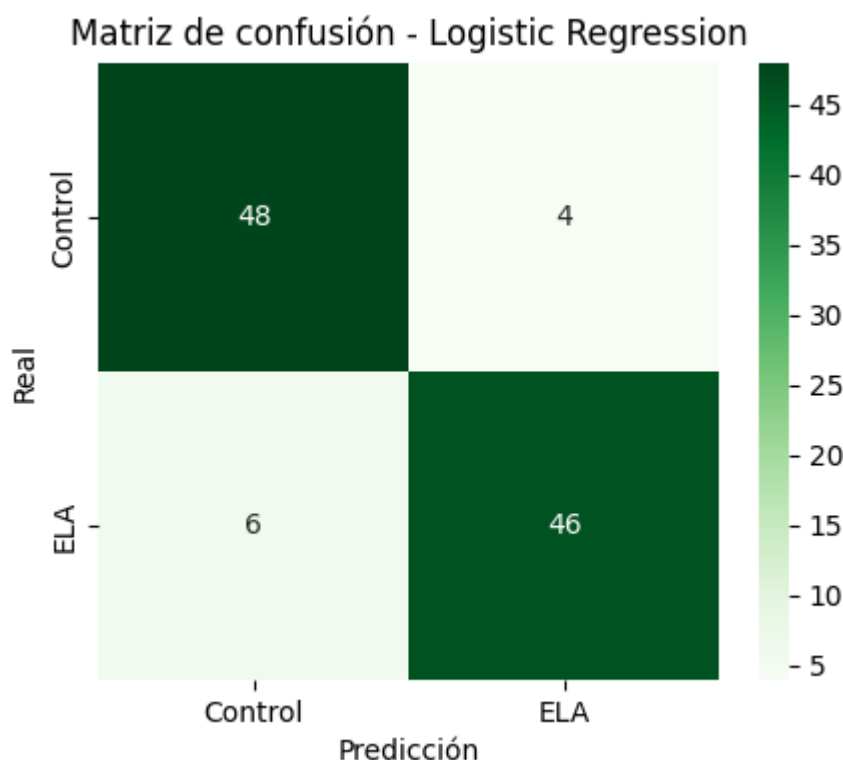


Figura 7. Matriz de confusión del modelo de Regresión Logística utilizando el conjunto completo de variables

Es importante analizar estos errores desde el punto de vista clínico [7,11]. Los 6 falsos negativos representan muestras con ELA que el modelo no detectaría, lo que podría retrasar el diagnóstico si el sistema se utilizara como herramienta de cribado. Sin embargo, en un entorno clínico real, estos casos también serían evaluados mediante otras pruebas y valoraciones médicas.

Por otro lado, los 4 falsos positivos representan registros sanos que el modelo clasificaría incorrectamente como ELA. Esto podría generar preocupación innecesaria, aunque posteriormente estos casos probablemente serían descartados mediante evaluación clínica completa.

La distribución bastante equilibrada entre falsos positivos y falsos negativos indica que el modelo no está claramente sesgado hacia ninguna de las dos clases, lo que sugiere que el uso del parámetro `class_weight="balanced"` funcionó correctamente durante el entrenamiento [36].

4.4.4 CURVA ROC Y AUC

En la Figura 8 se muestra la curva *ROC* del mejor modelo, que representa gráficamente la relación entre la sensibilidad y la tasa de falsos positivos para todos los posibles umbrales de decisión. El área bajo esta curva (*AUC*) fue de 0.973, lo que indica una capacidad discriminativa excelente y muy superior al valor 0.5, que correspondería a un clasificador aleatorio [45].

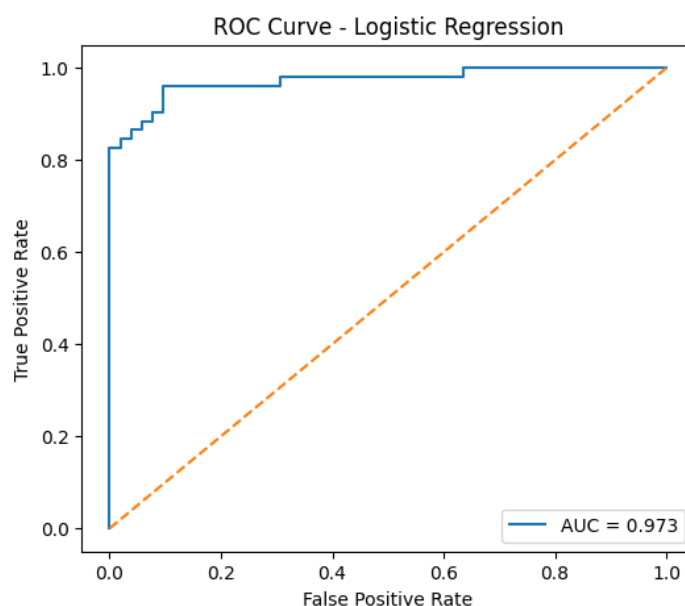


Figura 8. Curva ROC del modelo de Regresión Logística utilizando el conjunto completo de variables

Este valor de *AUC* tiene una interpretación muy clara: si se eligiera al azar un paciente con ELA y un sujeto sano del *dataset*, el modelo asignaría una probabilidad mayor de ELA al paciente enfermo en el 97.3% de los casos. Este resultado confirma la gran capacidad del modelo para diferenciar correctamente entre ambas clases.

Además, el valor de AUC obtenido es muy parecido al reportado en estudios anteriores de Martínez-Payá et al. (2017), donde se alcanzaron valores cercanos a 0.98 utilizando métodos estadísticos clásicos para analizar el músculo cuádriceps [16]. Esto demuestra que los modelos de *Machine Learning* son capaces de reproducir y confirmar los buenos resultados descritos previamente en la literatura científica.

La forma de la curva ROC también muestra que el modelo mantiene sensibilidades muy altas con tasas bajas de falsos positivos, algo especialmente importante en herramientas diagnósticas [45]. Además, modificando el umbral de decisión, el sistema podría configurarse para priorizar la sensibilidad y detectar el mayor número posible de pacientes con ELA, aunque eso implicara aceptar algunos falsos positivos adicionales. Esto podría ser especialmente útil en situaciones de cribado, donde es más importante evitar dejar pacientes sin diagnosticar [7].

4.5 COMPARACIÓN CON EL ESTADO DEL ARTE

4.5.1 CONTEXTO DE LA COMPARACIÓN

Uno de los aspectos más importantes de este trabajo fue comparar los resultados obtenidos con otros estudios publicados anteriormente sobre biomarcadores ecográficos e Inteligencia Artificial aplicada a la esclerosis lateral amiotrófica [16,17,46]. Esta comparación permite comprobar si el enfoque utilizado realmente aporta información nueva y útil, o si simplemente repite resultados ya conocidos sin mejoras metodológicas o de rendimiento.

En los últimos años, distintos estudios han demostrado que la ecografía muscular y el análisis computacional avanzado pueden ser útiles para detectar alteraciones relacionadas con la degeneración de las motoneuronas [16,17,23]. Sin embargo, muchos de estos trabajos tienen limitaciones importantes, como tamaños de muestras pequeños, falta de validación externa, diferencias metodológicas entre estudios y problemas de reproducibilidad [46]. Todo esto dificulta comparar directamente los resultados y confirmar si realmente pueden aplicarse a otros pacientes y hospitales.

4.5.2 COMPARACIÓN DE RENDIMIENTOS

Los resultados obtenidos en este trabajo muestran un rendimiento muy parecido al descrito en estudios anteriores, especialmente en métricas como *accuracy*, sensibilidad y *AUC* [16,17].

En concreto, el estudio de Martínez-Payá et al. (2017) [16], que es la referencia más cercana porque utilizó el mismo *dataset*, obtuvo valores de *AUC* entre 0.95 y 0.98 para distintos músculos utilizando regresión logística univariante y métodos estadísticos clásicos. En este trabajo, los modelos desarrollados consiguieron resultados dentro de ese mismo rango, alcanzando un *AUC* de 0.973 para el mejor modelo global y de 0.977 en el análisis específico del músculo cuádriceps.

Estos resultados son importantes por dos motivos principales. En primer lugar, confirman que los biomarcadores ecográficos utilizados contienen información fisiopatológica sólida y reproducible, ya que métodos diferentes (estadística clásica y *Machine Learning*) consiguen resultados muy similares utilizando los mismos datos [16,17].

En segundo lugar, muestran que el enfoque basado en *Machine Learning* con validación cruzada estratificada de 10 particiones es metodológicamente más riguroso que algunos análisis estadísticos clásicos sin validación cruzada [44]. Esto permite obtener estimaciones del rendimiento más fiables y menos influenciadas por el azar.

Además, los resultados apoyan la idea, descrita también en estudios recientes, de que combinar varios biomarcadores proporciona mejores resultados que utilizar una sola variable de forma aislada [16,32]. Los descriptores texturales aportaron información complementaria respecto a los biomarcadores ecográficos clásicos, algo coherente con investigadores que indican que los cambios microestructurales del músculo pueden detectarse antes de que aparezcan alteraciones visibles más evidentes [17,28].

4.5.3 VENTAJAS DEL ENFOQUE PROPUESTO FRENTE A ESTUDIOS PREVIOS

Una de las principales diferencias metodológicas de este trabajo respecto a muchos estudios anteriores es la comparación sistemática de varios algoritmos utilizando exactamente las mismas condiciones experimentales [36,43]. Aunque esta práctica es muy recomendable en *Machine Learning* biomédico, normalmente solo se prueban uno o dos modelos de forma aislada.

Otra diferencia importante es el análisis del impacto de distintos conjuntos de variables [16,28]. Esto permitió estudiar cuánto aporta cada tipo de biomarcador al rendimiento diagnóstico y obtener información útil para futuros trabajos. Muchos estudios anteriores utilizan un único conjunto de variables sin comparar diferentes combinaciones de forma sistemática.

Por último, también es importante destacar que algunos trabajos previos utilizan modelos muy complejos basados en *Deep Learning* [39,50]. Aunque estos modelos pueden conseguir resultados muy altos, normalmente necesitan grandes cantidades de datos, requieren mucha capacidad computacional y son más difíciles de interpretar clínicamente [48,49].

En este trabajo se observó que modelos relativamente simples, como la Regresión Logística o el *SVM*, pueden alcanzar resultados muy competitivos con el tamaño de *dataset* disponible [41,38]. Esto es especialmente importante pensando en una futura aplicación clínica real, ya que son modelos más fáciles de implementar, interpretar y utilizar en hospitales [48].

4.6 ANÁLISIS POR MÚSCULO: EL CASO DEL CUÁDRICEPS FEMORAL

4.6.1 MOTIVACIÓN DEL ANÁLISIS ESPECÍFICO

Además del análisis general utilizando todas las variables disponibles, también se realizaron estudios centrados únicamente en el músculo cuádriceps femoral. El objetivo era analizar si este músculo, por sí solo, tenía suficiente capacidad para diferenciar pacientes con esclerosis lateral amiotrófica y compararlo con los resultados descritos en otros estudios [16].

La elección del cuádriceps femoral se debe a varias razones importantes. En primer lugar, estudios previos de Martínez-Payá et al. (2017) [16] obtuvieron los mejores resultados diagnósticos utilizando este músculo, alcanzando un *AUC* de 0.98. Además, el cuádriceps es el músculo más grande del cuerpo humano, lo que facilita obtener imágenes de muy buena calidad y con regiones amplias para el análisis [16,18]. Por último, este músculo suele verse afectado con frecuencia en la ELA, especialmente en pacientes cuya enfermedad comienza en los miembros inferiores [8,16].

Las variables analizadas del cuádriceps incluyeron todos los biomarcadores ecográficos disponibles para este músculo: grosor muscular (Cdr.Thick), ecogenicidad (Cdr.Ecog), ecotextura (Cdr.EcTx) y los cinco descriptores *GLCM* (ASM.Cdr, CON.Cdr, COR.Cdr, IDM.Cdr y ENT.Cdr) [16,28].

4.6.2 RESULTADOS DEL ANÁLISIS DEL CUÁDRICEPS

Los resultados obtenidos con el modelo de Regresión Logística utilizando únicamente las variables del cuádriceps fueron muy buenos, como se muestra en la Tabla 3.

	Precisión	Recall	F1-score	Soporte
Control	0.89	0.94	0.92	52
ELA	0.94	0.88	0.91	52
Accuracy global			0.91	104

Tabla 3. Resultados del modelo específico del músculo cuádriceps femoral

El modelo alcanzó una *accuracy* del 91% utilizando solamente biomarcadores de un único músculo, superando ligeramente el 90% obtenido con el conjunto completo de variables [41]. Además, el *AUC* fue de 0.977 como se puede analizar en la Figura 9, también ligeramente superior al *AUC* de 0.973 obtenido con todas las variables [45]. Esto confirma la gran importancia del cuádriceps como músculo de referencia en el diagnóstico ecográfico de la ELA [16].

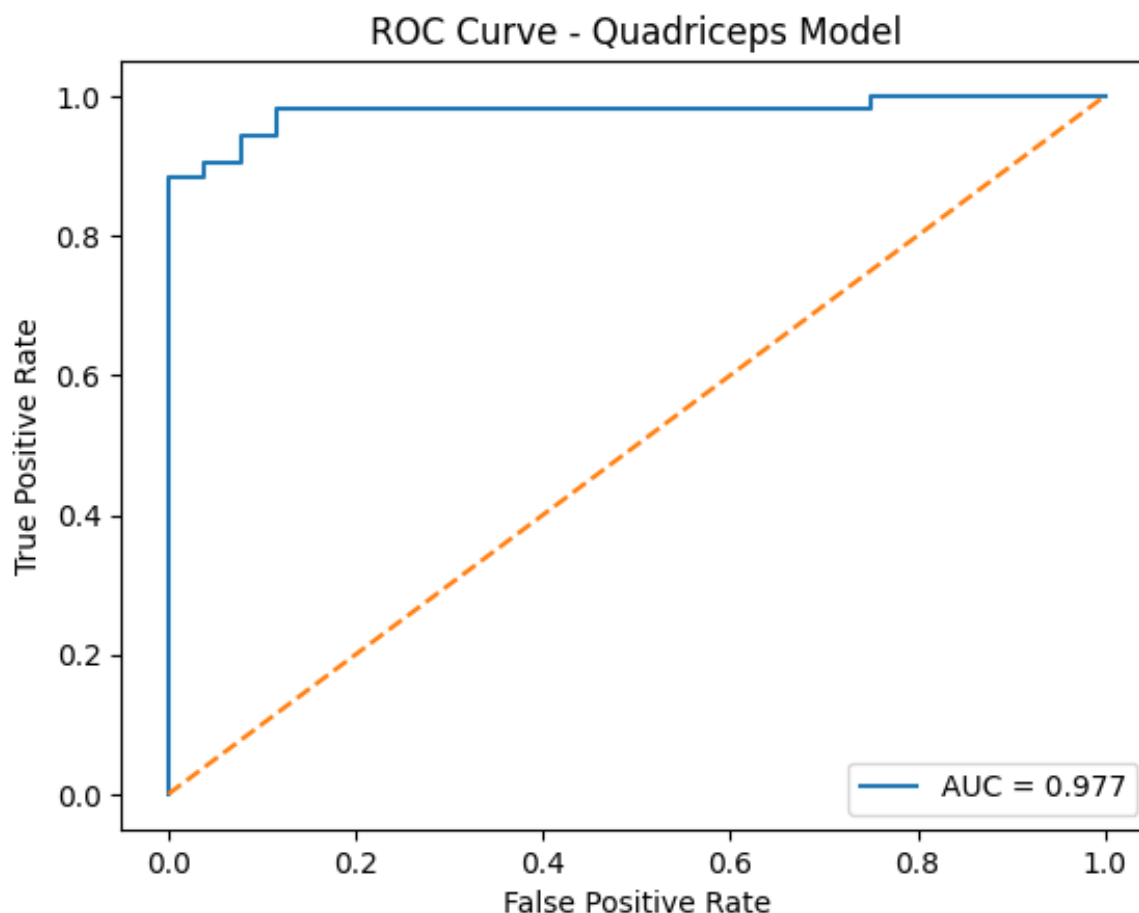


Figura 9. Curva ROC del modelo basado en biomarcadores ecográficos del músculo cuádriceps

La matriz de confusión de la Figura 10 mostró 49 verdaderos negativos, 46 verdaderos positivos, 3 falsos positivos y 6 falsos negativos. Comparado con el modelo global, el análisis de cuádriceps redujo los falsos positivos de 4 a 3 y mantuvo el mismo número de falsos negativos. Esto indica que el cuádriceps proporciona una señal diagnóstica especialmente clara para identificar sujetos sanos y reduce el número de controles clasificados erróneamente como pacientes con ELA [16,17].

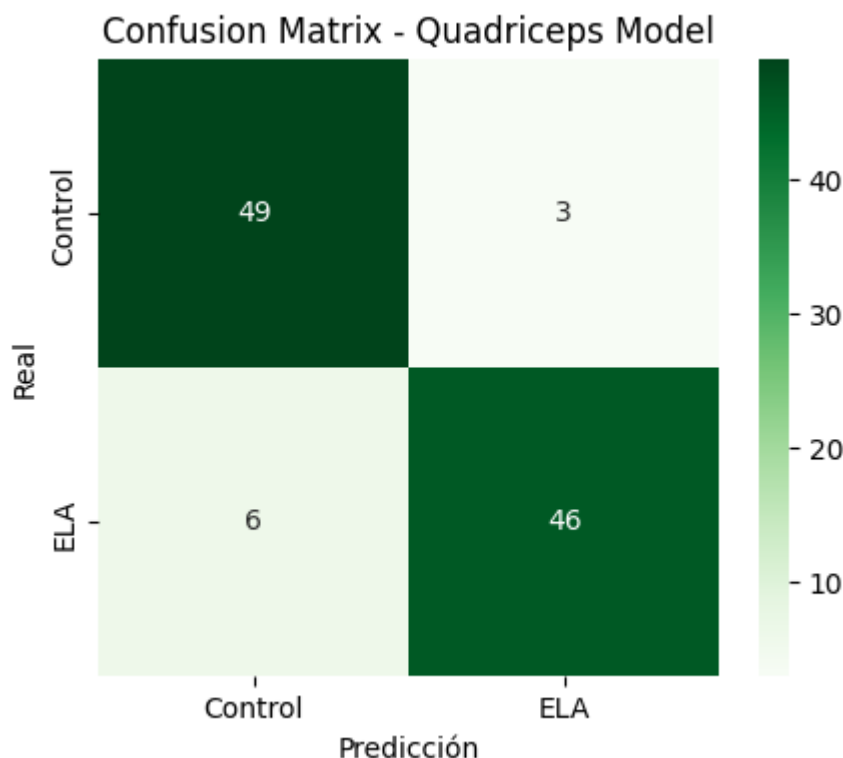


Figura 10. Matriz de confusión del modelo basado en biomarcadores ecográficos del músculo cuádriceps

El hecho de que un solo músculo consiga resultados tan similares, e incluso ligeramente mejores, que el modelo completo tiene una explicación clínica importante. Si en la práctica clínica fuese necesario reducir el tiempo de exploración, estudiar únicamente el cuádriceps podría aportar gran parte de la información diagnóstica obtenida con el protocolo completo de cuatro músculos [16].

4.6.3 EXPLICACIÓN FISIOPATOLÓGICA DE LA RELEVANCIA DEL CUÁDRICEPS

La gran capacidad del cuádriceps para diferenciar pacientes con ELA puede explicarse por varios factores fisiopatológicos [3,16]. Este músculo está innervado principalmente por el nervio femoral y por segmentos medulares L2-L4. Además, tiene una gran cantidad de unidades motoras y una superficie muscular muy amplia, lo que hace que los cambios degenerativos producidos por la denervación sean más evidentes y fáciles de detectar mediante ecografía [16,27].

Además, el cuádriceps suele verse afectado relativamente pronto en muchos pacientes con ELA, especialmente en las formas de inicio espinal en miembros inferiores [8,16]. Sus cambios estructurales también son muy sensibles a los descriptores texturales *GLCM* debido a la organización de sus fibras musculares y a su gran tamaño, que permite analizar regiones amplias y representativas del tejido muscular [16,28].

Los descriptores texturales del cuádriceps mostraron cambios muy consistentes en pacientes con ELA: aumentos del contraste y la entropía, y disminuciones de la homogeneidad y la energía [16,17]. Estos cambios reflejan que el músculo degenerado presenta una estructura más irregular y desorganizada.

4.6.4 LIMITACIONES DEL ANÁLISIS POR MÚSCULO ÚNICO

Aun así, también se observó que utilizar solo el cuádriceps no siempre mejoraba claramente el rendimiento respecto a usar varios músculos combinados en todos los experimentos realizados. Esto probablemente se debe a la gran heterogeneidad clínica de la ELA [8,15]. La enfermedad no afecta igual a todos los pacientes, y cada persona puede presentar diferentes patrones de afectación muscular dependiendo de la fase de la enfermedad, el tipo clínico y la región donde comenzaron los síntomas [8].

Por tanto, aunque el cuádriceps es un biomarcador muy potente en muchos pacientes, combinar información de varios músculos permite crear modelos más robustos y con mejor capacidad de generalización ante la gran variedad de presentaciones clínicas de la enfermedad [16]. Además del cuádriceps, también se observaron aportaciones importantes del tibial anterior, el bíceps braquial y los flexores del antebrazo, músculos que suelen afectarse de forma temprana en diferentes variantes clínicas de la ELA [16,17].

4.7 ANÁLISIS DE IMPORTANCIA DE VARIABLES

4.7.1 METODOLOGÍA DEL ANÁLISIS

Además de evaluar la capacidad de los distintos modelos de *Machine Learning* para clasificar pacientes, también es importante estudiar qué variables influyen más en las

decisiones de los algoritmos [48,49]. En aplicaciones médicas no solo interesa obtener modelos con buenos resultados, sino también entender qué biomarcadores o características están contribuyendo a las predicciones. Esto ayuda a aumentar la confianza en los modelos y facilita su posible uso en la práctica clínica [48].

Con este objetivo, se realizó un análisis de importancia de variables utilizando el modelo de Regresión Logística, ya que fue el algoritmo que obtuvo el mejor rendimiento global [41,42]. Una de las ventajas de este modelo es que permite interpretar fácilmente la influencia de cada variable mediante el estudio de sus coeficientes.

Antes de entrenar el modelo, todas las variables se normalizaron utilizando la técnica *StandardScaler* [36]. Este paso es necesario para evitar que las diferencias en las unidades o escalas de medida influyan en la interpretación de los coeficientes. Una vez entrenado el modelo, se obtuvieron los coeficientes asociados a cada variable y se calculó su valor absoluto. Cuanto mayor es este valor, mayor es la influencia de esa variable en la clasificación entre pacientes con ELA y sanos [41,42].

Este análisis permite identificar cuáles son los biomarcadores más útiles para diferenciar ambos grupos y comprobar si los resultados coinciden con lo que se conoce actualmente sobre la enfermedad [16,17].

4.7.2 VARIABLES MÁS RELEVANTES IDENTIFICADAS

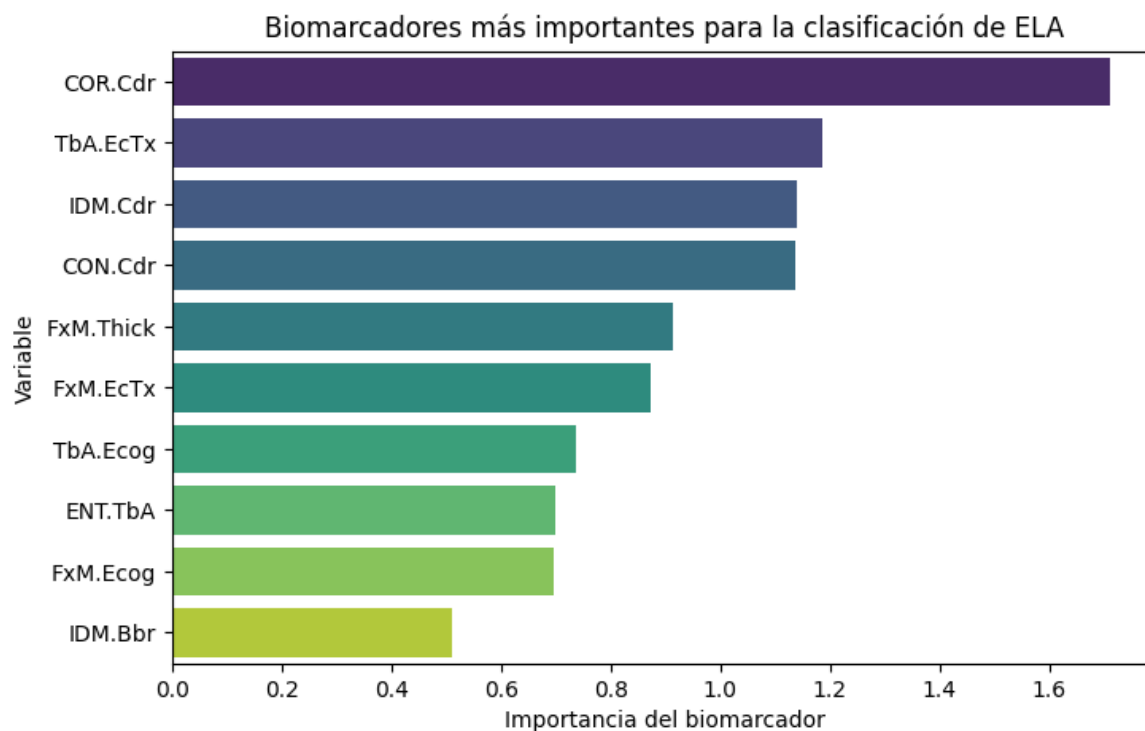


Figura 11. Ranking de los biomarcadores más importantes para clasificar pacientes con ELA según el modelo de Regresión Logística

La Figura 11 muestra las diez variables con mayor importancia dentro del modelo final de Regresión Logística. Los resultados indican que varios de los biomarcadores más importantes proceden del análisis de textura realizado sobre las imágenes ecográficas musculares.

La variable más relevante fue COR.Cdr, que corresponde al parámetro de correlación obtenido mediante matrices *GLCM* en el músculo cuádriceps [16,28]. Después aparecen TbA.EcTx, IDM.Cdr y CON.Cdr, que también tuvieron gran influencia en la capacidad de clasificación del modelo.

Es interesante observar que tres de las cuatro variables más importantes pertenecen al músculo cuádriceps. Este resultado coincide con los experimentos realizados sobre este músculo, donde ya se había observado su gran capacidad para diferenciar pacientes con ELA de sujetos sanos [16]. La presencia de biomarcadores procedentes del cuádriceps indica que

este músculo contiene información muy relevante sobre los cambios producidos por la enfermedad.

Además de los biomarcadores del cuádriceps, también aparecen variables de otros músculos estudiados. Entre ellas destacan TbA.EcTx y TbA.Ecog, correspondientes al tibial anterior, así como FxM.Thick, FxM.EcTx y FxM.Ecog, obtenidas de los flexores del antebrazo [16,17]. También aparece IDM.Bbr, que indica que el bíceps braquial aporta información útil para la clasificación.

Otro resultado importante es que muchas de las variables seleccionadas son descriptores de textura obtenidos mediante análisis *GLCM* [28,35]. Entre ellos se encuentran parámetros como COR (*Correlation*), IDM (*Inverse Difference Moment*), CON (*Contrast*) y ENT (*Entropy*). Estos descriptores permiten medir cómo se distribuyen los niveles de gris dentro de la imagen ecográfica y aportan información sobre la organización interna del tejido muscular [16,28].

A diferencia de biomarcadores más tradicionales, como el grosor muscular o la ecogenicidad media, los descriptores texturales pueden detectar patrones complejos y cambios microestructurales que no siempre son visibles a simple vista en la ecografía [16,32]. Esto explica que varios de ellos aparezcan entre las variables más importantes del modelo.

En conjunto, los resultados sugieren que combinar biomarcadores ecográficos clásicos con parámetros avanzados de textura proporciona una visión más completa del estado muscular de los pacientes y mejora la capacidad para diferenciar entre personas sanas y pacientes con ELA [16,17,28].

4.7.3 INTERPRETACIÓN CLÍNICA DE LOS RESULTADOS

Los resultados obtenidos son coherentes con lo que se conoce actualmente sobre la fisiopatología de la ELA [3,8]. Esta enfermedad provoca una degeneración progresiva de las motoneuronas superiores e inferiores, lo que produce denervación muscular, pérdida de fibras musculares funcionales, atrofia e infiltración de grasa y tejido conectivo [3,27].

Todos estos cambios modifican la estructura interna del músculo. A medida que la enfermedad avanza, las fibras musculares son sustituidas progresivamente por grasa y

fibrosis, haciendo que el tejido muscular sea cada vez más irregular y desorganizado [17,27]. Como consecuencia, también cambia la distribución de intensidades observada en las imágenes ecográficas.

Los descriptores *GLCM* identificados como más importantes por el modelo son especialmente sensibles a este tipo de cambios [16,28]. Esto indica que los cambios en la textura muscular contienen información diagnóstica muy valiosa y complementaria a la proporcionada por los biomarcadores ecográficos tradicionales. Este resultado demuestra que el análisis computacional avanzado de imágenes ecográficas puede detectar alteraciones asociadas a la ELA que no siempre son evidentes mediante una evaluación visual convencional [16,32].

Desde el punto de vista clínico, estos resultados son especialmente prometedores porque identificar los músculos y biomarcadores más informativos podría ayudar en el futuro a diseñar protocolos ecográficos más eficientes, centrados en las zonas que aportan más información [7,16].

En conclusión, el análisis de importancia de variables no solo permite entender mejor cómo funciona el modelo, sino también aporta nuevas evidencias sobre el potencial de los biomarcadores ecográficos avanzados para la detección y el estudio de la ELA.

4.8 INCORPORACIÓN DE VARIABLES

4.8.1 JUSTIFICACIÓN DE LA INCORPORACIÓN

Además de los biomarcadores ecográficos y los descriptores texturales, también se estudió la incorporación de variables clínicas para analizar si podían mejorar la capacidad predictiva de los modelos. Esta idea se basa en que el diagnóstico de la esclerosis lateral amiotrófica depende de muchos factores diferentes, por lo que combinar información de distintas fuentes puede ayudar a representar mejor la complejidad de la enfermedad [8].

Entre las variables clínicas incorporadas se incluyeron la edad (Age), el sexo (Sex), el peso (Weight), la altura (Height), el índice de masa corporal (*BMI*) y la presencia de fasciculaciones musculares observadas en los músculos estudiados [16,17]. Estas variables

se eligieron porque tienen relevancia clínica en la ELA y estaban disponibles en el *dataset* utilizado.

La incorporación de variables clínicas se basa en la idea de combinar diferentes tipos de información para obtener una visión más completa de la enfermedad, algo que se utiliza cada vez más en la Inteligencia Artificial aplicada a la medicina [39,51]. Este enfoque consiste en utilizar varias fuentes de datos que aportan información diferente pero complementaria. Mientras que los biomarcadores ecográficos permiten estudiar los cambios de la estructura y la textura del músculo, las variables clínicas aportan información sobre las características del paciente y los signos observados durante la exploración médica [16].

La edad es importante porque la ELA aparece con más frecuencia en adultos entre 50 y 70 años, por lo que puede aportar información útil sobre la probabilidad de desarrollar la enfermedad [4,12]. El sexo también es relevante, ya que la ELA es ligeramente más frecuente en hombres que en mujeres [12]. El *BMI* puede influir en los biomarcadores ecográficos porque el grosor muscular y la ecogenicidad cambian según la composición corporal de cada persona [16,18]. Por último, las fasciculaciones son uno de los signos más característicos de la afectación de motoneurona inferior en la ELA [11,25].

4.8.2 RESULTADOS DE LA INCORPORACIÓN DE VARIABLES CLÍNICAS

Los resultados mostraron que añadir variables clínicas mejoraba ligeramente el rendimiento global de algunos modelos, especialmente en *accuracy*, sensibilidad y estabilidad entre las distintas particiones de validación cruzada [46,51].

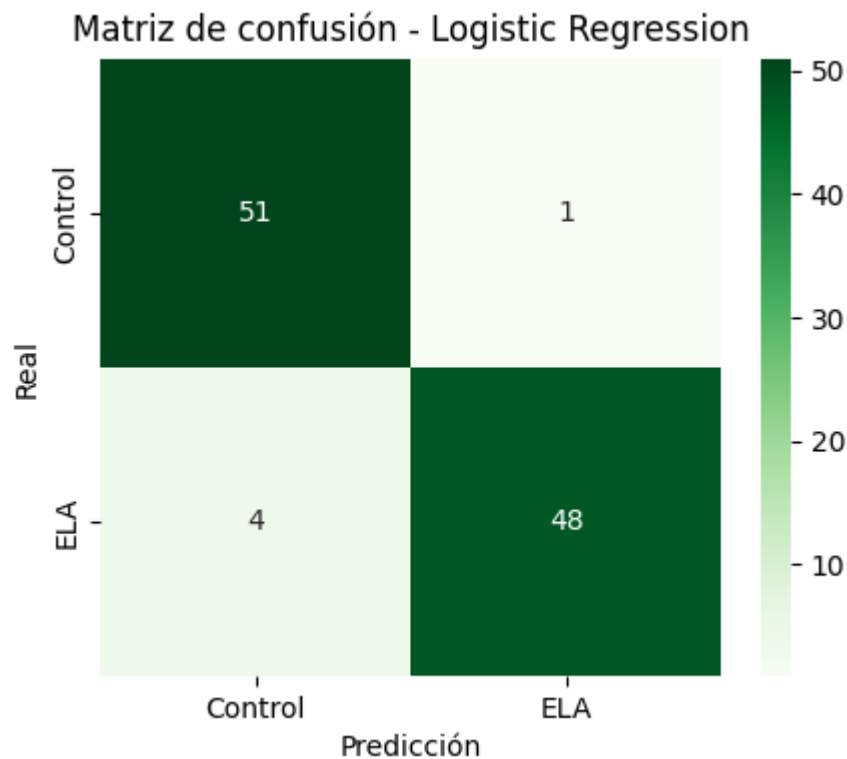


Figura 12. Matriz de confusión obtenida mediante Regresión Logística tras incorporación de variables clínicas y fasciculaciones

La incorporación de variables clínicas permitió alcanzar una *accuracy* global del 95%, mejorando el rendimiento obtenido utilizando únicamente biomarcadores ecográficos. La matriz de confusión de la Figura 12 mostró que el modelo clasificó correctamente 51 de los 52 registros de control y 48 de los 52 registros con ELA, produciéndose únicamente cinco errores de clasificación sobre un total de 104 muestras analizadas.

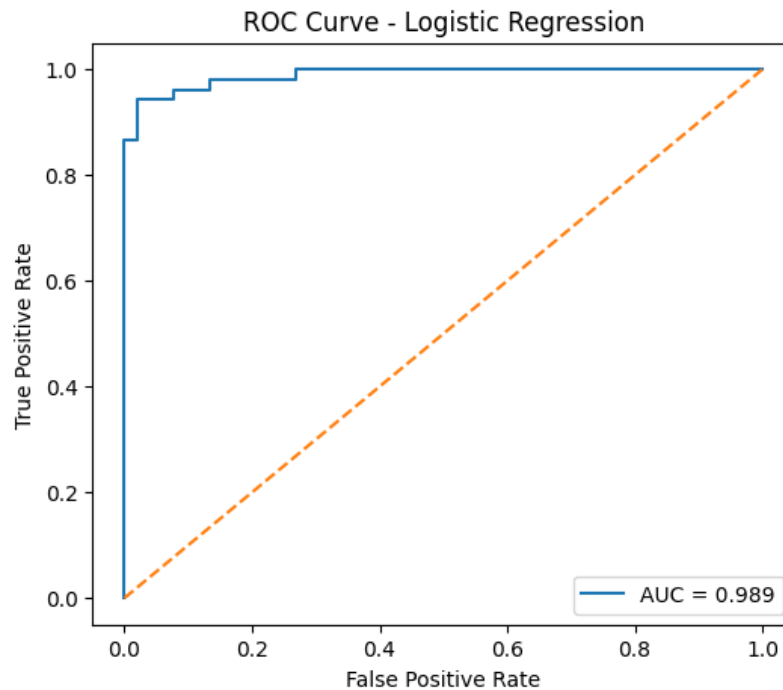


Figura 13. Curva ROC del modelo de Regresión Logística tras incorporar variables clínicas y fasciculaciones

En la Figura 13 se muestra la curva *ROC* que obtuvo un área bajo la curva (*AUC*) de 0.989 [45], lo que indica que el modelo tiene una capacidad muy alta para diferenciar entre pacientes con ELA y personas sanas. Este resultado sugiere que el modelo clasifica correctamente a ambos grupos en casi todos los casos analizados. Además, la mejora obtenida respecto al experimento en el que solo se utilizaron biomarcadores ecográficos muestra que añadir información clínica ayuda a mejorar la capacidad de clasificación del modelo [16,46] como se puede analizar en la Tabla 4.

Modelo	Accuracy	Recall	F1-score	AUC
Biomarcadores ecográficos	0.90	0.88	0.89	0.973
Biomarcadores + variables clínicas	0.95	0.92	0.95	0.989

Tabla 4. Comparación entre el modelo basado únicamente en biomarcadores ecográficos y el modelo que incorpora variables clínicas

4.8.3 IMPORTANCIA DE LAS VARIABLES CLÍNICAS

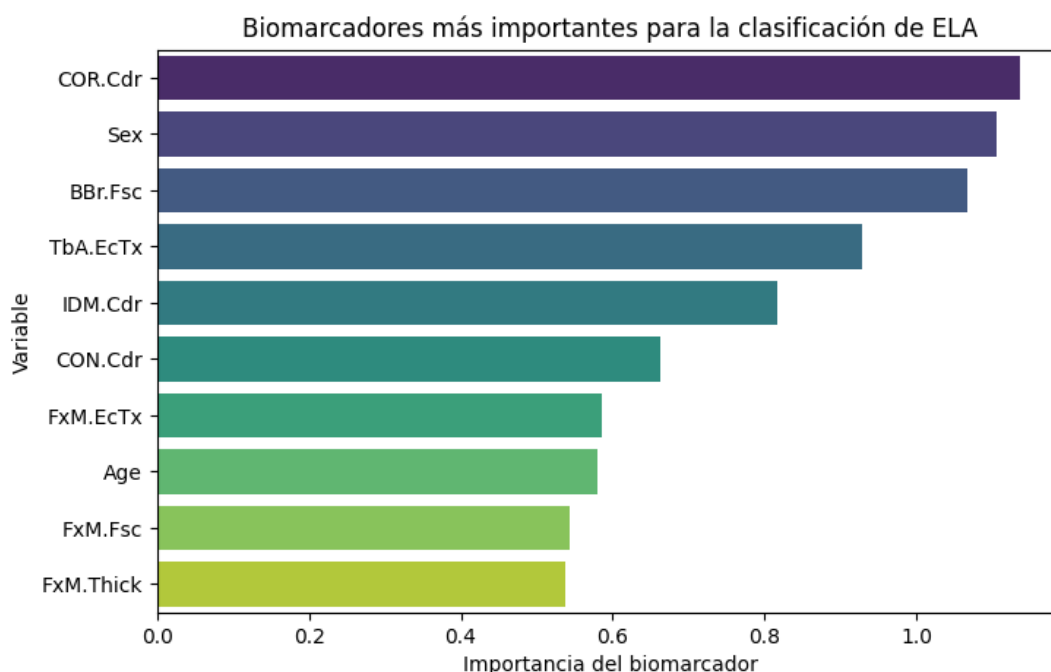


Figura 14. Variables clínicas con mayor importancia en el modelo de Regresión Logística tras incorporar biomarcadores ecográficos y variables clínicas

Con el objetivo de entender mejor cómo toma decisiones el modelo, se realizó un análisis de la importancia de las variables utilizando los coeficientes de la Regresión Logística. Como todas las variables se habían normalizado previamente mediante *StandardScaler*, el tamaño de estos coeficientes puede utilizarse para estimar cuánto influye cada variable en la clasificación final.

Los resultados de la Figura 14 mostraron que COR.Cdr siguió siendo la variable más importante para diferenciar entre pacientes con ELA y sanos [16,28]. Esto confirma que los descriptores de textura obtenidos mediante el análisis *GLCM* del músculo cuádriceps contienen información muy útil para el diagnóstico.

Sin embargo, al añadir la información clínica aparecieron nuevas variables entre las más relevantes del modelo. Destaca especialmente la variable sexo (Sex), que pasó a ser la segunda más importante [12]. Además, las fasciculaciones observadas en el bíceps braquial (BBr.Fsc) y en los flexores del antebrazo (FxM.Fsc) también se situaron entre las variables

con mayor influencia en la clasificación [11,25]. Por otro lado, la edad (Age) apareció igualmente entre las variables más relevantes, algo que coincide con lo que se conoce sobre la enfermedad, ya que la ELA es más frecuente en determinados grupos de edad [4,8,12].

Estos resultados indican que las variables clínicas aportan información adicional a la obtenida mediante la ecografía muscular. En lugar de reemplazar a los biomarcadores ecográficos, los complementan, proporcionando más información al modelo y ayudando a mejorar su capacidad para distinguir entre pacientes con ELA y personas sanas [16,51].

4.8.4 LIMITACIONES Y PERSPECTIVAS FUTURAS

Aunque los resultados fueron prometedores, añadir variables clínicas también presenta algunas limitaciones importantes [46]. El número de variables clínicas disponibles en el *dataset* era relativamente pequeño en comparación con la cantidad de biomarcadores ecográficos. Además, las variables clínicas disponibles eran, principalmente, información demográfica y la presencia de fasciculaciones, sin incluir otros parámetros funcionales o respiratorios que podrían aportar información relevante sobre la enfermedad [10,17].

Además, algunas variables clínicas pueden introducir sesgos si no se controlan correctamente durante el entrenamiento del modelo [44]. Por ejemplo, si en el *dataset* hay diferencias importantes de edad o proporción de hombres y mujeres entre el grupo ELA y el grupo control, el modelo podría aprender a utilizar esas diferencias demográficas como una forma indirecta de identificar la enfermedad. Esto podría causar problemas cuando se aplicase el modelo a otros *datasets* con distribuciones demográficas diferentes.

En futuros estudios sería importante trabajar con bases de datos más grandes e incluir variables clínicas recogidas a lo largo del tiempo. Por ejemplo, sería útil disponer de escalas funcionales completas como la ALSFRS-R en distintos momentos [10], medidas de fuerza muscular mediante la escala MRC [17] y parámetros respiratorios. Tener este tipo de información longitudinal permitiría desarrollar modelos mucho más potentes para seguir la progresión de la enfermedad, superando las limitaciones de los análisis transversales utilizados actualmente [17,46].

4.9 ANÁLISIS PRELIMINAR DE PROGRESIÓN DE LA ENFERMEDAD

4.9.1 PLANTEAMIENTO DEL PROBLEMA

Además del análisis principal para diferenciar pacientes con esclerosis lateral amiotrófica y sujetos sanos, también se realizó un estudio preliminar para analizar si los biomarcadores ecográficos podían servir para predecir la progresión de la enfermedad [17,46]. La idea principal era comprobar si estos biomarcadores no solo ayudan en el diagnóstico, sino también en estimar la gravedad de la enfermedad o la velocidad a la que progresa. Esto sería muy útil para el seguimiento clínico de los pacientes [7,17].

La progresión de la ELA es muy diferente entre pacientes, y esta variabilidad es uno de los principales problemas clínicos de la enfermedad [8,15]. Para estudiar este aspecto, se utilizó la escala ALSFRS como medida del estado funcional del paciente en el momento de la exploración ecográfica [10,17]. Se estableció un límite de 30 puntos para dividir a los pacientes en dos grupos: pacientes con ALSFRS mayor de 30 fueron considerados casos leves ($\text{Severity} = 0$) y pacientes con ALSFRS menor o igual a 30 fueron considerados casos avanzados ($\text{Severity} = 1$).

Este análisis se realizó únicamente con los 52 registros del grupo ELA, ya que la escala ALSFRS no se utiliza en sujetos sanos. Como en este caso el número de muestras era menor, se utilizó validación cruzada estratificada de 5 particiones en lugar de 10, para asegurar que cada partición tuviera suficientes pacientes de ambos grupos de severidad [44].

4.9.2 RESULTADOS PRELIMINARES

Los resultados obtenidos en este análisis fueron más modestos que en los experimentos de clasificación entre ELA y controles. Las *accuracies* obtenidas estuvieron aproximadamente entre el 70% y el 75%, dependiendo del modelo y de las variables utilizadas. Aunque estos resultados son inferiores a los obtenidos en el diagnóstico ELA/Control, esto no debe interpretarse como un mal resultado, sino como una muestra de que predecir la progresión de la enfermedad es un problema mucho más complejo [8,46].

La Regresión Logística utilizando biomarcadores clásicos y descriptores *GLCM* obtuvo el mejor resultado, con una *accuracy* aproximada de 0.73 [41]. Después se situaron el *SVM* con 0.71 [38] y el *Random Forest* con 0.70 [37]. Los modelos más complejos como el *Gradient Boosting*, mostraron más variabilidad entre las distintas particiones de validación cruzada, probablemente debido al tamaño reducido del *dataset* utilizado en este análisis [40,44].

El análisis de las puntuaciones ALSFRS mostró una gran variabilidad entre los pacientes, con valores aproximados entre 15 y 45 puntos. Esto coincide con la gran heterogeneidad clínica conocida de la ELA [8,10] y demuestra la dificultad de intentar predecir la progresión de la enfermedad utilizando únicamente una medición realizada en un momento concreto.

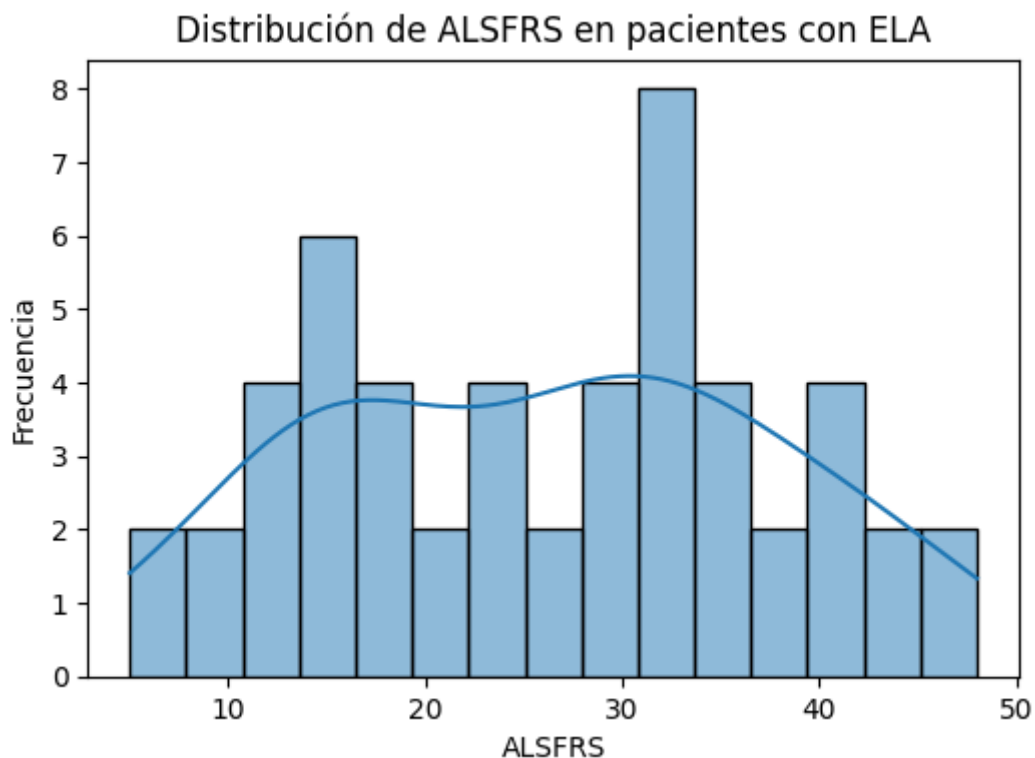


Figura 15. Distribución de las puntuaciones ALSFRS en los pacientes con ELA

La Figura 15 muestra la distribución de las puntuaciones ALSFRS observadas en los 52 registros con ELA. Se aprecia una amplia dispersión de valores, lo que refleja la heterogeneidad clínica característica de la enfermedad.

4.9.3 CORRELACIONES ENTRE BIOMARCADORES Y ESCALA ALSFRS

El análisis de correlaciones entre los biomarcadores ecográficos y la escala ALSFRS mostró que varios parámetros tenían asociaciones estadísticamente significativas con la gravedad funcional de los pacientes [16,17].

Los biomarcadores con mayor relación con la puntuación ALSFRS fueron grosor muscular del cuádriceps y la entropía *GLCM* del cuádriceps [16,17]. El grosor muscular mostró una correlación positiva: los pacientes con puntuaciones ALSFRS más altas, es decir, con mejor estado funcional, tendían a tener un mayor grosor muscular [20,21]. Por otro lado, la entropía *GLCM* mostró una correlación negativa: los pacientes con una enfermedad más avanzada presentaban un mayor desorden textural en la imagen ecográfica [17,28].

Estos resultados tienen sentido desde el punto de vista biológico. A medida que la ELA progresa, el músculo pierde masa muscular y se vuelve más irregular internamente debido a la degeneración, la infiltración grasa y la fibrosis [3,17,27]. Estos cambios se reflejan tanto en la reducción del grosor muscular como en el aumento de la entropía de la imagen ecográfica.

4.9.4 LIMITACIONES Y PERSPECTIVAS DEL ANÁLISIS DE PROGRESIÓN

Los resultados de este análisis deben considerarse un estudio inicial y exploratorio con varias limitaciones importantes. La principal limitación es que el *dataset* utilizado es transversal, es decir, cada paciente solo tiene una medición ecográfica. Esto impide estudiar directamente cómo cambian los biomarcadores a lo largo del tiempo. La progresión de una enfermedad neurodegenerativa es un proceso dinámico y, para analizarlo correctamente, serían necesarios datos longitudinales con múltiples mediciones por paciente [17,46].

Para obtener resultados más sólidos y útiles clínicamente, sería necesario disponer de estudios longitudinales similares al trabajo de seguimiento de Martínez-Payá et al. (2018) [17], donde se realizaron mediciones repetidas durante 20 semanas. Con este tipo de información sería posible desarrollar modelos capaces no solo de estimar la gravedad actual

de la enfermedad, sino también de predecir la velocidad de progresión y la evolución futura del paciente [7,17].

A pesar de estas limitaciones, este análisis preliminar confirma que existe relación real entre los biomarcadores ecográficos y la gravedad clínica de la ELA [16,17]. Esto apoya la idea de continuar investigando en esta línea utilizando *datasets* más amplios y estudios longitudinales en el futuro.

4.10 DISCUSIÓN DE RESULTADOS

4.10.1 SÍNTESIS DE LOS HALLAZGOS PRINCIPALES

Los resultados obtenidos en este trabajo muestran de forma bastante clara que las técnicas de *Machine Learning* aplicadas a biomarcadores ecográficos musculares tienen un gran potencial para clasificar pacientes con esclerosis lateral amiotrófica [16,36,39]. Los modelos desarrollados alcanzaron *accuracies* superiores al 88% en los mejores casos y valores de *AUC* cercanos a 0.97-0.98, lo que indica una capacidad muy alta para diferenciar correctamente entre pacientes con ELA y sujetos sanos.

Uno de los resultados más importantes fue comprobar que combinar diferentes tipos de biomarcadores mejora de forma constante el rendimiento de los modelos en comparación con utilizar solo un tipo de variable [16,28]. Esto tiene una aplicación práctica importante: un sistema diagnóstico basado en ecografía muscular debería combinar tanto biomarcadores clásicos, como el grosor muscular y la ecogenicidad, como descriptores texturales avanzados, para conseguir la mejor capacidad diagnóstica posible [16,32].

Otro hallazgo relevante fue confirmar la gran importancia del músculo cuádriceps femoral como biomarcador de referencia. Utilizando únicamente las variables ecográficas de este músculo se obtuvo un *AUC* de 0.977. Este resultado coincide con estudios previos de Martínez-Payá et al. [16] y refuerza la idea de que el cuádriceps es el músculo más útil para el diagnóstico ecográfico de la ELA dentro del protocolo de cuatro músculos utilizado en este trabajo.

4.10.2 RELEVANCIA CLÍNICA DE LOS RESULTADOS

Los resultados obtenidos tienen una relevancia clínica importante desde varios puntos de vista.

Desde la perspectiva diagnóstica, un sistema con una *accuracy* del 90% y un *AUC* de 0.973 podría ser una herramienta de apoyo al diagnóstico muy útil, especialmente en fases iniciales de la enfermedad, donde la electromiografía puede dar resultados poco concluyentes y el diagnóstico diferencial resulta más complicado [7,11,16].

Además, desde el punto de vista de la accesibilidad, es importante destacar que estos resultados se obtuvieron utilizando ecografía muscular, una técnica relativamente barata, rápida y disponible en muchos hospitales [18,22]. Esto supone una ventaja importante frente a otras herramientas diagnósticas más precisas, pero también más costosas e invasivas.

Si estos resultados se confirman en estudios multicéntricos con un mayor número de pacientes, el sistema podría utilizarse como herramienta de cribado o apoyo al diagnóstico en consultas de neurología de distintos niveles de especialización [7,16,46].

4.10.3 LIMITACIONES IDENTIFICADAS

Los resultados del trabajo deben interpretarse teniendo en cuenta varias limitaciones importantes.

La principal limitación es el tamaño relativamente pequeño del *dataset*, formado por 104 registros [12,46]. Aunque este tamaño es adecuado para un Trabajo Fin de Grado utilizando validación cruzada [44], no es suficiente para asegurar completamente que los modelos funcionen igual de bien en toda la diversidad clínica real de la ELA en la población general. Para confirmar estos resultados serían necesarios estudios multicéntricos con cientos o incluso miles de pacientes.

Otra limitación importante es la falta de validación externa [16,17]. Esto significa que los modelos todavía no se han probado utilizando otro *dataset* completamente independiente

procedente de otro hospital y obtenido con diferentes equipos de ecografía. Esta validación es imprescindible antes de pensar en una aplicación clínica real.

Por último, el *dataset* utilizado es transversal, es decir, cada paciente solo tiene una medición en un momento concreto. Esto limita mucho el estudio de la progresión de la enfermedad, que es un aspecto clínicamente muy importante y que requiere datos longitudinales recogidos a lo largo del tiempo [17,46].

4.10.4 PERSPECTIVAS DE TRABAJO FUTURO

Los resultados obtenidos abren varias líneas interesantes para futuros trabajos.

Una de las más importantes es incorporar técnicas de *Deep Learning*, especialmente redes neuronales convolucionales (*CNN*), aplicadas directamente a las imágenes ecográficas sin necesidad de calcular manualmente los descriptores texturales [39,50]. Este sería el siguiente paso lógico en esta línea de investigación.

También será muy importante realizar validaciones multicéntricas utilizando *datasets* más grandes para comprobar que los modelos mantienen un buen rendimiento en distintos hospitales y condiciones clínicas [16,46].

Además, disponer de datos longitudinales con varias mediciones por paciente permitiría estudiar de forma mucho más precisa la progresión de la enfermedad y desarrollar modelos capaces de predecir la evolución futura de cada paciente [17,44].

Por último, una de las líneas más prometedoras es combinar biomarcadores de diferentes tipos (ecográficos, clínicos, genéticos y neurofisiológicos) en modelos multimodales [39,51]. Esta integración podría mejorar todavía más la capacidad diagnóstica y pronóstica de los sistemas basados en Inteligencia Artificial para la ELA.

Capítulo 5. CONCLUSIONES

5.1 SÍNTESIS DEL TRABAJO REALIZADO

En este Trabajo Fin de Grado se ha realizado un estudio centrado en el uso de técnicas de *Machine Learning* para analizar biomarcadores ecográficos relacionados con la esclerosis lateral amiotrófica [16,36]. El objetivo principal fue evaluar si distintos modelos de Inteligencia Artificial eran capaces de diferenciar pacientes con ELA de sujetos sanos utilizando información obtenida mediante ecografía muscular cuantitativa y variables clínicas complementarias [16,17].

La principal motivación de este trabajo surge de la necesidad de desarrollar herramientas diagnósticas objetivas, precisas y no invasivas para una enfermedad neurodegenerativa tan grave y compleja como la ELA [7,8]. Actualmente, el diagnóstico sigue siendo principalmente clínico y suele requerir un largo proceso de seguimiento y descarte de otras enfermedades neurológicas [1,7]. Esto provoca retrasos diagnósticos de entre 10 y 18 meses desde la aparición de los primeros síntomas [2,5]. Este retraso afecta directamente al bienestar del paciente, a la organización de los cuidados y al tiempo disponible para iniciar tratamientos que puedan ayudar a frenar la enfermedad [9].

En este contexto, la ecografía muscular cuantitativa aparece como una alternativa muy interesante frente a otras técnicas diagnósticas más caras o invasivas [18,22]. Su bajo coste, facilidad de acceso, ausencia de radiación y posibilidad de repetirse con frecuencia la convierte en una herramienta muy útil para el seguimiento de pacientes con enfermedades neurodegenerativas [17,22]. Sin embargo, para aprovechar realmente toda la información de las imágenes ecográficas, es necesario ir más allá de la simple observación visual y utilizar métodos computacionales que permitan medir de forma objetiva los cambios estructurales del músculo [16,24].

En este punto es donde las técnicas de Inteligencia Artificial y aprendizaje automático aportan un valor muy importante. Durante los últimos años, los algoritmos de *Machine Learning* han demostrado un gran potencial en medicina, especialmente en tareas relacionadas con el diagnóstico de enfermedades, el análisis de imágenes médicas, la extracción automática de biomarcadores y la predicción clínica [36,39]. Su capacidad para combinar muchas variables diferentes y detectar patrones complejos y no lineales los hace especialmente adecuados para estudiar enfermedades tan complejas biológicamente como la ELA [51].

A lo largo del trabajo se desarrolló una metodología completa y reproducible que incluyó todas las etapas necesarias para crear un sistema de clasificación basado en *Machine Learning* [36]. El proceso comenzó con la carga y exploración inicial del *dataset* y terminó con la evaluación final de los modelos. Concretamente, el *pipeline* incluyó el preprocesamiento y normalización de datos [36], el análisis de correlaciones entre variables y la detección de posibles problemas de *data leakage* [44], la creación y comparación de distintos conjuntos de variables, el entrenamiento de varios modelos de clasificación mediante *pipelines* estructurados, la validación cruzada estratificada de 10 particiones para asegurar la robustez de los resultados [44], la evaluación mediante diferentes métricas diagnósticas [45] y el análisis de la importancia de las variables más relevantes [48,49].

Los siete modelos de *Machine Learning* analizados (Regresión Logística, *SVM*, *Random Forest*, *KNN*, Árbol de Decisión, *Gradient Boosting* y *Naive Bayes*) fueron seleccionados para representar distintos tipos de algoritmos, desde modelos lineales simples hasta métodos más complejos [36,37,38,40,41,43]. Esta variedad permitió obtener conclusiones más sólidas sobre el tipo de relación existente entre los biomarcadores ecográficos y el diagnóstico de la ELA, así como sobre cuáles son los modelos más adecuados para este tipo de problema.

También se compararon distintos conjuntos de variables, incluyendo biomarcadores ecográficos clásicos, descriptores texturales *GLCM* [16,28], variables clínicas y combinaciones de todas ellas [46]. Esta comparación permitió medir la contribución

específica de cada tipo de información al rendimiento diagnóstico del sistema y aporta información útil para futuros protocolos clínicos.

Además, el trabajo incluyó una revisión extensa y actualizada del estado del arte sobre la ELA y sus biomarcadores, la ecografía muscular cuantitativa, las técnicas de análisis textural, la Inteligencia Artificial en imagen médica y el *Machine Learning* aplicado a enfermedades neurodegenerativas [7,8,16,17,36]. Esta revisión bibliográfica, basada en publicaciones recientes e importantes de la literatura científica internacional, permitió contextualizar correctamente los resultados obtenidos y compararlos con investigaciones previas, especialmente con los trabajos de referencia de Martínez-Payá et al. (2017, 2018) [16,17], que utilizan el mismo *dataset* y representan el punto de comparación más directo para este trabajo.

En conjunto, este trabajo integra de forma coherente conocimientos de ingeniería biomédica, preprocesamiento digital de imagen, Inteligencia Artificial, análisis estadístico y neurología clínica [16,36,39]. Esta combinación de disciplinas no solo aporta valor académico al trabajo, sino que también refleja cómo funciona actualmente la investigación biomédica moderna, donde muchos de los avances más importantes surgen precisamente de la colaboración entre diferentes áreas de conocimiento.

5.2 PRINCIPALES RESULTADOS OBTENIDOS

Los resultados obtenidos a lo largo del estudio permiten sacar varias conclusiones importantes tanto desde el punto de vista técnico como clínico.

El *Machine Learning* permite clasificar pacientes con esclerosis lateral amiotrófica con un rendimiento elevado utilizando biomarcadores ecográficos [16,36]. Después de aplicar las técnicas de selección de variables y control del sobreajuste explicados en la metodología [44], los modelos alcanzaron resultados realistas y coherentes con los descritos en la literatura científica. El mejor modelo fue la Regresión Logística entrenada con el conjunto completo de variables [41,42], que obtuvo una *accuracy* del 90% y un *AUC* de 0.973. Esto

indica una excelente capacidad para diferenciar pacientes con ELA de sujetos sanos. Estos resultados son comparables a los obtenidos por Martínez-Payá et al. (2017) [16] usando métodos estadísticos clásicos, lo que demuestra que los modelos de *Machine Learning* pueden reproducir y reforzar resultados previos utilizando una metodología de validación más rigurosa.

La Regresión Logística fue el modelo más estable y equilibrado entre todos los algoritmos analizados [41,42]. Este resultado es importante porque indica que gran parte de la diferencia entre pacientes con ELA y controles pueden explicarse mediante relaciones relativamente lineales entre los biomarcadores musculares, sin necesidad de utilizar modelos extremadamente complejos [16]. Además, la Regresión Logística tiene la ventaja de ser fácil de interpretar, ya que permite analizar cuánto influye cada variable en la clasificación mediante sus coeficientes [41,48]. Esto la convierte en uno de los modelos más adecuados para una posible aplicación clínica futura.

La combinación de distintos tipos de biomarcadores mejoró de forma constante el rendimiento diagnóstico [16,17]. Los biomarcadores ecográficos clásicos y los descriptores texturales *GLCM* aportaron información complementaria sobre el estado del músculo. Al combinar ambos tipos de variables se obtiene una descripción más completa y útil del proceso neurodegenerativo que utilizando cada grupo de biomarcadores por separado. Además, los descriptores texturales demostraron ser especialmente sensibles para detectar alteraciones microestructurales del músculo que pueden aparecer en fases tempranas de la enfermedad, incluso antes de que el grosor muscular o la ecogenicidad presenten cambios evidentes [16,32].

El músculo cuádriceps femoral destacó como el biomarcador muscular más informativo [16]. El análisis específico de este músculo alcanzó una *accuracy* del 91% y un *AUC* de 0.977 utilizando únicamente sus parámetros ecográficos. Este resultado confirma y amplía los hallazgos de estudios previos [16,17]. Además, tiene una implicación clínica importante: en situaciones donde el tiempo de exploración sea limitado, estudiar solo el cuádriceps

podría aportar gran parte de la información diagnóstica obtenida con el protocolo completo de cuatro músculos.

La incorporación de variables clínicas también mejoró el rendimiento de los modelos [46,51]. Combinar biomarcadores ecográficos con variables como la edad, el sexo, el índice de masa corporal y la presencia de fasciculaciones musculares [11,12,25] permitió alcanzar *accuracies* cercanas al 95% en algunas configuraciones experimentales. Esto apoya la idea de que integrar información de diferentes tipos es una estrategia muy útil para aumentar la capacidad diagnóstica del sistema [39,51].

La detección y corrección del sobreajuste fue uno de los aspectos metodológicos más importantes del trabajo [44]. En los primeros experimentos aparecieron *accuracies* artificialmente cercanas al 100%, debido a correlaciones excesivas entre variables y posibles problemas de *data leakage*. La identificación y corrección de estos problemas, mediante el análisis detallado de correlaciones, la eliminación de variables problemáticas y el uso correcto de *pipelines* con validación cruzada [36,44], fue fundamental para conseguir modelos con capacidad predictiva real. Este proceso muestra claramente uno de los principales riesgos en el desarrollo de modelos biomédicos de *Machine Learning*: resultados aparentemente excelentes pueden deberse en realidad a errores metodológicos difíciles de detectar [44,46].

El análisis preliminar de progresión de la enfermedad mostró resultados más modestos, aunque prometedores. Los modelos entrenados para predecir la gravedad clínica según la escala ALSFRS [10,17] alcanzaron *accuracies* aproximadas del 70-75%. Esto refleja la dificultad de predecir la progresión, aunque los resultados demuestran que existe una relación real entre la información ecográfica y el estado funcional del paciente [17]. Esto abre una línea de investigación interesante para futuros estudios longitudinales.

5.3 INTERPRETACIÓN CLÍNICA DE LOS RESULTADOS

Uno de los aspectos más importantes de este trabajo ha sido la interpretación clínica de los resultados obtenidos. Esto significa que no solo se analizaron las métricas numéricas de rendimiento de los modelos, sino también cómo esos resultados se relacionan con la fisiopatología conocida de la esclerosis lateral amiotrófica y con su posible utilidad en la práctica clínica real [16,48].

La ELA es una enfermedad muy heterogénea tanto desde el punto de vista clínico como fisiopatológico [7,8]. Los pacientes pueden tener velocidades de progresión muy diferentes, distintos patrones de afectación muscular, diferentes zonas de inicio de la enfermedad y distintos niveles de deterioro funcional [8,15]. Esta variabilidad es uno de los principales problemas a la hora de desarrollar biomarcadores universales [7,46]. Por eso, los resultados obtenidos son especialmente relevantes: el hecho de que los modelos fueran capaces de clasificar correctamente alrededor del 90% de las muestras, a pesar de toda esta variabilidad, sugiere que los biomarcadores ecográficos están captando señales biológicas importantes y comunes del proceso neurodegenerativo, más allá de las diferencias individuales entre pacientes [16,17].

Desde el punto de vista fisiopatológico, los resultados son coherentes con los mecanismos conocidos de degeneración muscular en la ELA [3,27]. La pérdida progresiva de motoneuronas provoca atrofia muscular, infiltración grasa y fibrosis, cambios que modifican la estructura interna del músculo y que pueden detectarse mediante biomarcadores ecográficos [16,27]. La disminución del grosor muscular refleja la pérdida de masa muscular causada por la denervación [18,20]. El aumento de la ecogenicidad indica la sustitución de fibras musculares por tejido adiposo y conectivo [27]. Además, los cambios observados en los descriptores texturales *GLCM*, como el aumento del contraste y la entropía y la disminución de la homogeneidad y la energía, reflejan la mayor irregularidad y desorganización de la microestructura del músculo generada [16,17,28]. El hecho de que precisamente estos descriptores hayan sido los más importantes en el análisis de variables refuerza la validez biológica del enfoque utilizado.

La importancia especial del músculo cuádriceps femoral también puede explicarse desde una perspectiva fisiopatológica [16]. Este músculo tiene una gran masa muscular y una alta densidad de unidades motoras, por lo que es especialmente sensible a la pérdida de inervación [16,17]. Como consecuencia, los cambios degenerativos pueden detectarse relativamente pronto mediante ecografía en comparación con músculos más pequeños. Además, su estructura normal, formada por fibras largas y organizadas, hace que las alteraciones provocadas por la enfermedad sean especialmente visibles en los descriptores texturales *GLCM* [16,28].

La identificación de biomarcadores musculares capaces de diferenciar pacientes con alta precisión tiene implicaciones clínicas importantes. En primer lugar, sugiere que la ecografía muscular cuantitativa podría incorporarse al proceso diagnóstico de la ELA como una herramienta complementaria a la electromiografía y a la evaluación clínica [11,16], proporcionando información objetiva y reproducible que ayude a reducir el retraso diagnóstico [2,5]. En segundo lugar, la capacidad de estos biomarcadores para reflejar la gravedad clínica abre la posibilidad de utilizar la ecografía muscular para monitorizar la progresión de la enfermedad [17]. Esto podría facilitar el seguimiento de los pacientes y ayudar a evaluar la respuesta a nuevos tratamientos en ensayos clínicos [7,9].

Otro aspecto clínicamente importante es que los modelos interpretables, como la Regresión Logística, permiten saber exactamente qué variables influyen en las predicciones y cómo lo hacen [41,48]. Esto es fundamental en medicina, donde no solo importa que un modelo sea preciso, sino también que el médico pueda entender por qué el sistema propone un determinado diagnóstico [48,49]. Esta capacidad de explicación permite que el especialista pueda valorar críticamente las predicciones e integrarlas de forma responsable en su práctica clínica. Por este motivo, los llamados modelos de “caja negra”, aunque pueden conseguir precisiones altas, tienen más limitaciones para su uso clínico [48,49].

Los resultados obtenidos también tienen implicaciones para diseñar protocolos de exploración ecográfica más eficientes [16]. El análisis de importancia de variables permitió identificar qué músculos y qué parámetros aportaban más información diagnóstica [16,48].

Esto podría servir en el futuro para crear protocolos más rápidos y eficientes, centrados en las regiones y biomarcadores más útiles, reduciendo tanto el tiempo de exploración como el esfuerzo computacional sin perder capacidad diagnóstica.

Por último, es importante destacar que los modelos desarrollados en este trabajo no pretenden sustituir el criterio clínico del especialista. Su función es actuar como herramientas de apoyo que complementen la evaluación clínica tradicional aportando información cuantitativa y objetiva [39,48]. El diagnóstico de la ELA sigue siendo un proceso complejo que combina información clínica, neurofisiológica y de imagen [1,7,11]. En este contexto, el sistema desarrollado en este trabajo serviría como una fuente adicional de información que ayude al especialista en la toma de decisiones clínicas.

5.4 LIMITACIONES DEL ESTUDIO

Aunque los resultados obtenidos en este trabajo son prometedores, existen varias limitaciones importantes que deben tenerse en cuenta al interpretar las conclusiones y valorar su posible aplicación clínica.

La principal limitación del estudio es el tamaño reducido de la muestra. El *dataset* utilizado incluye un total de 52 sujetos, de los cuales 26 son pacientes con ELA y 26 controles sanos, con un total de 104 registros debido a las mediciones bilaterales. Para desarrollar y validar modelos de *Machine Learning*, este tamaño sigue siendo relativamente pequeño [44,46]. La ELA es una enfermedad rara, con una prevalencia aproximada de 5-7 casos por cada 100.000 habitantes [5,12], lo que dificulta mucho conseguir grupos grandes de pacientes.

El tamaño reducido de la muestra tiene varias consecuencias metodológicas importantes [44]. En primer lugar, aumenta el riesgo de sobreajuste de los modelos. Además, provoca una mayor variabilidad estadística en las métricas obtenidas mediante validación cruzada y limita la capacidad de los modelos para aprender toda la diversidad clínica de la enfermedad [46]. Aunque el uso de validación cruzada estratificada de 10 particiones ayuda a reducir este problema [44], no lo elimina completamente.

Otra limitación importante es que el análisis se realizó utilizando variables extraídas de las imágenes ecográficas, pero sin acceso directo a las imágenes originales. Aunque trabajar con biomarcadores cuantitativos calculados a partir de las imágenes es una metodología válida y coherente con el estado del arte [16,28], disponer de las imágenes originales habría permitido explorar enfoques más avanzados. Entre ellos destacan las técnicas de *Deep Learning* basadas en redes neuronales convolucionales (*CNN*), capaces de aprender automáticamente patrones complejos directamente a partir de los píxeles de las imágenes [39,50]. En muchos estudios de imagen médica, estos métodos han demostrado obtener mejores resultados que las técnicas basadas en variables calculadas manualmente, por lo que representan una línea de investigación futura muy interesante para la ELA.

El diseño del estudio también es principalmente transversal. La mayor parte del análisis se basa en una única medición ecográfica de cada paciente, sin seguimiento a lo largo del tiempo. Sin embargo, la progresión de la ELA es un proceso dinámico que evoluciona durante meses o años [17]. Para estudiar correctamente cómo cambian los biomarcadores con el tiempo serían necesarios estudios longitudinales con múltiples mediciones realizadas en distintos momentos [17,46]. Los estudios transversales permiten comparar grupos en un instante concreto, pero no permiten analizar directamente la evolución de la enfermedad ni desarrollar modelos capaces de predecir cómo progresará cada paciente individualmente.

Otra limitación importante es la ausencia de validación externa multicéntrica [16,17]. Los modelos desarrollados se entrenaron y evaluaron utilizando únicamente datos procedentes del Centro Universitario San Rafael de la Universidad Pontificia Comillas. Esto significa que todavía no se ha comprobado si los modelos funcionarían igual de bien en pacientes de otros hospitales, utilizando diferentes equipos de ecografía, distintos protocolos de adquisición o poblaciones con características demográficas diferentes.

Esta validación externa es imprescindible antes de pensar en una aplicación clínica real del sistema. Además, parámetros como la ecogenicidad y los descriptores texturales pueden variar bastante según el equipo de ecografía utilizado y la configuración de adquisición

[16,18], lo que puede afectar a la capacidad de generalización de los modelos entrenados con datos de un único centro.

También hay que tener en cuenta que el número de variables clínicas disponibles era relativamente limitado. Aunque incluir variables como la edad, el sexo, el índice de masa corporal (*BMI*) y la presencia de fasciculaciones musculares [11,12,25] mejoró el rendimiento de los modelos, disponer de más información clínica habría permitido estudiar con mayor profundidad el potencial de los modelos multimodales. Por ejemplo, habría sido útil contar con escalas funcionales completas [10], medidas de fuerza muscular segmentaria [17], parámetros respiratorios o información sobre el tiempo de evolución desde el inicio de los síntomas [2,5]. Además, algunos pacientes tenían valores ausentes en ciertas variables clínicas, lo que redujo el tamaño efectivo del *dataset* en algunos experimentos.

Por último, la interpretación fisiopatológica completa de algunos biomarcadores todavía necesita más investigación [3,16]. Aunque los resultados obtenidos son coherentes con los mecanismos conocidos de la ELA, la interpretación exacta de algunos descriptores texturales en relación con procesos biológicos concretos (como la denervación, la fibrosis o la infiltración de grasa) no siempre es directa [28,32]. Para entender mejor estas relaciones serían necesarios estudios adicionales utilizando técnicas de imagen de mayor resolución, como la resonancia magnética muscular, o estudios histológicos mediante biopsias musculares, que no estaban disponibles en este trabajo.

5.5 LÍNEAS FUTURAS DE TRABAJO

A partir de los resultados obtenidos y de las limitaciones encontradas de este trabajo, se plantean varias líneas de investigación futura que podrían ampliar y reforzar los hallazgos alcanzados.

En primer lugar, una de las necesidades más importantes es aumentar el tamaño de las bases de datos mediante estudios multicéntricos e internacionales [12,46]. Para ello sería necesario combinar datos de distintos hospitales y centros de investigación utilizando protocolos de

adquisición estandarizados [16,34]. Este tipo de colaboración permitiría entrenar modelos más robustos, reducir el riesgo de sobreajuste, mejorar la capacidad de generalización y validar los biomarcadores en muestras más representativas de la diversidad clínica real de la esclerosis lateral amiotrófica. Redes multicéntricas como *European Network to Cure ALS* o *ProAct* podrían servir como plataformas para desarrollar este tipo de proyectos colaborativos [44,46].

Además, disponer de las imágenes ecográficas originales permitiría aplicar técnicas de *Deep Learning* directamente sobre las imágenes. El siguiente paso lógico sería utilizar redes neuronales convolucionales (*CNN*) para analizar automáticamente las ecografías musculares [39,50]. Estas redes pueden aprender patrones complejos directamente a partir de los píxeles de la imagen, sin necesidad de calcular manualmente descriptores texturales. Esto podría permitir detectar información diagnóstica adicional que los descriptores *GLCM* tradicionales no consiguen captar [28,50]. Modelos como *ResNet*, *EfficientNet* o *Vision Transformers*, previamente entrenados mediante *transfer learning* para ecografías musculares de ELA, representan opciones especialmente prometedoras para futuras investigaciones.

También sería muy importante desarrollar estudios longitudinales con varias mediciones realizadas al mismo paciente a lo largo del tiempo [17,46]. Esto permitiría crear modelos capaces de predecir la velocidad de progresión de la enfermedad, estimar cuándo podrían aparecer eventos clínicos importantes (como la necesidad de ventilación no invasiva o la pérdida de la capacidad de caminar) y evaluar la respuesta a tratamientos [9,17]. Para este tipo de análisis podrían utilizarse modelos de series temporales, como redes neuronales recurrentes (*LSTM*) o modelos de supervivencia basados en *Machine Learning*. Además, disponer de datos longitudinales permitiría estudiar mucho mejor la progresión de la enfermedad, algo que en este trabajo solo pudo analizarse de forma preliminar y transversal [17].

Por otro lado, una de las líneas más prometedoras consiste en combinar biomarcadores ecográficos con otras fuentes de información, como resonancia magnética muscular, PET, electromiografía cuantitativa (*EMG*), biomarcadores moleculares (como neurofilamentos en

plasma y líquido cefalorraquídeo) [7], datos genéticos [3,14] y escalas clínicas funcionales [10]. Los modelos multimodales, que integran diferentes tipos de datos, han demostrado obtener mejores resultados que los modelos basados en una única fuente de información en muchas enfermedades neurodegenerativas [39,51], y probablemente también podrían mejorar el diagnóstico y pronóstico de la ELA. El desarrollo de sistemas capaces de combinar correctamente información tan diferente es actualmente uno de los grandes retos dentro de la Inteligencia Artificial biomédica.

Asimismo, desde el punto de vista clínico, uno de los objetivos más ambiciosos sería desarrollar software capaz de analizar automáticamente las imágenes ecográficas e integrar modelos de clasificación directamente en los equipos de ecografía utilizados en hospitales y consultas [16,39]. Este tipo de sistema permitiría que el proceso de exploración fuese prácticamente automático. Al finalizar la ecografía, el médico podría recibir en tiempo real una evaluación cuantitativa del estado muscular y una estimación objetiva de la probabilidad de ELA. Para conseguirlo sería necesario desarrollar interfaces fáciles de usar, integrar el sistema con las historias clínicas electrónicas y realizar estudios de validación en condiciones reales de uso clínico [48,49].

Además, sería interesante aplicar técnicas avanzadas de interpretabilidad, como *SHAP* (*SHapley Additive exPlanations*) o *LIME* (*Local Interpretable Model-agnostic Explanations*), a los modelos desarrollados [48,49]. Estas técnicas permitirían no solo identificar qué biomarcadores son más importantes en general, sino también explicar individualmente por qué el modelo toma una determinada decisión para cada paciente concreto. Este tipo de explicaciones son especialmente útiles en medicina porque ayudan al médico a entender qué variables influyeron más en el diagnóstico sugerido por el sistema y permiten valorar críticamente esa recomendación utilizando su experiencia clínica [48,49].

Por último, aunque este trabajo se ha centrado únicamente en la ELA, la metodología desarrollada podría adaptarse fácilmente al estudio de otras enfermedades neuromusculares, como la atrofia muscular espinal (AME), neuropatías periféricas, miopatías inflamatorias o esclerosis múltiple [22,36]. Reutilizar el *pipeline* de análisis y los modelos de *Machine*

Learning en diferentes enfermedades aumentaría mucho el impacto del trabajo realizado y podría contribuir al desarrollo de herramientas diagnósticas más generales y útiles para distintas patologías neuromusculares.

5.6 CONCLUSIÓN FINAL

En conclusión, este Trabajo de Fin de Grado demuestra de forma clara que las técnicas de *Machine Learning* aplicadas a biomarcadores ecográficos musculares pueden ser una herramienta muy prometedora y útil para clasificar pacientes con esclerosis lateral amiotrófica [16,36,39].

Los resultados obtenidos muestran que los modelos desarrollados son capaces de detectar patrones relacionados con la enfermedad con niveles altos de precisión, alcanzando *accuracies* cercanas al 90% y valores de *AUC* próximos a 0.97-0.98. Estos resultados son similares a los mejores publicados en la literatura científica [16,17] y, además, se han obtenido utilizando una metodología de validación más rigurosa que la empleada en muchos estudios anteriores [44].

El trabajo también demuestra la importancia de combinar distintos tipos de información, como biomarcadores ecográficos clásicos, análisis textural avanzado [16,28], variables clínicas [46] e Inteligencia Artificial [36,39], para construir modelos diagnósticos más completos y robustos que aquellos basados en una sola fuente de información. Aunque en este estudio solo se pudieron utilizar las variables disponibles en el *dataset*, este enfoque apunta hacia el desarrollo de una medicina de precisión basada en datos, donde la integración de diferentes tipos de información permite una caracterización mucho más completa e individualizada de la enfermedad [51].

Más allá del rendimiento técnico de los modelos, uno de los aspectos más importantes de este trabajo ha sido la interpretación clínica y biológica de los resultados [3,16,48]. Esto ha permitido relacionar los hallazgos obtenidos mediante *Machine Learning* con los procesos neurodegenerativos de la enfermedad y con las necesidades reales de pacientes y

profesionales sanitarios. Además, la identificación de biomarcadores musculares especialmente relevantes, como algunos descriptores texturales del músculo cuádriceps femoral [16,28], proporciona una base sólida para futuros estudios y para el desarrollo de protocolos de exploración clínica más eficientes.

Aunque todavía existen limitaciones importantes que deberían abordarse en futuros trabajos, especialmente relacionados con el tamaño reducido de la muestra [12,46], la falta de validación multicéntrica [16,17] y la ausencia de estudios longitudinales [17], los resultados obtenidos refuerzan el gran potencial de la Inteligencia Artificial aplicada a enfermedades neurodegenerativas [39,51] y contribuyen a un área de investigación con aplicaciones clínicas muy relevantes.

En conjunto, este trabajo contribuye al desarrollo de herramientas cuantitativas, objetivas y accesibles que en el futuro podrían ayudar a mejorar el diagnóstico precoz de la ELA, reduciendo el retraso diagnóstico y sus consecuencias para los pacientes [2,5,7]. También podrían facilitar una mejor monitorización de la progresión de la enfermedad mediante biomarcadores objetivos y reproducibles [17], mejorar el diseño de ensayos clínicos gracias a una mejor clasificación de pacientes según sus biomarcadores [7,9], favorecer el desarrollo de tratamientos personalizados adaptados al perfil de cada persona [51] y avanzar hacia una medicina más precisa, eficiente y centrada en las necesidades de los pacientes con ELA.

En definitiva, la combinación entre ecografía muscular e Inteligencia Artificial estudiada en este trabajo representa una de las líneas de investigación más prometedoras dentro del estudio de las enfermedades neurodegenerativas [16,17,39]. Los resultados obtenidos suponen una aportación concreta y rigurosa a este campo y podrían servir de ayuda para futuras investigaciones y, en última instancia, para mejorar la atención de los pacientes con ELA y de sus familias.

Por último, cabe destacar que el grupo de investigación del Centro Universitario San Rafael tiene previsto publicar los resultados de este trabajo en una revista científica, con el objetivo de difundir los hallazgos obtenidos y contribuir al avance del diagnóstico de la esclerosis lateral amiotrófica.

Capítulo 6. BIBLIOGRAFÍA

- [1] Brooks BR, Miller RG, Swah M, Munsat TL; World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*. 2000;1(5):293-299. <https://doi.org/10.1080/146608200300079536>
- [2] Rodríguez de Rivera FJ, Oreja Guevara C, Sanz Gallego I, San José Valiente B, Santiago Recuerda A, Gómez Mendieta MA, Arpa J, Díez Tejedor E. Evolución de pacientes con esclerosis lateral amiotrófica atendidos en una unidad multidisciplinar. *Neurología*. 2011; 26(8):455-460. <https://doi.org/10.1016/j.nrl.2011.01.021>
- [3] Riancho J, Gonzalo I, Ruiz-Soto M, Berciano J. ¿Por qué degeneran las motoneuronas? Actualización en la patogenia de la esclerosis lateral amiotrófica. *Neurología*. 2019; 34(1):27-37. <https://doi.org/10.1016/j.nrl.2015.12.001>
- [4] Fallas Sanabria M. Esclerosis lateral amiotrófica. *Revista Médica de Costa Rica y Centroamérica*. 2010; LXVII(591):89-92.
- [5] Castro-Rodríguez E, Azagra R, Gómez-Batiste X, Povedano M. La esclerosis lateral amiotrófica (ELA) desde la Atención Primaria. Epidemiología y características clínico-asistenciales. *Atención Primaria*. 2021;53:102158. <https://doi.org/10.1016/j.aprim.2021.102158>
- [6] Rowland LP, Shneider NA. Amyotrophic lateral sclerosis. *New England Journal of Medicine*. 2001;344(22):1688-1700. <https://doi.org/10.1056/NEJM200105313442207>
- [7] Turner MR, Hardiman O, Benatar M, Brooks BR, Chio A, de Carvalho M, Couratier P, Shaw PJ, Veldink JH, van den Berg LH. Controversies and priorities in amyotrophic lateral sclerosis. *The Lancet Neurology*. 2013;12(3):310-322. [https://doi.org/10.1016/S1474-4422\(13\)70036-X](https://doi.org/10.1016/S1474-4422(13)70036-X)
- [8] Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W, Shaw PJ, Simmons Z, van den Berg LH. Amyotrophic lateral sclerosis. *Nature Reviews Disease Primers*. 2017;3:17071. <https://doi.org/10.1038/nrdp.2017.85>
- [9] Bensimon G, Lacomblez L, Meininger V; ALS/Riluzole Study Group. A controlled trial of riluzole in amyotrophic lateral sclerosis. *New England Journal of Medicine*. 1994;330(9):585-591. <https://doi.org/10.1056/NEJM199403033300901>

- [10] Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, Nakanishi A. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*. 1999;169(1-2):13-21.
[https://doi.org/10.1016/S0022-510X\(99\)00210-5](https://doi.org/10.1016/S0022-510X(99)00210-5)
- [11] De Carvalho M, Dengler R, Eisen A, England JD, Kaji R, Kimura J, Mills K, Mitsumoto H, Nodera H, Shefner J, Swash M. Electrodiagnostic criteria for diagnosis of ALS. *Clinical Neurophysiology*. 2008;119(3):497-503.
<https://doi.org/10.1016/j.clinph.2007.09.143>
- [12] Marin B, Boumediene F, Logroscino G, Couratier P, Babron MC, Leutenegger AL, Coppetti M, Preux PM, Esquirol M. Variation in worldwide incidence of amyotrophic lateral sclerosis: a meta-analysis. *International Journal of Epidemiology*. 2017;46(1):57-74. <https://doi.org/10.1093/ije/dyw061>
- [13] Van Es MA, Van Vught PW, Blauw HM, Franke L, Saris CG, Andersen PM, van den Berg LH, Veldink JH. Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. *Nature Genetics*. 2008;40(1):29-31.
<https://doi.org/10.1038/ng.2007.52>
- [14] DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, Kouri N. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. 2011;72(2):245-256. <https://doi.org/10.1016/j.neuron.2011.09.011>
- [15] Wijesekera LC, Leigh PN. Amyotrophic lateral sclerosis. *Orphanet Journal of Rare Diseases*. 2009;4:3. <https://doi.org/10.1186/1750-1172-4-3>
- [16] Martínez-Payá JJ, Ríos-Díaz J, del Baño-Aledo ME, Tembl-Ferrairó JI, Vázquez-Costa JF, Medina-Mirapeix F. Quantitative muscle ultrasonography using textural analysis in amyotrophic lateral sclerosis. *Ultrasonic Imaging*. 2017;39(6):357-368.
<https://doi.org/10.1177/0161734617711370>
- [17] Martínez-Payá JJ, Ríos-Díaz J, Medina-Mirapeix F, Vázquez-Costa JF, del Baño-Aledo ME. Monitoring progression of amyotrophic lateral sclerosis using ultrasound morpho-textural muscle biomarkers: a pilot study. *Ultrasound in Medicine and Biology*. 2018;44(1):102-109. <https://doi.org/10.1016/j.ultrasmedbio.2017.09.013>
- [18] Arts IMP, van Rooij FG, Overeem S, Pillen S, Janssen HMHA, Schelhaas HJ, Zwarts MJ. Quantitative muscle ultrasonography in amyotrophic lateral sclerosis. *Ultrasound in*

- Medicine and Biology*. 2008;34(3):354-361.
<https://doi.org/10.1016/j.ultrasmedbio.2007.08.013>
- [19] Arts IMP, Overeem S, Pillen S, Kleine BU, Boeckstein WA, Zwarts MJ, Schelhaas HJ. Muscle ultrasonography: a diagnostic tool for amyotrophic lateral sclerosis. *Clinical Neurophysiology*. 2012;123(8):1662-1667. <https://doi.org/10.1016/j.clinph.2011.11.262>
- [20] Arts IMP, Overeem S, Pillen S, Schelhaas HJ, Zwarts MJ. Muscle changes in amyotrophic lateral sclerosis: a longitudinal ultrasonography study. *Clinical Neurophysiology*. 2011;122(3):623-628. <https://doi.org/10.1016/j.clinph.2010.07.023>
- [21] Arts IMP, Overeem S, Pillen S, Schelhaas HJ, Zwarts MJ. Muscle ultrasonography to predict survival in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*. 2011;82(5):552-554. <https://doi.org/10.1136/jnnp.2009.200519>
- [22] Pillen S, Arts IMP, Zwarts MJ. Muscle ultrasound in neuromuscular disorders. *Muscle and Nerve*. 2008;37(6):679-693. <https://doi.org/10.1002/mus.21015>
- [23] Grimm A, Prell T, Décard BF, Schumacher U, Witte OW, Axer H. Muscle ultrasonography as an additional diagnostic tool for the diagnosis of amyotrophic lateral sclerosis. *Clinical Neurophysiology*. 2015;126(4):820-827.
<https://doi.org/10.1016/j.clinph.2014.06.052>
- [24] Martínez-Payá JJ, del Baño-Aledo ME, Ríos-Díaz J, Tembl-Ferrairó JI, Vázquez-Costa JF, Medina-Mirapeix F. Muscular echovariation: a new biomarker in amyotrophic lateral sclerosis. *Ultrasound in Medicine and Biology*. 2017;43(6):1153-1162.
<https://doi.org/10.1016/j.ultrasmedbio.2017.02.002>
- [25] Misowa S, Noto Y, Shibuya K, Iose S, Sekiguchi Y, Nasu S, Kuwabara S. Ultrasonographic detection of fasciculations markedly increases diagnostic sensitivity of ALS. *Neurology*. 2011;77(16):1532-1537. <https://doi.org/10.1212/wnl.0b013e318233b36a>
- [26] Van Alfen N, Gilhuis HJ, Keijzers JP, Pillen S, van Engelen BG. Quantitative facial muscle ultrasound: feasibility and reproducibility. *Muscle and Nerve*. 2011;43(2):175-181.
<https://doi.org/10.1002/mus.23769>
- [27] Pillen S, Tak RO, Zwarts MJ, Lammens MM, Verrijp KN, Arts IMP, van der Laak JA, Hoogerbrugge PM, van Engelen BG, Verrips A. Skeletal muscle ultrasound: correlation between fibrous tissue and echo intensity. *Ultrasound in Medicine and Biology*. 2009;35(3):443-446. <https://doi.org/10.1016/j.ultrasmedbio.2008.09.016>

- [28] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*. 1973;3(6):610-621.
<https://doi.org/10.1109/TSMC.1973.4309314>
- [29] Galloway MM. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*. 1975;4(2):172-179. [https://doi.org/10.1016/S0146-664X\(75\)80008-6](https://doi.org/10.1016/S0146-664X(75)80008-6)
- [30] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002;24(7):971-987.
<https://doi.org/10.1109/TPAMI.2002.1017623>
- [31] Nanni L, Brahnam S, Ghidoni S, Menegatti E, Barrier T. Different approaches for extracting information from the co-occurrence matrix. *PLoS ONE*. 2013;8(12):e83554.
<https://doi.org/10.1371/journal.pone.0083554>
- [32] Molinari F, Caresio C, Acharya UR, Mookiah MRK, Minetto MA. Advances in quantitative muscle ultrasonography using texture analysis of ultrasound images. *Ultrasound in Medicine and Biology*. 2015;41(9):2520-2532.
<https://doi.org/10.1016/j.ultrasmedbio.2015.04.021>
- [33] Gdynia HJ, Müller HP, Ludolph AC, Königer H, Huber R. Quantitative Muscle Ultrasound in neuromuscular disorders using the parameters ‘intensity’, ‘entropy’, and ‘fractal dimension’. *European Journal of Neurology*. 2009;16(10):1151-1158.
<https://doi.org/10.1111/j.1468-1331.2009.02663.x>
- [34] Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, Löck S. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328-338.
<https://doi.org/10.1148/radiol.2020191145>
- [35] Aggarwal N, Agrawal RK. First and second order statistics features for classification of magnetic resonance brain images. *Journal of Signal and Information PROCessing*. 2012;3(2):146-153. <http://doi.org/10.4236/jsip.2012.32019>
- [36] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-2830.
- [37] Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32.
<https://doi.org/10.1023/A:1010933404324>

- [38] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;20(3):273-297.
<https://doi.org/10.1007/BF00994018>
- [39] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak J, van Ginneken B, Sánchez CI. A survey on Deep Learning in medical image analysis. *Medical Image Analysis*. 2017;42:60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- [40] Dietterich TG. Ensemble methods in Machine Learning. *Lecture Notes in Computer Science*. 2000;1875:1-15. https://doi.org/10.1007/3-540-45014-9_1
- [41] Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York: John Wiley and Sons; 2000.
- [42] Kleinbaum DG, Klein M. *Logistic Regression: A Self-Learning Text*. 3rd ed. New York: Springer; 2010.
- [43] Gerón A. *Hands-On Machine Learning with Scikit-learn, Keras, and TensorFlow*. 2nd ed. Sebastopol: O'Reilly Media; 2019.
- [44] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. 1995;2:1137-1143.
- [45] Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006;27(8):861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- [46] Pinto M, Barroso S, de Carvalho M. Quantitative Muscle Ultrasound and electromyography in patients with ALS: a prospective study. *Clinical Neurophysiology*. 2021; 132(8):1904-1910.
- [47] Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. *Deep Learning* for visual understanding: a review. *Neurocomputing*. 2016;187:27-48.
<https://doi.org/10.1016/j.neucom.2015.09.116>
- [48] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*. 2017;30:4765-4774.
- [49] Ribeiro MT, Singh S, Guestrin C. “Why should I trust you?”: explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference of Knowledge Discovery and Data Mining*. 2016:1135-1144.
<https://doi.org/10.1145/2939672.2939778>
- [50] LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature*. 2015;521(7553):436-444.
<https://doi.org/10.1038/nature14539>

- [51] Iniesta R, Stahl D, McGuffin P. Machine Learning, statistical learning and the future of biological research in psychiatry. *Psychological Medicine*. 2016;46(12):2455-2465.
<https://doi.org/10.1017/S0033291716001367>

ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS

Este Trabajo Fin de Grado está relacionado principalmente con el **ODS 3**, cuyo objetivo es mejorar la salud y el bienestar de las personas.

La Esclerosis Lateral Amiotrófica (ELA) es una enfermedad neurodegenerativa que provoca una pérdida progresiva de la movilidad y afecta de forma importante a la calidad de vida de los pacientes y de sus familias. Uno de los principales problemas de esta enfermedad es que no existe una prueba única que permita diagnosticarla de forma rápida y precisa.

Para ayudar a resolver este problema, en este proyecto se ha desarrollado un sistema basado en Inteligencia Artificial capaz de analizar información obtenida mediante ecografía muscular para diferenciar entre personas con ELA y personas sanas. Este tipo de herramientas puede ayudar a detectar la enfermedad antes y apoyar a los médicos en el proceso de diagnóstico.

Además, la ecografía muscular es una técnica sencilla, no invasiva, económica y disponible en muchos centros sanitarios. Por ello, este trabajo puede contribuir al desarrollo de herramientas que ayuden a mejorar la atención y el seguimiento de los pacientes con enfermedades neurodegenerativas.

Este proyecto también está relacionado con el **ODS 9**, que promueve la innovación y el desarrollo tecnológico.

El proyecto combina ecografía muscular e Inteligencia Artificial para desarrollar una herramienta capaz de ayudar en la identificación de patrones relacionados con la ELA. Esta combinación de tecnologías representa una aplicación innovadora dentro de la ingeniería biomédica.

Además, fomenta el uso de herramientas tecnológicas para analizar datos médicos y contribuye al desarrollo de nuevas soluciones que pueden mejorar el diagnóstico de enfermedades y apoyar futuras investigaciones en el ámbito de la salud.

En resumen, este Trabajo Fin de Grado contribuye principalmente al ODS 3 (Salud y Bienestar), ya que busca mejorar el diagnóstico de la ELA mediante nuevas herramientas de apoyo. También contribuye al ODS 9 (Industria, Innovación e Infraestructura) gracias al uso de técnicas innovadoras de Inteligencia Artificial y análisis de imágenes médicas.