

Facultat Internacional de Comerç i Economia Digital La Salle

Trabajo Final de Máster

Máster Universitario en Ciencia de los Datos / Data Science

**BENCHMARKING DE TÉCNICAS DE
PROCESAMIENTO DE DATOS Y REGRESIÓN
PARA PREDICIR EL PROGRESO DE
REHABILITACIÓN EN PACIENTES CON
TRASTORNOS MUSCULOESQUELÉTICOS EN
EL TOBILLO**

Alumno

Bernat Sort Rufat

Profesor Ponente

Roger Mallol Parera

ACTA DEL EXAMEN DEL TRABAJO FINAL DE MÁSTER

Reunido el Tribunal calificador en el día de la fecha, el alumno

Bernat Sort Rufat

Expuso su Trabajo de Final de Máster, el cual trató sobre el tema siguiente:

**BENCHMARKING DE TÉCNICAS DE PROCESAMIENTO DE DATOS Y
REGRESIÓN PARA PREDICTAR EL PROGRESO DE REHABILITACIÓN EN
PACIENTES CON TRASTORNOS MUSCULOESQUELÉTICOS EN EL TOBILLO**

Acabada la exposición y contestadas por parte del alumno las objeciones formuladas por los miembros del tribunal, este valoró el mencionado Trabajo con la calificación de

Barcelona,

VOCAL DEL TRIBUNAL

VOCAL DEL TRIBUNAL

PRESIDENTE DEL TRIBUNAL

Resumen

A pesar del progreso realizado en los últimos años en el campo de la salud y del análisis clínico de la marcha, actualmente existe una brecha en la comprensión y predicción del progreso de los pacientes con deficiencias musculoesqueléticas a lo largo de sus procesos terapéuticos y de rehabilitación. Este proyecto de investigación pretende utilizar técnicas de análisis de datos y aprendizaje automático para predecir del progreso y la evolución de pacientes con deficiencias musculoesqueléticas en el tobillo durante los procesos de rehabilitación.

Se utilizó el conjunto de datos público GaitRec, que contía datos anotados correspondientes a las fuerzas de reacción del suelo (GRF) bilaterales de pacientes con diversas discapacidades musculoesqueléticas. Se realizó un análisis exhaustivo para los datos de fuerzas de reacción del suelo en la dirección anteroposterior durante la fase de apoyo del ciclo de la marcha. Se exploraron diversas técnicas de procesamiento de datos, y se desarrollaron y evaluaron diferentes modelos de aprendizaje automático de regresión multi-salida, como la regresión lineal múltiple, la regresión de mínimos cuadrados parciales (PLS), la regresión de vectores de soporte (SVR), árboles de decisión, bosques aleatorios y *eXtreme Gradient Boosting* (XGBoost). Los modelos se evaluaron mediante métricas de regresión como la raíz del error cuadrático medio (RMSE), el coeficiente de determinación (R^2), y el coeficiente de correlación de Pearson. Se examinaron tanto los valores individuales para cada variable objetivo como sus promedios. También se aplicó el método *target-specific ranks*. Esto permitió una evaluación más completa de los modelos.

Los resultados mostraron que aplicar la reducción de dimensionalidad mediante PCA en los datos de la primera sesión de rehabilitación mejoró el rendimiento del modelo de regresión lineal múltiple. En general, todos los modelos mostraron una fuerte correlación entre las predicciones y las curvas reales correspondientes a la última sesión de rehabilitación. Además, lograron predecir la fase de apoyo con una sola pierna, obteniendo valores bajos de RMSE en estos puntos de la curva. Sin embargo, se observó un aumento en los valores de RMSE en los picos de máxima frenada y propulsión. El modelo PLS entrenado con una configuración óptima de 8 componentes, y el modelo SVR entrenado y evaluado en datos originales, mostraron un rendimiento superior en estos picos en comparación con los otros modelos. Sin embargo, ningún modelo logró predecir de manera precisa los valores máximos de frenada y propulsión durante la fase de apoyo del ciclo de la marcha.

Los modelos de aprendizaje automático, en particular el PLS y el SVR, logran predecir la evolución de los pacientes con deficiencias musculoesqueléticas en el tobillo durante la mayor parte de la fase del apoyo del ciclo de la marcha. No obstante, se necesitan investigaciones adicionales para mejorar la predicción de valores máximos de frenada y propulsión en esta fase.

Palabras clave: biomecánica, análisis clínico de la marcha, ciclo de la marcha, fuerzas de reacción del suelo (GRF), desórdenes musculoesqueléticos, rehabilitación, conjunto de datos GaitRec, regresión, aprendizaje automático.

Abstract

Despite the progress made in recent years in the field of health and clinical gait analysis, there is currently a gap in understanding and predicting the progress of patients with musculoskeletal deficiencies throughout their therapeutic and rehabilitation processes. This research project aims to use data analysis and machine learning techniques to predict the progress and evolution of patients with musculoskeletal ankle impairments during the rehabilitation process.

We used the publicly available GaitRec dataset, which includes annotated data related to bilateral ground reaction forces (GRF) obtained from patients with several musculoskeletal impairments. A comprehensive analysis was conducted for the anteroposterior ground reaction force data during the stance phase of the gait cycle. Several data processing techniques were explored, and different multi-output regression machine learning models were developed and evaluated, including multiple linear regression, partial least squares (PLS) regression, support vector regression (SVR), decision trees, random forests, and eXtreme Gradient Boosting (XGBoost). The models were evaluated using regression metrics such as root mean square error (RMSE), coefficient of determination (R^2), and Pearson correlation coefficient. Both individual values for each target variable and their averages were examined. The *target-specific ranks* method was also applied, allowing for a more comprehensive model evaluation.

Results demonstrated that applying dimensionality reduction using PCA on data from the first rehabilitation session improved the performance of the multiple linear regression model. Overall, all models exhibited a strong correlation between predictions and actual curves corresponding to the last rehabilitation session. They also successfully predicted the single-leg stance phase, achieving low RMSE values at these points in the curve. However, an increase in RMSE values was observed at the peaks of maximum braking and propulsion. The PLS model trained with an optimal configuration of 8 components and the SVR model trained and evaluated on original data demonstrated superior performance at these peaks compared to other models. Nevertheless, none of the models accurately predicted the maximum braking and propulsion values during the stance phase of the gait cycle.

Machine learning models, particularly PLS and SVR, can predict the evolution of patients with musculoskeletal ankle disorders during most of the stance phase of the gait cycle. However, further research is needed to improve the prediction of maximum braking and propulsion values in this phase.

Keywords: biomechanics, clinical gait analysis, gait cycle, ground reaction forces (GRF), musculoskeletal impairments, rehabilitation, GaitRec dataset, regression, machine learning.

Tabla de Contenidos

1	Introducción.....	1
1.1	Contexto	1
1.2	Motivación y justificación	1
1.3	Objetivos.....	2
1.4	Hipótesis	2
1.5	Estructura del documento	2
2	Estado del arte	4
3	Marco teórico	6
3.1	Biomecánica y fundamentos del análisis de la marcha	6
3.1.1	Biomecánica	6
3.1.2	Concepto de marcha	6
3.1.3	Análisis de la marcha	6
3.1.4	El ciclo de la marcha	7
3.1.5	Fuerzas de reacción del suelo (GRF)	7
3.1.6	La componente anteroposterior.....	8
3.1.7	Plataformas de fuerza	10
3.2	Análisis de Componentes Principales (PCA).....	10
3.3	Aprendizaje automático y regresión	11
3.4	Algoritmos de aprendizaje automático para problemas de regresión	11
3.4.1	Mínimos cuadrados ordinarios (OLS).....	11
3.4.2	Regresión de mínimos cuadrados parciales (PLS).....	13
3.4.3	Árboles de decisión	14
3.4.4	Bosques aleatorios.....	15
3.4.5	Extreme Gradient Boosting (XGBoost).....	17
3.4.6	Regresión de vectores de soporte (SVR)	18
3.5	Métricas de evaluación de los modelos de regresión	20
3.5.1	Raíz del error cuadrático medio (RMSE)	20
3.5.2	Coeficiente de determinación (R^2).....	20
3.5.3	Coeficiente de correlación de Pearson.....	21
4	Métodos.....	22
4.1	Herramientas y software	22
4.2	El conjunto de datos GaitRec.....	23
4.2.1	Introducción al conjunto de datos GaitRec	23
4.2.2	Preparación y procesamiento	23
4.2.3	Descripción y anotaciones	24
4.2.4	Estructura de los datos.....	26

4.3	Construcción de los conjuntos de datos	27
4.4	Análisis y comprobaciones iniciales	31
4.4.1	Exploración inicial de los datos generales	31
4.4.2	Análisis inicial de las componentes de las fuerzas de reacción del suelo ...	34
4.5	Estudio de la población y cohorte de pacientes	36
4.6	Análisis exploratorio de los datos	36
4.6.1	Exploración inicial del conjunto de entrenamiento	37
4.6.2	Distribución de las variables numéricas	37
4.6.3	Comprobación de la normalidad de las variables	37
4.6.4	Análisis de correlación	38
4.7	División de los datos	38
4.8	Preprocesamiento de los datos	38
4.8.1	Primera aproximación: datos sin preprocesar	39
4.8.2	Segunda Aproximación: PCA en los datos de entrada (X)	39
4.8.3	Tercera Aproximación: PCA en los datos de salida (Y)	39
4.8.4	Cuarta aproximación: PCA en los datos de entrada y salida (X e Y)	39
4.9	Análisis multivariante	40
4.9.1	Tipo de problema y algoritmos utilizados	40
4.9.2	Pipeline general de entrenamiento y evaluación de los modelos	40
4.9.3	Regresión Lineal Múltiple (MLR)	42
4.9.4	Regresión de mínimos cuadrados parciales (PLS <i>regression</i>)	44
4.9.5	Árboles de decisión	45
4.9.6	Random Forest	47
4.9.7	Extreme Gradient Boosting (XGBoost)	48
4.9.8	Support Vector Regression (SVR)	49
5	Resultados	52
5.1	Análisis global de los resultados de los modelos	52
5.2	Selección y análisis de los modelos de más alto y bajo rendimiento	56
6	Discusión	76
7	Consideraciones éticas	78
8	Coste económico	79
9	Limitaciones del estudio	80
10	Conclusiones	81
11	Líneas futuras	82
12	Referencias	83
13	Apéndice	88
13.1	Apéndice 1. Repositorio del proyecto	88
13.2	Apéndice 2. Resultados	89

13.2.1	Distribución de las variables numéricas.....	89
13.2.2	Análisis de correlación.....	95
13.2.3	Análisis multivariante	96

1 Introducción

1.1 Contexto

En los últimos años, el panorama de la investigación en el ámbito de la salud ha experimentado notables transformaciones debido a la creciente disponibilidad de grandes conjuntos de datos, provocando un creciente interés en la aplicación de métodos de aprendizaje automático y aprendizaje profundo [1].

Esta tendencia también se ha extendido en el campo del análisis clínico de la marcha (ACM) [2], [3]. Este campo interdisciplinar desempeña un papel fundamental en la comprensión de la biomecánica del movimiento humano, especialmente en personas con discapacidades musculoesqueléticas. El análisis clínico de la marcha tiene el potencial de ofrecer una visión profunda de las limitaciones funcionales y las alteraciones biomecánicas que experimentan estos pacientes, guiando así el desarrollo de intervenciones terapéuticas y estrategias de rehabilitación eficaces.

La biomecánica, el análisis de la marcha y la rehabilitación constituyen componentes integrales del panorama médico, cuyo objetivo es mejorar el bienestar y la capacidad funcional de las personas con deficiencias musculoesqueléticas. Los trastornos musculoesqueléticos abarcan una amplia gama de afecciones que afectan a huesos, músculos, ligamentos y articulaciones, y que a menudo provocan alteraciones en los patrones de movimiento y reducen la movilidad. Abordar estas deficiencias mediante terapias y regímenes de rehabilitación adaptados es fundamental no sólo para recuperar la función física, sino también para mejorar la calidad de vida general de las personas afectadas.

Sin embargo, hasta la fecha no se ha encontrado una forma de garantizar que la rehabilitación sea eficaz para lograr un cambio significativo en la vida del paciente, lo que a veces supone una grave pérdida de tiempo y dinero tanto para el paciente como para el sistema sanitario. Aunque la rehabilitación es muy prometedora a la hora de restablecer las funciones y mejorar el bienestar de las personas con deficiencias musculoesqueléticas, su eficacia varía mucho debido a la compleja interacción de factores que influyen en los resultados de la terapia. Esta imprevisibilidad supone un reto importante tanto para los médicos como para los pacientes, que invierten un tiempo, un esfuerzo y unos recursos considerables en programas de rehabilitación con resultados inciertos.

Es en este contexto donde la importancia de esta investigación emerge con mayor claridad. La laguna existente entre el impacto potencial de la rehabilitación y la realidad de unos resultados inciertos exige un cambio de paradigma. La integración de técnicas de aprendizaje automático para predecir el progreso de la terapia en pacientes con deficiencias musculoesqueléticas promete transformar el panorama de la rehabilitación aprovechando el valor de los datos.

1.2 Motivación y justificación

En este contexto, la base de esta investigación se fundamenta en una de las propuestas presentadas por Brian Horsak et al. en el artículo titulado *GaitRec: A Large-Scale Ground Reaction Force Dataset of Healthy and Impaired Gait* [4]. Este artículo proporciona un conjunto de datos exhaustivo y plantea cuestiones y áreas de investigación que se presentan como un terreno fértil para explorar y profundizar.

Esta investigación encuentra su motivación en la necesidad de reducir la brecha entre los objetivos de la rehabilitación y los resultados concretos alcanzados por individuos con deficiencias musculoesqueléticas. No obstante, las implicaciones de evaluar y predecir eficazmente el progreso de la terapia van más allá del ámbito clínico. Dotar a los médicos de herramientas para adaptar las intervenciones en función de indicadores objetivos de progreso

es una innovación que puede redefinir el panorama del tratamiento. Los pacientes, equipados con una comprensión más clara de su proceso de rehabilitación, pueden embarcarse en el camino de la recuperación con un sentido de propósito y dirección. Además, la posibilidad de optimizar la asignación de recursos dentro de los sistemas sanitarios, junto con el potencial para mitigar el coste emocional, físico y económico de los pacientes, guía la dirección de esta investigación.

La incertidumbre que rodea la efectividad de las intervenciones de rehabilitación no solo plantea un desafío práctico, sino también una preocupación profundamente humanitaria. A medida que las personas emprenden el camino hacia la recuperación, invirtiendo esperanzas, energía y recursos, la incapacidad para anticipar la evolución del progreso puede resultar en desilusión y frustración.

Desde una perspectiva social, las ineficiencias derivadas de programas de rehabilitación ineficaces o mal adaptados pueden ejercer una carga excesiva sobre los sistemas de atención sanitaria. La motivación subyacente de este proyecto surge en respuesta a estas inquietudes, como un llamada a tomar medidas con el fin de trazar una ruta más sólida y fundamentada hacia el éxito en la rehabilitación.

La convergencia de la ciencia de datos y la biomecánica ofrece una oportunidad que resuena directamente con el núcleo de la motivación de esta investigación. Aprovechando el valor de los datos, este estudio busca elevar el proceso de toma de decisiones terapéuticas desde el ámbito de la intuición y la experiencia hacia la evidencia y el análisis riguroso. En última instancia, la finalidad es mejorar la calidad de vida de las personas afectadas por deficiencias musculoesqueléticas y contribuir al desarrollo de enfoques de rehabilitación más eficaces y fundamentados.

1.3 Objetivos

El objetivo principal de este proyecto es analizar y predecir, mediante el desarrollo y la evaluación de distintos modelos de aprendizaje automático, el progreso y la evolución de los pacientes con deficiencias musculoesqueléticas en el tobillo sometidos a dos sesiones de terapia o rehabilitación.

1.4 Hipótesis

Se parte de la hipótesis de que los modelos de aprendizaje automático desarrollados y entrenados utilizando un amplio conjunto de datos anotados, que contiene paseos o *trials* de marcha con fuerzas de reacción del suelo (GRF) bilaterales de pacientes que presentan diversas discapacidades musculoesqueléticas [4], demostrarán la capacidad de predecir de manera precisa el progreso y la evolución de los pacientes sometidos a terapia o rehabilitación.

1.5 Estructura del documento

En la introducción se presenta el contexto necesario para la investigación y se sientan las bases de nuestro estudio. Se discute la relevancia de comprender y evaluar los patrones de la marcha en individuos con deficiencias musculoesqueléticas y el impacto potencial en su calidad de vida. Se explican los retos que plantea la incertidumbre de los resultados de la rehabilitación y las posibles ineficiencias de los sistemas sanitarios. Se plantean los objetivos de nuestra investigación, así como la hipótesis que guiará la investigación.

Seguidamente, se profundiza en el conjunto de conocimientos existentes relacionados con nuestro tema de investigación. Se realiza una revisión de estudios similares relacionados con la evaluación de terapias e intervenciones de individuos con deficiencias musculoesqueléticas, así como de las aplicaciones del aprendizaje automático en la predicción del progreso y evolución de los pacientes. Finalmente, se destacan algunas de las limitaciones del panorama actual.

A continuación, se establecen las bases conceptuales y teóricas sobre las cuales se desarrolla nuestro estudio. Se exploran conceptos clave relacionados con la biomecánica, el análisis de la marcha, técnicas de procesamiento de datos, y algoritmos y métricas de regresión.

En el apartado de métodos, se presentan los procedimientos adoptados para la realización de la investigación. Se mencionan las herramientas y el software utilizado, y se describe exhaustivamente el conjunto de datos GaitRec. A continuación, se explica con qué criterios se construyó el conjunto de datos específico para esta investigación, se presenta el análisis exploratorio de los datos, y se detalla el proceso de división de los datos en vectores de entrada y salida para la fase de entrenamiento y prueba. Posteriormente, se explican las diferentes aproximaciones de preprocesamiento de los datos. Finalmente, se describe en profundidad el análisis multivariante.

En el apartado de resultados, se muestran los resultados obtenidos por los modelos de aprendizaje automático, y se analizan en profundidad los modelos con mayor y menor rendimiento.

A continuación, se discuten los resultados del estudio, tanto del análisis exploratorio de los datos como del análisis multivariante.

En las siguientes secciones, se explican los aspectos éticos considerados, se detallan los costes aproximados asociados a este proyecto de investigación, y se muestran varias de las limitaciones del estudio.

Se presentan de manera concisa las conclusiones más destacadas y las contribuciones significativas derivadas de nuestra investigación. Se revisan los objetivos e hipótesis planteadas en el proyecto con el propósito de evaluar si se han alcanzado satisfactoriamente, y se destaca la importancia del trabajo realizado en el avance del campo de la evaluación del progreso de la rehabilitación y su posible impacto en la atención al paciente y los sistemas sanitarios.

Finalmente, se proponen posibles líneas futuras de investigación y desarrollo en este campo. Se reflexiona sobre cómo nuestro trabajo puede ampliarse o mejorarse, para en última instancia, acabar contribuyendo a la atención personalizada de los pacientes y a la optimización de las estrategias de rehabilitación.

2 Estado del arte

En los últimos años, las técnicas de aprendizaje automático y aprendizaje profundo han demostrado una notable eficacia en el ámbito del análisis clínico de la marcha (ACM). Estos métodos han demostrado ser exitosos en la categorización de cohortes de pacientes [2], [3] abarcando condiciones como el ictus [5], casos de parálisis cerebral [6], casos de osteoartritis [7], esclerosis múltiple [8], Parkinson [9], e individuos lidiando con diversos trastornos funcionales de la marcha [10].

En el ámbito del análisis predictivo de la marcha, los investigadores han explorado la aplicación de técnicas de aprendizaje automático para predecir los resultados postoperatorios en individuos con trastornos del movimiento. Además, en el ámbito de las intervenciones ortopédicas se han realizado contribuciones destinadas a mejorar la toma de decisiones y los resultados del tratamiento.

No obstante, a pesar de la proliferación de estudios que emplean técnicas de aprendizaje automático para analizar patrones y anomalías de la marcha, sigue existiendo un vacío a la hora de comprender y predecir la evolución de los pacientes con deficiencias musculoesqueléticas a lo largo de los procesos terapéuticos y de rehabilitación. Es por ello por lo que, aunque los artículos en los que nos hemos basado para realizar nuestra investigación no están directamente relacionados con la evaluación y predicción del progreso de la terapia de la marcha, si presentan similitudes en cuanto al uso del aprendizaje automático para predecir resultados relacionados con los patrones de la marcha.

El artículo científico titulado *Predicting Postoperative Knee Flexion during Gait of Cerebral Palsy Children*, de Omar A. Galarraga C. et al. [11] se centra en el uso del aprendizaje automático para predecir la flexión postoperatoria de la rodilla durante la marcha en niños con parálisis cerebral. El estudio aborda los retos que plantean trastornos neurológicos como la parálisis cerebral, que afectan negativamente a la marcha y la movilidad. El estudio emplea el análisis de componentes principales (PCA) para reducir la dimensionalidad de las características de entrada. Utilizando un modelo de regresión lineal múltiple, los autores predicen la flexión de la rodilla durante la marcha postoperatoria, teniendo en cuenta los datos cinemáticos preoperatorios proyectados en un espacio de dimensión inferior mediante el análisis de componentes principales (PCA), los resultados de la exploración física y el tipo de cirugía. Mediante la consideración de varias dimensiones, los investigadores logran predicciones precisas de la flexión de la rodilla. La incorporación de un código de cirugía mejora la capacidad predictiva del modelo, que abarca varias categorías quirúrgicas y sus combinaciones. Los resultados se presentan con intervalos de confianza, lo que proporciona una medida fiable de la certeza de la predicción, y evalúan el rendimiento de la predicción mediante la raíz del error cuadrático medio (RMSE). Además, se destaca la posible utilidad clínica de los modelos predictivos para ayudar en la toma de decisiones sobre intervenciones quirúrgicas y ofrecer a los pacientes información sobre los resultados probables.

El artículo científico *Estimation of Postoperative Knee Flexion at Initial Contact of Cerebral Palsy Children using Neural Networks* de Omar A. Galarraga C. et al. [12], profundiza en la estimación de la flexión postoperatoria de la rodilla, un parámetro crucial en el contexto de la parálisis cerebral y las intervenciones ortopédicas. El estudio aborda el reto de predecir la flexión de la rodilla tras la cirugía de alargamiento de los isquiotibiales, una intervención ortopédica habitual para aliviar dichas irregularidades de la marcha. Esta investigación aprovecha el potencial de las redes neuronales para el ajuste no lineal de datos, demostrando la viabilidad de predecir la flexión postoperatoria de la rodilla basándose en datos preoperatorios de la marcha, la exploración física y el tipo de cirugía. Los autores emplean una única arquitectura de red neuronal que toma múltiples variables de entrada relacionadas con el estado del paciente y la cirugía, y produce como salida los ángulos de flexión de rodilla

predichos en el contacto inicial. La medida del error considerada para cada paciente es la raíz del error cuadrático medio (RMSE). Mediante técnicas de aprendizaje supervisado, el estudio establece un simulador capaz de estimar los resultados de la flexión postoperatoria de la rodilla con un encomiable nivel de precisión.

El artículo científico titulado *Predicting Postoperative Gait in Cerebral Palsy*, de Omar A. Galarraga C. et al. [13], presenta un enfoque innovador para predecir la cinemática postoperatoria de las extremidades inferiores en pacientes con parálisis cerebral (PC). El estudio aprovecha la cinemática preoperatoria, los datos de la exploración física y la información quirúrgica para predecir los patrones cinemáticos tras la cirugía multinivel de evento único (SEMLS). Los autores emplean técnicas estadísticas de aprendizaje automático para desarrollar un sistema de predicción capaz de proporcionar estimaciones precisas de los ángulos de marcha postoperatorios para una diversa gama de procedimientos quirúrgicos y patrones de marcha. Por tanto, mientras que el primer artículo descrito en esta sección [11] se centra en la predicción de los ángulos de flexión de la rodilla, este nuevo estudio adopta un enfoque más exhaustivo para predecir los patrones generales de la marcha. No obstante, de manera similar a la investigación anteriormente mencionada [11], y con el objetivo de mejorar la precisión de la predicción y reducir la redundancia de datos, los autores utilizan el análisis de componentes principales (PCA) para reducir la dimensionalidad de las variables independientes, es decir, de los datos preoperatorios. A continuación, utilizan un único modelo de regresión lineal múltiple para predecir simultáneamente las múltiples variables cinemáticas postoperatorias. Es decir, este modelo de regresión tiene en cuenta una combinación de datos cinemáticos preoperatorios proyectados en un espacio de dimensión inferior mediante el análisis de componentes principales (PCA), datos del examen físico preoperatorio e información quirúrgica como características de entrada, y predice los resultados cinemáticos postoperatorios para una serie de ángulos cinemáticos. Finalmente, evalúan el rendimiento de la predicción mediante la raíz del error cuadrático medio (RMSE) para cada paciente del conjunto de prueba. Aunque la media postoperatoria resulta ser un predictor fiable para los ángulos del plano frontal, mostrando errores de predicción medios y desviaciones estándar mínimos, las predicciones para la dorsiflexión del tobillo no están bien estimadas y tienden a coincidir estrechamente con la cinemática media postoperatoria. Esto indica que el rendimiento del modelo para predecir la dorsiflexión del tobillo es menos preciso en comparación con otros parámetros cinemáticos, especialmente en comparación con los ángulos del plano sagital.

Los artículos analizados [11], [12], [13] ponen de manifiesto que la aplicación del aprendizaje automático para predecir resultados relacionados con los patrones de la marcha es muy prometedora, y destacan el poder de estas técnicas para proporcionar información valiosa sobre la planificación del tratamiento y la toma de decisiones. No obstante, también destacan la presencia de algunos de los desafíos existentes: la estimación precisa de la dorsiflexión del tobillo, que es un parámetro crucial en el análisis de la marcha, sigue siendo un reto importante, lo que plantea áreas con potencial para una investigación adicional.

Sin embargo, si bien ninguna de las investigaciones anteriores se enfoca directamente en la predicción del progreso de la terapia, sientan un precedente en el uso del aprendizaje automático para predecir el progreso y la evolución de los pacientes con deficiencias musculoesqueléticas sometidos a rehabilitación.

3 Marco teórico

En esta sección, presentamos el marco teórico que sustenta la investigación realizada. El objetivo de este apartado es establecer las bases conceptuales y teóricas sobre las cuales se desarrolla nuestro estudio. Se exploran conceptos clave relacionados con la biomecánica, técnicas de procesamiento de datos como el PCA, y métodos de regresión utilizados en la investigación.

3.1 Biomecánica y fundamentos del análisis de la marcha

3.1.1 Biomecánica

La biomecánica es un campo interdisciplinar que combina los principios de la biología y la física para comprender cómo las fuerzas externas influyen e intervienen en la estructura y función de los sistemas biológicos [14].

La biomecánica abarca desde los aspectos microscópicos, como la contracción muscular a nivel celular, hasta los movimientos macroscópicos de todo el cuerpo, como la acción de caminar. Sin embargo, en esta investigación la biomecánica se enfoca en el estudio de la marcha humana y sus alteraciones, concretamente los trastornos musculoesqueléticos en el tobillo.

3.1.2 Concepto de marcha

La marcha es el movimiento cíclico de las extremidades inferiores y superiores que tiene por objeto desplazar el cuerpo hacia delante. Durante la marcha, el cuerpo se mueve hacia adelante utilizando una extremidad de apoyo mientras la otra avanza hacia la siguiente posición de apoyo.

Esto lo logra con una serie de movimientos coordinados de las extremidades inferiores y el tronco para mantener la estabilidad y el avance constante mientras se soporta y transfiere el peso del cuerpo [14].

3.1.3 Análisis de la marcha

El conjunto de métodos empleados para observar, registrar, analizar e interpretar los patrones de movimiento inherentes a la habilidad de caminar se conoce como análisis de la marcha. El análisis de la marcha se convierte en análisis clínico cuando se busca discernir los efectos de las condiciones médicas y disfunciones para detectar deficiencias y limitaciones funcionales que contribuyen a la discapacidad en el contexto de la locomoción [15], [16].

Según Baker (2006) el análisis de la marcha pretende obtener información para (1) diagnosticar trastornos del movimiento, como una enfermedad o lesión específica, (2) evaluar su severidad, (3) monitorear el progreso de una intervención, y (4) predecir el resultado de una intervención [17]. Es por ello por lo que el análisis de la marcha podría ser útil para valorar los planes de tratamiento y diseñar un plan eficaz de rehabilitación.

No obstante, es esencial emplear una terminología clara y coherente para llevar a cabo de la manera más precisa este tipo de análisis. Para ello, debemos comprender el ciclo de la marcha y sus diferentes fases.

3.1.4 El ciclo de la marcha

El ciclo de la marcha se refiere al período completo desde el momento en que un pie toca el suelo hasta el siguiente contacto del mismo pie. En la mayoría de los casos, el ciclo de la marcha se divide en dos períodos o fases principales: la fase de apoyo (*stance phase*) y la fase de balanceo (*swing phase*). La fase de apoyo representa aproximadamente el primer 60% del ciclo de la marcha, mientras que la fase de balanceo el último 40% [18].

Durante la fase de apoyo, el pie está en contacto con el suelo y soporta el peso del cuerpo, mientras que, durante la fase de balanceo, el mismo pie se levanta del suelo y se mueve hacia adelante para preparar el próximo contacto [14].

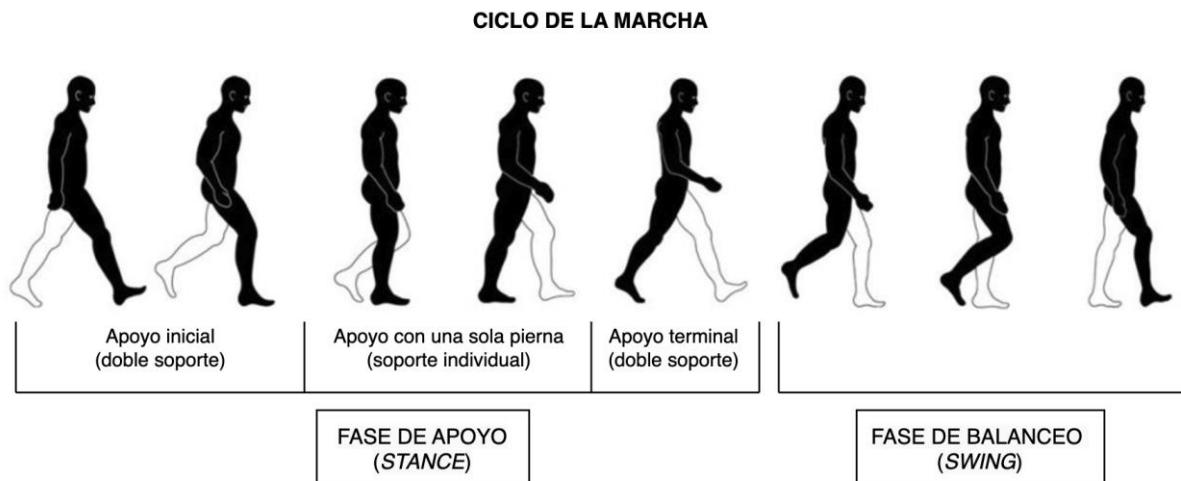


Figura 1. Ciclo y fases de la marcha humana [19].

No obstante, el período de apoyo puede dividirse en tres subperiodos. El primer subperiodo hace referencia al apoyo inicial con dos piernas durante el cual ambos pies están en contacto con el suelo. El segundo subperiodo se refiere al apoyo con una sola pierna durante el cual el pie opuesto o contralateral está en el aire. Finalmente, el tercer subperiodo, llamado apoyo terminal con dos piernas, ocurre cuando ambos pies están de nuevo en contacto con el suelo [20]. El ciclo de la marcha se representa en la Fig. 1.

3.1.5 Fuerzas de reacción del suelo (GRF)

En el análisis biomecánico de la marcha, la cinética juega un papel fundamental al examinar las fuerzas que actúan sobre el cuerpo humano durante el movimiento. La cinética es un subcampo de la mecánica que, en este caso, se centra en comprender cómo las fuerzas externas, como la gravedad y la presión, interactúan con las estructuras internas del cuerpo, como músculos, articulaciones y huesos [14].

Una medida esencial en este análisis es la fuerza de reacción del suelo o *ground reaction forces* (GRF) en inglés, que representa la respuesta de la superficie del suelo a la fuerza que el pie ejerce sobre ella durante la marcha. Es decir, la fuerza que ejerce el suelo sobre un cuerpo que está en contacto con él. Cuando el cuerpo está en reposo, esta magnitud corresponde al peso del cuerpo. Sin embargo, cuando el cuerpo está en movimiento, esta fuerza experimenta variaciones debido a las fuerzas de aceleración [21].

Cuantificar las fuerzas de reacción del suelo es una herramienta útil para que los médicos puedan medir de manera objetiva la locomoción humana, así como describir y analizar detalladamente el rendimiento de la marcha de un paciente [4].

3.1.6 La componente anteroposterior

En este estudio, nos enfocamos en analizar las fuerzas de reacción del suelo en la dirección anteroposterior (AP-GRF), es decir, las fuerzas de frenado y propulsión durante la marcha.

La fuerza de frenado, también se conoce como la fuerza anteroposterior en el apoyo del talón [22], mientras que la fuerza de propulsión es conocida también como fuerza de aceleración, impulso, o despegue [23].

La componente anteroposterior se representa con una curva que, en su inicio, indica la desaceleración que se produce cuando el pie entra en contacto con el suelo. Esta curva alcanza su máximo valor en la fase de doble apoyo. A continuación, la fuerza disminuye hasta llegar a cero cuando el apoyo es con una sola pierna, concentrando el peso en un solo pie. Posteriormente, se observa un nuevo pico en la gráfica que alcanza su valor máximo al comenzar la etapa de doble soporte, donde ambos pies están en contacto con el suelo [24], [25].

Por tanto, las fuerzas anteroposteriores representan tanto la desaceleración (fuerza de frenado) como la aceleración (fuerza de propulsión) que ocurren durante el impacto y el despegue del talón, respectivamente. Estas fuerzas se manifiestan con un pico negativo y otro positivo en la curva. En otras palabras, el punto de máxima desaceleración corresponde al valor mínimo en la gráfica, mientras que el punto de aceleración máxima se corresponde con el valor máximo [26] tal y como se muestra en la Fig. 2.

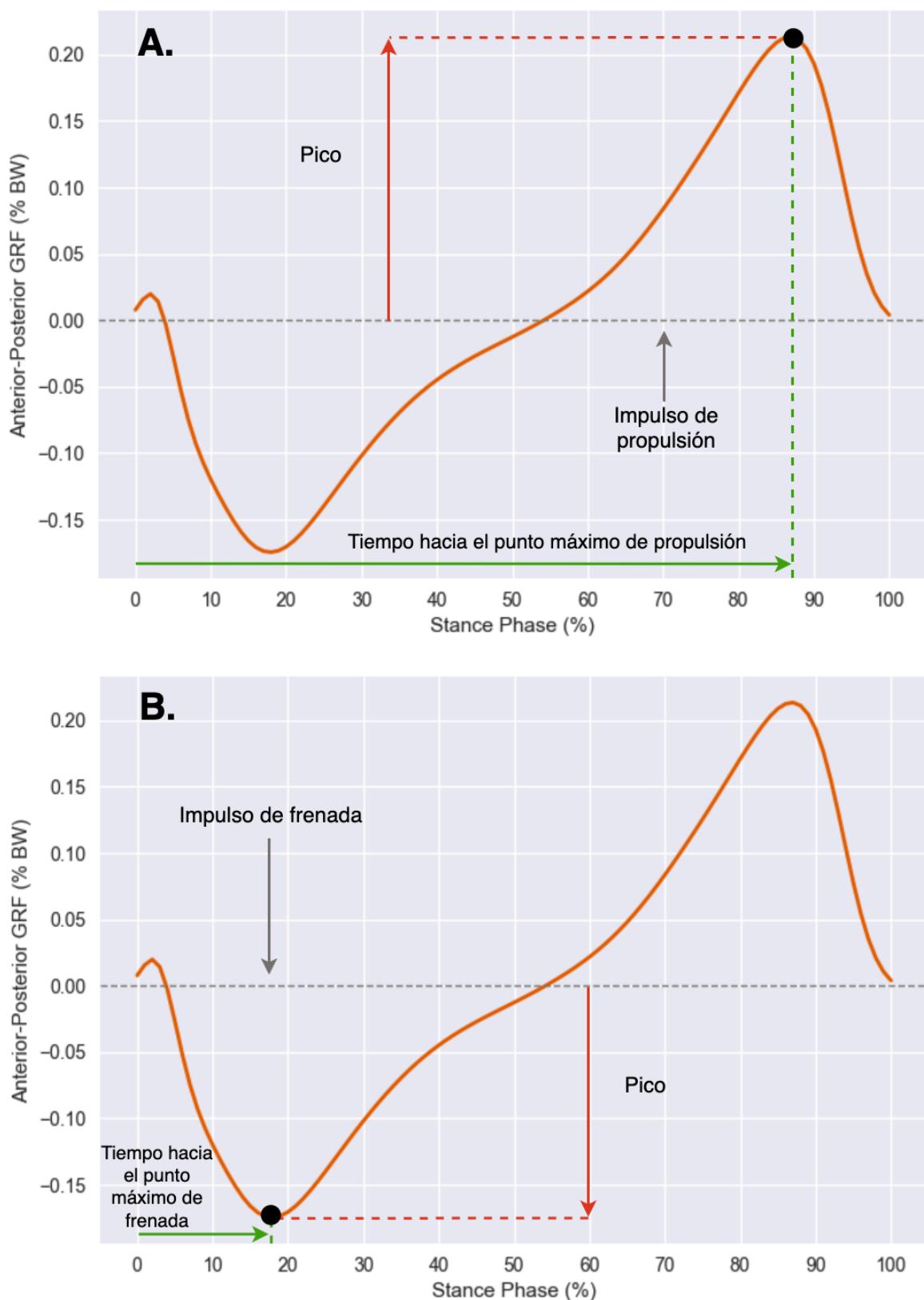


Figura 2. Representación gráfica de las fuerzas de reacción del suelo en la dirección anteroposterior (AP-GRF). A) Fuerzas de frenado durante la marcha. B) Fuerzas de propulsión durante la marcha.

3.1.7 Plataformas de fuerza

Para poder medir las fuerzas de reacción del suelo durante la marcha y analizar la componente anteroposterior, se utilizan las plataformas de fuerza (ver la Fig. 3) [26]. Actualmente, estas plataformas son el método de referencia para la medición de las fuerzas de reacción del suelo en dirección anteroposterior generadas durante la marcha [27].

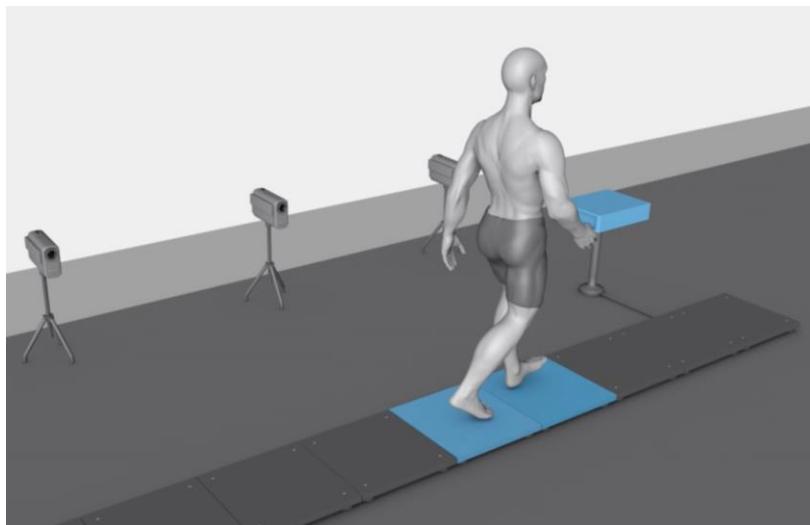


Figura 3. Plataformas de fuerza [28].

En estas plataformas, los pacientes realizan los paseos o *trials* durante las sesiones de evaluación. Los paseos son el conjunto de pasos que realiza un individuo durante una sesión de evaluación o grabación de la marcha, los cuales pueden estar compuestos por varios ciclos de la marcha. El centro de presión y las fuerzas de reacción del suelo se capturan en el sistema de coordenadas de la plataforma de fuerza [4]. De esta manera, se obtiene la gran cantidad de datos necesarios para llevar a cabo investigaciones en este ámbito.

3.2 Análisis de Componentes Principales (PCA)

El análisis de componentes principales (en inglés, PCA), es una técnica estadística utilizada para reducir la dimensionalidad de un conjunto de datos mientras se mantiene la mayor parte de su información original. Esta técnica fue creada por Pearson (1901) y desarrollada de forma independiente por Hotelling (1933), y se basa en la idea de que, en un conjunto de datos multivariado, las variables están correlacionadas entre sí, y estas correlaciones pueden describirse mediante una combinación lineal de las variables originales. El PCA busca transformar estas variables originales en un nuevo conjunto de variables, los componentes principales, que no están correlacionados. Los componentes principales están ordenados de forma que los primeros conservan la mayor parte de la variación presente en todas las variables originales [29].

Los objetivos del PCA son varios. Por un lado, busca identificar y extraer la información más importante o relevante contenida en un conjunto de datos. Al reducir la dimensionalidad de los datos, se eliminan las características menos significativas, lo que permite enfocarse en las relaciones más importantes entre las variables. Otra meta importante del PCA es comprimir el tamaño del conjunto de datos. Al transformar las variables originales en componentes principales, se crea una representación más compacta de los datos. Esto es especialmente útil cuando se trabaja con conjuntos de datos grandes, ya que reduce la carga computacional y el espacio de almacenamiento necesario. Además, el PCA simplifica la descripción de un conjunto

de datos al reemplazar un conjunto de variables correlacionadas por un conjunto más pequeño de componentes principales no correlacionados. Esto facilita la interpretación y el análisis de los datos, ya que se reduce la complejidad. También se utiliza para explorar y analizar la estructura de las observaciones y las variables en un conjunto de datos. Al identificar las direcciones principales en las que se extienden los datos, se muestran patrones y relaciones ocultas que pueden no ser evidentes en las variables originales. Finalmente, uno de los objetivos más destacados y utilizados del PCA es la reducción de dimensiones. Al representar los datos en un espacio de menor dimensión, se simplifica su representación sin una pérdida significativa de información. Esto es especialmente útil en visualización de datos y en la construcción de modelos de aprendizaje automático, donde la reducción de las dimensiones puede mejorar el rendimiento y reducir el riesgo de sobreajuste [29].

En problemas de regresión, la multicolinealidad (alta correlación entre las variables de entrada) puede afectar negativamente la precisión de los modelos. El PCA ayuda a eliminar la multicolinealidad al reemplazar las características originales por componentes principales no correlacionados. Esto mejora la estabilidad de los modelos de regresión al reducir la interferencia entre las variables de entrada.

3.3 Aprendizaje automático y regresión

El aprendizaje automático, es el proceso de resolver un problema práctico mediante la recopilación de un conjunto de datos y la construcción algorítmica de un modelo estadístico basado en ese conjunto de datos [30]. Este aprendizaje puede ser supervisado, semisupervisado, no supervisado y de refuerzo.

El aprendizaje supervisado utiliza un conjunto de datos etiquetados que consta de ejemplos (x, y). Cada ejemplo tiene un vector de características (x) y una etiqueta (y) [30].

El aprendizaje supervisado en problemas de clasificación implica predecir la pertenencia a categorías, como determinar si un correo es spam o no. En problemas de regresión, se predicen valores numéricos, como estimar el precio de una casa. Dado que nuestro estudio utiliza datos con etiquetas numéricas para predecir el progreso de pacientes, nos centraremos en el aprendizaje supervisado y en la regresión.

La regresión es un tipo de aprendizaje supervisado que busca predecir un valor numérico (llamado objetivo) a partir de un ejemplo etiquetado. La regresión con múltiples entradas y múltiples salidas es una extensión más compleja en la que se trata de predecir varios valores numéricos (objetivos) a partir de múltiples variables de entrada. Esto implica que el modelo debe capturar relaciones más sofisticadas entre las variables de entrada y las salidas. Por ejemplo, en el contexto de nuestra investigación, debemos intentar predecir la última sesión de rehabilitación de un paciente en base a la primera sesión de rehabilitación. La entrada (primera sesión de rehabilitación) y la salida (última sesión de rehabilitación) constan de 101 variables que representan los datos de fuerza de reacción del suelo en la dirección anteroposterior durante la fase de apoyo del ciclo de la marcha.

3.4 Algoritmos de aprendizaje automático para problemas de regresión

Existen varios algoritmos de aprendizaje automático para tareas de regresión. A continuación, se describirán los distintos algoritmos utilizados en esta investigación.

3.4.1 Mínimos cuadrados ordinarios (OLS)

Mínimos cuadrados ordinarios (en inglés, *Ordinary Least Square Regression* o OLS) es una técnica utilizada para estimar los parámetros (coeficientes) de un modelo de regresión

lineal. Se utiliza principalmente cuando se dispone de un conjunto de datos con una o más variables independientes y una única variable dependiente [30].

El objetivo principal de OLS es encontrar la relación lineal que mejor se ajuste entre las variables independientes (características) y la variable dependiente (objetivo) de tal forma que minimice la suma de las diferencias al cuadrado entre los valores observados y los valores predichos.

No obstante, existen unos supuestos que deben cumplirse. OLS asume que la relación entre las variables independientes y la variable dependiente es lineal. Esto significa que los cambios en las variables independientes conducen a cambios proporcionales en la variable dependiente. Además, OLS asume que existe un término de error aleatorio asociado a cada *data point*. Este término de error representa la discrepancia entre el valor observado y el valor predicho. Se supone que los errores se distribuyen normalmente con una media de cero.

El método OLS trata de encontrar los coeficientes (parámetros) de la ecuación lineal que mejor se ajusta a los datos. Para una regresión lineal simple con una variable independiente, la ecuación es la siguiente (1):

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

En la ecuación (1), Y es la variable dependiente (variable objetivo), y X es la variable independiente (característica). Por otro lado, β_0 es el intercepto, que representa el valor predicho de Y cuando X es cero, y β_1 es la pendiente, que representa el cambio en Y para un cambio de una unidad en X. Finalmente, ε es el término de error, que representa la diferencia entre los valores observados y predichos.

Los coeficientes se interpretan de la siguiente manera. Por un lado, β_1 representa el cambio en la variable dependiente para un cambio de una unidad en la variable independiente. Por otro lado, β_0 representa el valor previsto de la variable dependiente cuando la variable independiente es cero (el intercepto). Una vez estimados los coeficientes β_0 y β_1 , se utilizan para predecir la variable dependiente Y basándose en nuevos valores de la variable independiente X.

Para resumir, OLS es un método para ajustar una relación lineal entre variables minimizando la suma de las diferencias al cuadrado entre los valores observados y los predichos. Esta técnica estadística proporciona estimaciones del intercepto y la pendiente de la ecuación lineal que mejor describe los datos.

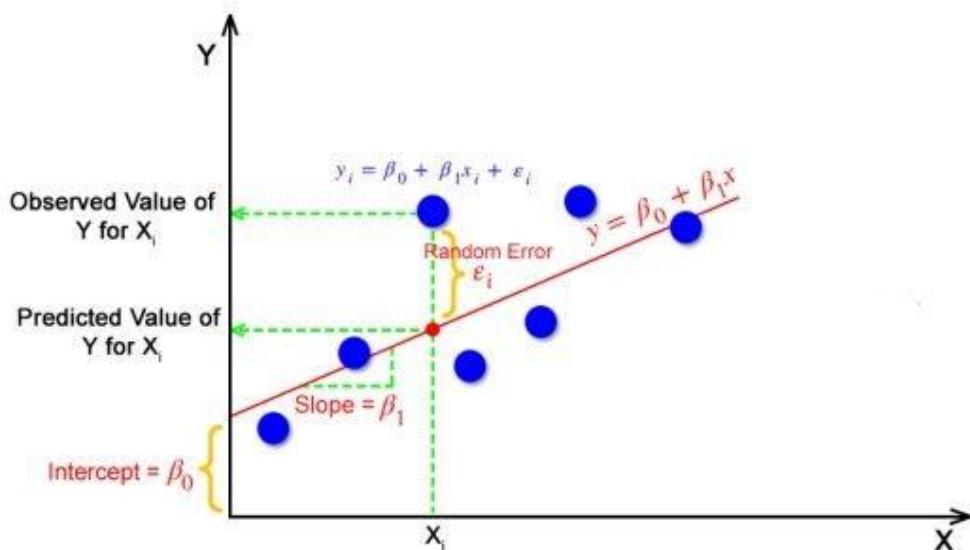


Figura 4. Representación gráfica de la regresión lineal [31].

3.4.2 Regresión de mínimos cuadrados parciales (PLS)

La regresión de mínimos cuadrados parciales (en inglés, *Partial Least Squares Regression* o PLS) es una técnica estadística utilizada para modelizar la relación entre un conjunto de variables independientes (características o predictores) y un conjunto de variables dependientes (objetivos) en situaciones en las que existen fuertes correlaciones y una posible multicolinealidad entre las características. El PLS resulta especialmente útil cuando se dispone de datos de gran dimensión y se desea reducir la dimensionalidad manteniendo la capacidad de predicción. Este método fue desarrollado por Herman Wold en 1966, inicialmente aplicado en ciencias sociales y posteriormente popularizado en quimiometría y evaluación sensorial debido a su hijo, Svante [32], [33].

El PLS introduce el concepto de variables latentes, también conocidas como componentes. Estas variables latentes son combinaciones lineales de las características originales. El objetivo del PLS es crear un conjunto de estas variables latentes que maximicen la covarianza entre las variables independientes y dependientes.

Esta técnica estadística realiza un proceso iterativo que empieza por encontrar la primera variable latente, que explica la mayor varianza tanto en las variables independientes como en las dependientes. A continuación, identifica la segunda variable latente que explica la mayor parte de la covarianza restante entre las variables, y así sucesivamente. El número de variables latentes a retener en el modelo es un hiperparámetro que puede ajustarse, por ejemplo mediante la validación cruzada (*cross-validation*), para conseguir un mejor rendimiento del modelo. Una vez obtenidas las variables latentes, el PLS construye un modelo de regresión lineal que relaciona las variables latentes con las variables dependientes. El modelo es similar a la regresión lineal múltiple, pero utiliza las variables latentes como predictores [34].

Aplicar la técnica del PLS tiene una serie de ventajas cuando se aplica en grandes conjuntos de datos, ya que reduce la dimensionalidad del espacio de características proyectando los datos sobre un conjunto más pequeño de variables latentes. Por otro lado, es robusto en el manejo de la multicolinealidad, que se produce cuando las variables independientes están altamente correlacionadas. Mediante la creación de variables latentes que capturan la varianza compartida entre las características, el PLS puede manejar la multicolinealidad más eficazmente que las técnicas de regresión tradicionales. Además, esta

técnica puede captar relaciones complejas entre los predictores y las variables objetivo, y realiza implícitamente la selección de características asignando menor importancia a las variables latentes con bajo poder explicativo. A diferencia de algunas técnicas de regresión que asumen independencia entre predictores, el PLS no hace esta asunción, haciéndolo adecuado para los datos que no cumplen esta condición [35].

3.4.3 Árboles de decisión

Los árboles de decisión (en inglés, *decision trees*) son modelos de aprendizaje automático que utilizan una estructura en forma de árbol para realizar predicciones de manera efectiva. A pesar de su asociación con problemas de clasificación, los árboles de decisión son igualmente útiles para problemas de regresión, donde la variable objetivo es un número real en lugar de una categoría.

Una característica fundamental de los árboles de decisión es su enfoque en la generalización en lugar de ajustarse estrictamente a los datos de entrenamiento. Esto se debe a su naturaleza de bajo sesgo y alta varianza, lo que significa que no hacen suposiciones significativas sobre la función objetivo. En cambio, se esfuerzan por identificar patrones y relaciones clave entre las características de entrada y la variable objetivo. Para lograr esto, los árboles de decisión toman decisiones jerárquicas basadas en las características de entrada, dividiendo los datos en subconjuntos basados en las características más relevantes [36].

La estructura de un árbol de decisión se asemeja a un diagrama de flujo (Fig. 5). Comienza con un nodo raíz que representa la población total y se ramifica en nodos de decisión basados en características específicas. Cada rama representa un resultado de la decisión y conduce a nodos principales que contienen predicciones o valores finales. Esta estructura es altamente interpretable y puede explicarse fácilmente incluso a personas que no sean técnicas.

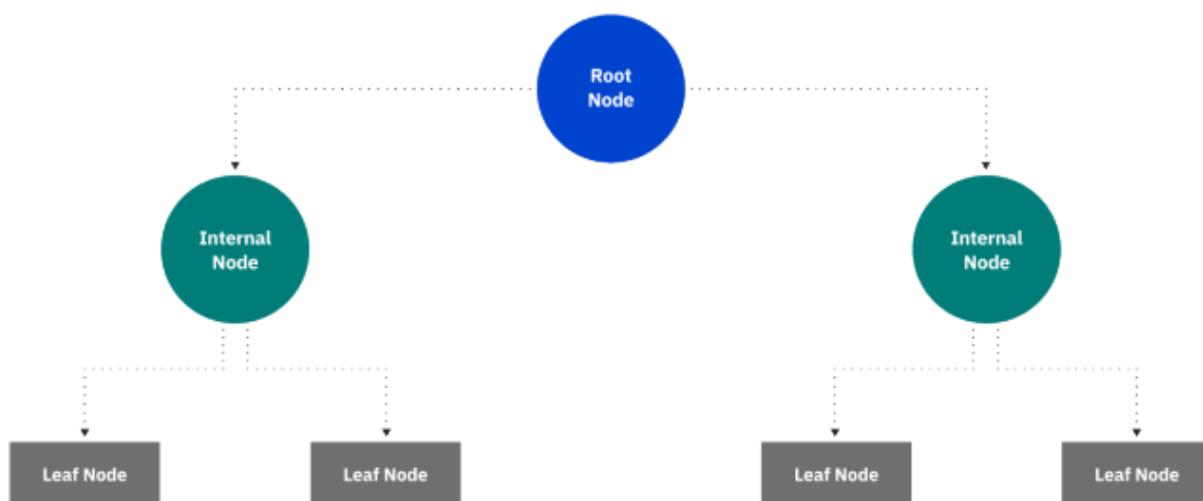


Figura 5. Estructura jerárquica del árbol de decisión [37].

Para construir un árbol de decisión, se utilizan diversas técnicas, como el índice de Gini, la ganancia de información, la prueba Chi-cuadrado y la entropía. Estas técnicas ayudan a determinar la relevancia de las características y a establecer las reglas que guían las divisiones en el árbol [37]. Por ejemplo, el índice de Gini mide la impureza de un conjunto de datos, donde un valor más bajo indica mayor pureza. Por otro lado, la ganancia de información se refiere a cuánta información se gana al dividir un conjunto de datos en función de una

característica específica. Otra métrica clave en los árboles de decisión es la entropía, que mide el nivel de incertidumbre o impureza en un conjunto de datos. Cuanto menor sea la entropía en un nodo, más puro será el conjunto de datos en ese punto, lo que significa que las muestras en ese nodo son más homogéneas en términos de la variable objetivo. Estas técnicas ayudan al árbol a decidir cuál es la mejor manera de separar los datos y crear ramificaciones en el árbol.

La poda o *pruning* es un proceso importante en la construcción de árboles de decisión. Implica la eliminación de subárboles no críticos para evitar el sobreajuste a los datos de entrenamiento. La poda puede realizarse durante la construcción del árbol (*pre-pruning*) o después de su construcción (*post-pruning*).

Los hiperparámetros más comunes que se pueden ajustar para optimizar el rendimiento de los modelos de árboles de decisión son la profundidad máxima del árbol de decisión (*max_depth*), el número mínimo de muestras necesarias para dividir un nodo interno (*min_samples_split*), y el número mínimo de muestras necesarias para estar en un nodo hoja (*min_samples_leaf*). La profundidad máxima del árbol de decisión controla la profundidad del árbol y, en consecuencia, su complejidad. Un árbol más profundo puede ajustarse mejor a los datos de entrenamiento, pero puede sobreajustarse. El número mínimo de muestras necesarias para dividir un nodo interno ayuda a controlar el sobreajuste especificando cuántos *data points* son necesarios para realizar una división. No obstante, un valor alto puede hacer que el árbol sea más conservador en las divisiones. Finalmente, el número mínimo de muestras necesarias para estar en un nodo hoja ayuda a controlar el sobreajuste asegurando que cada hoja tiene un número mínimo de *data points* [38].

3.4.4 Bosques aleatorios

Los bosques aleatorios, también conocidos como *random forests*, son una técnica de aprendizaje automático que se basa en la construcción de múltiples árboles de decisión individuales que trabajan en conjunto para lograr predicciones más precisas y robustas. Esta técnica se utiliza en problemas de clasificación y regresión, y se puede aplicar para la selección de características. En los problemas de clasificación, cada árbol del bosque aleatorio predice la etiqueta de clase, y la predicción final se determina por mayoría, es decir, la clase con más votos entre los árboles. En los problemas de regresión, cada árbol predice un valor numérico, y la predicción final suele ser la media (o media ponderada) de estos valores [39].

La idea clave detrás de los bosques aleatorios es reducir el sobreajuste y mejorar la capacidad de generalización del modelo. Esto se logra mediante un proceso de ensamblaje, donde se construye un conjunto de árboles de decisión, cada uno entrenado en un subconjunto aleatorio de datos y características. Esto se conoce como muestreo con reemplazo o *bootstrap sampling*, y permite que cada árbol se especialice en ciertas características y patrones de los datos. Además, en cada nodo de decisión de cada árbol, se realiza una aleatorización de características, lo que significa que se considera solo un subconjunto aleatorio de las características para tomar decisiones. Esto ayuda a reducir la correlación entre los árboles individuales y hace que el conjunto sea más robusto frente a diferentes tipos de datos. En el proceso de *bootstrap sampling* o muestreo con reemplazo, se toma un conjunto de datos original y se crea un conjunto de entrenamiento para cada árbol de decisión del bosque aleatorio. Esto significa que, al seleccionar muestras para un conjunto de entrenamiento específico, se permite que una misma muestra aparezca múltiples veces y, al mismo tiempo, algunas muestras pueden no ser seleccionadas. Este proceso crea diversidad en los conjuntos de entrenamiento para cada árbol. Cada conjunto de entrenamiento es ligeramente diferente y contiene una mezcla única de muestras del conjunto de datos original. Como resultado, cada árbol del bosque aleatorio se entrena en un subconjunto ligeramente diferente de datos y, por lo tanto, se especializa en ciertas características o patrones de ese conjunto de datos específico [40].

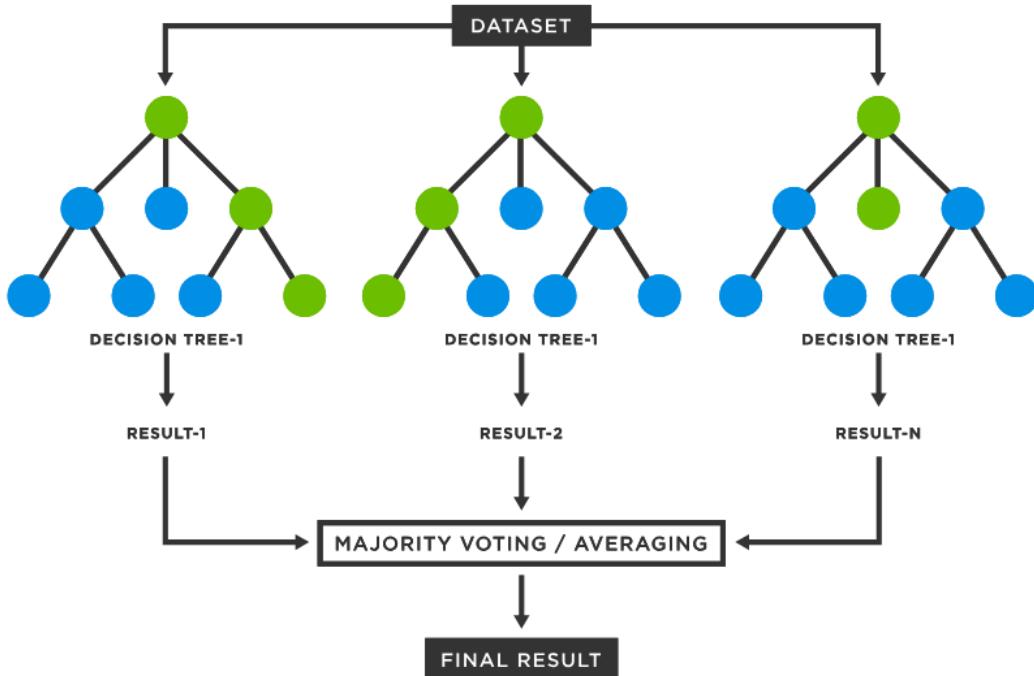


Figura 6. Esquema del bosque aleatorio [41].

Los bosques aleatorios ofrecen varias ventajas, como la reducción del sobreajuste gracias a la combinación de múltiples árboles, una mejor capacidad de generalización, robustez frente a datos con ruido y valores atípicos o *outliers*, y la capacidad de proporcionar información sobre la importancia de las características para la predicción.

Para optimizar el rendimiento de un modelo de bosque aleatorio, es posible ajustar varios hiperparámetros, como el número de árboles en el conjunto (*n_estimators*), la profundidad máxima de cada árbol (*max_depth*), el número mínimo de muestras requeridas para dividir un nodo (*min_samples_split*), el número mínimo de muestras requeridas para estar en un nodo hoja (*min_samples_leaf*), y el número de características a considerar al realizar divisiones (*max_features*). De la misma forma que en los árboles de decisión, cuanto mayor sea la profundidad máxima permitida para cada árbol de decisión en el bosque, más complejos pueden ser los árboles individuales, lo que puede llevar a un mayor riesgo de sobreajuste si no se controla. Por otro lado, una profundidad más baja puede producir árboles más simples y menos propensos al sobreajuste. Un valor bajo del número de muestras mínimas que se deben tener en un nodo hoja (terminal) de un árbol, permite nodos hoja con pocas muestras, lo que podría generar árboles más profundos y propensos al sobreajuste. Por el contrario, un valor alto garantiza que las hojas contengan más muestras, simplificando el árbol. Limitar el número de características puede ayudar a reducir la correlación entre árboles, lo que hace que el bosque sea más robusto [42].

Finalmente, es importante señalar que debido a que están formados por múltiples árboles de decisión, requieren más recursos computacionales en términos de memoria y tiempo de cálculo en comparación con un solo árbol. Su interpretación también puede ser más compleja que la de un solo árbol de decisión.

3.4.5 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) es un algoritmo de aprendizaje automático desarrollado por Tianqi Chen (2014) [43] utilizado tanto en problemas de clasificación como de regresión. Como los bosques aleatorios, XGBoost es un algoritmo de ensamblado que combina múltiples modelos de aprendizaje automático para mejorar la precisión de las predicciones.

La idea principal detrás de XGBoost es construir un conjunto de modelos de árboles de decisión débiles (generalmente árboles poco profundos) y combinar sus predicciones para obtener una predicción final más precisa y robusta. A diferencia de los bosques aleatorios, que construyen árboles de decisión de manera paralela e independiente, XGBoost utiliza un proceso de construcción secuencial, lo que lo hace extremadamente efectivo en la reducción del error y el aumento de la precisión.

El algoritmo XGBoost funciona de la siguiente manera. Este algoritmo utiliza árboles de decisión como estimadores base, donde cada árbol se crea de manera secuencial para corregir los errores cometidos por los árboles anteriores. Estos árboles son generalmente árboles poco profundos para evitar el sobreajuste. XGBoost utiliza una función de pérdida (*loss function*) específica para el problema que se está abordando, ya sea regresión o clasificación. La función de pérdida mide cuán lejos están las predicciones del modelo de los valores reales. El objetivo, es minimizar esta función de pérdida ajustando los parámetros del modelo. En la función de pérdida se incluyen los términos de regularización, que penalizan los modelos muy complejos y ayudan a mantener los árboles poco profundos para evitar el sobreajuste. Como en los árboles de decisión, XGBoost utiliza criterios de división (como la ganancia de información) para determinar cómo dividir los nodos de los árboles en cada iteración. No obstante, XGBoost utiliza el enfoque de *Gradient Boosting* (Fig. 7), que ajusta los árboles en función de los errores residuales de los árboles anteriores. Esto permite que cada nuevo árbol se enfoque en las áreas del espacio de características donde los modelos anteriores cometieron errores [44], [45].

Debido a las características mencionadas anteriormente, este algoritmo ofrece una serie de ventajas, como su alto rendimiento, su robustez frente a datos con ruido y valores atípicos gracias a la regularización, y su eficiencia computacional incluso en grandes conjuntos de datos.

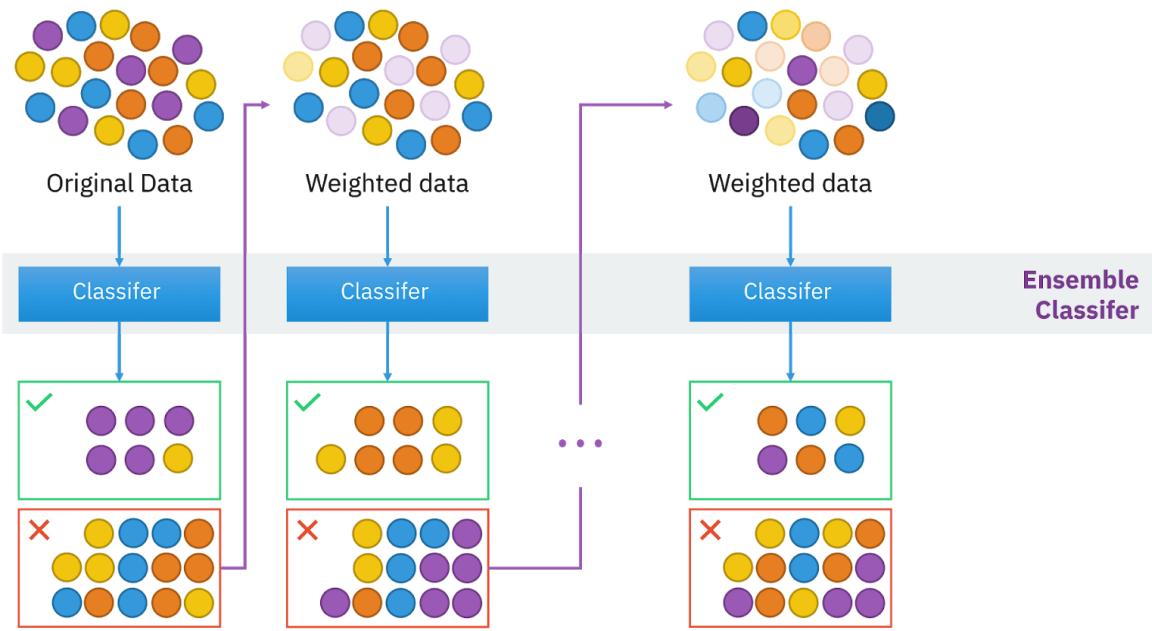


Figura 7. Esquema del algoritmo de *boosting* [44].

Para optimizar el rendimiento de un modelo XGBoost, se pueden ajustar varios hiperparámetros, como el número de estimadores o árboles en el ensamble (*n_estimators*), la tasa de aprendizaje (*learning_rate*), el peso mínimo requerido en un nodo hijo para continuar dividiendo el nodo (*min_child_weight*), la profundidad máxima de cada árbol (*max_depth*), la proporción de muestras utilizadas en cada iteración (*subsample*), la proporción de características utilizadas en cada árbol (*colsample_bytree*), la proporción de características utilizadas en cada nivel del árbol (*colsample_bylevel*), el parámetro de regularización L2 o Ridge (*reg_lambda*), el parámetro de regularización L1 o Lasso (*reg_alpha*), y el parámetro que controla cuánta reducción de pérdida es necesaria para realizar una partición (*gamma*) [42]. Sin embargo, aunque es un algoritmo muy efectivo, es más sensible a la selección de estos hiperparámetros y menos interpretable que otros modelos más simples.

3.4.6 Regresión de vectores de soporte (SVR)

La regresión de vectores de soporte (en inglés, *Support Vector Regression* o SVR) es un algoritmo de aprendizaje automático utilizado para resolver problemas de regresión, que se basa en el concepto de máquinas de vectores de soporte (en inglés, *Support Vector Machines* o SVM). Las máquinas de vectores de soporte, introducidas en 1995 por Vapnik et al. [46], originalmente se desarrollaron para problemas de clasificación. Sin embargo, posteriormente, se creó una nueva versión para los problemas de regresión [47].

El objetivo principal de SVR es encontrar una función de regresión que se ajuste a los datos de entrenamiento mientras mantiene un margen máximo a ambos lados de la función de regresión. A diferencia del SVM para clasificación, en SVR no se trabaja con ejemplos de diferentes clases, sino con variables de entrada y variables objetivo (etiquetas numéricas). La función de regresión busca minimizar la pérdida y al mismo tiempo asegurar que la mayoría de los *data points* de entrenamiento (no solo los más cercanos) estén dentro del margen especificado. Los *data points* que están dentro del margen se denominan "vectores soporte".

El algoritmo SVR empieza transformando los datos de entrenamiento para que se ajusten a un rango específico. Esto se hace utilizando una función de transformación llamada kernel, que puede ser lineal, polinómica, sigmoidal, o radial. Esta última es útil para datos no lineales.

La elección de la función kernel y sus parámetros afecta significativamente el rendimiento del modelo SVR y su capacidad para adaptarse a patrones complejos en los datos.

Se establece un margen de tolerancia alrededor de la función objetivo, que consiste en una banda o tubo. El objetivo es que la mayoría de las muestras de entrenamiento se encuentren dentro de este tubo. Los vectores de soporte son las muestras de entrenamiento que se encuentran en el límite del tubo o dentro de él. Estos puntos son críticos para la definición de la función de regresión. Como se ha mencionado anteriormente, el SVR busca encontrar la función de regresión que minimice el error, es decir, que permita que la mayoría de los vectores de soporte se encuentren dentro del margen de tolerancia. Esto se hace mediante técnicas de optimización, como la optimización cuadrática, con el objetivo de ajustar los parámetros de la función de manera óptima [48], [49].

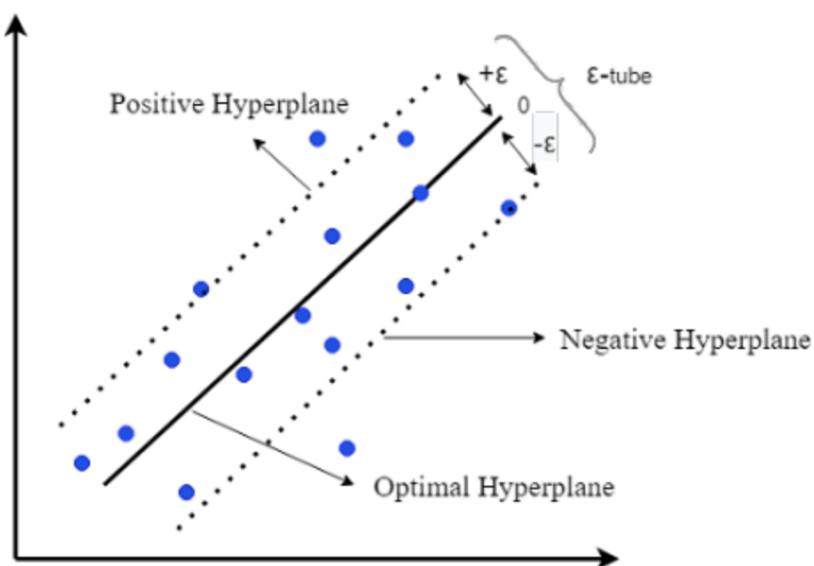


Figura 8. Parámetros de la regresión de vectores de soporte [50].

Debido a las características mencionadas anteriormente, este algoritmo ofrece una serie de ventajas. El algoritmo SVR tiene la capacidad de trabajar con datos no lineales, incluye términos de regularización que ayudan a controlar la complejidad del modelo y a evitar el sobreajuste, y es eficaz incluso en espacios de características de alta dimensión.

Para optimizar el rendimiento de un modelo de SVR, se pueden ajustar varios hiperparámetros, como el parámetro de regularización que controla la penalización por errores en la función de regresión (C), el parámetro que establece el ancho del margen de tolerancia alrededor de la función objetivo(ϵ), la función de *kernel* utilizada (*kernel*), el coeficiente del *kernel*, que controla la influencia de cada muestra en la función de regresión (*gamma*), y el grado del *kernel* polinómico (*degree*) [42].

Aunque el algoritmo SVR es capaz de manejar datos no lineales y es efectivo en problemas de regresión, se ha de señalar que es sensible a la elección de hiperparámetros y costoso computacionalmente, especialmente en grandes conjuntos de datos.

3.5 Métricas de evaluación de los modelos de regresión

La evaluación de modelos de regresión es esencial para determinar su rendimiento y su capacidad para hacer predicciones precisas. Para medir la calidad y el rendimiento de un modelo de regresión, se utilizan diversas métricas de evaluación. En este apartado, se describen tres de las métricas más comunes utilizadas para evaluar modelos de regresión: la raíz del error cuadrático medio (RMSE), el coeficiente de determinación (R^2) y el coeficiente de correlación de Pearson.

3.5.1 Raíz del error cuadrático medio (RMSE)

La raíz del error cuadrático Medio (en inglés, *Root Mean Square Error* o RMSE) mide la precisión de las predicciones de un modelo de regresión al calcular la raíz cuadrada de la media de los errores cuadráticos. Se expresa matemáticamente de la siguiente manera:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (2)$$

En la fórmula (2), n es el número de muestras en el conjunto de datos de prueba, y_i representa el valor real de la variable objetivo para la i -ésima muestra, e \hat{y}_i es la predicción del modelo para la i -ésima muestra [51].

El RMSE cuantifica la diferencia entre las predicciones del modelo y los valores reales. Cuanto menor sea el valor del RMSE, mejor será el ajuste del modelo a los datos. Es decir, un RMSE más bajo indica que el modelo tiene un mejor rendimiento en términos de precisión.

3.5.2 Coeficiente de determinación (R^2)

El coeficiente de determinación (R^2), proporciona información sobre la capacidad del modelo para explicar la variabilidad de los datos. Es decir, se interpreta como la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes [52]. El R^2 se calcula de la siguiente manera:

$$R^2 = \sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{(y_i - \bar{y})^2} \quad (3)$$

En la fórmula (3), n es el número de muestras en el conjunto de datos de prueba, y_i representa el valor real de la variable objetivo para la i -ésima muestra, \hat{y}_i es la predicción del modelo para la i -ésima muestra, e \bar{y} es la media de los valores reales y_i en el conjunto de prueba [51].

El coeficiente de determinación tiene como límite superior el valor 1, mientras que no tiene límite inferior [52]. Un valor de R^2 de 1 indica un ajuste perfecto, mientras que un valor de R^2 cercano a 1 indica que el modelo es capaz de explicar una gran parte de la variabilidad de los datos, indicando un buen ajuste del modelo. Por otro lado, un R^2 cercano a 0 indica que el modelo no puede explicar bien la variabilidad de los datos y tiene un mal rendimiento. Si el R^2 es negativo, significa que los valores predichos por el modelo son peores que si utilizáramos la media como valor predicho.

3.5.3 Coeficiente de correlación de Pearson

El coeficiente de correlación de Pearson (r), evalúa la relación lineal entre las variables predictoras y la variable objetivo es decir, mide la fuerza y la dirección de la correlación entre estas variables.

El coeficiente de correlación de Pearson varía entre -1 y 1. Un valor de cercano a 1 indica una fuerte correlación positiva entre las variables, lo que significa que cuando una variable aumenta, la otra también tiende a aumentar. Por otro lado, un valor de cercano a -1 indica una fuerte correlación negativa, lo que significa que cuando una variable aumenta, la otra tiende a disminuir. Un valor cercano a 0 indica una correlación débil o inexistente entre las variables [53].

4 Métodos

En esta sección se presentan los procedimientos que se han llevado a cabo para realizar este proyecto y dar respuesta a los objetivos planteados. Se mencionan las herramientas y el software utilizado, y se describe exhaustivamente el conjunto de datos GaitRec. A continuación, se explica con qué criterios se construyó el conjunto de datos específico para esta investigación, se presenta el análisis exploratorio de los datos, y se detalla el proceso de división de los datos en vectores de entrada y salida para la fase de entrenamiento y prueba. Posteriormente, se explican las diferentes aproximaciones de preprocesamiento de los datos. Finalmente, se describe en profundidad el análisis multivariante.

4.1 Herramientas y software

En el presente estudio se empleó el lenguaje de programación Python (versión 3.9.16) para el procesamiento de datos, el análisis y el desarrollo de modelos de aprendizaje automático. Se eligió Python por su amplia gama de bibliotecas y herramientas adaptadas a las aplicaciones de la ciencia de datos y el aprendizaje automático.

Se utilizaron las siguientes bibliotecas clave para facilitar diversos aspectos de la investigación:

- Pandas (versión 2.0.2): Pandas se empleó para manipular datos de forma eficiente [54].
- NumPy (versión 1.25.0): la biblioteca NumPy desempeñó un papel fundamental en los cálculos numéricos y las manipulaciones de matrices ya que permite realizar operaciones en matrices y *arrays* multidimensionales, esenciales para diversas tareas de análisis de datos [55].
- Matplotlib (versión 3.7.1) y Seaborn (versión 0.12.2): estas bibliotecas de visualización se utilizaron para crear distintos tipos de gráficos para el análisis exploratorio de datos y la presentación de resultados. Matplotlib proporciona un marco flexible para la creación de visualizaciones estáticas, interactivas y de calidad [56], mientras que Seaborn simplifica la creación de gráficos haciéndolos más estéticos [57].
- SciPy (versión 1.11.1): la biblioteca SciPy se utilizó para realizar cálculos estadísticos [58].
- Statsmodels (versión 0.14.0): statsmodels se utilizó para la construcción de modelos, la evaluación de la heteroscedasticidad y la preparación de los datos para el análisis de regresión lineal [59].
- Scikit-Learn (versión 1.3.0): la biblioteca de aprendizaje automático Scikit-Learn ofrece una amplia gama de algoritmos regresión, reducción dimensional, etc [60]. Es por ello que se empleó para la selección, el entrenamiento y la evaluación de modelos, así como para la reducción de la dimensionalidad de los datos de entrada.
- XGBoost (versión 1.7.6): XGBoost es una biblioteca distribuida optimizada de *gradient boosting* diseñada para ser altamente eficiente, flexible y portable [43]. Es por ello que se utilizó para construir y ajustar modelos de regresión en el marco del *Gradient Boosting*.

A parte de las bibliotecas anteriormente mencionadas, se adoptaron herramientas y prácticas específicas para garantizar un flujo de trabajo eficaz y organizado. Se eligió Visual Studio Code como principal entorno de desarrollo integrado para la codificación y gestión del proyecto, ya que ofrece una interfaz fácil de usar y mejora el proceso de desarrollo. Para aislar las dependencias del proyecto y garantizar la coherencia entre distintos entornos, se creó un

entorno virtual utilizando el módulo *venv*. Esto ayudó a gestionar las versiones de las bibliotecas y paquetes necesarios para el proyecto, minimizando los posibles conflictos. Además, se empleó Git como sistema de control de versiones para realizar un seguimiento de los cambios, colaborar y gestionar las distintas versiones del proyecto [61]. GitHub, una plataforma de desarrollo colaborativo, se utilizó para alojar el repositorio del proyecto y facilitar el control de versiones [62].

4.2 El conjunto de datos GaitRec

4.2.1 Introducción al conjunto de datos GaitRec

El conjunto de datos GaitRec, presentado por Horsak et al. [4], es un *dataset* a gran escala, exhaustivo y ampliamente etiquetado, diseñado para facilitar la investigación en el análisis de las fuerzas de reacción del suelo (GRF) durante la locomoción humana. Este conjunto de datos proviene de una base de datos clínica de la marcha mantenida por un centro de rehabilitación austriaco. La recopilación de datos se llevó a cabo entre 2007 y 2018, con la participación de pacientes con diversas discapacidades musculoesqueléticas y controles sanos. Es importante destacar que, antes de publicar del conjunto de datos, se obtuvo la aprobación del Comité de Ética local de Baja Austria (GS1-EK-4/299-2014).

4.2.2 Preparación y procesamiento

Antes de llevar a cabo las sesiones de análisis de la marcha, cada participante se sometió a un riguroso examen físico realizado por un médico.

Para la obtención de los datos, se solicitó a los participantes que caminaran sin ayuda a una velocidad de marcha elegida por ellos mismos en una pasarela de aproximadamente 10 metros que contaba con dos plataformas de fuerza incrustadas en su centro. Estas plataformas se colocaron en orden secuencial y al ras del suelo. Cada paciente realizó múltiples paseos o *trials* durante cada sesión de evaluación o grabación de la marcha. Durante cada sesión, los sujetos caminaron hasta obtener un número mínimo de grabaciones válidas, que en la mayoría de los casos fue de diez grabaciones. Se consideraron grabaciones válidas cuando los pacientes caminaban de manera natural y había un claro impacto del pie en cada plataforma de fuerza. Los pacientes caminaron descalzos, con su calzado ortopédico o calzado normal, y en algunos casos con plantillas ortopédicas. En el caso de los controles sanos, caminaron descalzos o con su calzado normal.

Se capturaron tres señales analógicas de fuerza de reacción del suelo (componente vertical, anteroposterior y medio-lateral), así como el centro de presión (COP), que posteriormente se convirtieron en señales digitales. El centro de presión y las fuerzas de reacción del suelo se registraron en el sistema de coordenadas de la plataforma de fuerza [4].

Para facilitar el análisis, se estandarizó la orientación de las señales medio-lateral y anteroposterior en todos los datos, de manera que las fuerzas medial y anterior siempre se representaran como valores positivos.

Debido a los protocolos internos del centro, los datos en bruto presentaban ruido al principio y al final de las señales. Con el propósito solucionar este problema y mejorar la calidad de los datos, se aplicó un umbral de 25 newtons (N) a todos los datos de fuerza y luego se calculó el centro de presión. Estos datos son las señales de fuerza de reacción del suelo sin procesar (en bruto). Además, se generaron datos postprocesados *ready to use*. Para ello, el centro de presión solo se calculó cuando la fuerza vertical alcanzó los 80 N a fin de evitar imprecisiones en el cálculo del centro de presión en valores de fuerza pequeños [4].

A continuación, se centraron las coordenadas medio-laterales del centro de presión en el valor medio, mientras que las coordenadas anteroposteriores se centraron en cero. Las señales de fuerza procesadas se filtraron mediante un filtro de paso bajo de segundo orden con una frecuencia de corte de 20 Hz, a fin de reducir el ruido, y se normalizaron en el tiempo

al 100% de la fase de apoyo (o *stance* en inglés) del ciclo de la marcha, es decir, 101 puntos. Los valores de amplitud de las tres componentes de fuerza se expresaron como múltiplos del peso corporal (BW), dividiendo la fuerza por el producto de la masa corporal y la aceleración debida a la gravedad [4].

Tanto la normalización de la amplitud como la normalización temporal son operaciones necesarias para reducir los efectos de las covariables (como la antropometría) en las señales y reducir las diferencias temporales que dificultan las comparaciones de diferentes pasos [63], [63], [64].

Además, implementaron el algoritmo propuesto por Sangeux y Polak [65] para eliminar cualquier valor atípico antes de incluirlo en el conjunto de datos.

4.2.3 Descripción y anotaciones

El conjunto de datos GaitRec contiene mediciones anonimizadas de fuerzas de reacción del suelo de un total de 2085 pacientes con deficiencias musculoesqueléticas (*gait disorders*, GD) y 211 controles sanos (HC).

Cada registro incluye metadatos adicionales como el identificador del sujeto, el identificador de la sesión, etiquetas que proporcionan información sobre la afectación específica, la edad, el sexo, el tipo de calzado, la velocidad de la marcha, si ese registro pertenece al conjunto de entrenamiento (balanceado o no) o al conjunto de prueba, etc. La descripción de todos los metadatos se muestra en la Tabla 1.

Los individuos clasificados como controles sanos se seleccionaron bajo la condición de no experimentar dolor o molestias en las extremidades inferiores ni en la columna vertebral, además de no llevar ningún tipo de ortesis ni plantillas ortopédicas. Se excluyeron aquellos que tenían antecedentes de cirugía o traumatismo en la columna vertebral o las extremidades inferiores [4].

La duración habitual de la estancia de los pacientes en el centro de rehabilitación varía desde algunos días hasta varias semanas, influenciada por elementos como el diagnóstico, el tratamiento, la terapia, la cirugía realizada y el avance en la recuperación. Durante este período de tiempo, los pacientes se someten periódicamente a análisis de la marcha, normalmente una vez por semana. Cuando los pacientes ingresan en el centro de rehabilitación, se establecen los resultados esperados de la terapia en colaboración con el terapeuta. Una vez que los objetivos se logran total o parcialmente, se procede a dar de alta a los pacientes, dándoles la opción de ingresar de nuevo si fuese necesario.

El presente conjunto de datos contiene los datos recopilados durante toda la estancia o estancias de cada paciente, abarcando así todo el progreso de rehabilitación de los pacientes [4]. El *dataset* fue etiquetado manualmente por un fisioterapeuta experimentado basándose en el diagnóstico médico disponible de cada paciente, y se clasifica jerárquicamente en una clase de controles sanos (HC) y cuatro clases basadas en el nivel de afectación de la articulación anatómica: cadera (H), rodilla (K), tobillo (A) y calcáneo (C). Las etiquetas de anotación proporcionan información sobre la afectación específica y su localización anatómica. En la Tabla 2 se muestra el resumen demográfico del conjunto de datos y las clases predefinidas. Esta extensa anotación permite un análisis en profundidad y la clasificación de los patrones de la marcha a través de diferentes tipos de deficiencias.

Tabla 1. Descripción de la información guardada en el fichero de metadatos [4].

Categorías/Variables	Formato	Unidad	Descripción
Identificadores			
SUBJECT_ID	integer	—	Identificador único de un sujeto
SESSION_ID	integer	—	Identificador único de una sesión
Etiquetas			
CLASS_LABEL	string	—	Etiquetas de clase anotadas
CLASS_LABEL_DETAILED	string	—	Etiquetas de clase anotadas para subclases
Metadatos de los sujetos			
SEX	binary	—	femenino = 0 masculino = 1
AGE	integer	años	Edad en la fecha de registro
HEIGHT	integer	centímetro	Altura corporal en centímetros
BODY_WEIGHT	double	Kg·m s ²	Peso corporal en Newton
BODY_MASS	double	kg	Masa corporal
SHOE_SIZE	double	EU	Número de calzado en el Sistema Europeo Continental
AFFECTED_SIDE	integer	—	izquierda = 0 derecha = 1 ambas = 2
Metadata de los paseos o <i>trials</i>			
SHOD_CONDITION	integer	—	descalzo y calcetines = 0 zapato normal = 1 zapato ortopédico = 2
ORTHOPEDIC_INSOLE	binary	—	sin plantilla = 0 con plantilla = 1
SPEED	integer	—	Velocidad de la marcha: lenta = 1 autoseleccionada = 2 rápida = 3
READMISSION	integer	—	indica el número de reingresos = 0 ... n
SESSION_TYPE	integer	—	medición inicial = 1 medición de control = 2 medición inicial tras readmisión = 3
SESSION_DATE	string	—	fecha de la sesión de registro en el formato "DD-MM-YYYY "
Información de Train-Test Split			
TRAIN	binary	—	forma parte (=1) o no forma parte (=0) de TRAIN
TRAIN_BALANCED	binary	—	forma parte (=1) o no forma parte (=0) de TRAIN_BALANCED
TEST	binary	—	forma parte (=1) o no forma parte (=0) de TEST

Tabla 2. Resumen demográfico del conjunto de datos y las clases predefinidas [4].

Clases	N	Edad (años) Media (SD)	Masa corporal (kg) Media (SD)	Género (m/f)	Número de paseos o <i>trials</i> bilaterales
Control sano	211	34.7 (13.9)	73.9 (15.6)	104/107	7755
Cadera	450	42.6 (12.8)	73.9 (15.6)	373/77	12748
Rodilla	625	41.6 (12.0)	84.3 (18.6)	426/199	19873
Tobillo	627	41.6 (11.4)	87.0 (18.0)	498/129	21386
Calcáneo	382	43.5 (10.4)	84.0 (14.5)	339/43	13970
Total	2295	41.5 (12.1)	83.6 (17.3)	1740/555	75732

4.2.4 Estructura de los datos

Todos datos se encuentran disponibles en formato de valores separados por comas (CSV). Los veinte archivos de datos de fuerzas de reacción del suelo (GRF) están organizados siguiendo una convención de nomenclatura "GRF-tipo-procesamiento-lado.csv" como se muestra en la Tabla 3. Aquí, "tipo" indica si el archivo contiene series temporales verticales ("F_V"), anteroposteriores ("F_AP"), medio laterales ("F_ML"), o centro de presión (COP) anteroposteriores o medio laterales ("COP_AP", "COP_ML"). El término "procesamiento" señala si los archivos contienen datos en bruto ("RAW") o datos postprocesados ("PRO"). Finalmente, "lado" indica si los datos provienen del lado derecho o izquierdo del cuerpo. Por tanto, se proporcionan los datos en dos formatos. Por un lado, datos en formato original (datos brutos o crudos), y por otro lado, datos postprocesados sometidos a procesos de reducción del ruido y normalización. Esto da flexibilidad a futuros investigadores, permitiéndoles utilizar los datos crudos y aplicar sus propios procesamientos, o emplear los datos postprocesados directamente.

Los archivos "GRF-tipo-procesamiento-lado.csv" se estructuran de manera matricial con N filas por M columnas. En cada fila se encuentran los datos correspondientes a un sujeto y una sesión de grabación específica. La primera columna identifica al sujeto ("SUBJECT_ID"), la segunda a la sesión de grabación ("SESSION_ID"), y la tercera al paseo o *trial* individual dentro de dicha sesión ("TRIAL_ID"). Del mismo modo, el archivo de metadatos ("GRF-metadata.csv") que contiene anotaciones y datos adicionales relacionados con el tema, se estructura en una matriz con N filas × M columnas (véase la Tabla 3).

Con el propósito de poder evaluar con precisión el rendimiento de los modelos de aprendizaje automático, en el *dataset* viene incorporada una división predefinida y aleatoria en tres subconjuntos para su uso en el entrenamiento y evaluación. El conjunto de datos GaitRec se segmenta en un conjunto de entrenamiento desequilibrado (TRAIN) y un conjunto de prueba (TEST). El primero se emplea para entrenar y perfeccionar modelos de aprendizaje automático utilizando métodos como la validación cruzada, mientras que el segundo se destina a la evaluación definitiva. No obstante, la disparidad entre las clases en el conjunto de entrenamiento puede influir negativamente en la optimización de los modelos. Para abordar esto, los autores crearon un subconjunto equilibrado de entrenamiento, denominado TRAIN_BALANCED, que incluye exclusivamente los datos correspondientes a las evaluaciones iniciales (primera sesión de medición), en la cual se contabilizan al menos cinco paseos o *trials* para cada lado del cuerpo por sesión.

Tabla 3. Descripción de los datos almacenados en los ficheros "GRF_*.csv". n es el número de fotogramas durante un paso a través de la plataforma de fuerza para los datos sin procesar ("RAW") o un vector normalizado en el tiempo de 101 puntos para los datos postprocesados ("PRO") [4].

Variables	Fichero asociado	Formato	Dimensión	Unidad	Descripción
Vertical GRF	GRF_F_V-RAW_*.csv	<i>double</i>	$1 \times n$	Newton	Fuerza vertical de reacción al suelo bruta
Anterior-posterior GRF	GRF_F_AP-Raw_*.csv	<i>double</i>	$1 \times n$	Newton	Fuerza de frenado y propulsión bruta
Medio-lateral GRF	GRF_F_ML_RAW_*.csv	<i>double</i>	$1 \times n$	Newton	Fuerza medio-lateral bruta
COP anterior-posterior	GRF_COP_AP_RAW_*.csv	<i>double</i>	$1 \times n$	Centímetro	Coordenada COP bruta en dirección de marcha
COP medio-lateral	GRF_COP_ML_RAW_*.csv	<i>double</i>	$1 \times n$	Centímetro	Coordenada COP bruta en dirección mediolateral
Vertical GRF	GRF_F_V_PRO_*.csv	<i>double</i>	$1 \times n$	Múltiplo del peso corporal	Fuerza de reacción vertical al suelo postprocesada
Anterior-posterior GRF	GRF_F_AP_PRO_*.csv	<i>double</i>	$1 \times n$	Múltiplo del peso corporal	Fuerza de frenado y propulsión postprocesada
Medio-lateral GRF	GRF_F_ML_PRO_*.csv	<i>double</i>	$1 \times n$	Múltiplo del peso corporal	Fuerza medio-lateral postprocesada
COP anterior-posterior	GRF_COP_AP_PRO_*.csv	<i>double</i>	$1 \times n$	% apoyo (<i>stance</i>)	Coordenada COP postprocesada en la dirección de la marcha
COP medio-lateral	GRF_COP_ML_PRO_*.csv	<i>double</i>	$1 \times n$	% apoyo (<i>stance</i>)	Coordenada COP post-procesada en dirección medio-lateral

4.3 Construcción de los conjuntos de datos

Para poder realizar nuestra investigación, primero de todo tuvimos que crear, a partir de los diferentes archivos en formato CSV que proporciona GaitRec, nuestros conjuntos de datos personalizados para que se adapte a los objetivos de nuestro proyecto.

Decidimos trabajar con el conjunto de datos postprocesados ("PRO") sometidos a procedimientos de eliminación de ruido y normalización. Trabajar con el conjunto de datos postprocesado, más limpio y fiable, coherente y estandarizado, y sin valores atípicos ofrece una serie de ventajas. Algunas de ellas son el ahorro de recursos computacionales y tiempo, la facilidad de reproducir nuestro estudio, así como comparar y contrastar los resultados obtenidos con otras futuras investigaciones que utilicen los mismos métodos de preprocesamiento o similares.

Optamos por trabajar con el conjunto de entrenamiento desbalanceado ("TRAIN") en lugar de utilizar el subconjunto balanceado ("TRAIN_BALANCED"). El subconjunto balanceado ("TRAIN_BALANCED") comprende sólo los datos de las evaluaciones iniciales (primera sesión de medición), que contienen al menos cinco paseos o *trials* para cada lado del cuerpo por

sesión [4]. Nuestra elección se basa en la necesidad de considerar la variabilidad y la evolución completa de los datos a lo largo del proceso de terapia y rehabilitación.

A continuación, creamos un archivo CSV con los metadatos de los pacientes pertenecientes solamente al conjunto de entrenamiento no balanceado, y otro archivo CSV con los metadatos de los pacientes pertenecientes solamente al conjunto de prueba.

Los archivos "GRF_tipo_procesamiento_lado.csv", que contienen los datos postprocesados para cada una de las extremidades inferiores de los pacientes presentes en el conjunto de entrenamiento y prueba, se guardan en la carpeta "postprocessed" (véase la Fig. 9).

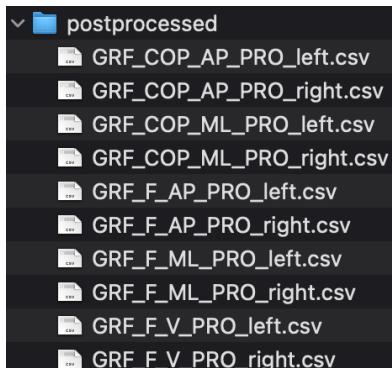


Figura 9. Contenido de la carpeta "postprocessed".

Con el objetivo de separar los pacientes pertenecientes al conjunto de entrenamiento de los pertenecientes al conjunto de prueba, creamos las carpetas "postprocessed_train" y "test_set" a partir de la carpeta "postprocessed". En estas carpetas guardamos los archivos "GRF_tipo_procesamiento_lado.csv" correspondientes a cada uno de los conjuntos de entrenamiento y prueba, respectivamente (véase la Fig. 10).

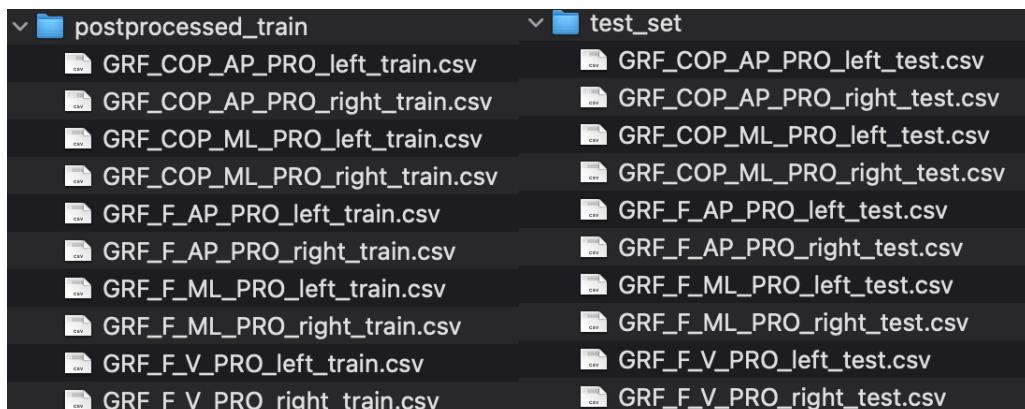


Figura 10. Contenido de las carpetas "postprocessed_train" y "test_set".

Con el fin de alcanzar el objetivo propuesto y poder realizar predicciones del progreso de la rehabilitación, se estableció un requisito mínimo de dos sesiones por paciente. En consecuencia, se excluyó a aquellos pacientes que habían participado en solo una sesión de terapia.

Cabe destacar también, la importancia de crear un subconjunto de individuos que fuese más o menos homogéneo, ya que esto contribuye a minimizar la variabilidad inherente en la muestra, lo que aumenta la fiabilidad y precisión de los análisis realizados. Además, al focalizar en grupos más homogéneos, se facilita la identificación de patrones y tendencias que podrían quedar disueltas en conjuntos más heterogéneos. Para ello, clasificamos los sujetos en función del número de sesiones y de la clase. Para tener los sujetos clasificados, filtramos en función del número deseado de sesiones y clase, y guardamos los datos filtrados en carpetas

específicas de grupos y sesiones para su posterior procesamiento o análisis. Inicialmente, seleccionamos todos los grupos (tobillo, calcáneo, cadera, controles sanos, y rodilla), así como los sujetos que habían realizado 2, 3 y 4 sesiones de terapia, ya que estos números de sesiones de rehabilitación eran los más comunes entre la mayoría de los sujetos. La estructura de las carpetas se muestra en la Fig. 11.

Finalmente, se tomó la decisión de comenzar por un enfoque específico, eligiendo el grupo "tobillo" (A) como punto de partida y empleando dos sesiones de terapia (véase la Tabla 4). Esta elección se basó en la necesidad de iniciar el análisis exploratorio de manera focalizada y controlada. Iniciar con un grupo y un número de sesiones determinados permitió una exploración detallada de los patrones y características presentes en esos datos, así como una evaluación de los procedimientos de análisis. Además, esta selección proporcionó una oportunidad para familiarizarse con los datos y comprender mejor sus particularidades, contribuyendo a la toma de decisiones informadas en la posterior adaptación de los métodos y enfoques. Esta elección permitió sentar un punto de partida práctico y manejable en el proceso investigativo.

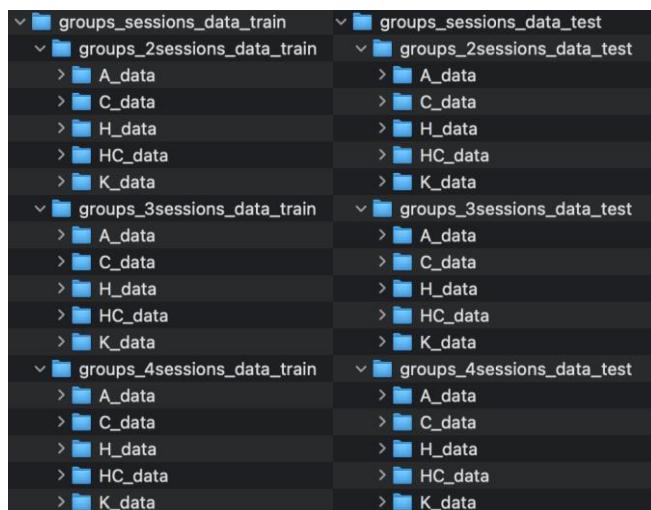


Figura 11. Carpetas y subcarpetas correspondientes al conjunto de entrenamiento (izquierda) y prueba (derecha) que contienen los datos de los sujetos con 2, 3 o 4 sesiones de rehabilitación para cada uno de los grupos (tobillo, calcáneo, cadera, rodilla, y controles sanos).

Tabla 4. Contenido de las carpetas "A_data" mostradas en la Fig. 11 para 2 sesiones de terapia.

Datos	N.º de sesiones	Grupo o clase	Ficheros	
TRAIN	2	Tobillo (A)	GRF_COP_AP_PRO_left_train_A.csv	
			GRF_COP_AP_PRO_right_train_A.csv	
			GRF_COP_ML_PRO_left_train_A.csv	
			GRF_COP_ML_PRO_right_train_A.csv	
			GRF_F_AP_PRO_left_train_A.csv	
	2		GRF_F_AP_PRO_right_train_A.csv	
			GRF_F_ML_PRO_left_train_A.csv	
			GRF_F_ML_PRO_right_train_A.csv	
			GRF_F_V_PRO_left_train_A.csv	
			GRF_F_V_PRO_right_train_A.csv	
TEST	2	Tobillo (A)	GRF_COP_AP_PRO_left_test_A.csv	
			GRF_COP_AP_PRO_right_test_A.csv	
			GRF_COP_ML_PRO_left_test_A.csv	
			GRF_COP_ML_PRO_right_test_A.csv	
			GRF_F_AP_PRO_left_test_A.csv	
	2		GRF_F_AP_PRO_right_test_A.csv	
			GRF_F_ML_PRO_left_test_A.csv	
			GRF_F_ML_PRO_right_test_A.csv	
			GRF_F_V_PRO_left_test_A.csv	
			GRF_F_V_PRO_right_test_A.csv	

El siguiente paso consistió en eliminar las sesiones intermedias y retener únicamente la primera y la última sesión de cada paciente. Para cada paciente y sesión, se calculó el valor promedio de los diferentes *trials* realizados en esa sesión. Esto permitió obtener una medida representativa de cada sesión para su posterior análisis.

Para realizar una mayor acotación en nuestro estudio, focalizamos nuestra atención en una variable concreta. En este sentido, decidimos centrarnos en la fuerza de reacción del suelo en dirección anteroposterior (fuerza de frenado y propulsión), es decir, los archivos "GRF_F_AP_PRO_right_train_A.csv" y "GRF_F_AP_PRO_left_train_A.csv". Una de las razones clave detrás de esta elección fue la naturaleza de las líneas promedio de la primera y última sesión de rehabilitación, que representan la evolución del promedio del peso corporal durante la fase de apoyo (porcentaje de la fase de apoyo o *stance*) en las sesiones de terapia. Estas líneas claramente exhiben un punto mínimo y uno máximo distintivos. Al centrarnos en estos valores extremos, logramos gráficos más comprensibles y con resultados visiblemente notorios. Elegimos esta variable con la intención de trabajar con datos que fueran más fáciles de interpretar y con un impacto visual más alto, lo que a su vez mejoró nuestra capacidad para analizar y extraer información valiosa.

Dado que disponíamos de dos archivos, uno correspondiente a cada extremidad inferior ("GRF_F_AP_PRO_right_train_A.csv" y "GRF_F_AP_PRO_left_train_A.csv"), fue necesario combinarlos. De esta manera unificamos los datos de ambas extremidades inferiores en un único conjunto de datos. Abordamos esta integración considerando cada pierna como una observación independiente. Así que optamos por unir el conjunto de datos de la pierna izquierda debajo del conjunto de datos de la pierna derecha (R), creando un patrón "RRRL". Esta aproximación nos permitió tratar cada extremidad inferior como si fuera un paciente diferente, ampliando así la cantidad de datos disponibles.

El resultado fueron dos conjuntos de datos separados, uno que contenía los datos de entrenamiento y otro los datos de prueba (véase la Fig. 12). Estos conjuntos de datos contienen el identificador único del sujeto ("SUBJECT_ID"), el identificador único de una sesión ("SESSION_ID"), y la media de los identificadores de los *trials* o paseos realizados en esa sesión ("TRIAL_ID"). Estas tres primeras columnas inicialmente se mantuvieron en la estructura inicial. Luego, se eliminaron debido a que contribuían al proceso de predicción. Las siguientes columnas (de "F_AP_PRO_1" a "F_AP_PRO_101") representan el valor promedio de los diferentes *trials* realizados en cada sesión por cada paciente para cada uno de los miembros inferiores (MMII). Es decir, representan los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior (fuerza de frenado y propulsión) en diferentes puntos durante la fase de apoyo del ciclo de la marcha, para todos los *trials* de esa sesión en particular para cada sujeto. Cada fila del *dataset* representa una observación de un sujeto en una sesión particular. La primera fila correspondiente a un paciente representa su primera sesión, mientras que la segunda fila representa la última sesión de ese mismo paciente. Además, los datos de cada fila están divididos en dos partes: la primera parte se refiere al miembro inferior derecho, mientras que la segunda parte se refiere al miembro inferior izquierdo. Es importante recordar que tomamos la decisión de unir los datos de la pierna izquierda (L) debajo de los datos de la pierna derecha (R), lo que resulta en un patrón "RRRL". Esto significa que primero se presentan los datos del miembro inferior derecho (R) y luego los datos del miembro inferior izquierdo (L).

	SUBJECT_ID	SESSION_ID	TRIAL_ID	F_AP_PRO_1	F_AP_PRO_2	F_AP_PRO_3	...	F_AP_PRO_99	F_AP_PRO_100	F_AP_PRO_101
MMII Derechos	0	1746	11976	5.500.000	-0.004533	-0.000129	0.005483	...	0.007602	-0.000017
	1	1746	12037	5.625.000	-0.006869	-0.000890	0.008038	...	0.013749	0.006494
	2	1748	18454	5.166.667	0.009228	0.018207	0.020520	...	0.017662	0.009402
	3	1748	18518	6.333.333	0.016802	0.027338	0.030654	...	0.009330	0.002944
	-0.000842
MMII Izquierdos	180	1746	11976	5.5	-0.004453	0.001465	0.010209	...	0.008095	-0.001383
	181	1746	12037	5.625	-0.001128	0.00584	0.014543	...	0.017820	0.006666
	182	1748	18454	5.166.667	0.011768	0.022459	0.026858	...	0.013584	0.003898
	183	1748	18518	6.333.333	0.013930	0.025193	0.027950	...	0.011538	0.002874

Figura 12. Estructura del conjunto de datos de entrenamiento y de prueba.

Cabe destacar que todos los pasos previamente descritos para construir los conjuntos de datos personalizados, de manera que se ajustaran a los objetivos específicos de nuestro proyecto, fueron unificados y encapsulados en funciones. Esta aproximación permitió la aplicación sistemática de dichos pasos tanto a los datos de entrenamiento como a los datos de prueba, otorgando una mayor eficiencia al proceso en su conjunto.

4.4 Análisis y comprobaciones iniciales

Mientras construíamos nuestro conjunto de datos, realizamos un análisis inicial previo y una serie de comprobaciones.

En el archivo CSV que contiene los metadatos de los pacientes pertenecientes solamente al conjunto de entrenamiento no balanceado, consultamos el número de pacientes por grupo o clase y el número de pacientes únicos para cada subclase.

En los archivos "GRF_tipo_procesamiento_lado.csv" de la carpeta "postprocessed_train", correspondientes al conjunto de entrenamiento, comprobamos si había pacientes que sólo tuvieran una sesión, o, por el contrario, que cada individuo tuviera varias sesiones. También consultamos el número máximo y mínimo de sesiones por paciente y visualizamos la distribución de los sujetos por número de sesiones. Por lo que respecta a los paseos o *trials*, comprobamos si había varios *trials* por paciente, consultamos el número máximo y mínimo de *trials* por paciente, y comprobamos que el número de *trials* para cada miembro inferior (derecho e izquierdo) fuese igual para cada paciente. A continuación, calculamos el número de *trials* por paciente y visualizamos los 50 sujetos con más *trials* y los 50 sujetos con menos *trials*. También visualizamos la distribución de los sujetos por número de *trials*.

Finalmente, visualizamos y comparamos los promedios de la primera y última sesión de pacientes con deficiencias musculoesqueléticas en el tobillo sometidos a 2 sesiones de rehabilitación en relación con los promedios de los controles sanos (HC). Para ello se utilizaron gráficos de líneas que representan la evolución de los promedios de peso corporal durante la fase de apoyo (% de la fase de apoyo o *stance*) en las sesiones de terapia. Esta visualización ayudó a identificar tendencias y patrones significativos en la evolución del peso corporal durante la fase de apoyo en diferentes sesiones de terapia.

4.4.1 Exploración inicial de los datos generales

En el conjunto de entrenamiento no balanceado, el 28% de los sujetos presentaban algún desorden en el tobillo (clase "A"), el 26.6% en la rodilla (clase "K"), el 18.5% en el calcáneo (clase "C"), y el 17.7% en la cadera (clase "H"). Por otro lado, los controles sanos (clase "HC") representaban el 9.2% de los sujetos del conjunto de datos (Fig. 13).

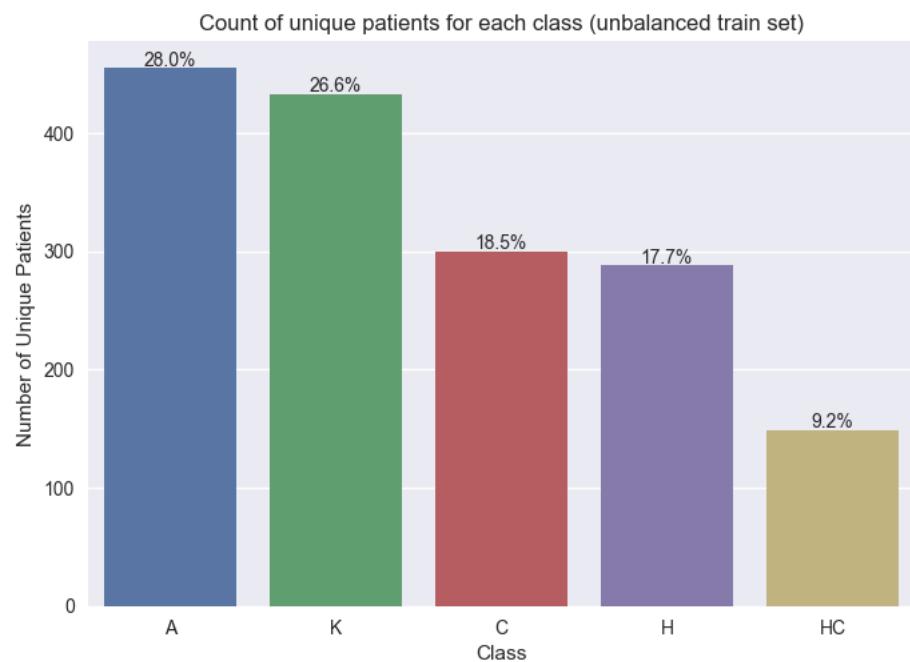


Figura 13. Número y porcentaje de sujetos en cada una de las distintas clases o grupos. La clase "A" hace referencia a los pacientes con alteraciones en el tobillo, la "K" a la rodilla, la "C" al calcáneo, la "H" a la cadera, y la clase "HC" hace referencia a los controles sanos.

En los archivos "GRF_tipo_procesamiento_lado.csv" de la carpeta "postprocessed_train" correspondientes al conjunto de entrenamiento, observamos que, aunque había 165 sujetos que solo habían realizado una sola sesión de terapia, el resto había asistido a varias sesiones durante su estancia en el centro de rehabilitación (Fig. 14). Predominantemente, los individuos participaron en dos sesiones de terapia (318 pacientes), tres sesiones (388 pacientes) y cuatro sesiones (252 pacientes).

Respecto al número de paseos o *trials* realizados por los sujetos, constatamos que cada sujeto había realizado varios paseos en cada sesión de terapia. Hubo 21 pacientes que realizaron tan solo 3 paseos, mientras que 2 sujetos realizaron 16 pasos en cada sesión. No obstante, como se observa en la Fig. 15, la mayoría de los sujetos (687 pacientes) realizaron 10 paseos por sesión de rehabilitación.

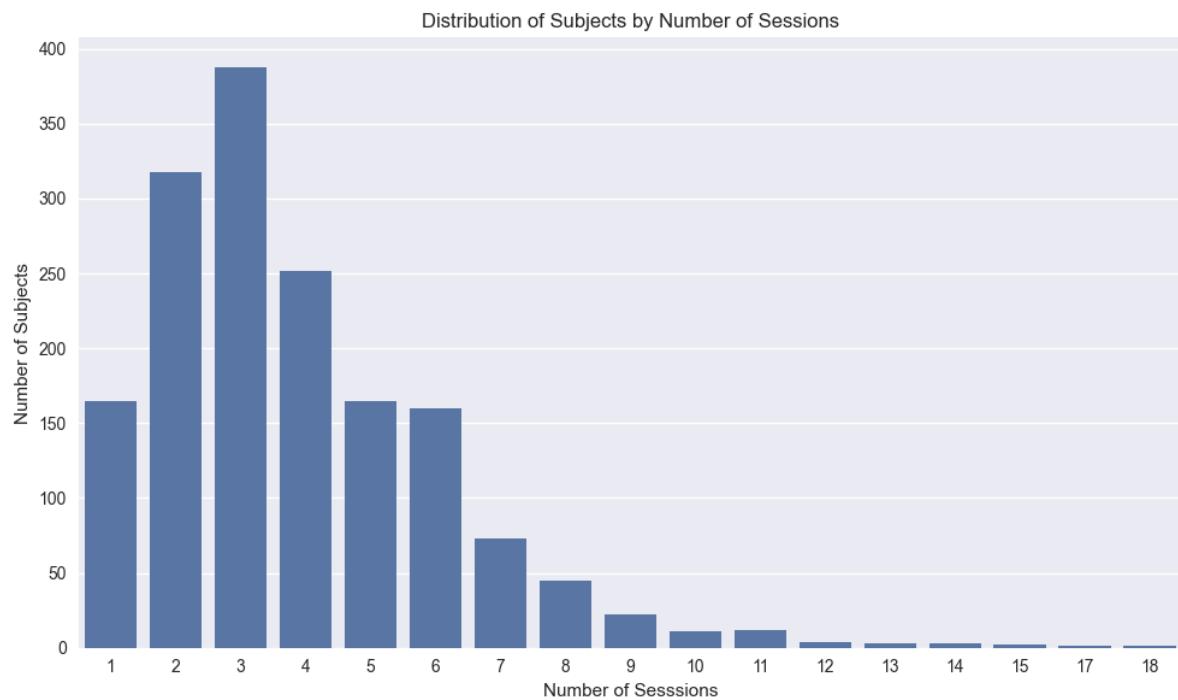


Figura 14. Distribución de los sujetos por número de sesiones en los archivos "GRF_tipo_procesamiento_lado.csv" de la carpeta "postprocessed_train", correspondientes al conjunto de entrenamiento.

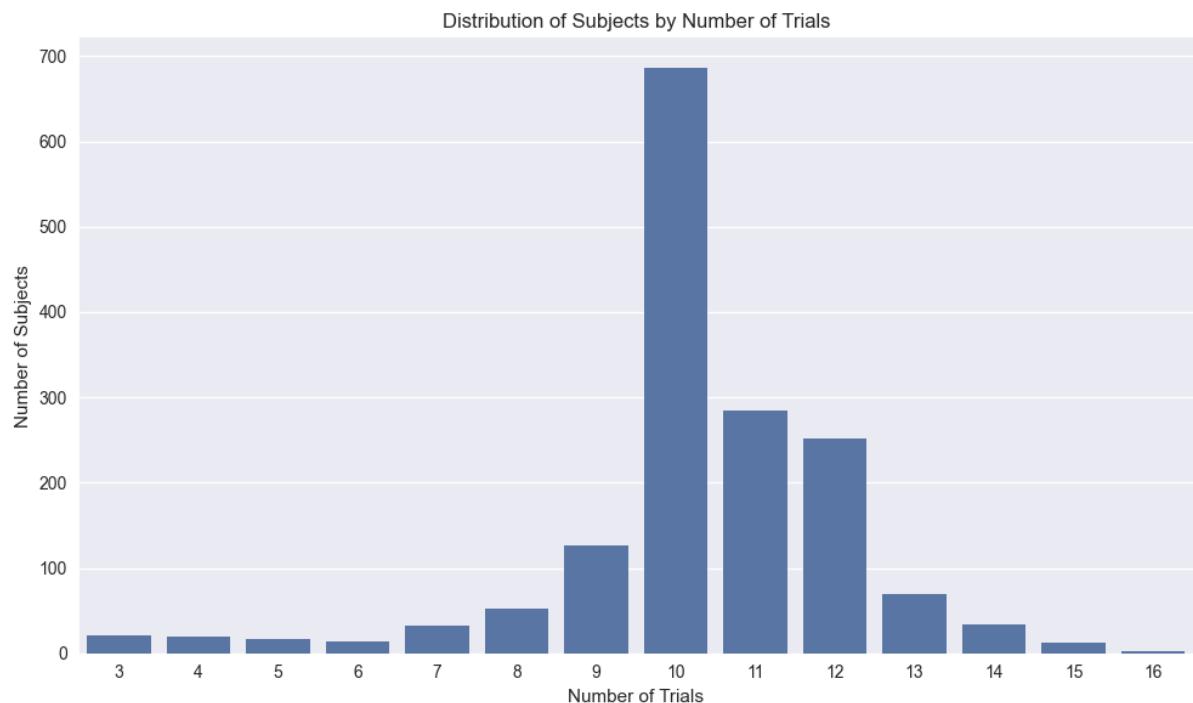


Figura 15. Distribución de los sujetos por número de paseos o *trials* en los archivos "GRF_tipo_procesamiento_lado.csv" de la carpeta "postprocessed_train", correspondientes al conjunto de entrenamiento.

4.4.2 Análisis inicial de las componentes de las fuerzas de reacción del suelo

Para los controles sanos y los sujetos con deficiencias musculoesqueléticas en el tobillo sometidos a 2 sesiones de terapia pertenecientes al conjunto de entrenamiento, obtuvimos las siguientes curvas que representan las fuerzas de reacción del suelo en la dirección vertical (Fig. 16a), anteroposterior (Fig. 16b) y medio-lateral (Fig. 16c) para cada uno de los miembros inferiores durante la marcha. Las líneas verdes y azules representan el promedio del peso corporal de los pacientes durante la fase de apoyo del ciclo de la marcha de la primera y última sesión, respectivamente. Por otro lado, la línea roja representa la evolución del promedio del peso corporal de los controles sanos ("HC") durante la fase de apoyo del ciclo de la marcha.

En las fuerzas verticales (Fig. 16a), para ambos miembros inferiores observamos un primer pico correspondiente al valor máximo de fuerza en el choque de talón. A continuación, se produjo un valle que indicaba el mínimo valor en la fase de apoyo total del pie. Finalmente, se observó un segundo pico correspondiente al máximo valor en la fase de despegue.

Para las fuerzas de reacción del suelo en la dirección anteroposterior, podemos ver las fuerzas de frenado (o de apoyo de talón) y propulsión (o despegue del talón) durante la marcha (Fig. 16b). Observamos que, al inicio, la curva descendía a causa de la desaceleración que se produce cuando el pie entra en contacto con el suelo. Esta curva alcanzó su máximo valor (el mínimo en la gráfica) en la fase de doble apoyo. A continuación, al disminuir la fuerza, la curva empezó a ascender hasta que la fuerza llegó a cero. Esto ocurrió cuando el apoyo es con una sola pierna. Posteriormente, se observó un nuevo pico en la gráfica que alcanzó su valor máximo cuando los dos pies se encontraban en contacto con el suelo (fase de doble soporte del ciclo de la marcha).

Para las fuerzas de reacción del suelo en la dirección medio-lateral, observamos lo siguiente (Fig. 16c). En la fase de apoyo inicial (0-6% del ciclo de la marcha), el pico de fuerza de reacción del suelo en el pie adelantado se dirigió lateralmente (por debajo del cero). Durante el apoyo con una sola extremidad inferior, el pico de fuerza de reacción del suelo en el pie de apoyo se dirigió medialmente (por encima del cero). De la misma forma, durante la fase de apoyo terminal o doble apoyo, el pico de fuerza de reacción del suelo en el pie de apoyo se dirigió medialmente.

En todos los casos, los valores promedio del peso corporal durante la fase de apoyo del ciclo de la marcha de la primera y última sesión, estuvieron casi solapados entre sí, distanciándose de lo que se consideraría una marcha saludable. La discrepancia más pronunciada se observó en los picos de las gráficas, es decir, durante la fase de apoyo o choque del talón contra el suelo.

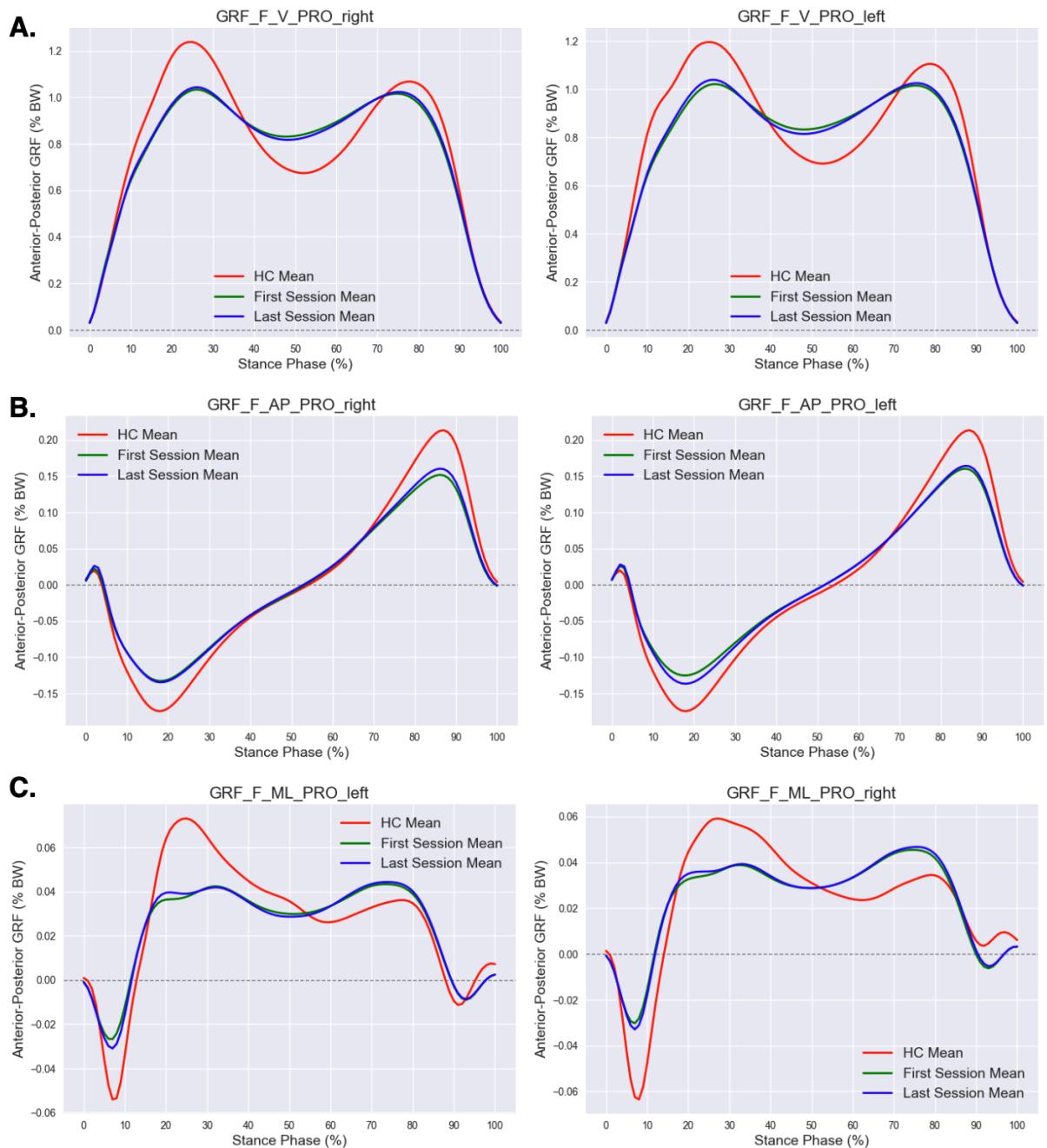


Figura 16. Representación gráfica de las fuerzas de reacción del suelo en la dirección vertical (A), anteroposterior (B) y medio-lateral (C) para cada uno de los miembros inferiores durante la marcha.

4.5 Estudio de la población y cohorte de pacientes

El conjunto de datos de entrenamiento creado para este estudio comprende un total de 90 pacientes únicos con deficiencias musculoesqueléticas en el tobillo, los cuales realizaron dos sesiones de terapia o rehabilitación. No obstante, consideramos cada miembro inferior como una observación independiente, lo que resulta en un total de 180 observaciones ($N=180$). Estas deficiencias musculoesqueléticas del tobillo incluyen fracturas de los maléolos, el astrágalo, la tibia o la parte inferior de la pierna, y roturas de ligamentos o del tendón de Aquiles, entre otros [4]. De entre los 90 pacientes, 73 son hombres y 17 mujeres, lo que refleja una proporción de sexo desigual. La edad media de la cohorte es de 42.82 años, con una masa corporal promedio de 89.96 kg. Los detalles sobre las características demográficas de la cohorte de pacientes se encuentran en la Tabla 5.

Adicionalmente, el estudio incluye 149 controles sanos. Como en el caso de los pacientes, tratamos cada miembro inferior como una observación independiente, lo que resultó en un total de 11126 observaciones ($N=11126$). Este elevado número de observaciones se debe a que cada paciente había hecho varios paseos o *trials* en cada sesión que se contaron como observaciones independientes. De estos 149 sujetos sanos, 67 son hombres y 82 mujeres, siendo esta una proporción de sexo más equilibrada en comparación con la cohorte de pacientes. La edad media de esta segunda cohorte es de 35.96 años, con una masa corporal promedio de 73.46 kg. Para más detalles sobre las características demográficas de la cohorte de controles sanos, consultar la Tabla 6.

Tabla 5. Resumen demográfico del conjunto de datos de entrenamiento de pacientes con deficiencias musculoesqueléticas en el tobillo.

Dataset	Clase	N.º sesiones	N.º pacientes	N	Edad media (años)	Masa corporal media (kg)	Género (m/f)
TRAIN	Tobillo (A)	2	90	180	42.82	89.96	73/17

Tabla 6. Resumen demográfico del conjunto de datos los controles sanos.

Dataset	Clase	N.º sujetos	N	Edad media (años)	Masa corporal media (kg)	Género (m/f)
TRAIN	Controles sanos (HC)	149	11126	35.96	73.46	67/82

4.6 Análisis exploratorio de los datos

Para comenzar el análisis exploratorio de datos, llevamos a cabo una exploración y una serie de pasos iniciales para entender la naturaleza y calidad de nuestro conjunto de datos de entrenamiento.

En primer lugar, realizamos una exploración visual del conjunto de datos para tener una visión general de su estructura. Posteriormente, comprobamos la presencia de valores nulos o faltantes, asegurándonos de que los datos estuvieran completos para un análisis adecuado. También, calculamos estadísticas descriptivas para las variables numéricas, lo que nos permitió obtener información sobre tendencias centrales y dispersión en los datos. Además, verificamos la existencia de duplicados en los registros para asegurarnos de la integridad del *dataset*. A continuación, nos enfocamos en comprender mejor la composición de las cohortes de pacientes y controles sanos. Realizamos un análisis específico para calcular el número total de pacientes y controles sanos incluidos en nuestro estudio. Al considerar cada miembro inferior

como una observación independiente, determinamos el número total de observaciones en ambas cohortes. Calculamos la edad media en años, la masa corporal promedio en kilogramos, y examinamos la distribución de sexo en ambas cohortes para obtener una idea más precisa de las características demográficas de los sujetos en cada cohorte.

Para obtener una comprensión más profunda de la estructura y relaciones presentes en el conjunto de datos realizamos los siguientes pasos. Visualizamos la distribución de cada una de las 101 variables numéricas mediante *boxplots* e histogramas para analizar la dispersión de los datos y detectar posibles valores atípicos o patrones de distribución inusuales. Posteriormente, evaluamos la normalidad de los datos utilizando la prueba de Anderson-Darling. Esta prueba nos permitió determinar si las variables seguían o no una distribución normal. Además, realizamos un análisis de correlación entre todas las variables utilizando el coeficiente de correlación de Pearson. Aunque generalmente se recomienda escalar las variables antes de llevar a cabo un análisis de correlación, en nuestro caso decidimos no hacerlo ya que al trabajar con los datos postprocesados proporcionados por GaitRec, ya habían sido normalizados previamente.

Todo este proceso fue encapsulado en funciones para garantizar la modularidad y reutilización en nuestro análisis.

En resumen, el análisis exploratorio nos proporcionó una comprensión sólida de la estructura, calidad, distribución y relaciones en nuestro conjunto de datos, lo que sentó las bases para aplicar análisis más avanzados, técnicas de procesamiento de datos y poder realizar los modelos de aprendizaje automático.

4.6.1 Exploración inicial del conjunto de entrenamiento

Ya en con los conjuntos de datos de entrenamiento y prueba construidos específicamente para nuestro estudio, realizamos una primera inspección general en el *dataset* de entrenamiento para comprender su estructura y tener una visión general del mismo.

El conjunto de datos contenía 360 instancias o filas y 104 variables o columnas. La estructura se define en la sección de "[Métodos](#)" y puede visualizarse en la Fig. 12. El conjunto de datos de entrenamiento no presentó ni valores nulos o faltantes ni valores duplicados.

4.6.2 Distribución de las variables numéricas

Los histogramas fueron aproximadamente simétricos alrededor de la media y en forma de campana para las 101 variables numéricas de entrada y salida del conjunto de entrenamiento. Los diagramas de caja o *boxplots* presentaron aproximadamente proporciones iguales entorno a la mediana. No obstante, se observaron algunos valores atípicos ([Fig. A1](#)). Por tanto, estas visualizaciones parecían indicar que las 101 variables se distribuían de manera normal.

Debemos recordar que estas 101 variables representaban los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior (expresada en porcentaje del peso corporal) durante los 101 puntos en los que se divide la fase de apoyo del ciclo de la marcha. En este caso, estas 101 variables incluían la primera y última sesión de rehabilitación para cada miembro inferior (considerados como observaciones independientes) de los pacientes con trastornos musculoesqueléticos en el tobillo que realizaron dos sesiones de rehabilitación.

4.6.3 Comprobación de la normalidad de las variables

Según los resultados de la prueba de Anderson-Darling, utilizada para verificar la normalidad de las 101 variables del conjunto de entrenamiento, se rechazó la hipótesis nula para un total de 57 variables, indicando que estas no se distribuían de manera normal. Para las 44 variables restantes, se aceptó la hipótesis nula, indicando que se distribuían normalmente.

4.6.4 Análisis de correlación

El análisis de correlación realizado entre las 101 variables del conjunto de entrenamiento utilizando el método de correlación de Pearson, proporcionó un extenso gráfico donde observamos varios patrones ([Fig. A2](#)).

Alrededor de la diagonal principal, el color rojo era intenso, indicando una fuerte correlación positiva entre las variables vecinas. No obstante, al alejarnos de la diagonal, se observó una gradación de colores, pasando de tonos rojos claros a azules. Este cambio de color indicó una disminución gradual en la fuerza de correlación entre las variables a medida que se alejaban unas de otras en el conjunto de datos. Las áreas con tonos azules indicaban una correlación débil, lo que podría indicar que esas variables no estaban muy relacionadas.

Un hecho destacado que se observó fue que, aunque en su mayor parte la diagonal era gruesa, al llegar a la variable 64 aproximadamente, se estrechó y se formó una concentración de color rojo. Esta concentración de tonos rojos intensos entre las variables 78 y 86 formaba una especie de "bulto" que alcanzaba su máximo en la variable 85. Esta forma abultada de tonalidad roja intensa indicó una fuerte correlación entre esas variables específicas, que correspondían a la fase posterior del despegue del talón después del punto máximo de propulsión. La presencia de esta alta correlación entre variables vecinas podría indicar que, en el momento de predecir una variable en particular, podría ser probable que estuviéramos utilizando información de variables cercanas debido a la relación y la dependencia entre ellas. En estos casos, también sería probable que existieran relaciones lineales significativas entre ellas.

4.7 División de los datos

En pasos previos, realizamos la separación de los datos de entrenamiento y prueba en conjuntos de datos distintos. Sin embargo, antes de introducir los datos en cualquier modelo de aprendizaje automático, se deben separar las características y las variables objetivo o de salida. Esto es necesario para garantizar la integridad de los modelos de aprendizaje automático, prevenir la fuga de datos (*data leakage*) y poder hacer una evaluación precisa del rendimiento del modelo en datos no vistos. En nuestro caso, las variables predictoras son los valores de la primera sesión de terapia, localizados en las filas impares, mientras que las variables objetivo son los valores de la última sesión, ubicados en las filas pares.

El proceso de separación de variables predictoras y variables objetivo o de salida lo ejecutamos mediante la implementación de una función específica para dicha tarea. Esta función excluye las columnas "SUBJECT_ID", "SESSION_ID" y "TRIAL_ID", y crea los *arrays* "X_train", "y_train", "X_test" y "y_test". Para llevarlo a cabo, extrae las filas impares del conjunto de entrenamiento y las asigna a "X_train". A continuación, extrae las filas pares del conjunto de entrenamiento y las asigna a "y_train". Posteriormente, realiza el mismo procedimiento para los datos del conjunto de prueba.

4.8 Preprocesamiento de los datos

Una vez tuvimos los datos en el formato adecuado para poder introducirlos en los modelos de aprendizaje automático, iniciamos la fase de preprocesamiento. En la etapa de preprocesamiento de los datos, implementamos varias aproximaciones con el objetivo de optimizar el rendimiento de los modelos de aprendizaje automático. Estas aproximaciones se aplicaron a los conjuntos de datos de entrada y salida, con el propósito de explorar cómo distintos preprocesamientos podrían impactar en las predicciones de los modelos.

4.8.1 Primera aproximación: datos sin preprocessar

En la primera aproximación no se aplicó ningún tipo de procesamiento adicional a los datos de entrada y salida. Las predicciones se realizaron utilizando los datos en su forma original, sin modificar la dimensionalidad ni el contenido de las variables predictoras y/o objetivo.

4.8.2 Segunda Aproximación: PCA en los datos de entrada (X)

La segunda aproximación consistió en aplicar el análisis de componentes principales (PCA) solo a los datos de entrada (X), que son los valores de la primera sesión de terapia. La idea detrás de esto es reducir la dimensionalidad de las variables predictoras para mejorar la precisión de las predicciones y eliminar redundancias. Aunque generalmente se recomienda escalar las variables antes de aplicar el PCA, en este caso no lo hicimos debido a que los datos postprocesados proporcionados por GaitRec ya habían sido normalizados previamente. Establecimos seleccionar los componentes principales que explicaran o retuvieran al menos el 90% de la varianza de los datos originales. La aplicación de esta segunda aproximación vino inspirada por el trabajo realizado por Omar A. Galarraga C. et al. descrito artículo científico titulado *Predicting Postoperative Gait in Cerebral Palsy* [13] donde tratando de resolver un problema similar al nuestro, utilizaron el análisis de componentes principales para reducir la dimensionalidad de las variables independientes o de entrada con el objetivo de mejorar la precisión de la predicción y reducir la redundancia de datos.

4.8.3 Tercera Aproximación: PCA en los datos de salida (Y)

En la tercera aproximación, aplicamos el PCA solo a los datos de salida (Y), correspondientes a los valores de la última sesión de terapia. De la misma forma que en la segunda aproximación, optamos por no escalar las variables antes de aplicar el PCA debido a la previa normalización de los datos, y establecimos nuevamente el objetivo de seleccionar los componentes principales que explicaran o retuvieran al menos el 90% de la varianza de los datos originales. Después de obtener las predicciones de los modelos de aprendizaje automático, transformamos esas predicciones de nuevo al espacio original de las variables objetivo utilizando el método *inverse_transform()* proporcionado por la librería Scikit-Learn. Para volver al espacio original, en lugar de utilizar los componentes principales originales, utilizamos los componentes principales predichos. Es decir, teníamos los datos de salida del conjunto de prueba original ("y_test"), hicimos un PCA, y nos fuimos a un espacio nuevo. Por tanto, podíamos volver al espacio original porque disponíamos de las combinaciones lineales. Con esas combinaciones lineales tuvimos unos componentes principales que posteriormente predecimos de nuevo.

4.8.4 Cuarta aproximación: PCA en los datos de entrada y salida (X e Y)

En la cuarta y última aproximación, aplicamos el PCA tanto a los datos de entrada (X) como a los de salida (Y) con el objetivo de reducir la dimensionalidad de los datos de la primera y última sesión de terapia. Establecimos nuevamente el objetivo de seleccionar los componentes principales que explicaran o retuvieran al menos el 90% de la varianza de los datos originales. De la misma forma que en la tercera aproximación, después de obtener las predicciones con los modelos de aprendizaje automático, revertimos la transformación del PCA transformando los datos de nuevo al espacio original de las variables objetivo utilizando el método *inverse_transform()* proporcionado por la librería Scikit-Learn.

Cada uno de los algoritmos de aprendizaje automático utilizados en este estudio pasó por estas cuatro aproximaciones o tipos de preprocessamiento. En otras palabras, generamos

cuatro modelos distintos para cada algoritmo, considerando cada uno de los tipos de preprocessamiento mencionados anteriormente. Esto nos permitió evaluar cómo cada aproximación afectaba al rendimiento de los modelos y nos permitió determinar cuál de ellas proporcionó las mejores predicciones. El proceso de generación y evaluación de estos modelos de aprendizaje automático se describe detalladamente en la siguiente sección.

4.9 Análisis multivariante

4.9.1 Tipo de problema y algoritmos utilizados

En esta investigación nos enfrentamos a un problema de regresión múltiple, ya el objetivo era predecir los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior (fuerza de frenado y propulsión) de la última sesión de terapia en función de las características de la primera sesión. Esto, sumado a que los conjuntos de datos estaban etiquetados, fue el motivo por el cual se utilizaron diversos algoritmos de regresión de aprendizaje automático supervisados.

Los algoritmos seleccionados para llevar a cabo las predicciones fueron la regresión lineal múltiple (MLR), la regresión de mínimos cuadrados parciales o *Partial Least Squares regression* (PLS regression), árboles de decisión (*Decision Trees*), bosques aleatorios o *Random Forests*, eXtreme Gradient Boosting (XGBoost) y regresión de vectores de soporte o *Support Vector Regression* (SVR).

4.9.2 Pipeline general de entrenamiento y evaluación de los modelos

Para cada uno de los algoritmos empleados, desarrollamos funciones específicas para entrenar y evaluar el rendimiento de cada modelo en los conjuntos de entrenamiento y prueba, respectivamente. No obstante, cabe mencionar que el conjunto de datos GaitRec incluía una partición del conjunto de datos para poder crear solo los conjuntos de entrenamiento y prueba, pero no para crear el conjunto de validación. Al no disponer de un número muy elevado de observaciones en el conjunto de entrenamiento ($N=180$), y siguiendo las recomendaciones especificadas en la sección de “Notas de uso” del conjunto de datos GaitRec [4], realizamos la optimización de los modelos mediante distintas técnicas de validación cruzada (*cross-validation*) sólo en el conjunto de entrenamiento, manteniendo el conjunto de prueba intacto hasta la evaluación final para evitar la fuga de datos o *data leakage*.

4.9.2.1 Pipeline general de entrenamiento de los modelos

En la fase de entrenamiento de la mayoría de los algoritmos, usamos técnicas como la validación cruzada para realizar el ajuste de hiperparámetros (*hyperparameter tuning*) y seleccionar la mejor configuración del modelo. Una vez completado el proceso de ajuste de hiperparámetros y seleccionada la mejor configuración del modelo a partir de los resultados de la validación cruzada, entrenamos el modelo final utilizando todo el conjunto de datos de entrenamiento con los hiperparámetros seleccionados. Este proceso específico de cada algoritmo se explicará en detalle más adelante.

4.9.2.2 Pipeline general de evaluación de los modelos en datos no vistos

En la fase de evaluación, evaluamos el rendimiento de los modelos en el conjunto de prueba, es decir, en los datos no vistos. Las métricas de evaluación como la raíz del error cuadrático medio (RMSE), el coeficiente de correlación de Pearson, y el coeficiente de determinación (R^2) para cada una de las variables objetivo, nos permitieron evaluar los modelos en los diferentes puntos de la curva predicha de manera más precisa. Por otro lado, el método *target-specific ranks* [66], en el cual se promedian los rangos en lugar de las

puntuaciones individuales de las métricas de regresión, resaltó la diferencia de rendimiento entre los distintos modelos permitiéndonos compararlos entre sí.

En el caso de la tercera y cuarta aproximación, al aplicar el PCA a los datos, reducimos su dimensionalidad. Sin embargo, para comprender completamente cómo se desempeñarían los modelos, era fundamental evaluar las predicciones en el espacio original de características. Es por ello por lo que evaluamos el rendimiento de los modelos después de haber realizado la transformación inversa de las predicciones al espacio de características original. Es decir, evaluamos los valores predichos dentro del espacio de características original utilizando las mismas métricas de evaluación mencionadas anteriormente. Este proceso fue fundamental para comprender el rendimiento de los modelos en las variables de salida reales, teniendo en cuenta las características del espacio original.

Además, realizamos una serie de visualizaciones que nos ayudaron a comprender e interpretar mejor los resultados obtenidos. Creamos gráficos de dispersión de los valores predichos frente a los valores reales para cada una de las 101 variable objetivo, añadiendo la línea de correlación perfecta como referencia (con una pendiente de 1 y una intersección en el origen), la línea de la regresión lineal, y la ecuación de la recta de regresión en cada gráfico. De manera similar, creamos otro gráfico de dispersión, pero en este caso con todos los puntos para comparar los valores reales y los valores predichos por los modelos de manera global. En el mismo gráfico trazamos línea de correlación perfecta que representa la relación ideal donde los valores reales y predichos son idénticos, la línea de regresión lineal que representa la relación lineal ajustada entre los valores reales y predichos, y la ecuación de la recta de regresión.

Finalmente, creamos un gráfico que resumía y presentaba de manera clara la predicción de los modelos y las métricas de evaluación empleadas. En este gráfico de líneas mostramos los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior (fuerza de frenado y propulsión) en 101 puntos durante la fase de apoyo del ciclo de la marcha. Se incluyeron para su comparación los valores promedio correspondientes a la primera sesión, a la última sesión real, y a la última sesión de rehabilitación predicha por los distintos modelos. Además, se incorporó como referencia la curva que representaba los valores promedio de los controles sanos. Para tener una comprensión mayor de la calidad de las predicciones, en el mismo gráfico incluimos las métricas de evaluación como la raíz del error cuadrático medio (RMSE), el coeficiente de correlación de Pearson, y el coeficiente de determinación (R^2) de los modelos.

Para complementar la evaluación de los resultados de los modelos, decidimos crear una serie de gráficos destinados a examinar y analizar la posible dependencia o relación entre el valor de las métricas de evaluación y los puntos de la curva predicha. Esto nos permitiría, por ejemplo, identificar si existían puntos de la curva más difíciles de predecir que otros, o si en los puntos críticos de la curva, como el máximo y el mínimo, podrían estar asociados con un mayor error. Para lograrlo, representamos gráficamente los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior (fuerza de frenado y propulsión) en 101 puntos durante la fase de apoyo del ciclo de la marcha de última sesión de rehabilitación predicha por los distintos modelos. Entonces, asignamos colores a los puntos de la curva en base a los valores de las tres métricas de evaluación: la raíz del error cuadrático medio (RMSE), el coeficiente de correlación de Pearson, y el coeficiente de determinación (R^2). Estas representaciones gráficas nos brindaron una visión más detallada y contextualizada de cómo las métricas de evaluación estaban relacionadas con los puntos individuales de la curva predicha, lo que proporcionó valiosos conocimientos sobre la calidad y las limitaciones de los modelos.

4.9.2.3 Validación de los modelos de regresión múltiple

Con el propósito de validar el ajuste de los modelos de regresión múltiple y saber si eran apropiados para nuestros datos, llevamos a cabo el análisis de los residuos o errores generados por los modelos.

Para realizar el análisis de los residuos, llevamos a cabo dos tipos de representaciones gráficas: el gráfico de residuos frente a los valores predichos por el modelo, y el histograma de los residuos. El gráfico de los residuos en función de los valores predichos nos permitió visualizar la dispersión de los residuos frente a los valores predichos. Es decir, identificar si los residuos se distribuían aleatoriamente alrededor del cero sin un patrón discernible, indicando un buen ajuste del modelo, o si mostraban algún tipo de patrón discernible, lo que podría indicar que el modelo no estaría capturando adecuadamente la variación en los datos. El segundo gráfico, el histograma de los residuos se utilizó para verificar si los errores seguían una distribución normal alrededor de cero, o si había alguna asimetría o sesgo en su distribución.

La validación de los modelos de regresión múltiple a través del análisis de los residuos proporcionó información sobre el ajuste de mismos a nuestros datos. Los gráficos empleados para el análisis nos permitieron evaluar tanto la dispersión como la distribución de los residuos, permitiéndonos comprender la calidad de las predicciones y la fiabilidad de los resultados obtenidos a partir de los modelos de regresión múltiple.

El proceso de evaluación descrito se encapsuló en funciones, creando un *pipeline* de evaluación apto para aplicarlo a todos los algoritmos de aprendizaje automático utilizados en el estudio y poder así, evaluar los modelos de manera automática. A continuación, se detallarán las particularidades y características propias de cada algoritmo y modelo de aprendizaje automático utilizando en este estudio.

4.9.3 Regresión Lineal Múltiple (MLR)

El primer algoritmo de aprendizaje automático supervisado que utilizamos fue la regresión lineal múltiple (MLR). La elección de este primer algoritmo vino inspirada por el trabajo realizado por Omar A. Galarraga C. et al. descrito artículo científico titulado *Predicting Postoperative Gait in Cerebral Palsy* [13] donde tratando de resolver un problema similar al nuestro, utilizaron continuación un modelo de regresión lineal múltiple para predecir simultáneamente múltiples variables cinemáticas postoperatorias.

Entrenamos y evaluamos los modelos de regresión lineal múltiple para cada una de cuatro las aproximaciones, con el propósito de explorar cómo distintos preprocesamientos podrían impactar en las predicciones de los modelos.

4.9.3.1 Entrenamiento y validación en el conjunto de entrenamiento

Se desarrolló una función encargada de entrenar modelos de regresión lineal múltiple utilizando la técnica de *Leave-One-Out Cross-Validation (LOOCV)* para evaluar el rendimiento de los modelos. Para cada una de las 101 variables objetivo, ajustamos un modelo lineal utilizando todas las observaciones excepto una, y luego evaluamos el rendimiento del modelo sobre la observación excluida. Repetimos esta evaluación para cada observación en el conjunto de entrenamiento, lo que garantizó una evaluación exhaustiva y redujo el riesgo de sobreajuste o *overfitting*. Los resultados obtenidos fueron los modelos ajustados para cada variable objetivo, las predicciones generadas durante el proceso de *LOOCV*, y los valores de la raíz del error cuadrático medio (RMSE), los valores del coeficiente de determinación (R^2) y los coeficientes de correlación de Pearson, tanto para cada variable objetivo como para el modelo en global.

Entrenamos y validamos los modelos de regresión lineal múltiple para cada una de cuatro las aproximaciones. Para la primera aproximación, el modelo se entrenó con los datos de

entrenamiento sin procesar. Para la segunda aproximación, aplicamos el análisis de componentes principales (PCA) solo a los datos de entrada ("X_train"). Es decir, entrenamos el modelo con los datos de la primera sesión de terapia reducidos y los de la última sesión sin reducir ("y_train"). Para la tercera aproximación, aplicamos el análisis de componentes principales (PCA) solo a los datos de salida ("y_train"). Es decir, entrenamos el modelo con los datos de la primera sesión de terapia sin reducir y los de la última sesión reducidos. Para la cuarta aproximación, aplicamos el PCA tanto a los datos de entrada ("X_train") como a los de salida ("y_train"). Es decir, entrenamos el modelo con los datos reducidos de la primera y última sesión de terapia. El proceso de entrenamiento para cada una de las cuatro aproximaciones se realizó de la misma forma para todos los algoritmos utilizados en el estudio, exceptuando la regresión de mínimos cuadrados parciales (PLS).

4.9.3.2 Evaluación en el conjunto de prueba

En esta etapa, evaluamos la calidad y el rendimiento de los modelos de regresión lineal múltiple entrenados en el conjunto de entrenamiento sobre datos no vistos, es decir, el conjunto de prueba.

Utilizando el *pipeline* de evaluación anteriormente descrito, obtuvimos de manera automática las métricas de evaluación como la raíz del error cuadrático medio (RMSE), el coeficiente de correlación de Pearson, y el coeficiente de determinación (R^2) globales del modelo. Además, en el caso de la tercera y cuarta aproximación, utilizando las mismas métricas evaluamos las predicciones en el espacio original después de la transformación inversa para comprender el rendimiento de los modelos en las variables reales.

Visualizamos los gráficos de dispersión de los valores predichos frente a los valores reales para cada una de las 101 variable objetivo, así como el gráfico de dispersión con todos los puntos para comparar los valores reales y los valores predichos por los modelos de manera global. Mostramos también el gráfico que resumía y presentaba de manera clara la predicción de los modelos y las métricas de evaluación empleadas. Para complementar la evaluación de los resultados de los modelos, representamos los distintos gráficos destinados a examinar y analizar la posible dependencia o relación entre el valor de las métricas de evaluación y los puntos de la curva predicha.

Finalmente, validamos el ajuste de los modelos de regresión múltiple mediante el análisis de los residuos o errores generados por los modelos. Para ello representamos los residuos frente a los valores predichos por el modelo, y el histograma de los residuos.

Evaluamos los modelos de regresión lineal múltiple en el conjunto de prueba para cada una de cuatro las aproximaciones. Para la primera aproximación, evaluamos el rendimiento del modelo en los datos de prueba sin procesar. Para la segunda aproximación, aplicamos el análisis de componentes principales (PCA) solo a los datos de entrada ("X_test"). Es decir, evaluamos el rendimiento del modelo en los datos de prueba de la primera sesión de terapia reducidos y en los de la última sesión sin reducir ("y_test"). Para la tercera aproximación, aplicamos el análisis de componentes principales (PCA) solo a los datos de salida ("y_test"). Es decir, evaluamos el rendimiento del modelo en los datos de prueba de la primera sesión de terapia sin reducir y en los de la última sesión reducidos. Para la cuarta aproximación, aplicamos el PCA tanto a los datos de entrada ("X_test") como a los de salida ("y_test"). Es decir, evaluamos el rendimiento del modelo en los datos de prueba reducidos de la primera y última sesión de terapia. En el caso de la tercera y cuarta aproximación, después de obtener las predicciones de los modelos de aprendizaje automático, llevamos a cabo una transformación inversa de esas predicciones. Esta transformación nos permitió regresar al espacio original de las variables objetivo. En estas dos aproximaciones específicas, el análisis de residuos se realizó utilizando las predicciones que habían sido reconstruidas al espacio original de características.

El proceso de evaluación en cada una de las cuatro aproximaciones se realizó de la misma forma para todos los algoritmos utilizados en el estudio, exceptuando la regresión de mínimos cuadrados parciales (PLS).

4.9.3.3 Comprobación de los supuestos

Para verificar si podíamos usar los modelos de regresión lineal múltiple y garantizar que fuesen válidos y fiables, comprobamos los supuestos más importantes: el supuesto de linealidad, de independencia de errores o residuos, de normalidad de los residuos, y de homocedasticidad.

Para comprobar el supuesto de linealidad, representamos gráficamente los residuos frente a los valores predichos para cada variable objetivo. Esto nos permitió visualizar la relación entre los residuos y las predicciones y evaluar si el supuesto de linealidad se cumplía para cada variable objetivo. Adicionalmente, creamos gráficos conjuntos (*joint plots*) entre cada predictor y variable objetivo, mostrando la relación existente entre ellos.

Para comprobar el supuesto de independencia de los residuos, examinamos el gráfico de residuos frente a los valores predichos para cada variable objetivo añadiendo el resultado la prueba estadística de Durbin-Watson en cada uno. Esto nos permitió detectar la posible presencia de autocorrelación (dependencia de los errores) en los residuos.

Para comprobar el supuesto de normalidad de los residuos utilizamos la prueba estadística de Anderson-Darling. Con esta prueba determinamos si los residuos seguían una distribución normal al nivel de significación del 5% para cada variable objetivo.

Finalmente, para comprobar el supuesto de homocedasticidad en los residuos, utilizamos la prueba estadística de Breusch-Pagan. Mediante el *p-value*, esta prueba nos permitió comprobar si varianza de los residuos era constante a lo largo de los valores predichos.

Después de comprobar los supuestos, en la fase de evaluación de la regresión lineal múltiple se crearon dos gráficos adicionales que representaban los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior (fuerza de frenado y propulsión) en 101 puntos durante la fase de apoyo del ciclo de la marcha de última sesión de rehabilitación predicha por los distintos modelos. En el primer gráfico, asignamos colores a los puntos de la curva en base a los valores de la prueba estadística de Anderson-Darling. En el segundo gráfico coloreamos de rojo los puntos de la curva predicha dónde los residuos no seguían una distribución normal.

4.9.4 Regresión de mínimos cuadrados parciales (PLS regression)

El segundo algoritmo de aprendizaje automático supervisado que utilizamos fue la regresión de mínimos cuadrados parciales (PLS regression). Elegimos el PLS debido a que crea componentes latentes que reduzcan la dimensionalidad manteniendo información relevante. Por tanto, es útil cuando se trabaja con conjuntos de datos donde hay muchas variables predictoras y existe una posible multicolinealidad. Además, puede manejar tanto relaciones lineales como no lineales en los datos.

4.9.4.1 Entrenamiento y validación en el conjunto de entrenamiento

Desarrollamos una función encargada de entrenar modelos de regresión de mínimos cuadrados parciales y evaluar su rendimiento en el conjunto de entrenamiento. Aunque generalmente se recomienda escalar las variables antes de llevar a cabo un PLS, en nuestro caso decidimos no hacerlo debido a que los datos postprocesados proporcionados por GaitRec

habían sido normalizados previamente. Para encontrar el número óptimo de componentes o variables latentes para el modelo PLS, realizamos el ajuste de hiperparámetros (*hyperparameter tuning*), que se llevó a cabo mediante *GridSearchCV*. Se utilizó una validación cruzada de 10 iteraciones (*10-fold cross-validation*) y la métrica de puntuación (*scoring*) se estableció en la raíz del error cuadrático medio negativo (RMSE) para facilitar la minimización del error. Representamos gráficamente los valores del RMSE correspondientes a diferentes números de componentes, con el mejor número de componentes señalado, lo que ayudó a visualizar y tener una referencia de la elección óptima. Sin embargo, con el objetivo de evitar aumentar la complejidad del modelo, elegimos el número de componentes en el que la diferencia entre un componente y el siguiente no superara el 2%. Una vez seleccionado el número óptimo de componentes, volvimos a entrenar el modelo de PLS utilizando la configuración optima de 8 componentes. Esto fue un paso esencial para obtener el modelo final con la configuración óptima. A continuación, realizamos la validación cruzada con 10 iteraciones (*10-fold cross-validation*) y obtuvimos las predicciones utilizando el modelo entrenado. Finalmente, los resultados obtenidos para el conjunto de entrenamiento fueron el modelo PLS, los valores de la raíz del error cuadrático medio (RMSE) y del coeficiente de determinación (R^2), y los coeficientes de correlación de Pearson para el modelo en global.

En el caso del PLS, seguimos un enfoque diferente en comparación con las aproximaciones previamente explicadas. Esto se debe a que el PLS es una técnica de reducción de la dimensionalidad supervisada más avanzada que el PCA, ya que considera de manera integrada la relación entre las variables predictoras (X) y las variables objetivo (y). El PLS crea variables latentes que maximizan la covarianza entre las matrices X e y. A diferencia del PCA, que se enfoca únicamente en reducir la dimensionalidad de las variables predictoras, el PLS toma en cuenta tanto las características de la entrada como la relación con la salida. Esta característica lo convierte en una técnica más completa y con un mayor potencial en términos de rendimiento predictivo. En este sentido, el PLS podría considerarse una evolución más avanzada de la segunda aproximación.

4.9.4.2 Evaluación en el conjunto de prueba

En esta etapa, evaluamos la calidad y el rendimiento del modelo de regresión de mínimos cuadrados parciales entrenado en el conjunto de entrenamiento sobre datos no vistos, es decir, sobre el conjunto de prueba.

Utilizando el *pipeline* de evaluación anteriormente descrito, obtuvimos de manera automática las métricas de evaluación como la raíz del error cuadrático medio (RMSE), el coeficiente de correlación de Pearson, y el coeficiente de determinación (R^2) globales del modelo. Visualizamos los gráficos de dispersión de los valores predichos frente a los valores reales para cada una de las 101 variable objetivo, así como el gráfico de dispersión con todos los puntos para comparar los valores reales y los valores predichos por los modelos de manera global. Mostramos también el gráfico que resumía y presentaba de manera clara la predicción de los modelos y las métricas de evaluación empleadas. Para complementar la evaluación de los resultados de los modelos, representamos los distintos gráficos destinados a examinar y analizar la posible dependencia o relación entre el valor de las métricas de evaluación y los puntos de la curva predicha. Finalmente, validamos el ajuste de los modelos de regresión múltiple mediante el análisis de los residuos o errores generados por los modelos. Para ello representamos los residuos frente a los valores predichos por el modelo, y el histograma de los residuos.

4.9.5 Árboles de decisión

El tercer algoritmo de aprendizaje automático supervisado que utilizamos fueron los árboles de decisión o *Decision Trees*. Este algoritmo se utilizó como un punto de partida o una

primera aproximación a las técnicas de *ensamble learning* más avanzadas utilizadas posteriormente, como los bosques aleatorios (*Random Forest*) y *Gradient Boosting*.

Los árboles de decisión, que son algoritmos no lineales de aprendizaje automático, podrían ser apropiados para nuestra investigación ya que pueden resolver problemas de regresión, las predicciones generadas son fáciles de interpretar y explicar, y son capaces de modelar relaciones complejas entre las variables de entrada y predecir múltiples variables de salida simultáneamente. No obstante, a diferencia de algoritmos lineales como la regresión lineal múltiple o el PLS, los árboles de decisión son propensos a la alta varianza.

Entrenamos y evaluamos los modelos de regresión lineal múltiple para cada una de cuatro las aproximaciones, con el propósito de explorar cómo distintos preprocesamientos podrían impactar en las predicciones de los modelos.

4.9.5.1 Entrenamiento y validación en el conjunto de entrenamiento

Desarrollamos una función encargada de entrenar los árboles de decisión y evaluar su rendimiento en el conjunto de entrenamiento.

Inicializamos el árbol de regresión y definimos una serie de hiperparámetros para ajustar el modelo.

Los hiperparámetros que incluimos en el proceso de *hyperparameter tuning* fueron el máximo de profundidad del árbol (*max_depth*), el número mínimo de muestras requeridas para dividir un nodo interno (*min_samples_split*) y el número mínimo de muestras requeridas en un nodo hoja (*min_samples_leaf*).

Los rangos de valores para los hiperparámetros se muestran en la Fig. 17. Utilizamos la técnica de *GridSearchCV* para explorar diferentes las combinaciones de hiperparámetros. Esta búsqueda se realizó utilizando una validación cruzada de 10 iteraciones para evitar el sobreajuste, y la métrica de puntuación (*scoring*) se estableció en la raíz del error cuadrático medio negativo (RMSE) para facilitar la minimización del error.

Identificamos los mejores hiperparámetros que generaron el menor RMSE en la validación cruzada. Estos parámetros optimizados se utilizaron para entrenar el modelo final en el conjunto de entrenamiento.

A continuación, realizamos la validación cruzada con 10 iteraciones (*10-fold cross-validation*) utilizando el modelo optimizado y obtuvimos las predicciones de las variables de salida.

Finalmente, los resultados obtenidos para el conjunto de entrenamiento fueron el modelo de árbol de decisión, los valores de la raíz del error cuadrático medio (RMSE) y del coeficiente de determinación (R^2), y los coeficientes de correlación de Pearson para el modelo en global.

```
# Define a range of hyperparameters for tuning the Decision Tree
param_grid = {
    'max_depth': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60],
    'min_samples_split': [2, 5, 10, 15, 20, 25, 30, 35, 40],
    'min_samples_leaf': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]
}
```

Figura 17. Rangos de valores para los hiperparámetros de los árboles de decisión.

Entrenamos y validamos el modelo de árbol de decisión para cada una de cuatro las aproximaciones, de la misma forma que con el algoritmo de regresión lineal múltiple. Como se mencionó en apartados anteriores, el proceso de entrenamiento para cada una de las cuatro aproximaciones se realizó de la misma forma para todos los algoritmos utilizados en el estudio, exceptuando la regresión de mínimos cuadrados parciales (PLS).

4.9.5.2 Evaluación en el conjunto de prueba

En esta etapa, evaluamos la calidad y el rendimiento del modelo de árbol de decisión entrenado en el conjunto de entrenamiento sobre datos no vistos, es decir, sobre el conjunto de prueba. Como se comentó en apartados anteriores, el proceso de evaluación se realizó de la misma forma para todos los algoritmos, exceptuando la regresión de mínimos cuadrados parciales (PLS).

4.9.6 Random Forest

El cuarto algoritmo de aprendizaje automático supervisado que utilizamos fue el bosque aleatorio o *random forest*. Dado que ya habíamos explorado los árboles de decisión, optamos llevar nuestra investigación a un nivel superior utilizando métodos de *ensamble learning* como el *random forest*. Los bosques aleatorios están formados por muchos árboles de decisión individuales, y utilizan la agregación *bootstrap* para combinar estos árboles y alcanzar una solución mucho mejor que la proporcionada por cualquiera de los árboles individuales.

4.9.6.1 Entrenamiento y validación en el conjunto de entrenamiento

Desarrollamos una función encargada de entrenar el modelo de *random forest* y evaluar su rendimiento en el conjunto de entrenamiento.

Inicializamos el bosque aleatorio y definimos una serie de hiperparámetros para ajustar el modelo.

Los hiperparámetros que utilizamos en el proceso de *hyperparameter tuning* incluyeron el número de árboles en el bosque aleatorio (*n_estimators*), el número de características a considerar en cada división (*max_features*), la profundidad máxima de cada árbol (*max_depth*), el número mínimo de muestras necesarias para dividir un nodo (*min_samples_split*), y el número mínimo de muestras necesarias en cada nodo hoja (*min_samples_leaf*). Los rangos de valores para los hiperparámetros se muestran en la Fig. 18.

Para explorar la amplia gama de combinaciones de hiperparámetros de manera más eficiente, optamos por utilizar la técnica de *RandomizedSearchCV* en lugar de *GridSearchCV* para el ajuste de hiperparámetros en el modelo de *random forest*. Dado que *GridSearchCV* evalúa todas las combinaciones posibles, se volvió computacionalmente costoso en nuestro conjunto de datos. *RandomizedSearchCV*, por otro lado, realiza una búsqueda aleatoria de hiperparámetros, lo que redujo significativamente el tiempo de ejecución y permitió la exploración de diversas configuraciones, lo cual fue esencial para lograr un equilibrio entre rendimiento y eficiencia.

La búsqueda se realizó utilizando una validación cruzada de 10 iteraciones para evitar el sobreajuste, y la métrica de puntuación (*scoring*) se estableció en la raíz del error cuadrático medio negativo (RMSE) para facilitar la minimización del error.

Identificamos los mejores hiperparámetros que generaron el menor RMSE en la validación cruzada. Estos parámetros optimizados se utilizaron para entrenar el modelo final en el conjunto de entrenamiento. A continuación, realizamos la validación cruzada con 10 iteraciones (*10-fold cross-validation*) utilizando el modelo optimizado y obtuvimos las predicciones de las variables de salida.

Finalmente, los resultados obtenidos para el conjunto de entrenamiento fueron el modelo de *random forest*, los valores de la raíz del error cuadrático medio (RMSE) y del coeficiente de determinación (R^2), y los coeficientes de correlación de Pearson para el modelo en global.

Entrenamos y validamos el modelo de bosque aleatorio para cada una de cuatro las aproximaciones, de la misma forma que el resto de los algoritmos utilizados en el estudio, exceptuando la regresión de mínimos cuadrados parciales (PLS).

```
# Define a range of hyperparameters for tuning the Random Forest:  
    # Number of trees in random forest  
n_estimators = [int(x) for x in np.linspace(start = 100, stop = 1200, num = 12)]  
    # Number of features to consider at every split  
max_features = ['sqrt', 'log2', None]  
    # Maximum number of levels in tree  
max_depth = [int(x) for x in np.linspace(5, 30, num = 6)]  
    # Minimum number of samples required to split a node  
min_samples_split = [2, 5, 10, 15, 20, 25, 30, 35, 40, 100]  
    # Minimum number of samples required at each leaf node  
min_samples_leaf = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]  
    # Method of selecting samples for training each tree  
  
# Create the random grid  
random_grid = {'n_estimators': n_estimators,  
               'max_features': max_features,  
               'max_depth': max_depth,  
               'min_samples_split': min_samples_split,  
               'min_samples_leaf': min_samples_leaf}
```

Figura 18. Rangos de valores para los hiperparámetros de los árboles aleatorios.

4.9.6.2 Evaluación en el conjunto de prueba

Finalmente, evaluamos la calidad y el rendimiento del modelo de *random forest* entrenado en el conjunto de entrenamiento sobre datos no vistos, es decir, sobre el conjunto de prueba. Como se comentó en apartados anteriores, el proceso de evaluación se realizó de la misma forma para todos los algoritmos, exceptuando la regresión de mínimos cuadrados parciales (PLS).

4.9.7 Extreme Gradient Boosting (XGBoost)

El quinto algoritmo de aprendizaje automático supervisado que utilizamos fue el *eXtreme Gradient Boosting (XGBoost)*. La elección de implementar el *XGBoost* se fundamenta en una estrategia progresiva y evolutiva en relación con los algoritmos basados en árboles que utilizamos previamente. Empezamos con los árboles de decisión, y luego seguimos con los bosques aleatorios. Por tanto, en este contexto *XGBoost* es el próximo paso. Esta técnica de *ensamble learning* también se basa en árboles, pero introduce mejoras significativas en términos de eficiencia y rendimiento. *XGBoost* optimiza la creación de árboles individuales mediante el uso de *gradient boosting* y técnicas de regularización. Esto ayuda a mitigar el sobreajuste o *overfitting* y a obtener predicciones más precisas en comparación con los algoritmos anteriores.

4.9.7.1 Entrenamiento y validación en el conjunto de entrenamiento

Para la implementación de *eXtreme Gradient Boosting* en nuestra investigación, desarrollamos una función encargada de entrenar el modelo de *XGBoost* y evaluar su rendimiento en el conjunto de entrenamiento. Entrenamos un único modelo con *hyperparameter tuning* y lo utilizamos para predecir todas las 101 variables objetivo simultáneamente. Este enfoque simplificó el proceso de modelado, pero cabía la posibilidad de sacrificar algo de interpretabilidad.

Inicializamos el *XGBoostRegressor* y definimos una serie de hiperparámetros para ajustar el modelo. Los hiperparámetros que utilizamos en el proceso de *hyperparameter tuning* incluyeron el número de estimadores o árboles en el ensamble (*n_estimators*), la tasa de aprendizaje (*learning_rate*), el peso mínimo requerido en un nodo hijo para continuar

dividiendo el nodo (*min_child_weight*), la profundidad máxima de cada árbol (*max_depth*), la proporción de muestras utilizadas en cada iteración (*subsample*), la proporción de características utilizadas en cada árbol (*colsample_bytree*), la proporción de características utilizadas en cada nivel del árbol (*colsample_bylevel*), el parámetro de regularización L2 o Ridge (*reg_lambda*), el parámetro de regularización L1 o Lasso (*reg_alpha*), y el parámetro que controla cuánta reducción de pérdida es necesaria para realizar una partición (*gamma*). Los rangos de valores para los hiperparámetros se muestran en la Fig. 19.

Para explorar la amplia gama de combinaciones de hiperparámetros de manera más eficiente, optamos por utilizar la técnica de *RandomizedSearchCV*. Este enfoque nos permitió obtener combinaciones óptimas en menos tiempo. La búsqueda se realizó utilizando una validación cruzada de 5 iteraciones para evitar el sobreajuste, y la métrica de puntuación (*scoring*) se estableció en la raíz del error cuadrático medio negativo (RMSE) para facilitar la minimización del error. Identificamos los mejores hiperparámetros que generaron el menor RMSE en la validación cruzada. Estos parámetros optimizados se utilizaron para entrenar el modelo final en el conjunto de entrenamiento. A continuación, realizamos la validación cruzada con 10 iteraciones (*10-fold cross-validation*) utilizando el modelo optimizado y obtuvimos las predicciones de las variables de salida.

Finalmente, al igual que con los algoritmos previos, los resultados obtenidos para el conjunto de entrenamiento fueron el modelo de *XGBoost*, los valores de la raíz del error cuadrático medio (RMSE) y del coeficiente de determinación (R^2), y los coeficientes de correlación de Pearson para el modelo en global.

Entrenamos y validamos el modelo de *XGBoost* para cada una de cuatro las aproximaciones, de la misma forma que el resto de los algoritmos utilizados en el estudio, exceptuando la regresión de mínimos cuadrados parciales (PLS).

```
# Define a range of hyperparameters for tuning the XGBRegressor:
param_dist = {
    'n_estimators': [10, 100, 300, 500, 800, 1000, 2000, 3000, 4000, 5000], # int ranging from 10 to 5000
    'learning_rate': [0.01, 0.05, 0.1, 0.2, 0.3, 1.0], # real ranging from 0.01 to 1.0
    'min_child_weight': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], # int between 1 and 10
    'max_depth': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50], # int between 1 and 50
    'max_delta_step': [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20], # int between 0 and 20
    'subsample': [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], # real from 0.1 to 1.0
    'colsample_bytree': [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], # real from 0.1 to 1.0
    'colsample_bylevel': [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], # real from 0.1 to 1.0
    'reg_lambda': [1e-9, 1e-8, 1e-7, 1e-6, 1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1.0,
                   10.0, 20.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0, 90.0, 100.0], # real between 1e-9 and 100.0
    'reg_alpha': [1e-9, 0.1, 0.7, 0.8, 0.9, 1.0, 2.0, 3.0, 4.0, 5.0, 10.0,
                  20.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0, 90.0, 100.0], # real between 1e-9 and 100.0
    'gamma': [1e-9, 0.1, 0.7, 0.8, 0.9, 1.0, 2.0, 3.0, 4.0, 5.0, 10.0,
              20.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0, 90.0, 100.0], # real between 1e-9 and 100.0
}
```

Figura 19. Rangos de valores para los hiperparámetros del *eXtreme Gradient Boosting* (XGBoost).

4.9.7.2 Evaluación en el conjunto de prueba

Finalmente, evaluamos la calidad y el rendimiento del modelo de XGBoost entrenado en el conjunto de entrenamiento sobre datos no vistos, es decir, sobre el conjunto de prueba. Como se comentó en apartados anteriores, el proceso de evaluación se realizó de la misma forma para todos los algoritmos, exceptuando la regresión de mínimos cuadrados parciales (PLS).

4.9.8 Support Vector Regression (SVR)

El sexto (y último) algoritmo de aprendizaje automático supervisado que utilizamos fue la regresión de vectores de soporte o *Support Vector Regression (SVR)*.

Teniendo en cuenta los algoritmos utilizados hasta ahora, la regresión de vectores de soporte proporcionó una perspectiva adicional a nuestro estudio.

4.9.8.1 Entrenamiento y validación en el conjunto de entrenamiento

Para la implementación de regresión de vectores de soporte en nuestra investigación, desarrollamos una función encargada de entrenar los modelos de SVR y evaluar su rendimiento en el conjunto de entrenamiento.

Sin embargo, cabe destacar que no todos los algoritmos de regresión admiten la regresión múltiple. La regresión de vectores de soporte es un ejemplo de ello. Por tanto, tuvimos que desarrollar un modelo independiente para cada valor numérico a predecir, lo que se conoce como enfoque o aproximación directa (*direct approach*). Aunque esto supone que las salidas son independientes entre sí, lo que podría no ser una suposición correcta.

Implementamos un proceso de entrenamiento y evaluación utilizando el concepto de *MultiOutputRegressor*. Este enfoque nos permitió manejar simultáneamente múltiples variables objetivo.

Realizamos *hyperparameter tuning* con el objetivo de encontrar la configuración óptima para el modelo SVR. Utilizamos la técnica de *RandomizedSearchCV* para explorar diversas combinaciones de hiperparámetros. Los hiperparámetros que utilizamos fueron el parámetro de regularización que controla la penalización por errores en la función de regresión (*C*), la tolerancia del error (*epsilon*), la función de *kernel* utilizada en el modelo SVR (*kernel*), el coeficiente del kernel, que afecta la forma de la función de decisión (*gamma*), y el grado del *kernel* polinómico (*degree*). Los rangos de valores para los hiperparámetros se muestran en la Fig. 20.

Para explorar la amplia gama de combinaciones de hiperparámetros de manera eficiente, optamos por utilizar la técnica de *RandomizedSearchCV*.

La búsqueda se realizó utilizando una validación cruzada de 5 iteraciones para evitar el sobreajuste, y la métrica de puntuación (*scoring*) se estableció en la raíz del error cuadrático medio negativo (RMSE) para facilitar la minimización del error.

Identificamos los mejores hiperparámetros que generaron el menor RMSE en la validación cruzada. Estos parámetros optimizados se utilizaron para entrenar el modelo final en el conjunto de entrenamiento. A continuación, realizamos la validación cruzada con 10 iteraciones (*10-fold cross-validation*) utilizando el modelo optimizado y obtuvimos las predicciones de las variables de salida.

Finalmente, al igual que con los algoritmos previos, los resultados obtenidos para el conjunto de entrenamiento fueron el modelo de *SVR*, los valores de la raíz del error cuadrático medio (RMSE) y del coeficiente de determinación (R^2), y los coeficientes de correlación de Pearson para el modelo en global.

Entrenamos y validamos el modelo de regresión de vectores de soporte para cada una de cuatro las aproximaciones, de la misma forma que el resto de los algoritmos utilizados en el estudio, exceptuando la regresión de mínimos cuadrados parciales (PLS).

```
# Define the parameter grid for hyperparameter tuning the SVR:
param_dist = {
    'estimator__C': np.logspace(-3, 3, num=10), # Regularization parameter
    'estimator__epsilon': np.logspace(-4, 2, 7),
    'estimator__kernel': ['linear', 'poly', 'rbf', 'sigmoid'],
    'estimator__gamma': ['scale', 'auto'],
    'estimator__degree': [2, 3, 4, 5]
}
```

Figura 20. Rangos de valores para los hiperparámetros de la regresión de vectores de soporte (SVR).

4.9.8.2 Evaluación en el conjunto de prueba

Finalmente, evaluamos la calidad y el rendimiento del modelo de regresión de vectores de soporte entrenado en el conjunto de entrenamiento sobre datos no vistos, es decir, sobre el conjunto de prueba. Como se comentó en apartados anteriores, el proceso de evaluación se realizó de la misma forma para todos los algoritmos, exceptuando la regresión de mínimos cuadrados parciales (PLS).

5 Resultados

El capítulo de los resultados de los diferentes algoritmos de aprendizaje automático se estructurará en dos secciones. En la primera sección, se llevará a cabo un análisis global de los resultados obtenidos por los distintos modelos. En la segunda sección, se seleccionarán los dos modelos con mayor rendimiento y los dos modelos con el rendimiento más bajo. Estos cuatro modelos se analizarán en detalle y se compararán entre sí.

5.1 Análisis global de los resultados de los modelos

En la Tabla 7 se muestran los resultados obtenidos por los distintos algoritmos de aprendizaje automático en cada una de las distintas aproximaciones para los conjuntos de entrenamiento y prueba. Por otra parte, en la Fig. 21 se muestra el promedio de los rangos para cada una de las métricas de evaluación (RMSE, R^2 , y coeficiente de correlación de Pearson) de cada modelo de aprendizaje automático evaluado en el conjunto de prueba.

Observando las métricas de evaluación globales de la Tabla 7, notamos que en general, la mayoría de los modelos presentaron resultados muy similares en términos de RMSE, coeficiente de determinación y coeficiente de correlación de Pearson. No obstante, aplicar el método de rangos específicos de las variables objetivo o *target-specific ranks* (Fig. 21), nos permitió resaltar las diferencias entre los distintos modelos.

En los modelos de regresión lineal múltiple, en general obtuvimos valores bajos de RMSE en los conjuntos de entrenamiento y prueba en todas las aproximaciones. En términos del coeficiente de correlación de Pearson, se observó una fuerte relación lineal entre los valores reales y predicciones de los modelos en los que se aplicó la aproximación 1, 2, y 4 en los datos de entrenamiento y prueba. No obstante, es importante señalar que, en la primera y tercera aproximación, el coeficiente de determinación (R^2) global en el conjunto de prueba fue negativo. Esto indicó que los valores predichos por los modelos eran peores que si utilizáramos la media como valor predicho, y que, por tanto, los modelos no se ajustaron ni generalizaron bien en los datos de prueba. Cabe destacar que el modelo de regresión lineal múltiple con los datos de entrada reducidos mediante el PCA (aproximación 2) proporcionó resultados notablemente buenos, situándose entre los modelos con mayor rendimiento.

En cuanto a los árboles de decisión, proporcionaron resultados similares a los modelos de regresión lineal múltiple en las aproximaciones 1, 3, y 4. En términos de correlación, los árboles de decisión fueron los que presentaron un coeficiente de correlación global más bajo, indicando que presentaron las predicciones más alejadas respecto los datos reales.

Los modelos de bosques aleatorios se ubicaron en un punto intermedio en términos de rendimiento en comparación con otros modelos. El hecho de aplicar distintos preprocesamientos a los datos de entrada y de salida no pareció afectar en gran medida a su rendimiento, pues las cuatro aproximaciones proporcionaron resultados similares en términos de RMSE, coeficiente de determinación, y coeficiente de correlación de Pearson.

Para XGBoost, obtuvimos valores de RMSE, R^2 , y coeficientes de correlación similares en los conjuntos de entrenamiento y prueba en las cuatro aproximaciones. A diferencia de los demás algoritmos, la reducción de la dimensionalidad de los datos de salida mediante el PCA (aproximación 3) proporcionó los mejores resultados de entre las cuatro aproximaciones, mientras que la reducción de los datos de entrada (aproximación 2) condujo el rendimiento más bajo en términos de RMSE, coeficiente de determinación, y coeficiente de correlación de Pearson.

El rendimiento de los modelos de regresión de vectores de soporte o *Support Vector Regression* (SVR) fue notablemente alto. Es interesante destacar que, a diferencia de los otros modelos, este algoritmo mostró un rendimiento elevado incluso cuando no se aplicó ninguna reducción de la dimensionalidad en los datos (aproximación 1). Junto al PLS, estos modelos

destacaron por tener los rendimientos más sólidos y consistentes en términos de RMSE, coeficiente de determinación, y coeficiente de correlación de Pearson en ambos conjuntos de datos.

Por último, la regresión de mínimos cuadrados parciales (PLS) destacó como el modelo más consistente y efectivo en términos de rendimiento, superando a todos los demás modelos al mantener resultados sólidos y coherentes tanto en los conjuntos de entrenamiento como en los de prueba.

En términos generales, la mayoría de los modelos mostraron coeficientes de determinación bajos y coeficientes de correlación altos en los conjuntos de entrenamiento y prueba. Además, a excepción de XGBoost y SVR, los modelos entrenados y evaluados con los datos de entrada reducidos, superaron en rendimiento a aquellos que utilizaron datos de salida reducidos o que no experimentaron ninguna reducción de la dimensionalidad.

Tabla 7. Comparación del rendimiento de los distintos algoritmos de aprendizaje automático en cada una de las distintas aproximaciones en los conjuntos de entrenamiento y prueba.

Algoritmo	Aproximación	Dataset	RMSE (red./orig.)	R² (red./orig.)	Corr. Pearson (red./orig.)
MLR	1	Train	0.023	0.860	0.996
		Test	0.020	-0.144	0.976
	2	Train	0.015	0.550	0.985
		Test	0.014	0.442	0.989
	3	Train	0.097	0.855	0.660
		Test	0.083/0.020	0.026/-0.100	0.613/0.977
	4	Train	0.060	0.607	0.824
		Test	0.055/0.014	0.547/0.428	0.785/0.988
PLS	-	Train	0.014	0.585	0.986
		Test	0.014	0.509	0.989
Decision Tree	1	Train	0.021	0.195	0.974
		Test	0.018	0.072	0.980
	2	Train	0.020	0.248	0.976
		Test	0.017	0.201	0.983
	3	Train	0.088	0.173	0.617
		Test	0.078/0.018	0.119/0.062	0.520/0.980
	4	Train	0.085	0.186	0.658
		Test	0.072/0.017	0.216/0.197	0.595/0.983
Random Forest	1	Train	0.018	0.397	0.981
		Test	0.015	0.326	0.987
	2	Train	0.018	0.405	0.981
		Test	0.015	0.350	0.987
	3	Train	0.074	0.364	0.752
		Test	0.064/0.016	0.366/0.304	0.699/0.986
	4	Train	0.073	0.404	0.765
		Test	0.061/0.015	0.419/0.340	0.727/0.987
XGBoost	1	Train	0.016	0.513	0.985
		Test	0.014	0.421	0.988
	2	Train	0.018	0.411	0.982
		Test	0.015	0.306	0.987
	3	Train	0.070	0.481	0.778
		Test	0.059/0.015	0.491/0.390	0.741/0.987
	4	Train	0.068	0.508	0.787
		Test	0.059/0.015	0.471/0.376	0.746/0.988
SVR	1	Train	0.015	0.598	0.987
		Test	0.014	0.509	0.989
	2	Train	0.016	0.496	0.985
		Test	0.014	0.430	0.988
	3	Train	0.063	0.583	0.825
		Test	0.056/0.014	0.541/0.412	0.775/0.988
	4	Train	0.063	0.566	0.823
		Test	0.057/0.015	0.518/0.389	0.772/0.988

Nota: las distintas aproximaciones se refieren a la forma en la que se procesaron los datos. La aproximación 1 se refiere a los datos sin procesar (en su forma original), la aproximación 2 implica la aplicación del PCA en los datos de entrada (X), la aproximación 3 implica la aplicación del PCA en los datos de salida (Y), y la aproximación 4 implica la aplicación del PCA en los datos de entrada (X) y de salida (Y). Con el PLS no se siguió este enfoque ya que se trata de una técnica más avanzada que el PCA.

La abreviación "red." hace referencia a la evaluación de las predicciones en el espacio reducido (aproximaciones 3 y 4), mientras que la abreviación "orig." hace referencia a la evaluación de las predicciones en el espacio original reconstruido (aproximaciones 3 y 4).

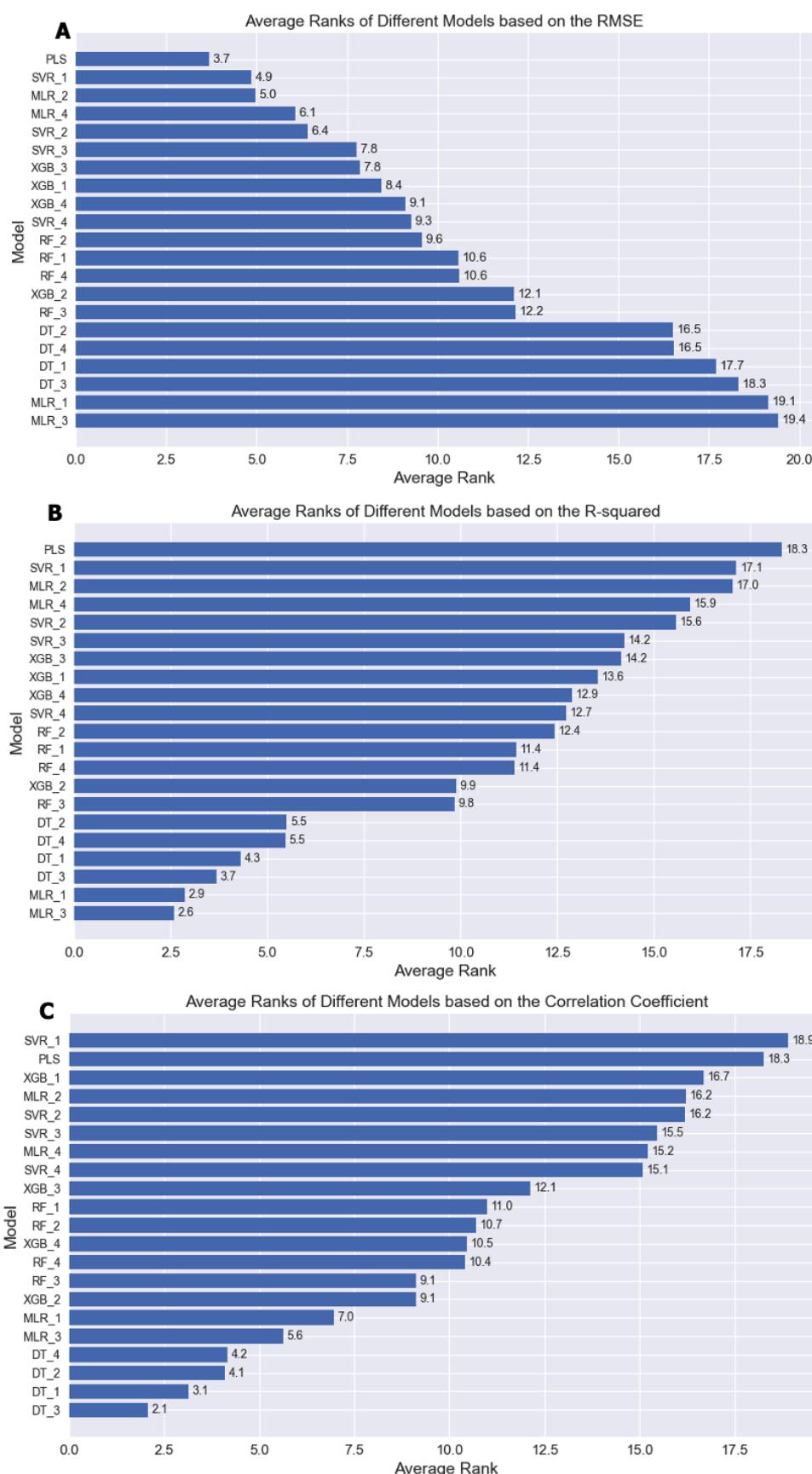


Figura 21. Promedio de los rangos para cada una de las métricas de evaluación de cada modelo de aprendizaje automático evaluado en el conjunto de prueba. A) Promedio de los rangos en base al RMSE. B) Promedio de los rangos en base al R^2 . C) Promedio de los rangos en base al coeficiente de correlación de Pearson.

5.2 Selección y análisis de los modelos de más alto y bajo rendimiento

Una vez analizados globalmente los resultados obtenidos por los modelos de aprendizaje automático en las distintas aproximaciones, llevaremos a cabo una evaluación detallada de los modelos que demostraron el mejor y el peor rendimiento a la hora de predecir los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior (fuerza de frenado y propulsión) de la última sesión de terapia en función de las características de la primera sesión en pacientes con deficiencias musculoesqueléticas en el tobillo.

Por un lado, se realizará un análisis detallado de dos modelos que destacaron por su alto rendimiento: la regresión de mínimos cuadrados parciales (PLS) y la regresión de vectores de soporte (SVR) (ver Fig. 21). Por otro lado, se analizarán y compararán dos modelos que mostraron un desempeño menos favorable: el modelo de regresión lineal múltiple entrenado y evaluado con los datos originales (aproximación 1) y el modelo de regresión lineal múltiple entrenado y evaluado con los datos de salida reducidos mediante el PCA (aproximación 3).

Los modelos se presentarán en orden descendente según su rendimiento y se analizarán y compararán durante el análisis de cada uno. Por tanto, se empezará con la regresión de mínimos cuadrados parciales (PLS) y el modelo de regresión de vectores de soporte (SVR) entrenado y evaluado en los datos originales (aproximación 1). A continuación, se procederá con el modelo de regresión lineal múltiple entrenado y evaluado con los datos originales (aproximación 1) y el modelo de regresión lineal múltiple entrenado y evaluado con los datos de salida reducidos mediante el PCA (aproximación 3).

5.2.1.1 Regresión de mínimos cuadrados parciales (PLS)

Para el modelo de regresión de mínimos cuadrados parciales (PLS) entrenado con una configuración óptima de 8 componentes, se obtuvieron los siguientes resultados.

En el conjunto de entrenamiento, el valor promedio de la raíz del error cuadrático medio (RMSE) del modelo, calculado a través de las 10 iteraciones de validación cruzada para cada una de las 101 variables objetivo, fue de 0.014. Este valor de RMSE, que representa el promedio de las discrepancias entre las predicciones del modelo y los valores reales en las variables objetivo, puede considerarse bajo en relación con la variabilidad de las fuerzas de reacción del suelo durante la fase de apoyo de la marcha. Por tanto, indica que, en promedio, las predicciones del modelo fueron precisas y se ajustaron bien a los datos de entrenamiento observados.

El valor promedio del coeficiente de determinación (R^2) del modelo, calculado a través de las 10 iteraciones de validación cruzada para cada variable objetivo, fue de 0.585. Esto significa que, en promedio, el 58.5% de la variabilidad en las variables objetivo fue explicada por el modelo, es decir, por las variables predictoras, en el conjunto de entrenamiento.

El valor promedio del coeficiente de correlación de Pearson global del modelo, calculado a través de las 10 iteraciones de validación cruzada para cada variable objetivo, fue de 0.986. Este valor promedio indica una correlación positiva fuerte entre las variables predictoras y las variables objetivo en el conjunto de entrenamiento.

En el conjunto de prueba, el valor promedio de la raíz del error cuadrático medio (RMSE) del modelo fue de 0.014. Esto significa que, en promedio, las predicciones del modelo presentaron una discrepancia de 0.014 unidades en comparación con los valores reales en las variables objetivo. Este valor de RMSE, que es igual al del conjunto de entrenamiento, parece indicar que el modelo generalizó con éxito en los nuevos datos del conjunto de prueba.

El valor promedio del coeficiente de determinación (R^2) en el conjunto de prueba fue de 0.509, similar al obtenido en el conjunto de entrenamiento. Este valor significa que, en promedio, el 50.9% de la variabilidad en las variables objetivo fue explicada por el modelo, es decir, por las variables predictoras, en el conjunto de prueba.

En el conjunto de prueba, el valor promedio del coeficiente de correlación de Pearson fue de 0.989, superando al del conjunto de entrenamiento. Este valor cercano a 1 indica una fuerte relación lineal entre las predicciones del modelo y los valores reales en las variables objetivo. Es decir, las predicciones del modelo siguieron una tendencia similar a los datos reales en el conjunto de prueba.

En la Fig. 22, se representa de manera global los valores reales frente a los valores predichos por el modelo en el conjunto de prueba. Podemos observar el valor promedio del coeficiente de correlación de Pearson de 0.989 mencionado anteriormente, la línea de correlación perfecta, la línea de regresión lineal que representa la relación lineal ajustada entre los valores reales y predichos, y la ecuación de la recta de regresión. La pendiente de la recta fue de 0.947, lo que significa que, en promedio, las predicciones tendieron a ser un poco más bajas que los valores reales del conjunto de prueba.

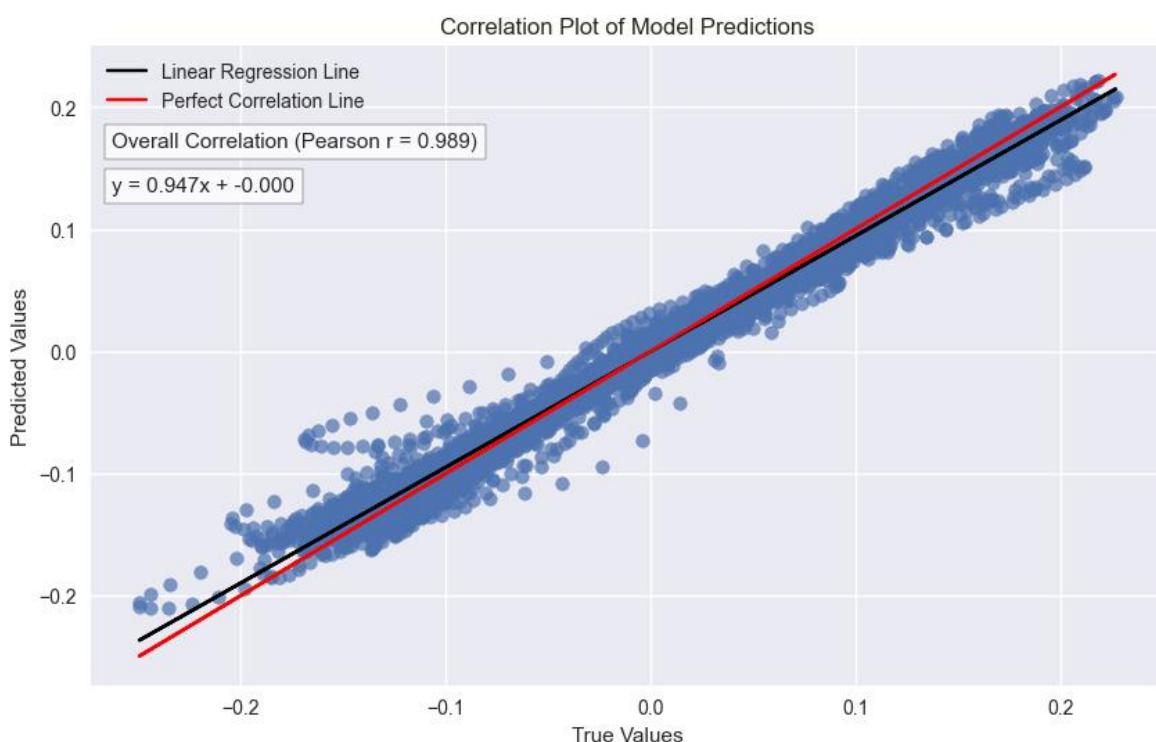


Figura 22. Comparación entre los valores reales y predichos por el modelo de regresión de mínimos cuadrados parciales (PLS). Se presenta el valor promedio del coeficiente de correlación de Pearson obtenido en el conjunto de prueba, la línea de correlación perfecta, la línea de regresión lineal ajustada y la ecuación de la recta de regresión.

En la Fig. 23 se muestra un gráfico que resume y presenta de manera clara la predicción del modelo de regresión de mínimos cuadrados parciales (PLS), así como los valores promedio de las métricas de evaluación empleadas. Se mostraron los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior (fuerza de frenado y propulsión) en los 101 puntos durante la fase de apoyo del ciclo de la marcha. Se puede notar una clara diferencia entre la curva perteneciente a los controles sanos (curva negra) frente a las curvas de los pacientes con desórdenes musculoesqueléticos en el tobillo, estando la primera por encima de las otras en los picos de máxima frenada y propulsión.

Es importante destacar que la curva correspondiente a la predicción de la última sesión de rehabilitación (curva azul) se situó entre las curvas correspondientes a la primera y última sesión real de rehabilitación (curva roja y verde, respectivamente) durante dos momentos específicos: el apoyo del talón (primer pico en el gráfico, asociado al pico máximo de frenado) y el despegue del talón (segundo pico en el gráfico, relacionado con el pico máximo de propulsión). Estos resultados indican que el modelo de PLS logra capturar patrones similares a los observados en las sesiones reales de rehabilitación. Sin embargo, no logra predecir la curva con total exactitud, especialmente en el pico máximo de frenada.

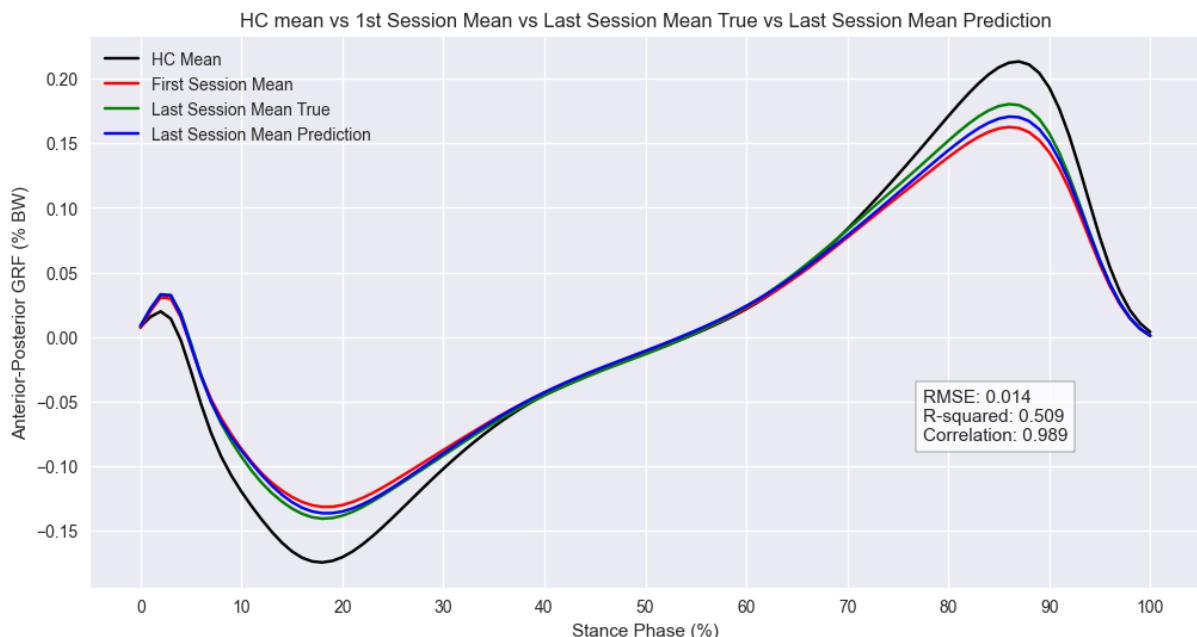


Figura 23. Comparación de los valores promedio correspondientes a la primera y última sesión real (en rojo y verde, respectivamente), y a la última sesión de rehabilitación predicha (en azul) por el modelo de regresión de mínimos cuadrados parciales (PLS). Se incorpora como referencia la curva que representa los valores promedio de los controles sanos (en negro), así como los valores de las distintas métricas de evaluación obtenidas por el modelo en los datos no vistos.

Las gráficas de codificación por colores (Fig. 24) nos mostraron que los valores de RMSE eran mayores en picos de la curva predicha, es decir, durante la fase de apoyo o choque del talón contra el suelo, tal y como se observa en la Fig. 24a. Esto coincide con lo que observamos en la Fig. 23. Por tanto, el modelo tendió a tener un rendimiento ligeramente inferior en esta fase del ciclo de la marcha.

En la Fig. 24b, podemos observar valores de R^2 de entre aproximadamente 0.30 y 0.35 en el pico de máxima frenada. Por otro lado, se observa que, durante la fase del despegue del talón, los valores del coeficiente de determinación fueron más elevados, llegando hasta 0.70 aproximadamente, indicando un mejor ajuste del modelo en esa fase del ciclo de la marcha.

Los valores de los coeficientes de correlación de Pearson fueron disminuyendo durante la fase de apoyo del talón, alcanzando los valores más bajos en el punto máximo de frenada. Por otro lado, antes y después del punto máximo de propulsión (fase de despegue del talón), los coeficientes de correlación son más altos (Fig. 24c). Esto indica que hubo una correlación positiva más fuerte entre los valores reales y los valores predichos en las fases de despegue del talón en comparación con la fase de apoyo del talón.

Al realizar el análisis de los residuos para validar el ajuste del modelo de regresión de mínimos cuadrados parciales (PLS), se generaron las gráficas presentadas en la Fig. 25. En el histograma, se aprecia una distribución que se parece a una distribución normal, con la excepción de la presencia de un valor atípico en la parte izquierda del gráfico. Por otro lado, en el gráfico de los residuos frente a los valores predichos, observamos que los residuos se distribuyen de forma relativamente uniforme alrededor del cero. Si embargo, se aprecia una ligera tendencia descendente hacia la derecha. Esta ligera tendencia podría significar que en los valores negativos, es decir, la fuerza de frenada, los errores tendieron a ser más elevados en comparación con otras partes de la curva. Es decir, el modelo cometió errores que sesgaron las predicciones hacia arriba, ya que la curva real se encontraba por debajo de las predicciones. Adicionalmente, podemos identificar el valor atípico en la esquina inferior izquierda, que coincide con el valor atípico identificado en el histograma.

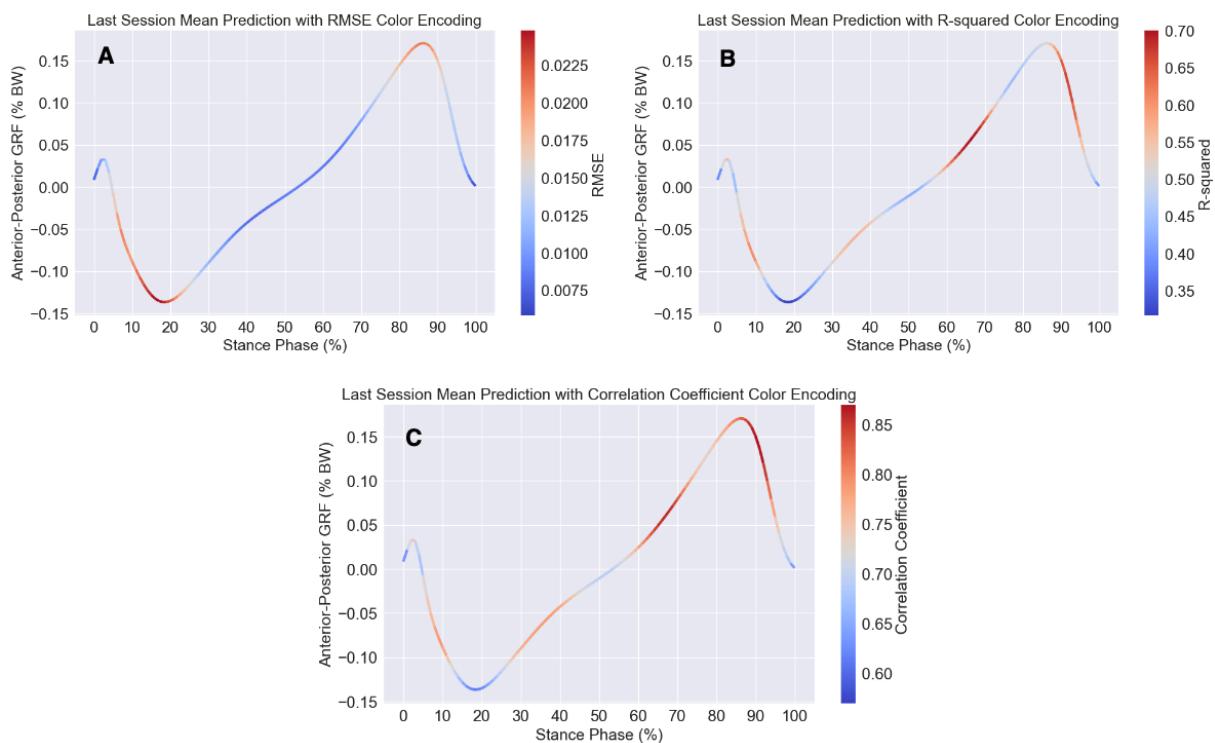


Figura 24. Curvas correspondientes a la última sesión predicha por el modelo de regresión de mínimos cuadrados parciales (PLS). Estas curvas están coloreadas en función de los valores de las diversas métricas de regresión obtenidas al evaluar el modelo en el conjunto de prueba. A) Puntos de la curva predicha coloreados en función del RMSE. B) Puntos de la curva predicha coloreados en función del R^2 . C) Puntos de la curva predicha coloreados en función del coeficiente de correlación de Pearson.

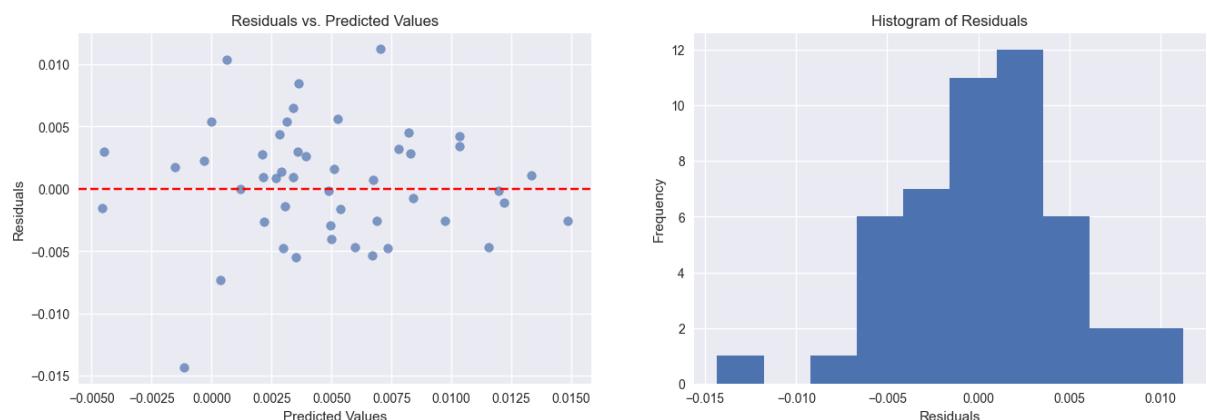


Figura 25. Gráfico de residuos frente a los valores predichos (izquierda) por el modelo de regresión de mínimos cuadrados parciales (PLS), e histograma de los residuos (derecha).

5.2.1.2 Regresión de vectores de soporte (SVR) en datos originales

Para el modelo de regresión de vectores de soporte (SVR) entrenado y evaluado en los datos originales, es decir, con los datos sin reducción de la dimensionalidad (aproximación 1), se obtuvieron los siguientes resultados.

El modelo se entrenó con la configuración óptima de hiperparámetros. Los mejores hiperparámetros obtenidos en el proceso de *hyperparameter tuning* utilizando la validación cruzada de 5 iteraciones fueron un parámetro de regularización “C” de 46.41, una tolerancia del error (*epsilon*) de 0.001, un kernel gaussiano (“rbf”), y un coeficiente del kernel (*gamma*) que utiliza 1/n_características (“auto”). El grado (*degree*) de la función *kernel* polinómica ('poly') se ignoró debido a que el kernel era gaussiano.

En el conjunto de entrenamiento, el valor promedio de la raíz del error cuadrático medio (RMSE) del modelo, calculado a través de las 10 iteraciones de validación cruzada para cada una de las 101 variables objetivo, fue de 0.015. Este valor de RMSE, que representa el promedio de las discrepancias entre las predicciones del modelo y los valores reales en las variables objetivo y que es ligeramente superior al obtenido en el modelo de PLS, puede considerarse bajo en relación con la variabilidad de las fuerzas de reacción del suelo durante la fase de apoyo de la marcha. Por tanto, indica que, en promedio, las predicciones del modelo fueron precisas y se ajustaron bien a los datos de entrenamiento observados.

El valor promedio del coeficiente de determinación (R^2) del modelo, calculado a través de las 10 iteraciones de validación cruzada para cada variable objetivo, fue de 0.598, ligeramente superior al del modelo PLS. Esto significa que, en promedio, el 59.8% de la variabilidad en las variables objetivo fue explicada por el modelo, es decir, por las variables predictoras, en el conjunto de entrenamiento.

El valor promedio del coeficiente de correlación de Pearson global del modelo, calculado a través de las 10 iteraciones de validación cruzada para cada variable objetivo, fue de 0.987, ligeramente superior al del PLS. Este valor promedio indica una correlación positiva fuerte entre las variables predictoras y las variables objetivo en el conjunto de entrenamiento.

En el conjunto de prueba, el valor promedio de la raíz del error cuadrático medio (RMSE) del modelo fue de 0.014, coincidiendo con el del modelo PLS en el conjunto de prueba. Esto significa que, en promedio, las predicciones del modelo SVR presentaron una discrepancia de 0.014 unidades en comparación con los valores reales en las variables objetivo. Este valor de RMSE es ligeramente inferior al observado en el conjunto de entrenamiento. No obstante, a

pesar de esta leve diferencia, parece que el modelo generalizó con éxito en los nuevos datos del conjunto de prueba.

El valor promedio del coeficiente de determinación (R^2) en el conjunto de prueba fue de 0.509. Este valor es ligeramente inferior al obtenido en el conjunto de entrenamiento e idéntico al obtenido por modelo PLS en el conjunto de prueba. Un coeficiente de determinación del 0.509 significa que, en promedio, el 50.9% de la variabilidad en las variables objetivo fue explicada por el modelo, es decir, por las variables predictoras, en el conjunto de prueba. Por tanto, el modelo logró mantener un rendimiento similar al del modelo PLS en la fase de prueba.

En el conjunto de prueba, observamos que el valor promedio del coeficiente de correlación de Pearson fue de 0.989, superando al del conjunto de entrenamiento y coincidiendo con el obtenido por el modelo PLS en el conjunto de prueba. Este valor cercano a 1 indica una fuerte relación lineal entre las predicciones del modelo y los valores reales en las variables objetivo. Es decir, las predicciones del modelo se ajustaron muy bien a la tendencia de los datos reales en el conjunto de prueba, lo que refleja una buena capacidad para capturar las relaciones entre las variables predictoras y las variables objetivo en el conjunto de datos de no vistos.

En la Fig. 26, se representa de manera global los valores reales frente a los valores predichos por el modelo de regresión de vectores de soporte (aproximación 1) en el conjunto de prueba. Podemos observar el valor promedio del coeficiente de correlación de Pearson de 0.989 mencionado anteriormente, la línea de correlación perfecta, la línea de regresión lineal que representa la relación lineal ajustada entre los valores reales y predichos, y la ecuación de la recta de regresión. La pendiente de la recta fue de 0.942, lo indica que, en promedio, las predicciones del modelo tendieron a ser un poco más bajas que los valores reales en el conjunto de prueba. Este valor de pendiente de la recta fue ligeramente inferior al obtenido por el modelo PLS, lo que significa que, en promedio, las predicciones del modelo SVR tendieron a ser un poco más bajas que los valores reales en el conjunto de prueba en comparación con el modelo PLS. Es decir, el modelo SVR tuvo una ligera tendencia a subestimar las respuestas en relación con el modelo PLS en el conjunto de prueba. Esto podría indicar que el modelo SVR tiende a ser ligeramente más conservador al hacer predicciones y no predice valores tan altos como el modelo PLS.

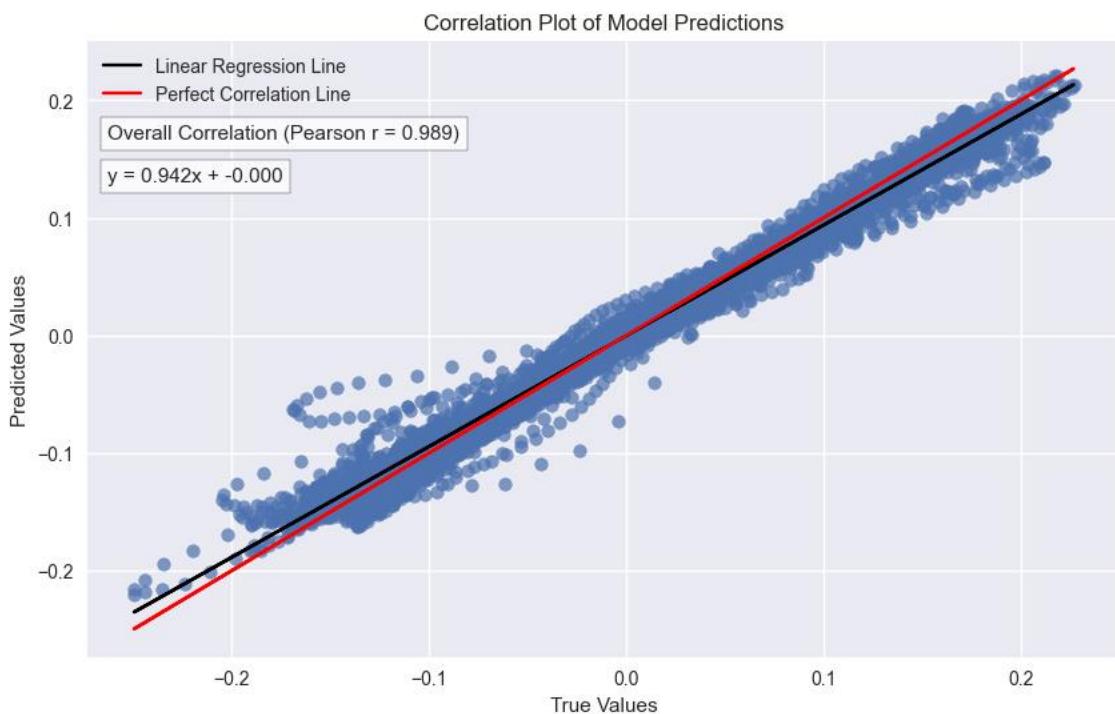


Figura 26. Comparación entre los valores reales y predichos por el modelo de regresión de vectores de soporte (SVR) entrenado y evaluado en los datos originales (aproximación 1: datos sin reducción de la dimensionalidad). Se presenta el valor promedio del coeficiente de correlación de Pearson obtenido en el conjunto de prueba, la línea de correlación perfecta, la línea de regresión lineal ajustada y la ecuación de la recta de regresión.

En la Fig. 27 se muestra un gráfico que resume y presenta de manera visual la predicción del modelo de regresión de vectores de soporte (SVR) entrenado y evaluado en los datos originales, junto con los valores promedio de las métricas de evaluación utilizadas. En este gráfico se muestran los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior (fuerza de frenado y propulsión) en los 101 puntos durante la fase de apoyo del ciclo de la marcha.

Como en el caso del modelo PLS, se puede notar una clara diferencia entre la curva perteneciente a los controles sanos (representada en negro) frente a las curvas de los pacientes con desórdenes musculoesqueléticos en el tobillo. La curva de los controles sanos se encuentra por encima de las otras en los picos de máxima frenada y propulsión.

Por un lado, observamos que en la primera parte de la fase de apoyo del talón y en el pico de máxima frenada, la curva correspondiente a la predicción de la última sesión de rehabilitación (representada en azul) se situó muy próxima y casi superpuesta a la curva correspondiente a la primera sesión de rehabilitación (en rojo). En la parte final de la fase de apoyo del talón, la curva azul se superpuso a la curva correspondiente a la última sesión de rehabilitación real.

Por otro lado, durante el despegue del talón (segundo pico en el gráfico, relacionado con el pico máximo de propulsión), la curva correspondiente a la predicción de la última sesión de rehabilitación (en azul) se situó ligeramente por encima de la curva correspondiente a la primera sesión de rehabilitación, quedando situada entre las curvas correspondientes a la primera y última sesión real de rehabilitación.

Aunque los valores promedio de las métricas de evaluación en el conjunto de prueba coincidan para los modelos SVR y PLS, se puede observar que en el caso del modelo SVR, durante la fase de apoyo y despegue del talón, las predicciones no se ajustan tan bien a la curva real en comparación con el modelo PLS.

Esto queda reflejado en las gráficas de codificación por colores (Fig. 28). Tal y como se muestra en la Fig. 28a, los valores de RMSE fueron mayores en los picos de la curva predicha, es decir, durante la fase de apoyo (o choque del talón contra el suelo), y durante la fase de despegue del talón. No obstante, estos errores fueron mayores que los errores de las predicciones generadas por el modelo PLS. Por tanto, el rendimiento del modelo SVR fue ligeramente inferior al del modelo PLS en estas fases del ciclo de la marcha. En la Fig. 28b, podemos observar que valores de R^2 a lo largo de la curva predicha coinciden con los valores obtenidos por el modelo PLS. Los valores de los coeficientes de correlación de Pearson fueron disminuyendo durante la fase de apoyo del talón, alcanzando los valores más bajos en el punto máximo de frenada. Por otro lado, antes y después del punto máximo de propulsión (correspondiente a la fase de despegue del talón), los coeficientes de correlación fueron más altos (Fig. 28c). Este patrón fue similar al obtenido por el modelo PLS. Sin embargo, para el modelo SVR se observaron puntos en la curva con una correlación más alta, llegando hasta un 90% de correlación en algunos puntos de la fase de despegue del talón. Por lo tanto, se evidenció una correlación positiva más fuerte entre los valores reales y las predicciones en las fases de despegue del talón en comparación con la fase de apoyo del talón.

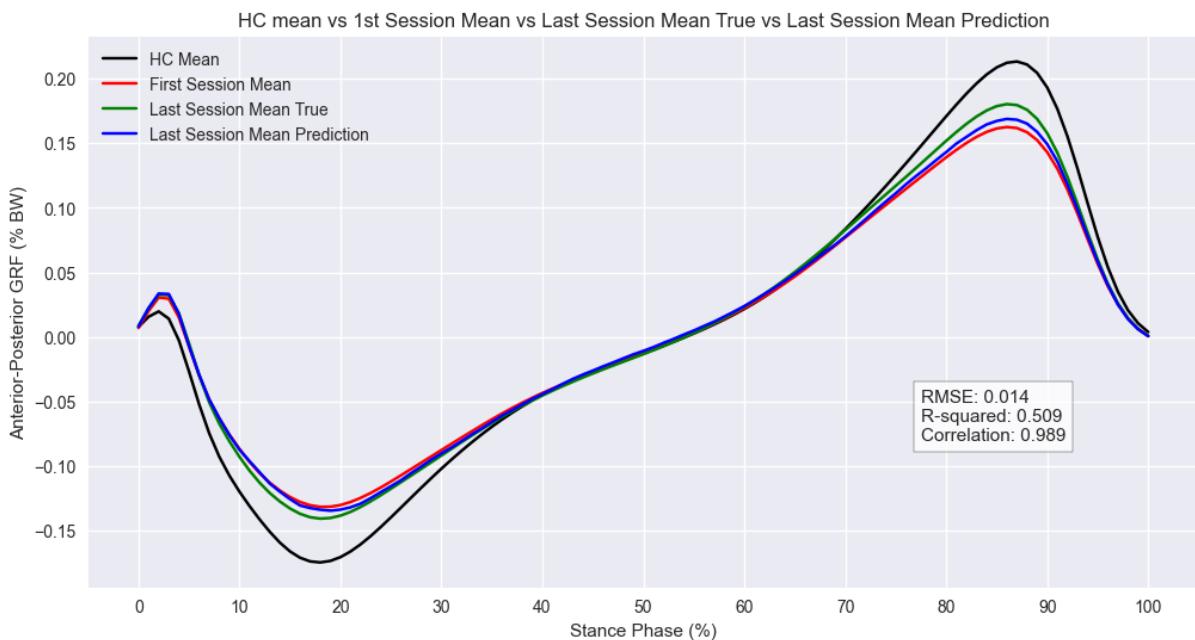


Figura 27. Comparación de los valores promedio correspondientes a la primera y última sesión real (en rojo y verde, respectivamente), y a la última sesión de rehabilitación predicha (en azul) por el modelo de regresión de vectores de soporte (SVR) entrenado y evaluado en los datos originales (aproximación 1: datos sin reducción de la dimensionalidad). Se incorpora como referencia la curva que representa los valores promedio de los controles sanos (en negro), así como los valores de las distintas métricas de evaluación obtenidas por el modelo en los datos no vistos.

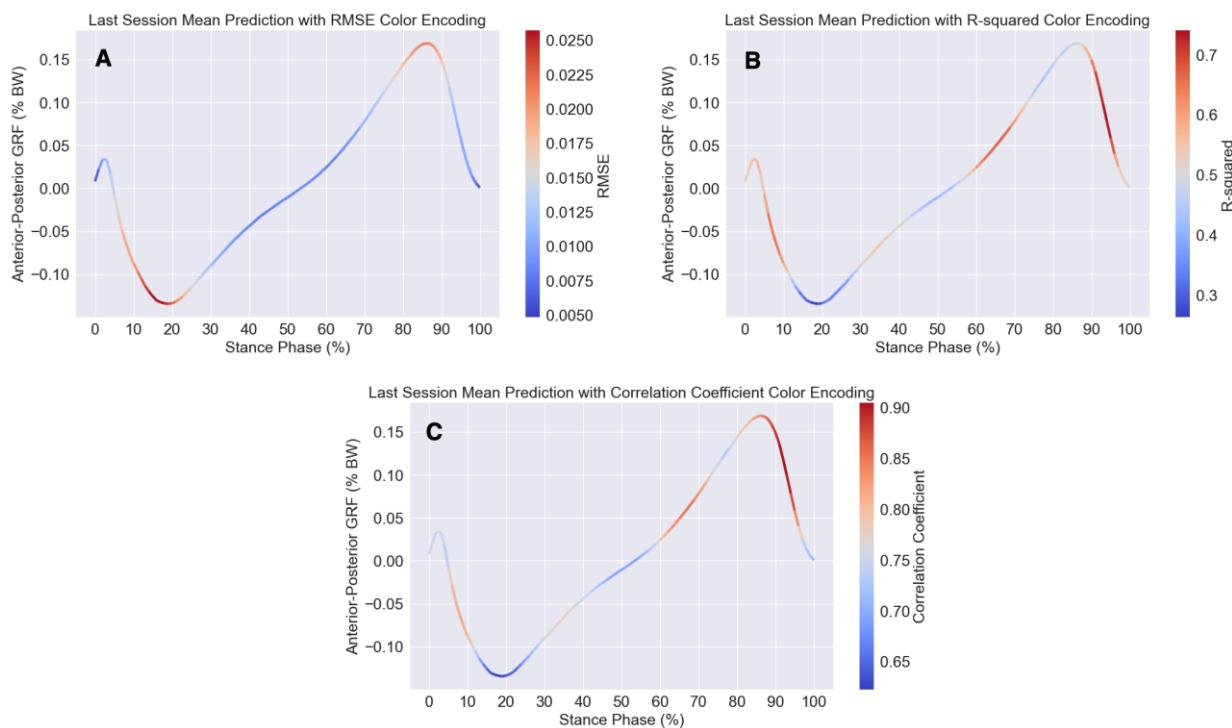


Figura 28. Curvas correspondientes a la última sesión predicha por el modelo de regresión de vectores de soporte (SVR) entrenado y evaluado en los datos originales (aproximación 1: datos sin reducción de la dimensionalidad). Estas curvas están coloreadas en función de los valores de las diversas métricas de regresión obtenidas al evaluar el modelo en el conjunto de prueba. A) Puntos de la curva predicha coloreados en función del RMSE. B) Puntos de la curva predicha coloreados en función del R^2 . C) Puntos de la curva predicha coloreados en función del coeficiente de correlación de Pearson.

El gráfico de los residuos frente a los valores predichos y el histograma de los residuos (Fig. 29) fueron muy similares a los del modelo PLS.

En el histograma, se aprecia una distribución que se parece a una distribución normal, con la excepción de la presencia de un valor atípico en la parte izquierda del gráfico, como en el caso del PLS. Por otro lado, en el gráfico de los residuos frente a los valores predichos, observamos que los residuos se distribuyen de forma relativamente uniforme alrededor del cero. Si embargo, como en el caso del PLS, se aprecia una ligera tendencia descendente hacia la derecha que podría significar que, en los valores negativos (correspondientes a la fuerza de frenada), los errores tendieron a ser más elevados en comparación con otras partes de la curva. Es decir, como en el caso del modelo PLS, el modelo actual (SVR) cometió errores que sesgaron las predicciones hacia arriba, ya que la curva real se encontraba por debajo de las predicciones. Además, podemos identificar el valor atípico en la esquina inferior izquierda, que coincide con el valor atípico identificado en el histograma.



Figura 29. A la izquierda se muestra el gráfico de residuos frente a los valores predichos por el modelo de regresión de vectores de soporte (SVR) entrenado y evaluado en los datos originales (aproximación 1: datos sin reducción de la dimensionalidad). A la derecha se muestra el histograma de los residuos.

5.2.1.3 Regresión lineal múltiple en datos sin preprocessar

Para el primer modelo de regresión lineal múltiple (MLR), entrenado y evaluado con los datos originales (aproximación 1), obtuvimos los siguientes resultados.

En el conjunto de entrenamiento, el valor promedio de la raíz del error cuadrático medio (RMSE) del modelo, calculado en todas las variables objetivo utilizando la técnica *Leave-One-Out Cross-Validation* (LOOCV), fue de 0.023. Este valor de RMSE, que representa el promedio de las discrepancias entre las predicciones del modelo y los valores reales en las variables objetivo, podría considerarse bajo en relación con la variabilidad de las fuerzas de reacción del suelo durante la fase de apoyo de la marcha. No obstante, es el mayor de los errores obtenidos en comparación con los modelos de SVR y PLS. El valor promedio del coeficiente de determinación (R^2) del modelo, calculado en todas las variables objetivo utilizando la técnica *Leave-One-Out Cross-Validation* (LOOCV), fue de 0.860, el mayor de entre todos los modelos del estudio. Esto significa que, en promedio, el 86% de la variabilidad en las variables objetivo fue explicada por el modelo (por las variables predictoras) en el conjunto de entrenamiento. El valor promedio del coeficiente de correlación de Pearson del modelo, calculado en todas las variables objetivo utilizando la técnica *Leave-One-Out Cross-Validation* (LOOCV), fue de 0.966. Este valor promedio indica una correlación positiva fuerte entre las variables predictoras y las variables objetivo en el conjunto de entrenamiento, siendo ligeramente mayor que el obtenido por los modelos de SVR y PLS en el mismo *dataset*.

En el conjunto de prueba, el valor promedio de la raíz del error cuadrático medio (RMSE) del modelo fue de 0.020. Esto significa que las predicciones del modelo presentaron una discrepancia promedio de 0.020 unidades en relación con los valores reales en las variables objetivo. Este valor de RMSE es ligeramente menor que el del conjunto de entrenamiento, pero es mayor que el RMSE obtenido en los modelos SVR y PLS.

El valor promedio del coeficiente de determinación (R^2) en el conjunto de prueba fue de -0.144. Un R^2 negativo indica que los valores predichos por el modelo son peores que si utilizáramos la media como valor predicho. Por tanto, esto significa que el modelo no se ajustó ni generalizó bien a los datos de prueba. El valor promedio del coeficiente de correlación de Pearson en el conjunto de prueba fue de 0.976. Este valor cercano a 1 indica una fuerte relación lineal entre las predicciones del modelo y los valores reales en las variables objetivo. Es decir, las predicciones del modelo siguieron una tendencia similar a los datos reales, como en el caso de los modelos anteriores (SVR y PLS).

En la [Fig. A3](#), se representan los valores predichos por el primer modelo de regresión lineal múltiple en el conjunto de prueba frente a los valores reales para cada una de las 101 variables objetivo. En cada gráfica se muestra el RMSE y el coeficiente de correlación de Pearson, la línea de correlación perfecta como referencia (con una pendiente de 1 y una intersección en el origen), la línea de la regresión lineal, y la ecuación de la recta de regresión en cada gráfico. Observamos que, en la mayoría de las gráficas, la línea de la regresión lineal se separó o torció respecto la línea de correlación perfecta. Esto fue debido a la presencia de valores atípicos o *outliers*, que influyeron negativamente la correlación.

En la Fig. 30, se representa de manera global los valores reales frente a los valores predichos por el modelo de regresión lineal múltiple (aproximación 1) en el conjunto de prueba. Podemos observar el valor promedio del coeficiente de correlación de Pearson mencionado anteriormente (0.976), la línea de correlación perfecta, la línea de regresión lineal que representa la relación lineal ajustada entre los valores reales y predichos, y la ecuación de la recta de regresión. La pendiente de la recta fue de 0.910, lo que significa que, en promedio, las predicciones del modelo tendieron a ser un poco más bajas que los valores reales en el conjunto de prueba. Este valor de pendiente de la recta fue ligeramente inferior al obtenido por los modelos PLS y SVR, lo que significa que, en promedio, las predicciones del actual modelo tendieron a ser más bajas que los valores reales en el conjunto de prueba en comparación con los modelos mencionados. Es decir, este modelo de regresión lineal múltiple tuvo una ligera tendencia a subestimar las respuestas en relación con el modelo PLS y SVR en el conjunto de prueba. Esto podría indicar que el modelo tiende a ser más conservador al hacer predicciones y no predice valores tan altos como los modelos PLS y SVR durante las fases de apoyo y despegue del talón.

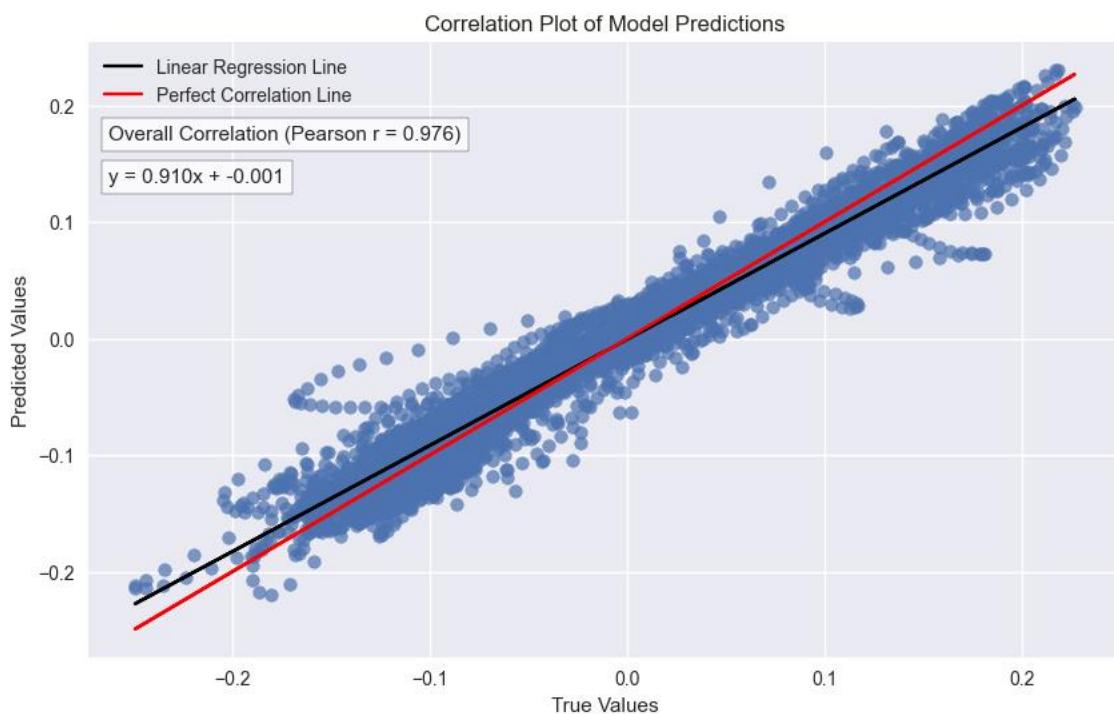


Figura 30. Comparación entre los valores reales y predichos por el primer modelo de regresión lineal múltiple entrenado y evaluado con los datos originales (aproximación 1). Se presenta valor promedio del coeficiente de correlación de Pearson obtenido en el conjunto de prueba, la línea de correlación perfecta, la línea de regresión lineal ajustada y la ecuación de la recta de regresión.

En la Fig. 31 se muestra lo que mencionamos anteriormente. Por un lado, observamos que, en el pico de máxima frenada, la curva correspondiente a la predicción de la última sesión de rehabilitación (representada en azul) se situó entre la primera y última sesión reales. No obstante, en la parte final del mismo pico, la curva predicha se superpuso a la curva correspondiente a la primera sesión de rehabilitación (en rojo).

Por otro lado, durante el despegue del talón (segundo pico en el gráfico correspondiente al pico máximo de propulsión), la curva correspondiente a la predicción de la última sesión de rehabilitación se situó ligeramente por debajo de la primera sesión de terapia.

En resumen, las predicciones en la fase de apoyo del talón y, sobre todo, durante la fase de despegue del talón, no se ajustaron adecuadamente a la curva correspondiente a la última sesión real (en verde); en su lugar, se situaron por debajo o se superpusieron con la curva roja correspondiente a la primera sesión de rehabilitación.

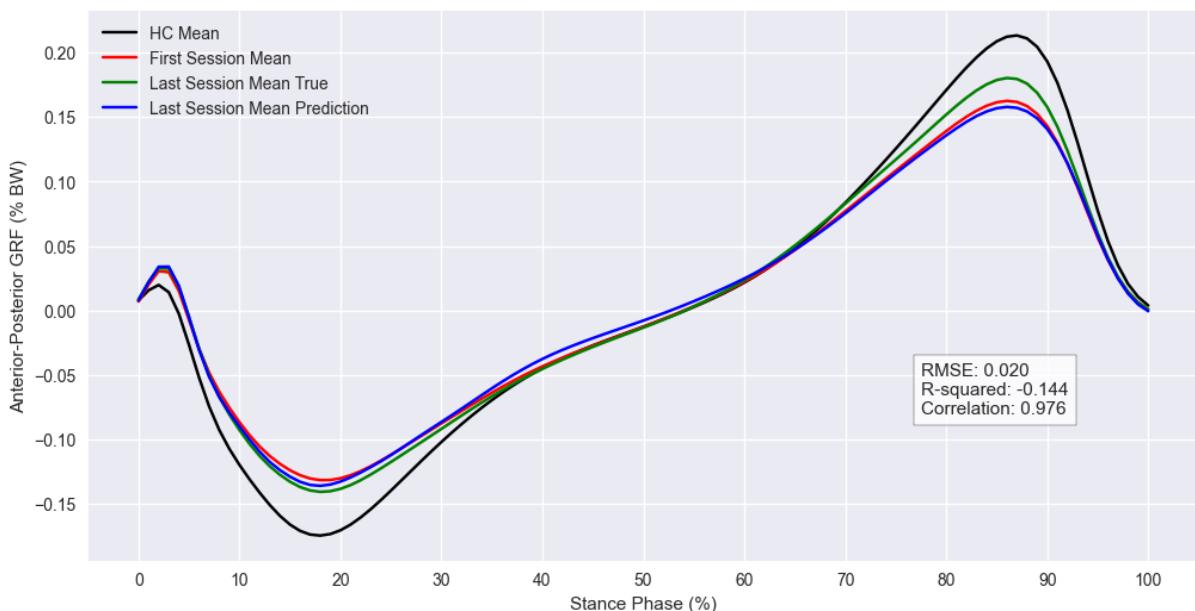


Figura 31. Comparación de los valores promedio correspondientes a la primera y última sesión real (en rojo y verde, respectivamente), y a la última sesión de rehabilitación predicha (en azul) por el primer modelo de regresión lineal múltiple entrenado y evaluado en los datos originales (aproximación 1). Se incorpora como referencia la curva que representa los valores promedio de los controles sanos (en negro), así como los valores de las distintas métricas de evaluación obtenidas en los datos no vistos por el modelo.

Tal y como se comentaba, en la Fig. 32a se muestra que los valores de RMSE fueron mayores en los picos de la curva predicha, es decir, durante la fase de apoyo (o choque del talón contra el suelo), y durante la fase de despegue del talón. Este patrón ha sido común en todos los modelos. No obstante, estos errores fueron mayores que los errores de las predicciones generadas por los modelos SVR y PLS. Por tanto, parece que el rendimiento del modelo fue inferior al de los modelos anteriormente mencionados.

En la Fig. 32b, podemos observar que valores de R^2 a lo largo de la curva predicha son inferiores a los de los modelos de SVR y PLS, siendo incluso negativos en varios puntos de la curva correspondientes a la fase final de apoyo del talón y a la fase inicial de despegue del talón.

Los valores de los coeficientes de correlación de Pearson fueron disminuyendo durante la fase de apoyo del talón, alcanzando los valores más bajos en el punto máximo de frenada. Por otro lado, antes y después del punto máximo de propulsión (correspondiente a la fase de despegue del talón), los coeficientes de correlación fueron más altos (Fig. 32c). Este patrón fue similar al obtenido por los modelos de SVR y PLS, aunque en el caso de este modelo de

regresión lineal múltiple, los valores de correlación fueron ligeramente inferiores en comparación con los otros dos modelos mencionados.

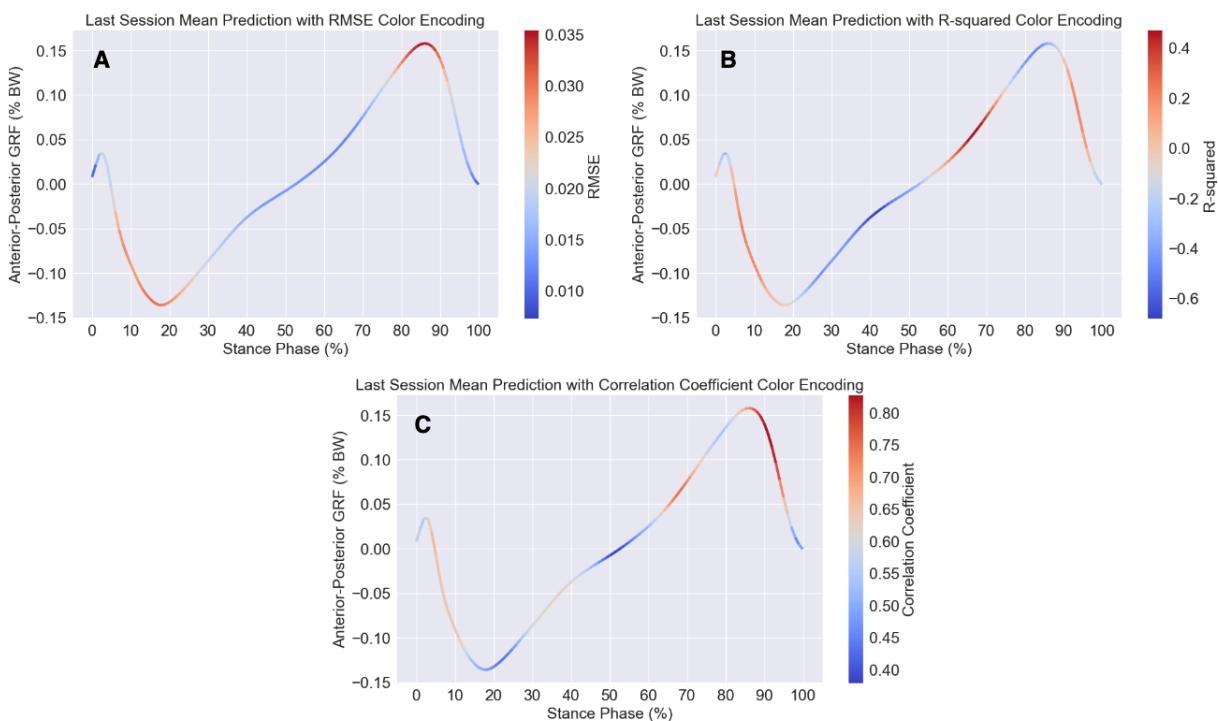


Figura 32. Curvas correspondientes a la última sesión predicha por el primer modelo de regresión lineal múltiple entrenado y evaluado en los datos originales (aproximación 1). Estas curvas están coloreadas en función de los valores de las diversas métricas de regresión obtenidas al evaluar el modelo en el conjunto de prueba. A) Puntos de la curva predicha coloreados en función del RMSE. B) Puntos de la curva predicha coloreados en función del R^2 . C) Puntos de la curva predicha coloreados en función del coeficiente de correlación de Pearson.

Para analizar los residuos con el objetivo validar el primer modelo de regresión lineal múltiple, obtuvimos las gráficas mostradas en la Fig. 33.

En el histograma, observamos lo que parece ser una distribución bimodal, indicando que había dos tipos de residuos: aquellos generados en la parte izquierda y aquellos en la parte derecha de la gráfica.

En el gráfico de residuos frente a los valores predichos, no observamos una distribución uniforme alrededor del cero; en cambio, los residuos mostraron una correlación inversa con los valores predichos, es decir, descendieron. Esto indica una relación lineal negativa entre los residuos y los valores predichos. Es decir, en la región de valores negativos (correspondiente a la fuerza de frenado), los errores fueron más grandes (indicando que las predicciones estaban por encima de los valores reales), mientras que en la región de valores positivos (correspondiente a la fuerza de propulsión), los errores fueron más pequeños (indicando que las predicciones estaban por debajo de los valores reales).

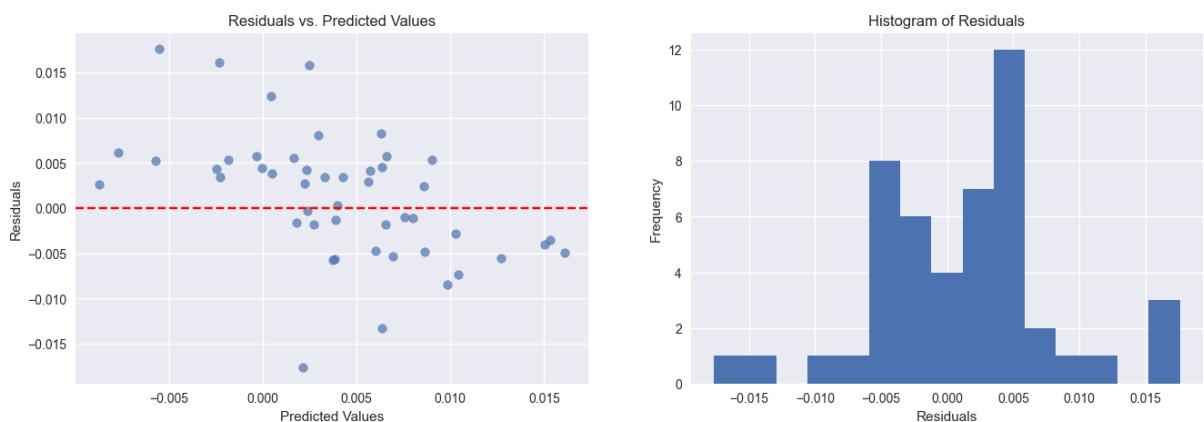


Figura 33. Gráfico de residuos frente a los valores predichos (izquierda) por el primer modelo de regresión lineal múltiple entrenado y evaluado en los datos originales (aproximación 1), e histograma de los residuos (derecha).

5.2.1.3.1 Comprobación de los supuestos de la regresión lineal

Linealidad

Los gráficos de los residuos frente a los valores predichos para cada variable objetivo mostraron que los puntos se distribuían aleatoriamente alrededor del cero ([Fig. A4](#)). Por otro lado, en los gráficos conjuntos se observó una relación lineal entre cada predictor y cada variable objetivo. Por tanto, estos dos tipos de visualizaciones indicaron que se cumplía el supuesto de linealidad.

Independencia de los residuos

Examinando la misma figura ([Fig. A4](#)), observamos que el gráfico de residuos no presentó ningún patrón o estructura discernible; los residuos aparecieron dispersos aleatoriamente alrededor de la línea horizontal. En la Prueba de Durbin-Watson los valores fueron cercanos a 2, indicando que no hubo autocorrelación (independencia de los errores). Por tanto, se cumplió el supuesto de independencia de los errores o residuos.

Normalidad de los residuos

La prueba estadística de Anderson-Darling reveló que, en 20 de las 101 variables objetivo, el estadístico Anderson-Darling superó el valor crítico correspondiente a un nivel de significancia del 5%. Es decir, para esas 20 variables se rechazó la hipótesis nula de normalidad al nivel de significancia del 5%, indicando que los residuos de esos modelos específicos no seguían una distribución normal. Las variables objetivo en cuestión fueron la número 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 89, 90, 91, 92, 93, 94, 95 y 96. Esto se representó en la Fig. 34, donde se codificaron los puntos de la curva según los valores de la prueba de Anderson-Darling. Los 20 puntos de la curva donde el estadístico Anderson-Darling superó el valor crítico para un nivel de significancia del 5% se destacan en colores distintos.

Homocedasticidad

La prueba estadística Breusch-Pagan reveló que todos los p-values fueron iguales o superiores al nivel de significancia del 5% ([Fig. A4](#)). Esto indica que se cumplió el supuesto de igualdad de varianzas (homocedasticidad), es decir, la varianza de los residuos se mantuvo constante a lo largo de los valores predichos.

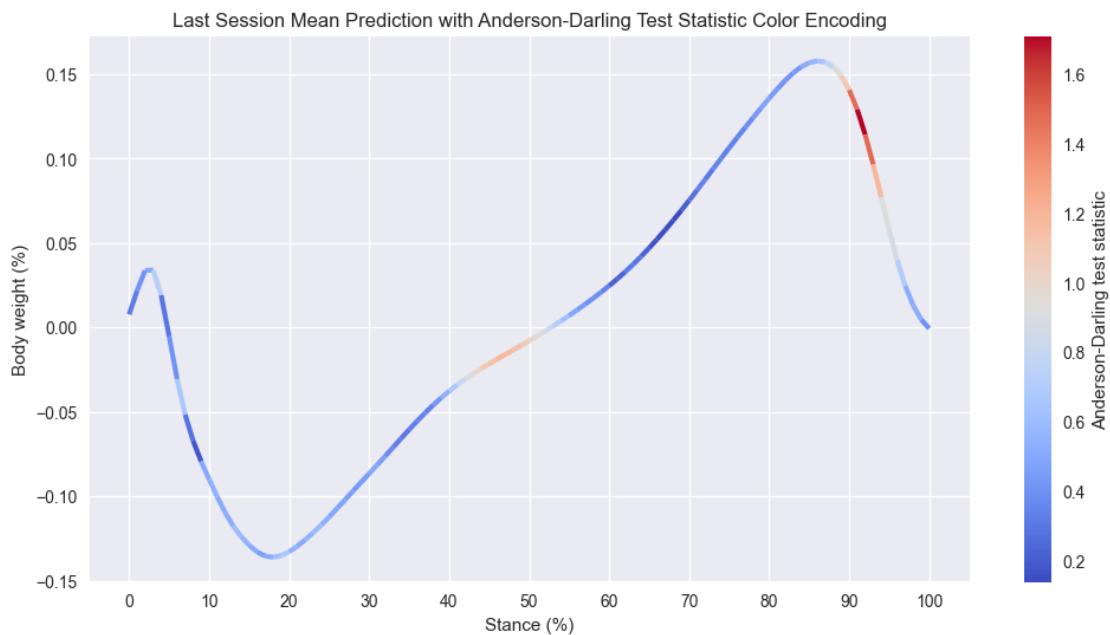


Figura 34. Puntos de la curva (correspondiente a la última sesión predicha) codificados en función de los valores de la prueba de Anderson-Darling. Se destacan los 20 puntos de la curva donde el estadístico Anderson-Darling superó el valor crítico para un nivel de significancia del 5%.

5.2.1.4 Regresión lineal múltiple en datos de salida reducidos mediante PCA

El tercer modelo de regresión lineal múltiple se entrenó y evaluó en los datos de salida reducidos mediante el PCA (aproximación 3). Al aplicar el PCA a los datos de salida correspondientes a la última sesión de rehabilitación (Y) en los conjuntos de entrenamiento y prueba, se logró reducir los datos a 5 componentes principales, los cuales conservaron el 90% de la varianza de los datos originales (Fig. 35).

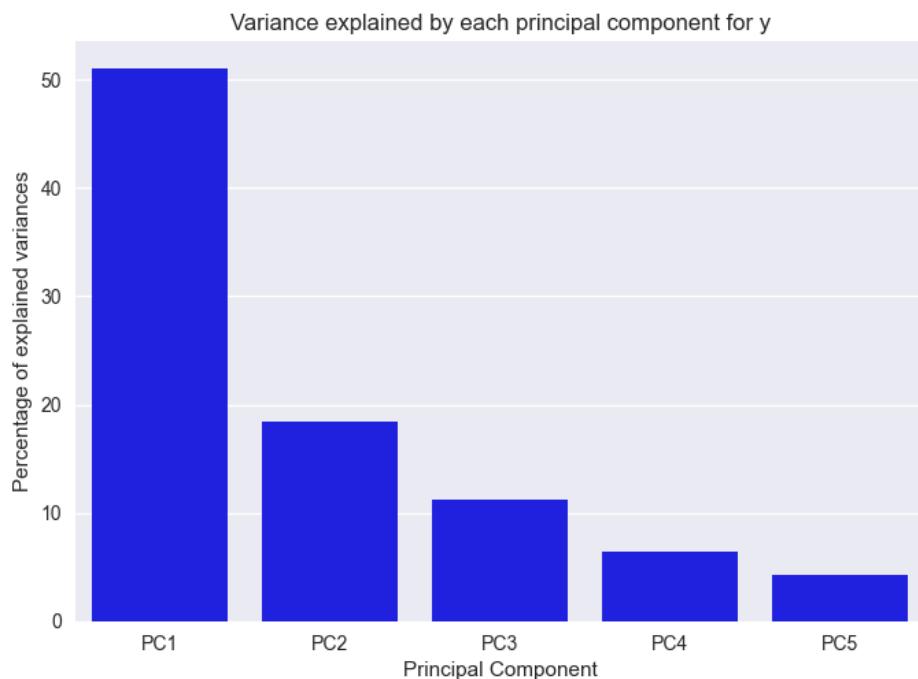


Figura 35. Porcentaje de la varianza explicada por los 5 componentes principales. El gráfico muestra cómo cada uno de los 5 componentes principales contribuye al porcentaje total de varianza explicada en los datos después de aplicar el PCA en los datos de salida correspondientes a la última sesión de rehabilitación.

Las predicciones en el conjunto de entrenamiento, realizadas en el espacio reducido de 5 componentes principales obtenidos mediante PCA, proporcionaron los resultados descritos a continuación.

El valor promedio de la raíz del error cuadrático medio (RMSE) del modelo, calculado en todas las variables objetivo utilizando la técnica *Leave-One-Out Cross-Validation* (LOOCV), fue de 0.097, el más alto de entre todos modelos utilizados en el estudio.

El valor promedio del coeficiente de determinación (R^2) del modelo, calculado en todas las variables objetivo utilizando la técnica *Leave-One-Out Cross-Validation* (LOOCV), fue de 0.855. Esto significa que, en promedio, el 85.5% de la variabilidad en las variables objetivo fue explicada por el modelo (por las variables predictoras) en el conjunto de entrenamiento. Este valor de coeficiente de determinación fue mayor que en los modelos de SVR y PLS, pero menor que en el modelo de regresión lineal múltiple en los datos originales (aproximación 1).

El coeficiente de correlación de Pearson global del modelo, calculado en todas las variables objetivo utilizando la técnica *Leave-One-Out Cross-Validation* (LOOCV), fue de 0.660. Este valor promedio indica una correlación positiva moderada entre las variables predictoras y las variables objetivo en el conjunto de entrenamiento, siendo uno de los más pequeños coeficientes de correlación de todos los modelos del estudio.

Por otro lado, en el conjunto de prueba, las predicciones en el espacio reducido se describen a continuación, junto con los resultados de las predicciones después de ser transformadas al espacio original.

El valor promedio de la raíz del error cuadrático medio (RMSE) del modelo fue de 0.083 para el espacio reducido y 0.020 para el espacio original tras la transformación inversa. Por tanto, después de aplicar la transformación inversa y regresar al espacio original de las variables objetivo, las predicciones mejoraron sustancialmente y presentaron una discrepancia promedio mucho menor con respecto a los valores reales. Aunque el RMSE coincide con el obtenido en el modelo regresión lineal múltiple entrenado y evaluado en los datos originales (aproximación 1), los modelos PLS y SVR (aproximación 3) proporcionaron RMSE menores.

En el espacio reducido (tras aplicar PCA) el valor promedio del coeficiente de determinación (R^2) en el conjunto de prueba fue de 0.026. Esto significa que, en promedio, solo el 2.6% de la variabilidad en las variables objetivo fue explicada por el modelo en este espacio. Este valor indica un ajuste muy pobre del modelo a los datos en el espacio reducido. Por otro lado, en el espacio original (tras la transformación inversa). El valor promedio del coeficiente de determinación en el conjunto de prueba fue aún más bajo: -0.100. Un valor negativo en R^2 indica que el modelo en este espacio no solo no explica la variabilidad de los datos, sino que realiza predicciones que están en promedio peor que simplemente tomar el valor promedio de las variables objetivo. Esto indica un rendimiento muy deficiente del modelo en el espacio original.

En el espacio reducido (tras aplicar PCA) el valor promedio del coeficiente de correlación de Pearson en el conjunto de prueba fue de 0.613. Este valor indica una correlación positiva moderada entre las variables predictoras y las variables objetivo en el espacio reducido.

Por otro lado, en el espacio original (tras la transformación inversa), el valor promedio del coeficiente de correlación de Pearson en el conjunto de prueba fue de 0.977. Este valor cercano a 1 indica una fuerte relación lineal entre las predicciones del modelo y los valores reales en las variables objetivo. Es decir, las predicciones del modelo siguieron una tendencia similar a los datos reales, como en el caso de los modelos anteriores (PLS, SVR y regresión linear múltiple en los datos originales).

Debemos tener en cuenta que las métricas en el espacio reducido proporcionan una visión parcial del rendimiento del modelo. Para comprender completamente el significado de

estas predicciones, fue necesario volver al espacio original, donde las variables objetivo conservan su forma original y, por lo tanto, los resultados son más comprensibles. A partir de este punto, los resultados presentados se refieren a las predicciones en el espacio original, después de llevar a cabo la transformación inversa.

En la Fig. 36, se representa de manera global los valores reales frente a los valores predichos por el modelo de regresión lineal múltiple (aproximación 3) en el conjunto de prueba. Podemos observar el valor promedio del coeficiente de correlación de Pearson mencionado anteriormente (0.977), la línea de correlación perfecta, la línea de regresión lineal que representa la relación lineal ajustada entre los valores reales y predichos, y la ecuación de la recta de regresión. La pendiente de la recta fue de 0.907, lo que significa que, en promedio, las predicciones del modelo tendieron a ser un poco más bajas que los valores reales en el conjunto de prueba. Este valor de pendiente de la recta fue inferior al obtenido por los modelos PLS y SVR, y similar al obtenido por la regresión lineal múltiple (aproximación 1). Esto significa que, en promedio, las predicciones del actual modelo tendieron a ser más bajas que los valores reales en el conjunto de prueba en comparación con los modelos mencionados. Es decir, este modelo de regresión lineal múltiple tuvo una ligera tendencia a subestimar las respuestas en relación con los otros modelos mencionados anteriormente. Esto podría indicar que el modelo tiende a ser aún más conservador que los anteriores al hacer predicciones, y no predice valores tan altos como los modelos PLS y SVR durante las fases de apoyo y despegue del talón.

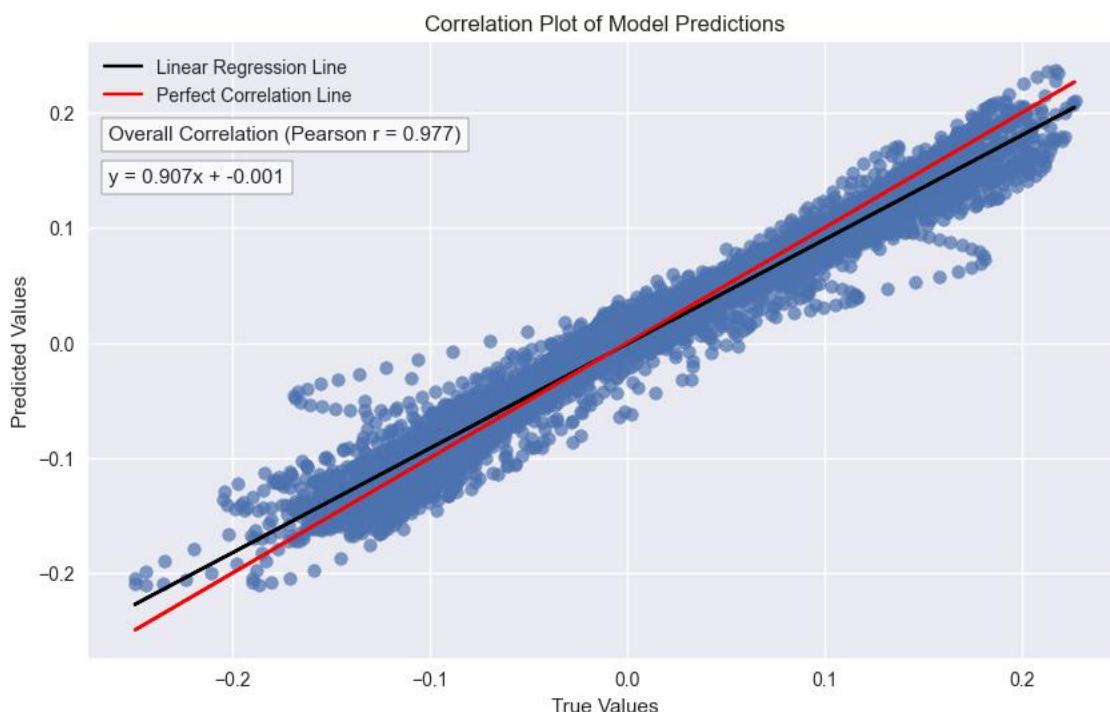


Figura 36. Comparación entre los valores reales y predichos por el tercer modelo de regresión lineal múltiple entrenado y evaluado con los datos de salida reducidos (aproximación 3). Se presenta el valor promedio del coeficiente de correlación de Pearson obtenido en el conjunto de prueba, la línea de correlación perfecta, la línea de regresión lineal ajustada y la ecuación de la recta de regresión.

En la Fig. 37 se muestra lo siguiente. Por un lado, observamos que, en el pico de máxima frenada, la curva correspondiente a la predicción de la última sesión de rehabilitación (representada en azul) se situó entre la primera y última sesión reales (representadas en rojo y verde, respectivamente), aunque muy cercana a la primera sesión. No obstante, en la parte final del mismo pico, la curva predicha se superpuso a la curva correspondiente a la primera

sesión de rehabilitación. El comportamiento en esta fase fue similar al de la regresión lineal múltiple en los datos originales (aproximación 1).

Por otro lado, durante el despegue del talón (segundo pico en el gráfico correspondiente al pico máximo de propulsión), la curva correspondiente a la predicción de la última sesión de rehabilitación se situó ligeramente por debajo de la primera sesión de terapia, solapándose en algunos puntos de la parte final de la fase de despegue del talón.

En resumen, de una forma similar al caso de la regresión lineal múltiple en los datos originales (aproximación 1), las predicciones en la fase de apoyo del talón y, sobre todo, durante la fase de despegue del talón, no se ajustaron adecuadamente a la curva correspondiente a la última sesión real (en verde). En su lugar, se situaron por debajo o se superpusieron con la curva roja correspondiente a la primera sesión de rehabilitación.

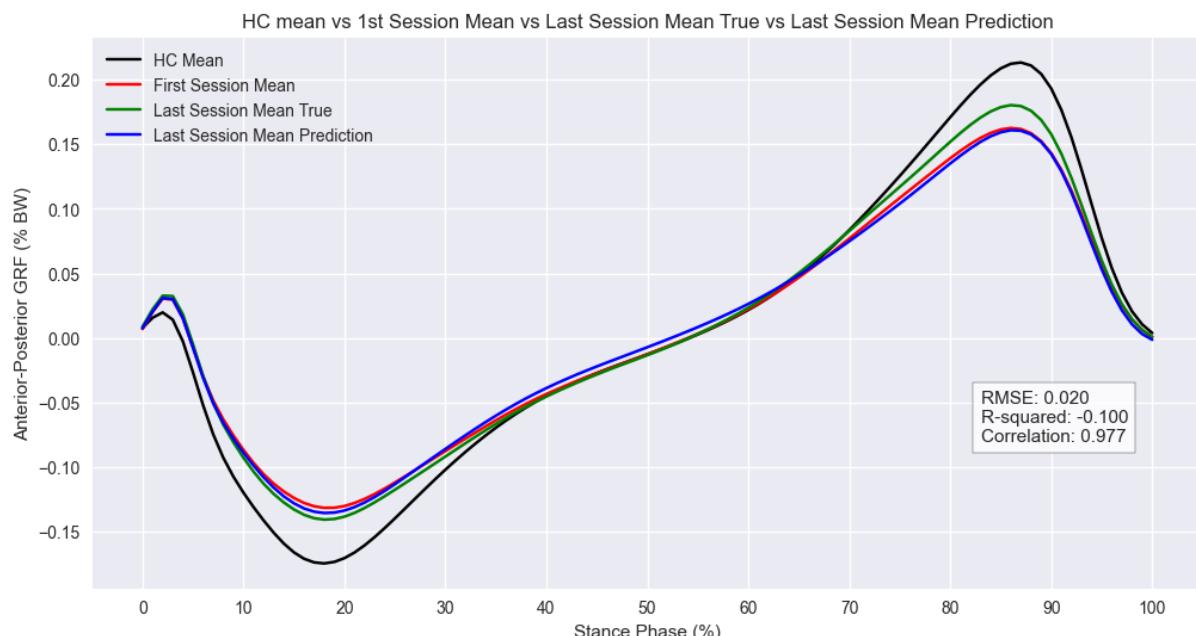


Figura 37. Comparación de los valores promedio correspondientes a la primera y última sesión real (en rojo y verde, respectivamente), y a la última sesión de rehabilitación predicha (en azul) por el tercer modelo de regresión lineal múltiple entrenado y evaluado con los datos de salida reducidos (aproximación 3). Se incorpora como referencia la curva que representa los valores promedio de los controles sanos (en negro), así como los valores de las distintas métricas de evaluación obtenidas por el modelo en los datos no vistos.

Las curvas correspondientes a la última sesión predicha, coloreadas en función de los valores de las diversas métricas de regresión obtenidas al evaluar el modelo en el conjunto de prueba, se muestran en la Fig. 38. Los resultados son muy parecidos al del modelo de regresión lineal múltiple en los datos originales (aproximación 1) para los valores de RMSE (Fig. 38a) y R^2 (Fig. 38b), y, en consecuencia, peores a los modelos PLS y SVR (aproximación 3).

Sin embargo, los valores de los coeficientes de correlación de Pearson se mantuvieron bastante uniformes a lo largo de la curva predicha (Fig. 38c).

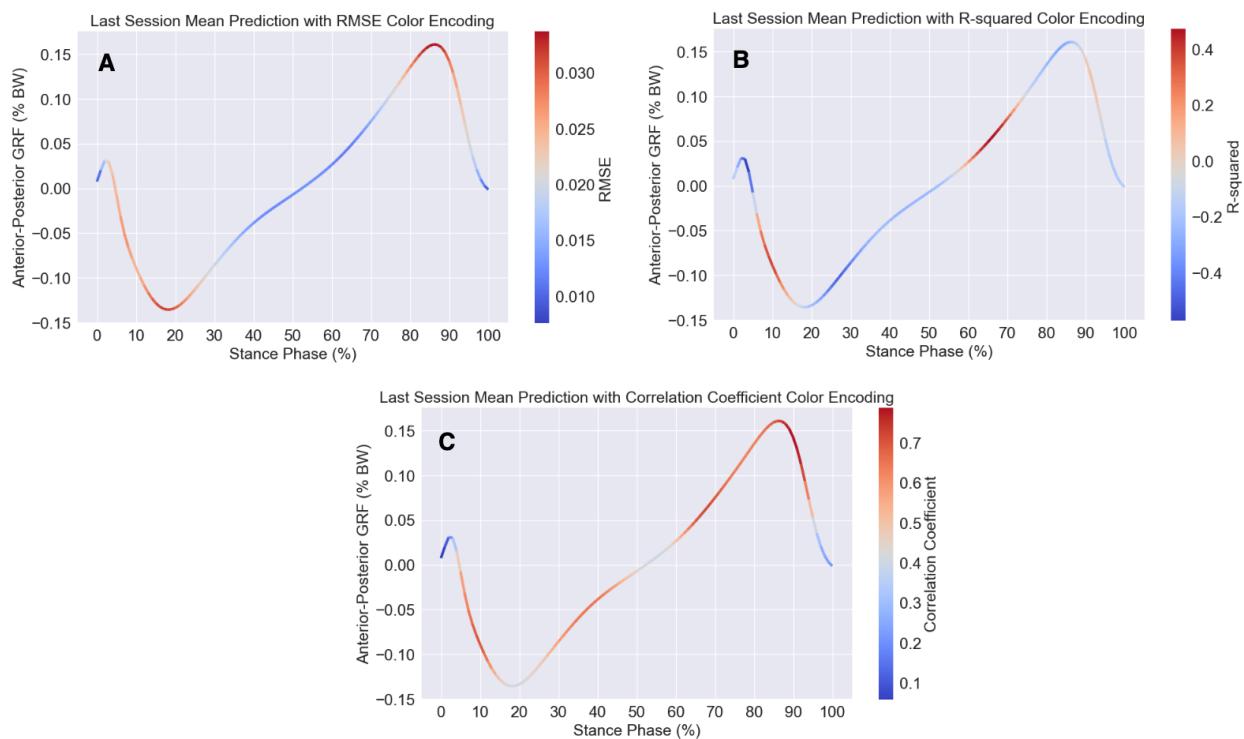


Figura 38. Curvas correspondientes a la última sesión predicha por el tercer modelo de regresión lineal múltiple entrenado y evaluado con los datos de salida reducidos (aproximación 3). Estas curvas están coloreadas en función de los valores de las diversas métricas de regresión obtenidas al evaluar el modelo en el conjunto de prueba. A) Puntos de la curva predicha coloreados en función del RMSE. B) Puntos de la curva predicha coloreados en función del R^2 . C) Puntos de la curva predicha coloreados en función del coeficiente de correlación de Pearson.

En el análisis de los residuos (Fig. 39) obtuvimos unos resultados muy similares a los de la regresión lineal múltiple en los datos originales (aproximación 1).

En el histograma, observamos lo que parece ser una distribución bimodal, indicando que había dos tipos de residuos: aquellos generados en la parte izquierda y aquellos en la parte derecha de la gráfica. También se observan valores atípicos en ambos lados del histograma.

En el gráfico de residuos frente a los valores predichos, no observamos una distribución uniforme alrededor del cero; en cambio, los residuos mostraron una correlación inversa con los valores predichos, es decir, descendieron. Esto indica una relación lineal negativa entre los residuos y los valores predichos, que parece ser más pronunciada incluso que la del modelo de regresión lineal múltiple en los datos originales (aproximación 1). Es decir, en la región de valores negativos (correspondiente a la fuerza de frenado), los errores fueron más grandes (indicando que las predicciones estaban por encima de los valores reales), mientras que en la región de valores positivos (correspondiente a la fuerza de propulsión), los errores fueron más pequeños (indicando que las predicciones estaban por debajo de los valores reales). Además, se pueden visualizar los valores atípicos observados en el histograma.

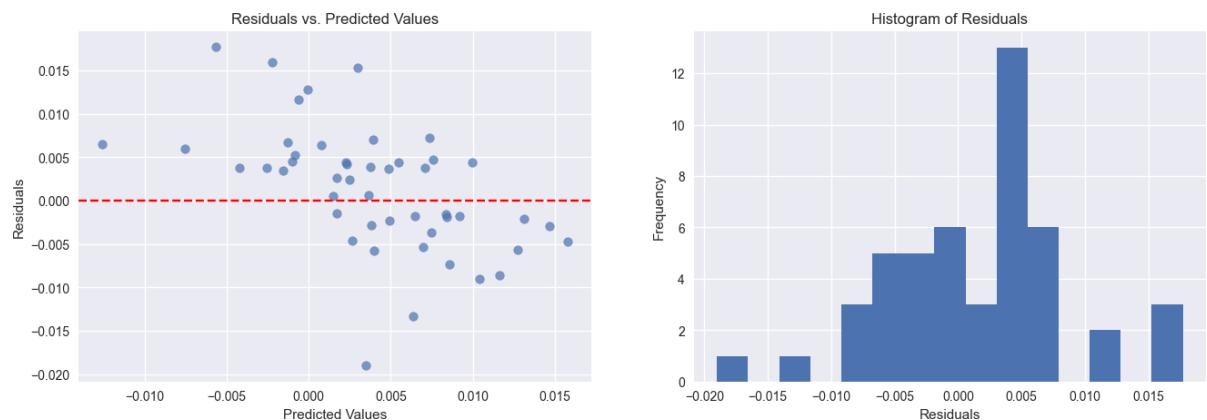


Figura 39. A la izquierda se muestra el gráfico de residuos frente a los valores predichos por el tercer modelo de regresión lineal múltiple entrenado y evaluado con los datos de salida reducidos (aproximación 3). A la derecha se muestra el histograma de los residuos.

6 Discusión

En el presente trabajo de investigación, hemos abordado un importante problema en el ámbito de la salud y en el campo del análisis clínico de la marcha utilizando técnicas de aprendizaje automático y análisis de datos. A pesar del progreso realizado en los últimos años, existe una brecha en la comprensión y predicción del progreso de los pacientes con deficiencias musculoesqueléticas a lo largo de sus procesos terapéuticos y de rehabilitación. Este trabajo se basa en investigaciones previas [11], [12], [13] que utilizan el aprendizaje automático para predecir resultados relacionados con los patrones de la marcha y sienta un precedente en el uso de estas técnicas para predecir la evolución de pacientes sometidos a rehabilitación.

A lo largo de este estudio, hemos explorado diferentes algoritmos de aprendizaje automático y hemos aplicado diferentes preprocesamientos a los datos con el objetivo de mejorar la precisión en la predicción de la última sesión de rehabilitación de los pacientes con desórdenes musculoesqueléticos en el tobillo, en base a la primera sesión de terapia realizada. En este capítulo, empezaremos discutiendo los descubrimientos que surgieron durante al análisis exploratorio de los datos, continuando con la elección y comparación de los modelos empleados en este estudio. Evaluaremos en detalle cómo se desempeñaron y discutiremos sus ventajas y limitaciones en relación con la predicción de la evolución de pacientes sometidos a rehabilitación en el tobillo. Además, exploraremos las distintas aproximaciones realizadas en el preprocesamiento de los datos y analizaremos su impacto en el rendimiento de los modelos. A continuación, profundizaremos en el rendimiento de los diversos algoritmos y modelos de aprendizaje automático utilizados en este trabajo, evaluando si se comportaron de acuerdo con nuestras expectativas y comentando sus resultados más destacados. Por último, dedicaremos una sección a discutir las posibles líneas futuras de investigación que se deriven de los descubrimientos y limitaciones de este estudio.

Durante el análisis exploratorio de los datos, observamos que algunas variables muestran una distribución normal, mientras que otras no siguen esta distribución. En concreto, las variables que no se distribuyen normalmente corresponden a la fase final del despegue del talón durante la fase de apoyo del ciclo de la marcha. También identificamos la presencia de algunos valores atípicos u *outliers* en nuestros conjuntos de datos de entrenamiento, lo que indica la existencia de valores extremos que pueden influir en nuestras predicciones. Mediante el análisis de correlación, pudimos detectar una alta correlación entre variables vecinas indica que existe una relación sistemática entre las medidas tomadas en diferentes momentos de la fase de apoyo de la marcha. Esto respalda la idea de que las diferentes etapas del ciclo de la marcha están interconectadas y, por lo tanto, las fuerzas medidas en momentos cercanos están influidas por las mismas tendencias y patrones. La correlación positiva más elevada se encuentra en la fase posterior del despegue del talón después del punto máximo de propulsión.

Nuestra selección de algoritmos incluye la regresión lineal múltiple (MLR), la regresión de mínimos cuadrados parciales (PLS), árboles de decisión (Decision Trees), bosques aleatorios (Random Forests), eXtreme Gradient Boosting (XGBoost) y regresión de vectores de soporte (SVR). En el contexto de un problema de regresión múltiple como el abordado en este estudio, la evaluación de los modelos a través de la aplicación del método *target-specific ranks* [66], en el cual se promedian los rangos en lugar de las puntuaciones individuales de las métricas de regresión, permite resaltar las diferencias entre los distintos modelos. Esto facilita su comparación en contraste con el uso de métricas de regresión promediadas entre todas las variables objetivo.

Los resultados muestran que el modelo PLS entrenado con una configuración óptima de 8 componentes y el modelo SVR entrenado y evaluado en los datos originales, presentan un mayor rendimiento a la hora de predecir la última sesión de rehabilitación a partir de la

primera sesión. Logra un rendimiento similar el modelo de regresión lineal múltiple, entrenado y evaluado con los datos de la primera sesión de terapia (la entrada X) sometidos a una reducción de dimensionalidad mediante PCA que retiene cinco componentes, explicando así el 90% de la varianza de los datos.

Por contra, los árboles de decisión, junto con el modelo de regresión lineal múltiple entrenado y evaluado en los datos originales, así como el modelo de regresión lineal múltiple entrenado y evaluado con los datos de la última sesión de terapia reducidos mediante PCA (5 componentes explican el 90% de la varianza de los datos) y posteriormente reconstruidos, muestran un rendimiento inferior en comparación con los otros modelos.

Por otro lado, los bosques aleatorios y los modelos XGBoost, no destacaron en términos de rendimiento, quedando situados en un punto intermedio en cuanto a la calidad de las predicciones.

Las distintas aproximaciones de preprocessamiento implementadas han demostrado jugar un papel importante en el rendimiento de los modelos.

Aunque el modelo de regresión de vectores de soporte se desempeñó mejor con los datos originales, en general, se observa una mejora en el rendimiento de ciertos modelos al reducir el tamaño de la entrada, es decir, los datos de la primera sesión de rehabilitación. Esto es especialmente evidente en el caso de la regresión lineal múltiple, donde la reducción del tamaño del vector de entrada, maximizando la varianza y eliminando redundancias, le permitió destacar en términos de rendimiento sobre otros modelos que utilizaron distintas aproximaciones. Esta mejora del rendimiento reduciendo los datos de entrada mediante el PCA coincide con las investigaciones de Omar A. Galarraga C. et al. [12], [13].

Sin embargo, el modelo PLS, al considerar de manera integrada la relación entre las variables predictoras (X) y las variables objetivo (y), superó en rendimiento a los modelos cuyos datos de entrenamiento y prueba fueron reducidos mediante PCA.

Por otro lado, cabe destacar que tanto la reducción de dimensionalidad aplicada solo a los datos de salida mediante PCA, como la reducción simultánea de la dimensionalidad de los datos de entrada y salida mediante PCA, no parecen tener un impacto positivo en el rendimiento de los modelos de regresión lineal múltiple, árboles de decisión, bosques aleatorios y XGBoost, debido a la pérdida de información.

En cuanto a las fuerzas de reacción del suelo en la dirección anteroposterior durante la fase de apoyo del ciclo de la marcha, todos los modelos son capaces de predecir la forma de la curva de manera adecuada, mostrando una correlación entre de los puntos de la curva predicha y la curva correspondiente a la última sesión real de rehabilitación.

En relación a la parte central de la curva, que representa la disminución gradual de la fuerza hasta llegar a cero y corresponde a la fase de apoyo con una sola pierna, los modelos logran predecirla con éxito. Se observa que para estos puntos de la curva, los valores de RMSE son bajos, lo que indica que las predicciones son precisas y se ajustan adecuadamente a los valores reales de la última sesión de rehabilitación.

No obstante, el RMSE aumenta en los puntos correspondientes a los picos de máxima frenada y propulsión. Por tanto, los modelos tienden a cometer errores más significativos en las áreas críticas de la curva, correspondientes a la fase de apoyo de talón en el suelo, y a la fase del despegue del talón. En consecuencia, la mayor desviación se observa en los puntos más altos de la curva, es decir, los picos de máxima frenada y propulsión.

La distribución de los residuos de los modelos de bajo rendimiento no sigue una distribución normal. Por el contrario, los modelos de mayor rendimiento muestran una distribución de residuos que se asemeja a una distribución normal, con residuos distribuidos de manera relativamente uniforme alrededor de cero. Sin embargo, la ligera tendencia descendente hacia la derecha en la gráfica de residuos frente a los valores predichos indica

que en los valores negativos (correspondientes a la fuerza de frenada), los errores tienden a ser más altos en comparación con otras partes de la curva. En otras palabras, estos modelos cometen errores que sesgan las predicciones hacia arriba, ya que la curva real se encuentra por debajo de las predicciones. Además, se identifican valores atípicos.

Por tanto, existen dos grupos de modelos: aquellos cuyas predicciones se superponen o se sitúan por debajo de la primera sesión de rehabilitación, y aquellos cuyas predicciones se sitúan entre la primera y la última sesión de rehabilitación. Dentro de este último grupo, los modelos son bastante similares entre sí, pero método *target-specific ranks*, permite resaltar las diferencias y evaluar su rendimiento de manera más precisa.

Finalmente, aunque es importante destacar que ninguno de los modelos logra predecir con precisión los valores máximos de frenada y propulsión durante la fase de apoyo del ciclo de la marcha, y aunque todavía queden algunos desarrollos y validaciones por hacer, este trabajo representa un primer avance hacia la predicción del resultado de la rehabilitación en pacientes con desordenes musculoesqueléticos en el tobillo. El modelo PLS, el modelo SVR entrenado y evaluado en los datos originales, y el modelo de regresión lineal múltiple, entrenado y evaluado con los datos de la primera sesión de terapia reducidos mediante el PCA, pueden servir como base sobre la cual construir investigaciones posteriores en este campo.

7 Consideraciones éticas

En esta sección, se explican los aspectos éticos considerados en este estudio.

En este proyecto de investigación, se utilizó el conjunto de datos GaitRec, proporcionado por Horsak et al. [4]. Es necesario destacar que, antes de la publicación de este conjunto de datos, se obtuvo la aprobación del Comité de Ética local de Baja Austria (GS1-EK-4/299-2014). Esta aprobación garantiza que la recopilación de datos se llevó a cabo de manera ética y bajo los principios de privacidad y consentimiento informado. El cumplimiento de las pautas éticas y legales en la obtención de los datos subyacentes es fundamental para garantizar su uso responsable y respetuoso.

No hay conflictos de intereses en relación con este estudio, es decir, no hubo influencia externa que pudiera sesgar los resultados o las conclusiones.

Se ha atribuido y reconocido adecuadamente la procedencia de las fuentes y las contribuciones de otros investigadores en este campo.

Se proporciona una descripción detallada de la metodología utilizada en el estudio, lo que incluye el procesamiento de datos y los modelos de aprendizaje automático. Esto permite la replicabilidad y la verificabilidad de los resultados por parte de otros investigadores. Además, para fomentar la transparencia y la colaboración con la comunidad científica, se ha compartido el repositorio que contiene código desarrollado para este proyecto ([Apéndice 1](#)).

Finalmente, debemos mencionar que la investigación se llevó a cabo con la intención de contribuir al campo de la salud y el análisis clínico de la marcha, con el propósito de mejorar la calidad de atención a pacientes con deficiencias musculoesqueléticas.

8 Coste económico

En esta sección, se detallan los costes aproximados asociados a este proyecto de investigación.

En primer lugar, es relevante destacar que los datos utilizados en este estudio no generaron ningún costo, ya que el conjunto de datos GaitRec, proporcionado por Horskak et al. [4], está disponible públicamente sin restricciones de pago ni acceso.

En términos de recursos humanos, el principal gasto corresponde al sueldo del científico de datos que realizó este proyecto. De acuerdo con la información proporcionada por "Glassdoor", una plataforma líder en empleo y reclutamiento a nivel mundial, el sueldo bruto promedio para el puesto de *junior data scientist* en España es de aproximadamente 28.772 € al año, considerando una jornada laboral de 40 horas a la semana [67]. Esto equivaldría aproximadamente a 13.82 € la hora.

El desarrollo del proyecto llevó un período de tres meses y medio, equivalente a aproximadamente 77 días laborables. Durante este período, se realizaron diversas tareas que incluyeron el estudio del dominio, la búsqueda bibliográfica, el desarrollo del código del proyecto y las reuniones semanales con el tutor. El desarrollo del código se dividió en la construcción de los conjuntos de datos, el análisis exploratorio de los datos, el procesamiento de los datos, y el desarrollo y evaluación de los modelos de aprendizaje automático. En promedio, se destinaron 7 horas diarias a estas actividades, lo que sumaría un total aproximado de 540 horas invertidas en la realización de este proyecto. En la Tabla 8 se puede observar las horas dedicadas a cada tarea y su coste asociado, así como el coste total del proyecto. Teniendo esto en cuenta, se estima que el costo total de este proyecto ascendería a aproximadamente 7460 €.

Tabla 8. Horas dedicadas a cada parte del proyecto y su coste económico asociado.

Parte del proyecto	Horas dedicadas (h)	Coste económico asociado (€)
Búsqueda bibliográfica y estudio del dominio	70	967
Construcción de los conjuntos de datos	97	1340
Ánálisis exploratorio de los datos	118	1630
Procesamiento de los datos	55	760
Desarrollo y evaluación de los modelos de aprendizaje automático	190	2625
Reuniones con el tutor	10	138
TOTAL:	540 h	7460 €

9 Limitaciones del estudio

En esta sección, se explican varias de las limitaciones que nos hemos encontrado y que afectan a la validez y generalización de los resultados.

El primer reto que nos encontramos fue la falta de investigaciones previas en el uso de técnicas de aprendizaje automático para la predicción de la evolución de pacientes con desórdenes musculoesqueléticos en el tobillo durante los procesos de rehabilitación. La falta de literatura existente en este campo nos limitó a la hora de tener puntos de referencia y poder realizar comparaciones con otros estudios.

Una de las limitaciones más evidentes de este estudio es el tamaño relativamente pequeño de la muestra de pacientes. El número de participantes incluidos en el estudio podría no ser representativo de la diversidad de los desórdenes musculoesqueléticos en el tobillo. Esto podría influir en la capacidad de generalizar los resultados a poblaciones más amplias.

Aunque se exploraron varios algoritmos de aprendizaje automático, es importante destacar que ninguno de los modelos logró predecir con precisión los valores máximos de frenada y propulsión durante la fase de apoyo del ciclo de la marcha. Esto indica que esos tramos del ciclo de la marcha no fueron completamente capturados.

Este estudio se centró en la componente anteroposterior de las fuerzas de reacción del suelo. Otras componentes, como las fuerzas verticales y medio-laterales, no se consideraron en este análisis, lo que podría limitar la comprensión completa de la marcha.

Finalmente, es importante señalar que los resultados de este estudio están basados en datos recopilados en un centro de rehabilitación austriaco. Esto limita la representatividad geográfica y demográfica de los pacientes, lo que podría restringir la generalización de los resultados a poblaciones de otras regiones.

10 Conclusiones

Este trabajo de investigación pretende abordar un existente desafío en el campo de la salud y el análisis clínico de la marcha, utilizando algoritmos de aprendizaje automático y análisis de datos para predecir la evolución de pacientes con deficiencias musculoesqueléticas en el tobillo durante sus procesos de rehabilitación.

A lo largo de la investigación, hemos realizado un análisis exhaustivo de los datos correspondientes a las fuerzas de reacción del suelo en la dirección anteroposterior durante la fase de apoyo del ciclo de la marcha. Asimismo, hemos explorado como distintas aproximaciones de procesamiento de los datos afectan al rendimiento de los modelos de aprendizaje automático a la hora de realizar las predicciones.

El método *target-specific ranks* nos ha permitido comparar el rendimiento de los distintos modelos entre sí, mientras que los valores de las métricas de evaluación de regresión como la raíz del error cuadrático medio (RMSE), el coeficiente de determinación (R^2), y el coeficiente de correlación de Pearson para cada una de las variables objetivo, nos han permitido evaluar los modelos en los diferentes puntos de la curva predicha de manera más precisa.

Nuestros resultados han resaltado la importancia de aplicar estrategias de reducción de la dimensionalidad (como el PCA) en los datos de entrada correspondientes a la primera sesión de rehabilitación en algoritmos como la regresión lineal múltiple. Este modelo, en conjunto con el modelo de regresión de mínimos cuadrados parciales (PLS), y el modelo de regresión de vectores de soporte (SVR) entrenado en datos originales muestran un rendimiento superior a los otros modelos de la investigación en la predicción de la última sesión de rehabilitación a partir de la primera sesión de terapia para pacientes con desórdenes musculoesqueléticos en el tobillo. Estos modelos han sido capaces de predecir la forma de la curva correspondiente a la última sesión de rehabilitación real de manera adecuada, así como la fase de apoyo con una sola pierna. No obstante, a pesar de los resultados obtenidos es importante señalar que ninguno de los modelos logró predecir con precisión los valores máximos de frenada (correspondiente a la fase de apoyo del talón) y propulsión (correspondiente a la fase de despegue del talón) durante la fase de apoyo del ciclo de la marcha.

En conclusión, este estudio proporciona una base para futuras investigaciones destinadas a predecir los resultados de la rehabilitación en pacientes con trastornos musculoesqueléticos en el tobillo. Sin embargo, aún existen retos por superar, principalmente en relación con la predicción en las fases de máxima frenada y propulsión en la fase de apoyo del ciclo de la marcha. Esto resalta la necesidad de realizar investigaciones y desarrollos adicionales para mejorar la capacidad predictiva de los modelos en estas áreas clave y, en última instancia, contribuir a una atención médica más efectiva y personalizada.

11 Líneas futuras

Una línea de mejora para futuras investigaciones es la optimización de los modelos de predicción, centrándose específicamente en los puntos clave de la curva de fuerzas de reacción del suelo durante la fase de apoyo, es decir, el punto máximo de frenada y el punto máximo de propulsión. En lugar de predecir toda la curva de manera continua, un enfoque más detallado podría centrarse en estos momentos críticos que son fundamentales para la comprensión de la marcha y la rehabilitación.

El tratamiento de los valores atípicos en los datos para mejorar la robustez de los modelos es también un punto de mejora en futuros estudios. Quizás el hecho de que la predicción presente un offset en los puntos máximos de la curva se debe a los valores atípicos, y hacer una recta de calibración podría ser un enfoque a tener en cuenta para solucionarlo.

Para garantizar la validez y aplicabilidad de los modelos, es necesario considerar la inclusión de un número más elevado de pacientes en futuras investigaciones. Un tamaño de muestra más grande permitirá una representación más completa de la diversidad de pacientes con desórdenes musculoesqueléticos en el tobillo.

Además de aumentar el tamaño de la muestra, se podría ampliar la investigación para incluir pacientes con una variedad más elevada de desórdenes musculoesqueléticos, como desórdenes en la rodilla, en la cadera o en el calcáneo. Esto proporcionaría una comprensión más completa de cómo los modelos se comportan en diferentes contextos clínicos, y permitiría evaluar la capacidad de los modelos para adaptarse a diversas situaciones clínicas, para en última instancia, proporcionar una herramienta más versátil.

Aunque este estudio se enfocó en la componente anteroposterior de las fuerzas de reacción del suelo, sería interesante extender la investigación para incluir otras componentes, como la vertical o la medio-lateral. Estas componentes desempeñan un papel significativo en la marcha y pueden ofrecer información adicional sobre el proceso de rehabilitación del paciente.

Estas mejoras propuestas para futuras investigaciones tienen como objetivo que, en última instancia, los modelos de aprendizaje automático se conviertan en herramientas de apoyo a la toma de decisiones para los médicos. Estas herramientas les ayudarán a evaluar con mayor precisión la efectividad de los tratamientos de rehabilitación y les proporcionarán información útil para tomar decisiones fundamentadas sobre si deben continuar, ajustar o finalizar la terapia de sus pacientes.

12 Referencias

- [1] Eric J Topol. 2019. *High-performance medicine: The convergence of human and artificial intelligence*. *Nature Medicine* 25, 1 (2019), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- [2] Joana Figueiredo, Cristina P. Santos, and Juan C. Moreno. 2018. *Automatic recognition of gait patterns in human motor disorders using machine learning: A review*. *Medical Engineering and Physics* 53 (2018), 1–12. <https://doi.org/10.1016/j.medengphy.2017.12.006>
- [3] Wolfgang I Schöllhorn. 2004. *Applications of artificial neural nets in clinical biomechanics*. *Clinical Biomechanics* 19, 9 (2004), 876–898. <https://doi.org/10.1016/j.clinbiomech.2004.04.005>
- [4] Horsak, B., Slijepcevic, D., Raberger, AM. et al. GaitRec, a large-scale ground reaction force dataset of healthy and impaired gait. *Sci Data* 7, 143 (2020). <https://doi.org/10.1038/s41597-020-0481-z>
- [5] Hong-yin Lau, Kai-yu Tong, and Hailong Zhu. 2009. Support vector machine for classification of walking conditions of persons after stroke with dropped foot. *Human Movement Science* 28, 4 (Aug. 2009), 504–514. <https://doi.org/10.1016/j.humov.2008.12.003>
- [6] Leen Van Gestel, Tinne De Laet, Enrico Di Lello, Herman Bruyninckx, Guy Molenaers, Anja Van Campenhout, Erwin Aertbeliën, Mike Schwartz, Hans Wambacq, Paul De Cock, and Kaat Desloovere. 2011. Probabilistic gait classification in children with cerebral palsy: A Bayesian approach. *Research in Developmental Disabilities* 32, 6 (Nov. 2011), 2542–2552. <https://doi.org/10.1016/j.ridd.2011.07.004>
- [7] Corina Nüesch, Victor Valderrabano, Cora Huber, Vinzenz von Tscharner, and Geert Pagenstert. 2012. Gait patterns of asymmetric ankle osteoarthritis patients. *Clinical Biomechanics* 27, 6 (July 2012), 613–618. <https://doi.org/10.1016/j.clinbiomech.2011.12.016>
- [8] Murad Alaqtash, Thompson Sarkodie-Gyan, Huiying Yu, Olac Fuentes, Richard Brower, and Amr Abdelgawad. 2011. Automatic classification of pathological gait patterns using ground reaction forces and machine learning algorithms. In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS). IEEE, 453–457. <https://doi.org/10.1109/IEMBS.2011.6090063>
- [9] Ferdous Wahid, Rezaul K Begg, Chris J Hass, Saman Halgamuge, and David C Ackland. 2015. Classification of Parkinson's disease gait using spatial-temporal gait features. *IEEE Journal of Biomedical and Health Informatics* 19, 6 (2015), 1794–1802.
- [10] Djordje Slijepcevic, Matthias Zeppelzauer, Anna-Maria Gorgas, Caterine Schwab, Michael Schüller, Arnold Baca, Christian Breiteneder, and Brian Horsak. 2017. Automatic classification of functional gait disorders. *IEEE Journal of Biomedical and Health Informatics* 22, 5 (2017), 1653–1661. <https://doi.org/10.1109/JBHI.2017.2785682>
- [11] O. A. Galarraga C., V. Vigneron, B. Dorizzi, N. Khouri & E. Desailly (2015) Predicting postoperative knee flexion during gait of cerebral palsy children, *Computer Methods in Biomechanics and Biomedical Engineering*, 18:sup1, 1940-1941, DOI: 10.1080/10255842.2015.1070583
- [12] Galarraga, Omar & Vigneron, Vincent & Dorizzi, Bernadette & Khouri, Nejib & Desailly, Eric. (2015). Estimation of Postoperative Knee Flexion at Initial Contact of Cerebral Palsy

- Children using Neural Networks. ICPRAM 2015 - 4th International Conference on Pattern Recognition Applications and Methods, Proceedings. 2. 10.5220/0005286503380342.
- [13] Galarraga C, O. A., Vigneron, V., Dorizzi, B., Khouri, N., & Desailly, E. (2017). Predicting postoperative gait in cerebral palsy. *Gait & posture*, 52, 45–51. <https://doi.org/10.1016/j.gaitpost.2016.11.012>
- [14] Biomechanics and Gait Analysis. (2020). Biomechanics and Gait Analysis. Elsevier. <https://doi.org/10.1016/c2016-0-04553-9>
- [15] Baker, R., Esquenazi, A., Benedetti, M. G., & Deslooverre, K. (2016). Gait analysis: Clinical facts. European Journal of Physical and Rehabilitation Medicine, 52(4), 560–574.
- [16] Pirker, W., & Katzenschlager, R. (2017, February 1). Gait disorders in adults and the elderly: A clinical guide. *Wiener Klinische Wochenschrift*. Springer-Verlag Wien. <https://doi.org/10.1007/s00508-016-1096-4>
- [17] Baker, R. (2006). Gait analysis methods in rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 3. <https://doi.org/10.1186/1743-0003-3-4>
- [18] Blanc, Y., Balmer, C., Landis, T., & Vingerhoets, F. (1999). Temporal parameters and patterns of the foot roll over during walking: Normative data for healthy adults. *Gait and Posture*, 10(2), 97–108. [https://doi.org/10.1016/S0966-6362\(99\)00019-3](https://doi.org/10.1016/S0966-6362(99)00019-3)
- [19] Stöckel, T., Jacksteit, R., Behrens, M., Skripitz, R., Bader, R., & Mau-Moeller, A. (2015). The mental representation of the human gait in young and older adults. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00943>
- [20] Linden, M. V. D. (2010). Gait Analysis, Normal and Pathological Function, 2nd ed. J. Perry, J.M. Burnfield, Slack Inc., 576 pages, ISBN 978-1-55642r-r766-4. Physiotherapy, 304(8), 907.
- [21] Marasovič, T., Cecić, M., & Zanchi, V. (2009). Analysis and interpretation of ground reaction forces in normal gait. *WSEAS Transactions on Systems*, 8(9), 1105–1114.
- [22] Sánchez Lacuesta, J. J., J.M., P. P., Hoyos Fuentes, J. V., Viosca Herreo, E., Soler Gracia, C., Comín Clavijo, M. (1999). Biomecánica de la Marcha Humana y Patológica (1 ed.). Valencia: Instituto de Biomecánica de Valencia (IBV)
- [23] Cámara, J. (2012). Fuerza de reacción del suelo durante la marcha. Componente vertical y antero-posterior. *Educación Física y Deporte*, 30(2), 607–625. <https://doi.org/10.17533/udea.efyd.11319>
- [24] San Gil Sorbet, M^a Amaya. Análisis dinámico de la marcha. estudio de los centros de presión sobre la huella plantar. influencia de los distintos calzados. Tesis Doctoral. Facultad de medicina de la Universidad de Alcalá de Henares, 1.991
- [25] Ramos Sánchez, Mabel. Utilidad del análisis tridimensional de la marcha como sistema evaluador del estado clínico y funcional de pacientes sometidos a artroplastia de rodilla. Tesis doctoral. Facultad de Medicina de la Universidad Complutense de Madrid. Madrid, 2000
- [26] Vázquez, S. C. (2004). Análisis de la marcha humana con plataformas dinamométricas: influencia del transporte de carga. Universidad Complutense de Madrid (pp. 1–335). Retrieved from <http://eprints.ucm.es/4401/>
- [27] Revi, D. A., Alvarez, A. M., Walsh, C. J., De Rossi, S. M. M., & Awad, L. N. (2020). Indirect measurement of anterior-posterior ground reaction forces using a minimal set of wearable inertial sensors: From healthy to hemiparetic walking. *Journal of*

- NeuroEngineering and Rehabilitation, 17(1). <https://doi.org/10.1186/s12984-020-00700-7>*
- [28] Gait analysis with Kistler Force Plates: Precision makes all the difference. Kistler. (n.d.). <https://www.kistler.com/US/en/gait-analysis/C00000024>
- [29] Mishra, S., Sarkar, U., Taraphder, S., Datta, S., Prasanna, D., Saikhom, R., ... Laishram, M. (2017). Multivariate Statistical Data Analysis- Principal Component Analysis (PCA). *International Journal of Livestock Research, 7*(5), 60–78.
- [30] Burkov, A. (2019). *The Hundred-Page Machine Learning Book-Andriy Burkov. Expert Systems* (Vol. 5, pp. 132–150). Retrieved from <http://doi.wiley.com/10.1111/j.1468-0394.1988.tb00341.x>
- [31] Mali, K. (2023, August 9). Everything you need to know about linear regression!. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/10/everything-you-need-to-know-about-linear-regression/>
- [32] Helland, I.S. (1990). Partial least squares regression and statistical models. *Scandinavian Journal of Statistics, 17*, 97-114.
- [33] Abdi, H. (2007). Partial least square regression pls-regression. In N. Salkind (Ed.), Encyclopedia of measurement and statistics (pp. 1–13). Sage. Retrieved from <http://www.utd.edu/~Herve/Abdi-PLSR2007-pretty.pdf>
- [34] Pelliccia, D. (2022, December 8). *Partial least squares regression in python*. NIRPY Research. <https://nirpyresearch.com/partial-least-squares-regression-python/>
- [35] Abdi, H., & Williams, L. J. (2013). Partial least squares methods: partial least squares correlation and partial least square regression. *Methods in molecular biology (Clifton, N.J.), 930*, 549–579. https://doi.org/10.1007/978-1-62703-059-5_23
- [36] Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1*(1), 14–23. <https://doi.org/10.1002/widm.8>
- [37] IBM. Árboles de decisión. Retrieved from <https://www.ibm.com/es-es/topics/decision-trees>
- [38] L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth, Belmont, CA, 1984.
- [39] Liaw, Andy & Wiener, Matthew. (2001). Classification and Regression by RandomForest. Forest. 23.
- [40] Lee, Tae Hwy, Ullah, Aman and Wang, Ran, (2019), Bootstrap Aggregating and Random Forest, No 201918, Working Papers, University of California at Riverside, Department of Economics, <https://EconPapers.repec.org/RePEc:ucr:wpaper:201918>.
- [41] *What is a random forest?*. TIBCO Software. <https://www.tibco.com/reference-center/what-is-a-random-forest>
- [42] Banachewicz, K., Massaron, L., & Goldbloom, A. (2022). Chapter 8: Hyperparameter Optimization. In *The Kaggle Book: Data Analysis and Machine Learning for Competitive Data Science*. essay, Packt.
- [43] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 96 (Vol. 13-17-August-2016, pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- [44] Wikipedia contributors. (2023, August 13). Boosting (machine learning). In *Wikipedia, The Free Encyclopedia*. Retrieved 11:14, September 4, 2023,

- from [https://en.wikipedia.org/w/index.php?title=Boosting_\(machine_learning\)&oldid=170081755](https://en.wikipedia.org/w/index.php?title=Boosting_(machine_learning)&oldid=170081755)
- [45] Friedman, Jerome. (2002). Stochastic Gradient Boosting. *Computational Statistics & Data Analysis*, 38, 367–378. 10.1016/S0167-9473(01)00065-2.
- [46] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273–297. <http://doi.org/10.1023/A:1022627411411>
- [47] Vapnik, V., Golowich, S. E., & Smola, A. (1997). Support vector method for function approximation, regression estimation, and signal processing. In *Advances in Neural Information Processing Systems* (pp. 281–287). Neural information processing systems foundation.
- [48] Drucker, H., Surges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. In *Advances in Neural Information Processing Systems* (pp. 155–161). Neural information processing systems foundation.
- [49] Smola, A. J., & Schölkopf, B. (2004, August). A tutorial on support vector regression. *Statistics and Computing*. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- [50] Sidharth. (2022, November 30). *Support Vector Regression: A comprehensive guide with example*. PyCodeMates. <https://www.pycodemates.com/2022/11/support-vector-regression-a-complete-guide-with-example.html>
- [51] Banachewicz, K., Massaron, L., & Goldbloom, A. (2022). Chapter 5: Competition Tasks and Metrics. In *The Kaggle Book: Data Analysis and Machine Learning for Competitive Data Science*. essay, Packt.
- [52] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ. Computer science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- [53] Sedgwick, P. (2012, July 7). Pearson's correlation coefficient. *BMJ (Online)*. <https://doi.org/10.1136/bmj.e4483>
- [54] McKinney, W., & others. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).
- [55] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- [56] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95.
- [57] Waskom, M., Botvinnik, O., Kane, D., Hobson, P., Lukauskas, S., Gemperline, D., C., Qalieh, A. (2017). *mwaskom/seaborn: v0.8.1* (September 2017). Zenodo. <https://doi.org/10.5281/zenodo.883859>
- [58] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- [59] Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

- [60] Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.
- [61] Chacon, S., & Straub, B. (2014). *Pro git*. Apress.
- [62] github. (2020). *Github*. Retrieved from <https://github.com/>
- [63] Mullineaux, D. R., Milner, C. E., Davis, I. S. & Hamill, J. Normalization of ground reaction forces. *J. Appl. Biomech.* 22, 230–233 (2006).
- [64] Helwig, N. E., Hong, S., Hsiao-Wecksler, E. T. & Polk, J. D. Methods to temporally align gait cycle data. *J. Biomech.* 44, 561–566 (2011).
- [65] Sangeux, M. & Polak, J. A simple method to choose the most representative stride and detect outliers. *Gait Posture* 41, 726–730 (2015).
- [66] Korneva, E., & Blockeel, H. (2020). Towards Better Evaluation of Multi-target Regression Models. In *Communications in Computer and Information Science* (Vol. 1323, pp. 353–362). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-65965-3_23
- [67] Glasdoor. (Agosto, 2023). Sueldos para el puesto de Junio Data Scientist en España. Retrieved from https://www.glassdoor.es/Sueldos/junior-data-scientist-sueldo-SRCH_K00,21.htm

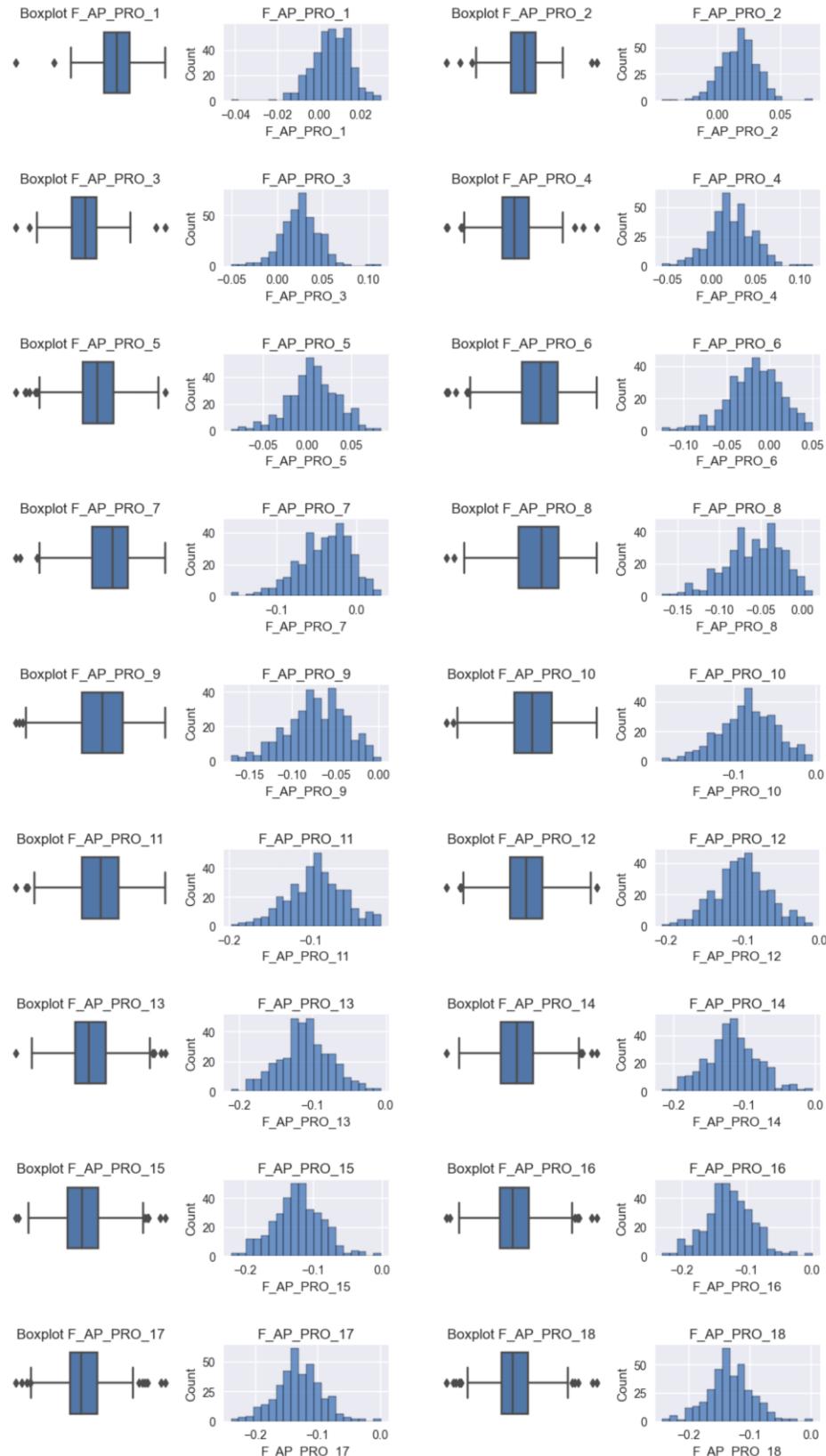
13 Apéndice

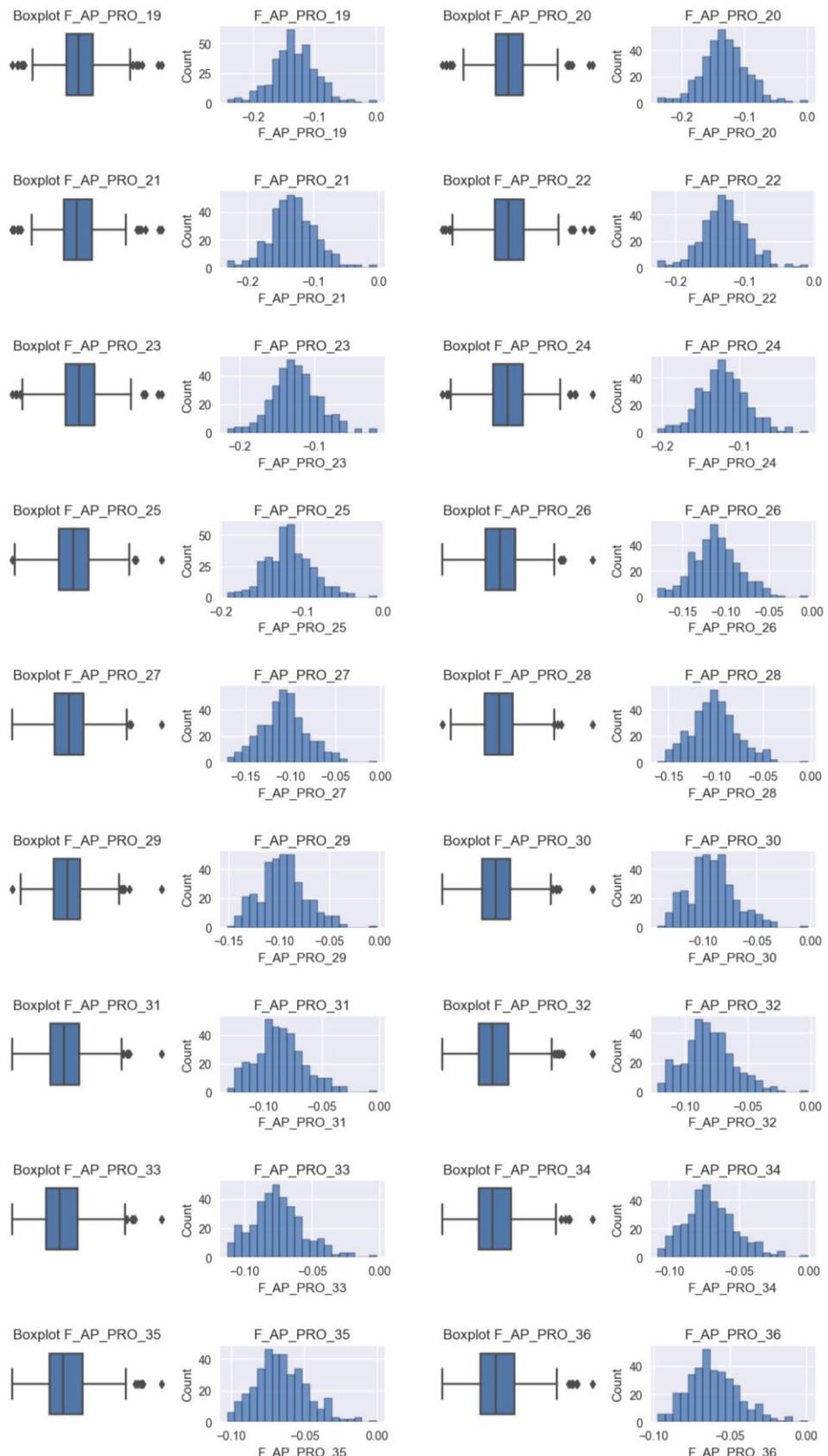
13.1 Apéndice 1. Repositorio del proyecto

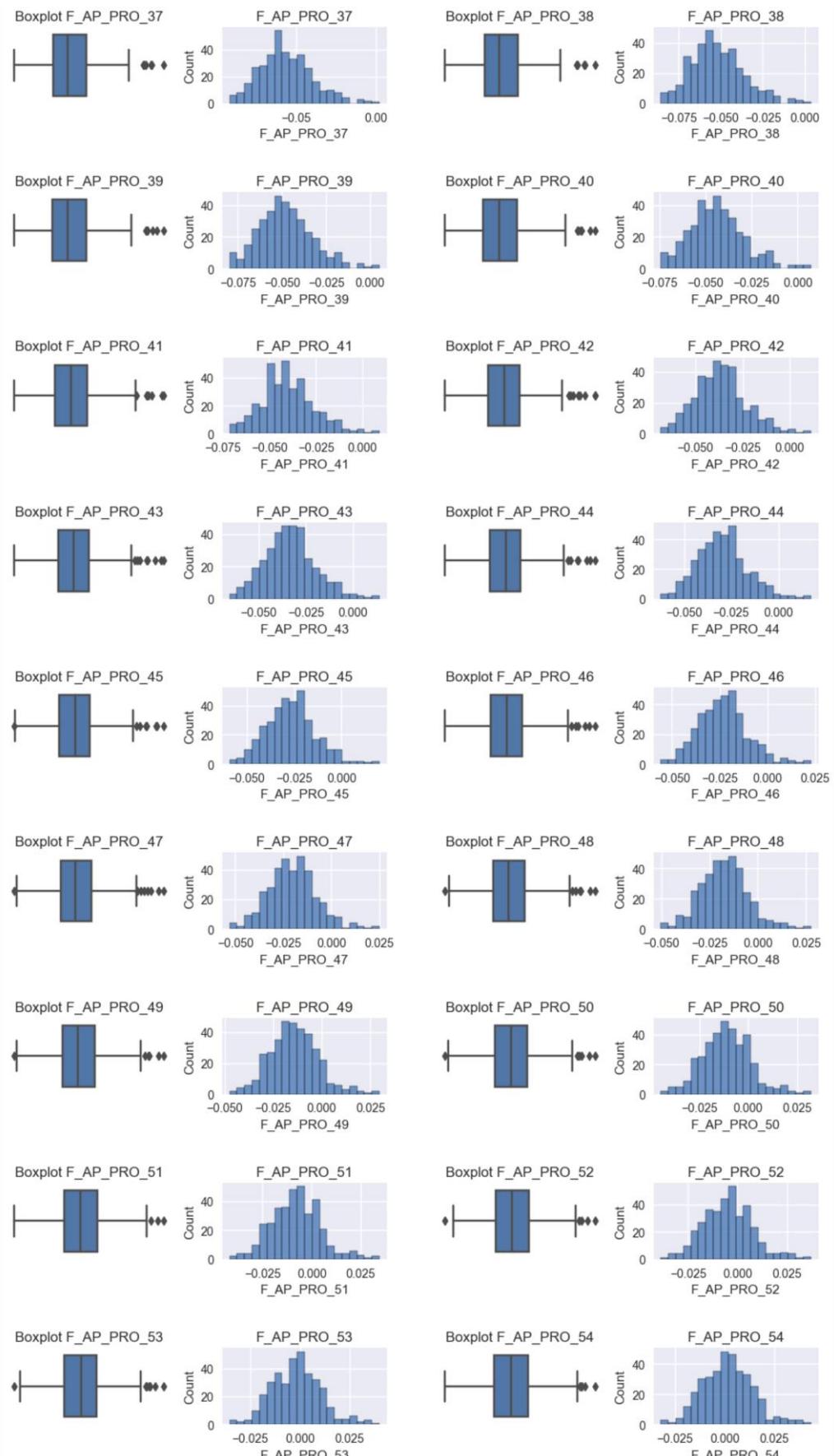
El código fuente desarrollado para este proyecto está disponible en el repositorio GitHub:
<https://github.com/bernatso/Terapy-Progress-Gait-Prediction.git>

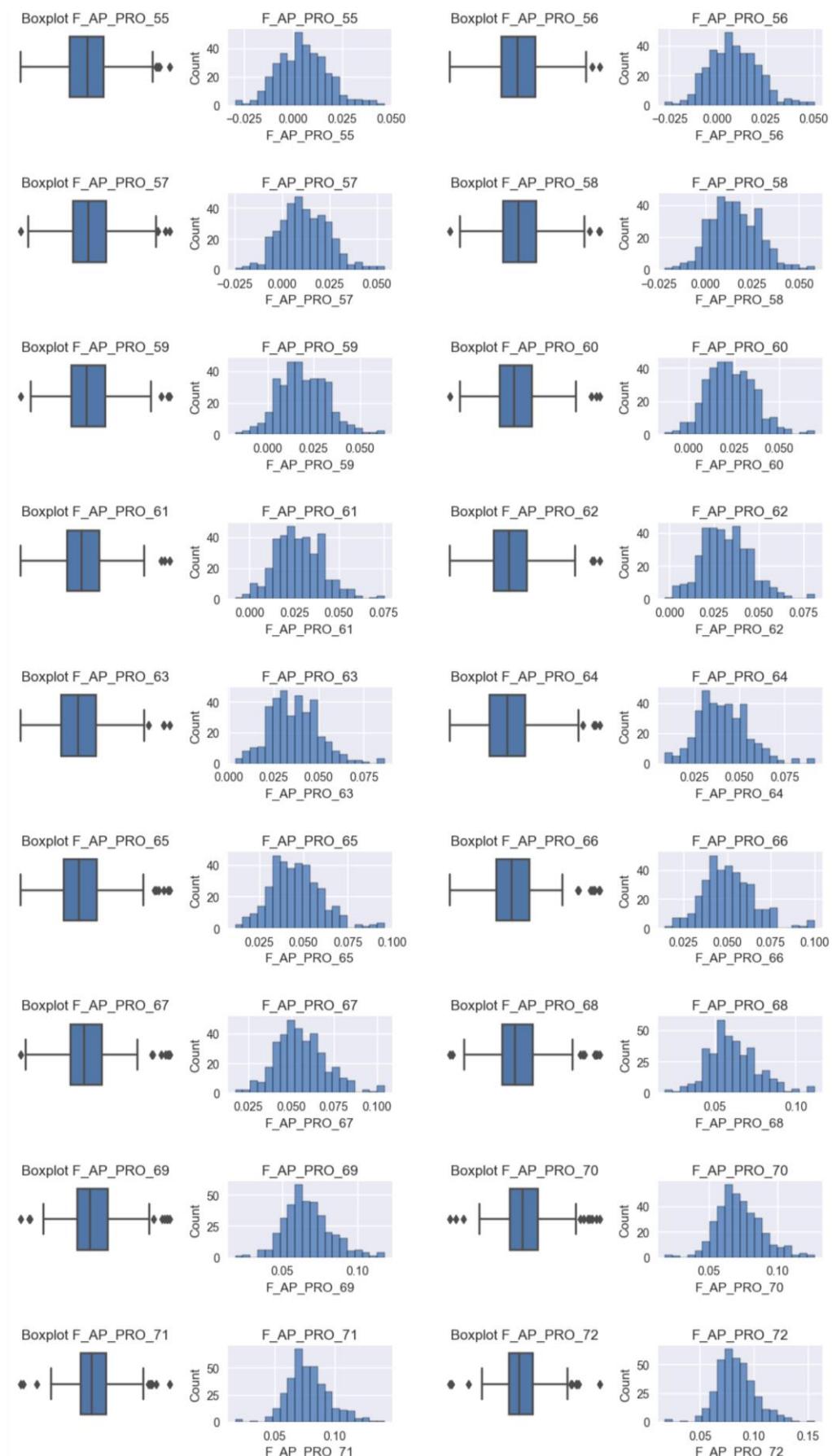
13.2 Apéndice 2. Resultados

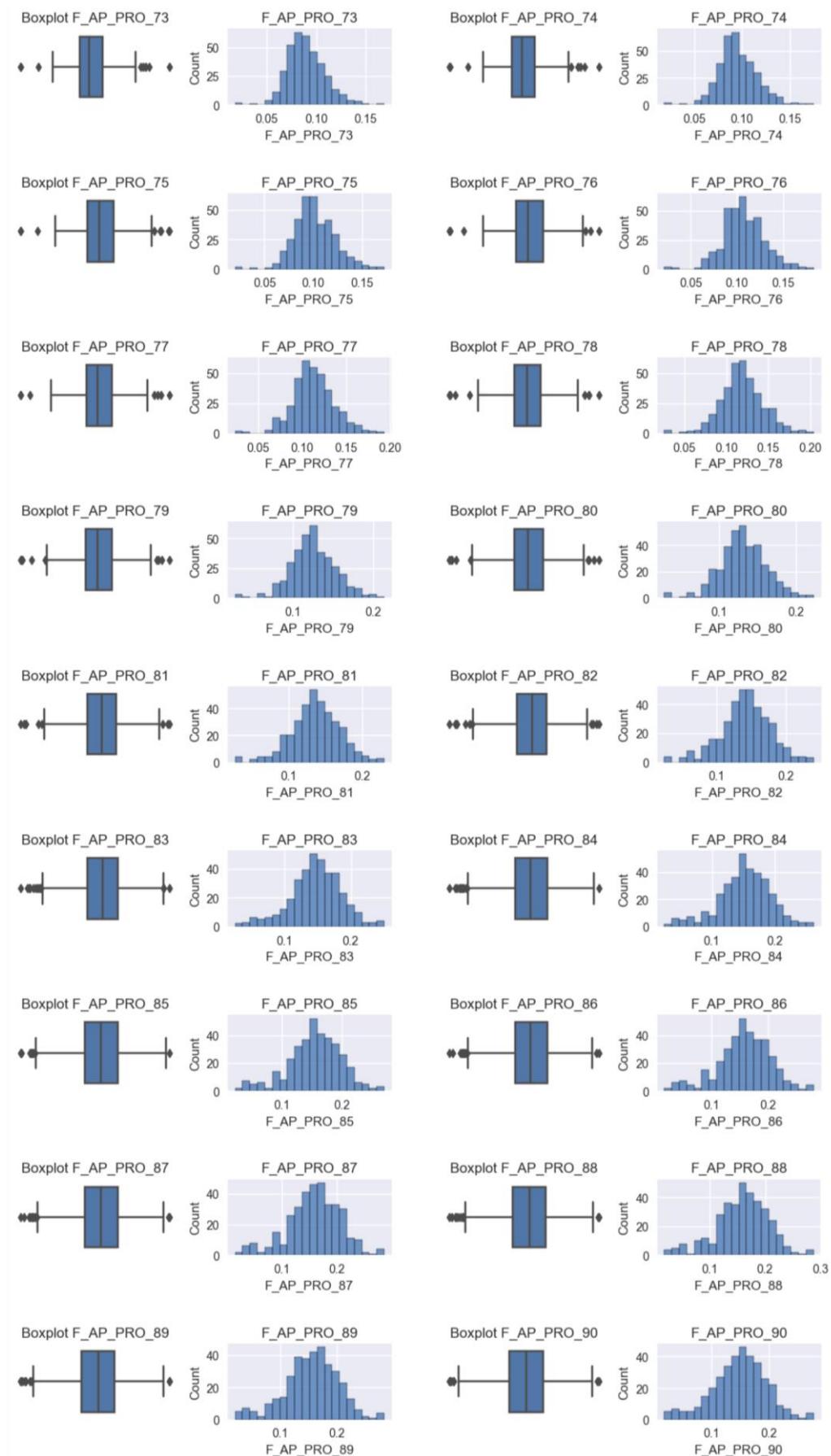
13.2.1 Distribución de las variables numéricas











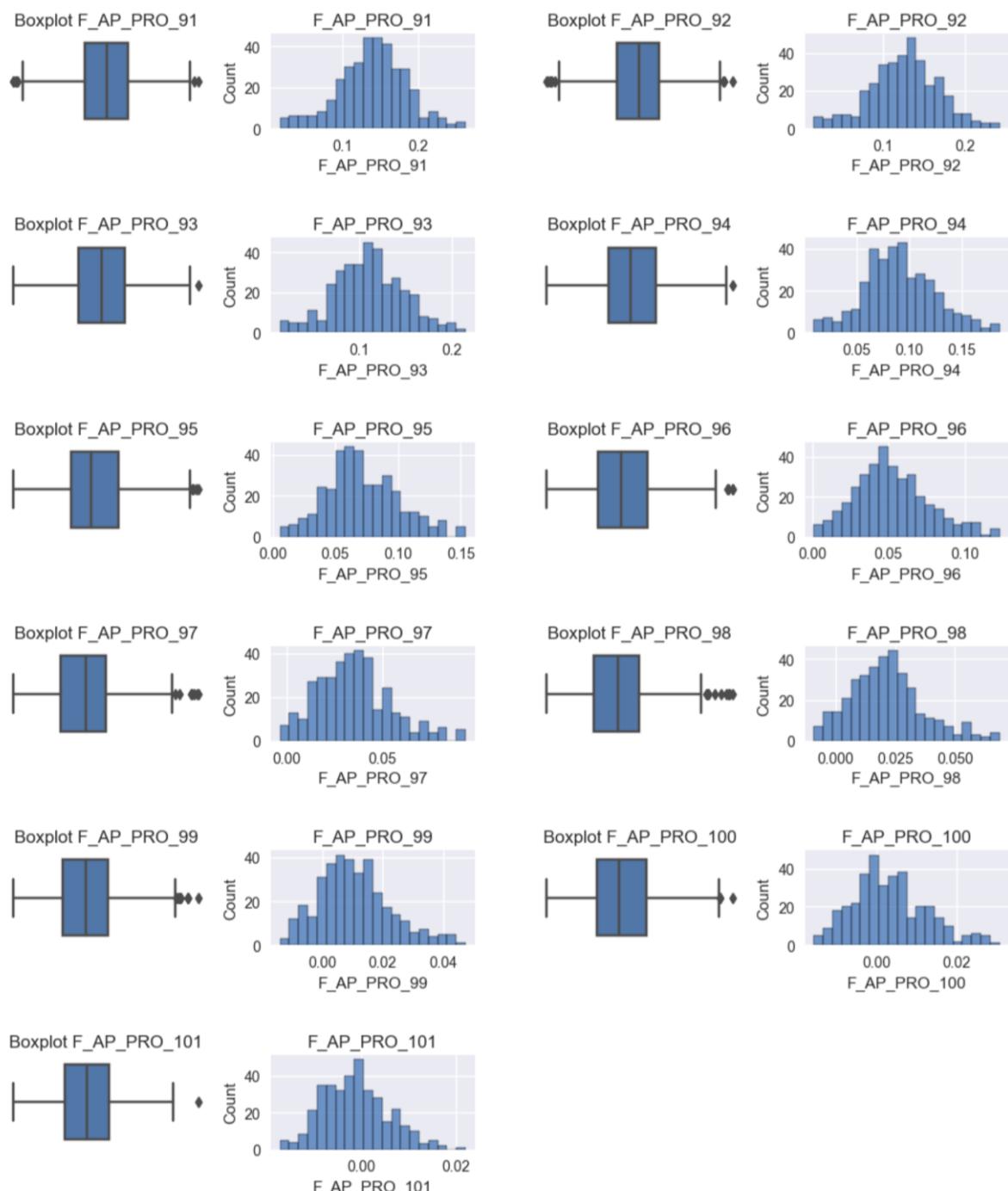


Figura A1. Distribución de las 101 variables del conjunto de entrenamiento mediante histogramas y diagramas de caja o *boxplots*. Estas 101 variables representan los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior durante los 101 puntos en los que se divide la fase de apoyo del ciclo de la marcha. En este caso, estas 101 variables incluían la primera y última sesión de rehabilitación para cada miembro inferior (considerados como observaciones independientes) de los pacientes con trastornos musculoesqueléticos en el tobillo que realizaron dos sesiones de rehabilitación.

13.2.2 Análisis de correlación

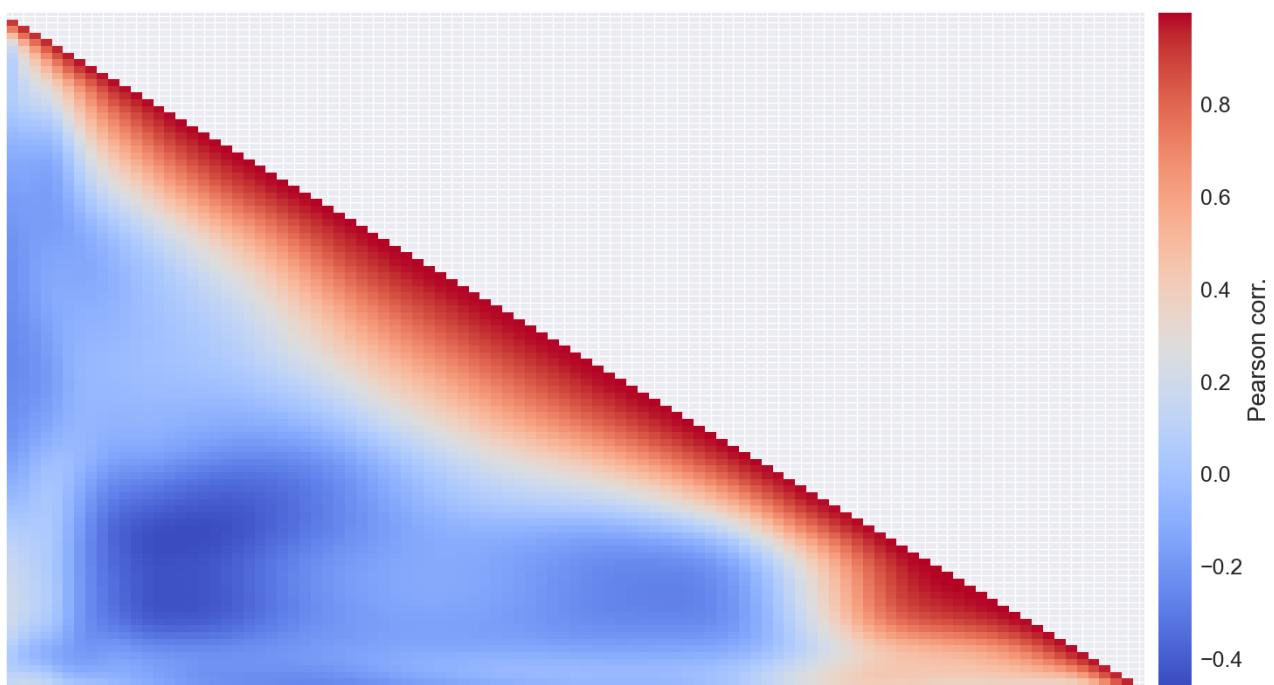
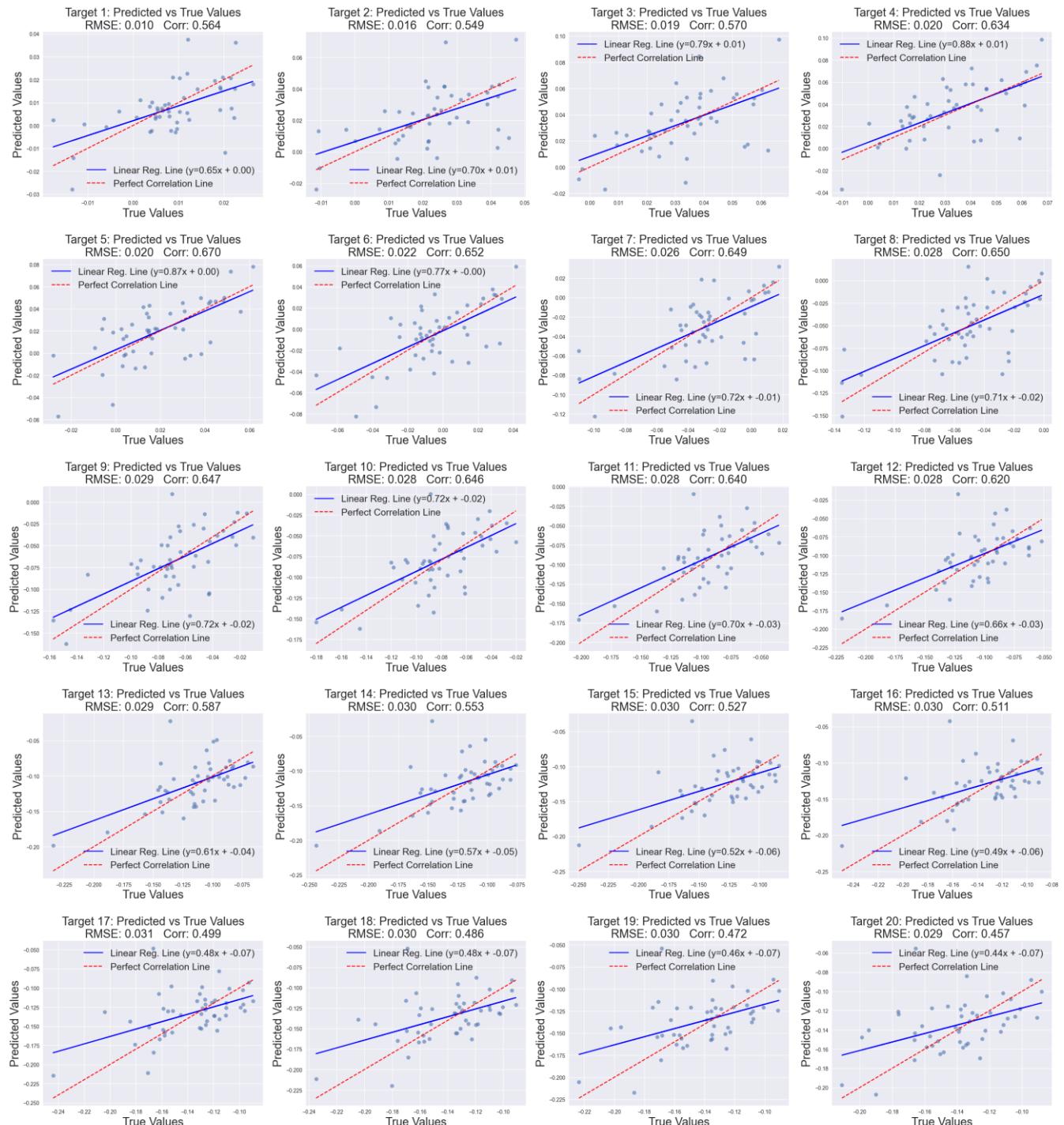
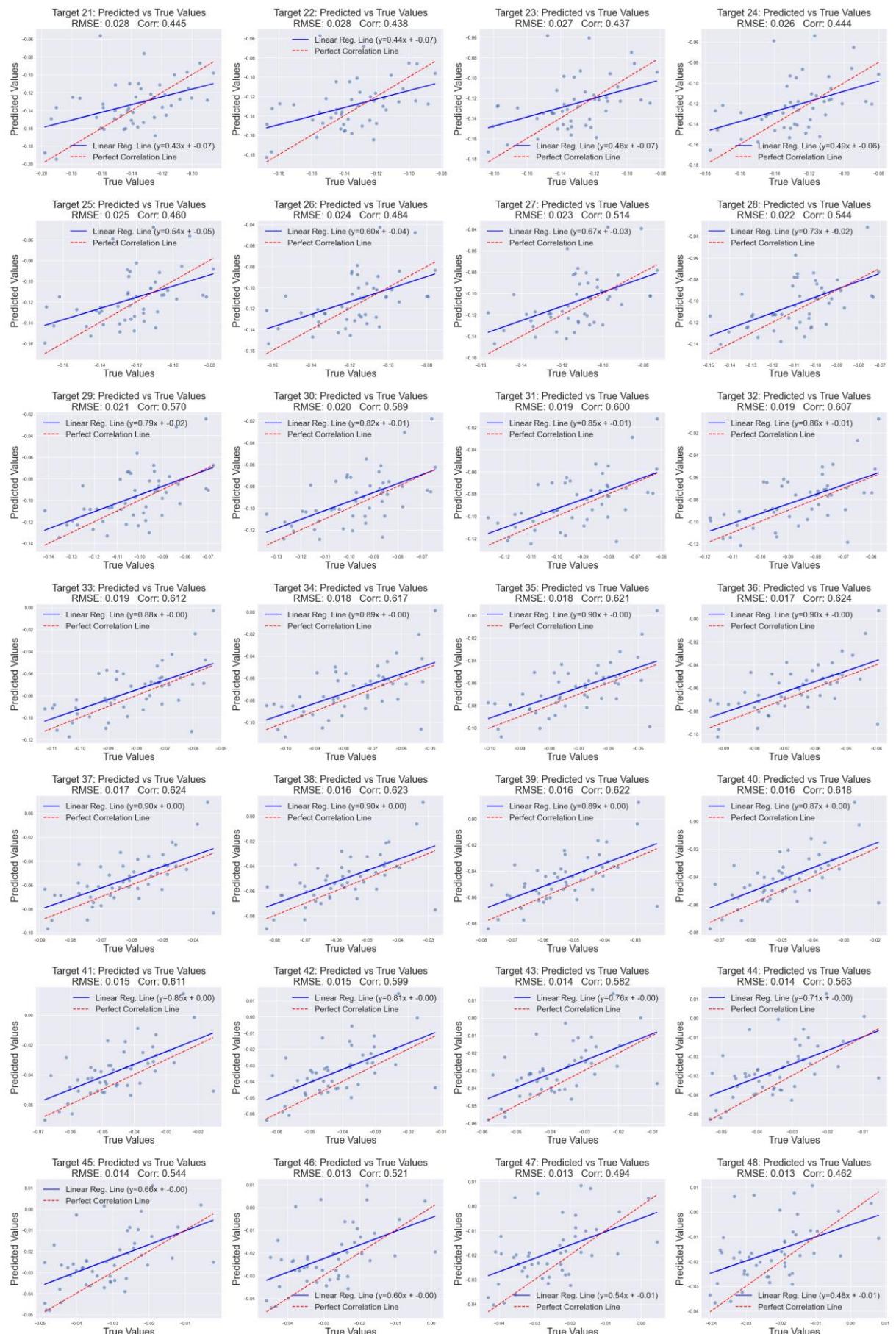
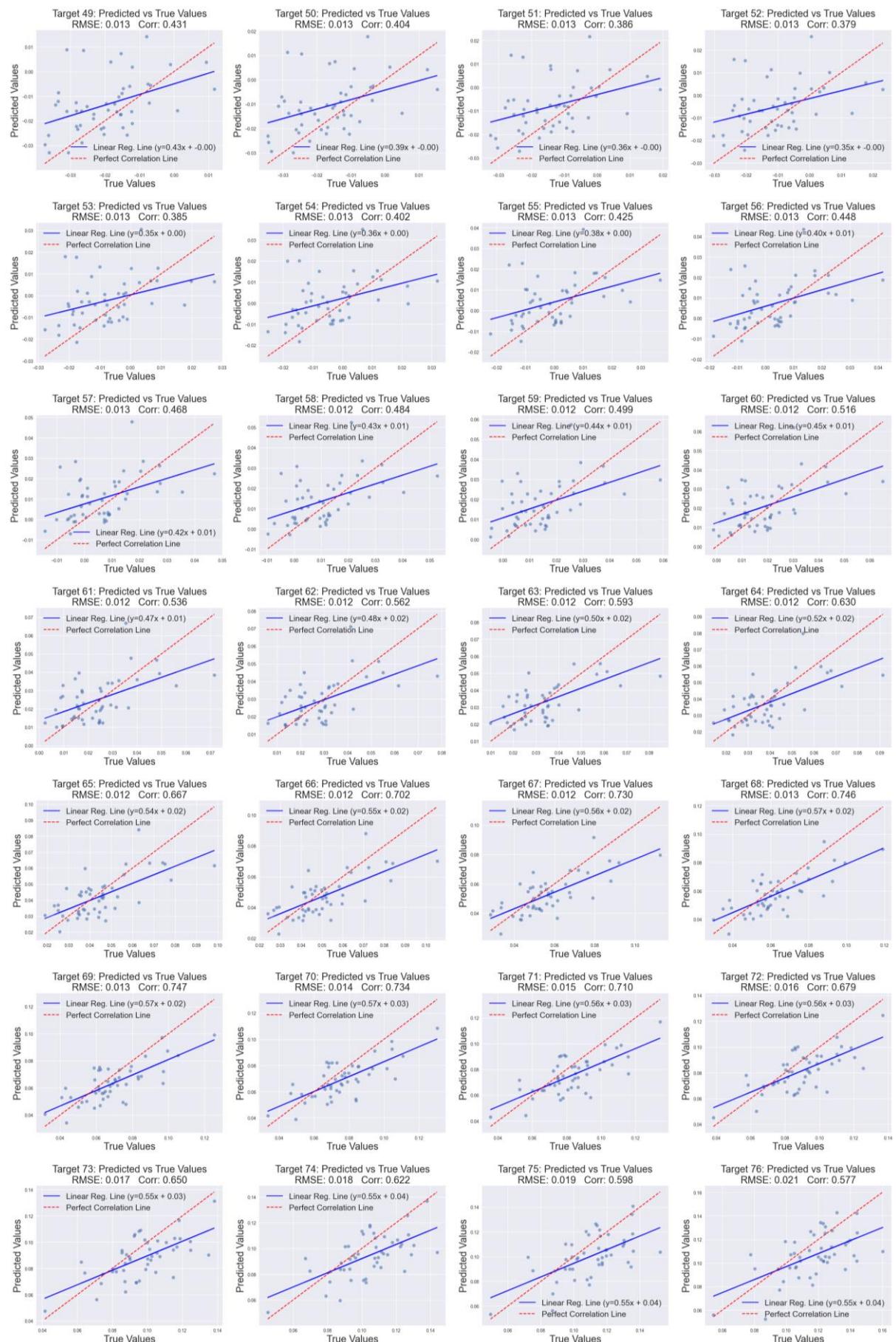


Figura A2. Gráfico de correlación entre las 101 variables del conjunto de entrenamiento utilizando el método de correlación de Pearson. El eje horizontal y vertical representan las 101 variables que corresponden con los valores promedio de la fuerza de reacción del suelo en dirección anteroposterior durante los 101 puntos en los que se divide la fase de apoyo del ciclo de la marcha. En este caso, estas 101 variables incluían la primera y última sesión de rehabilitación para cada miembro inferior (considerados como observaciones independientes) de los pacientes con trastornos musculoesqueléticos en el tobillo que realizaron dos sesiones de rehabilitación.

13.2.3 Análisis multivariante







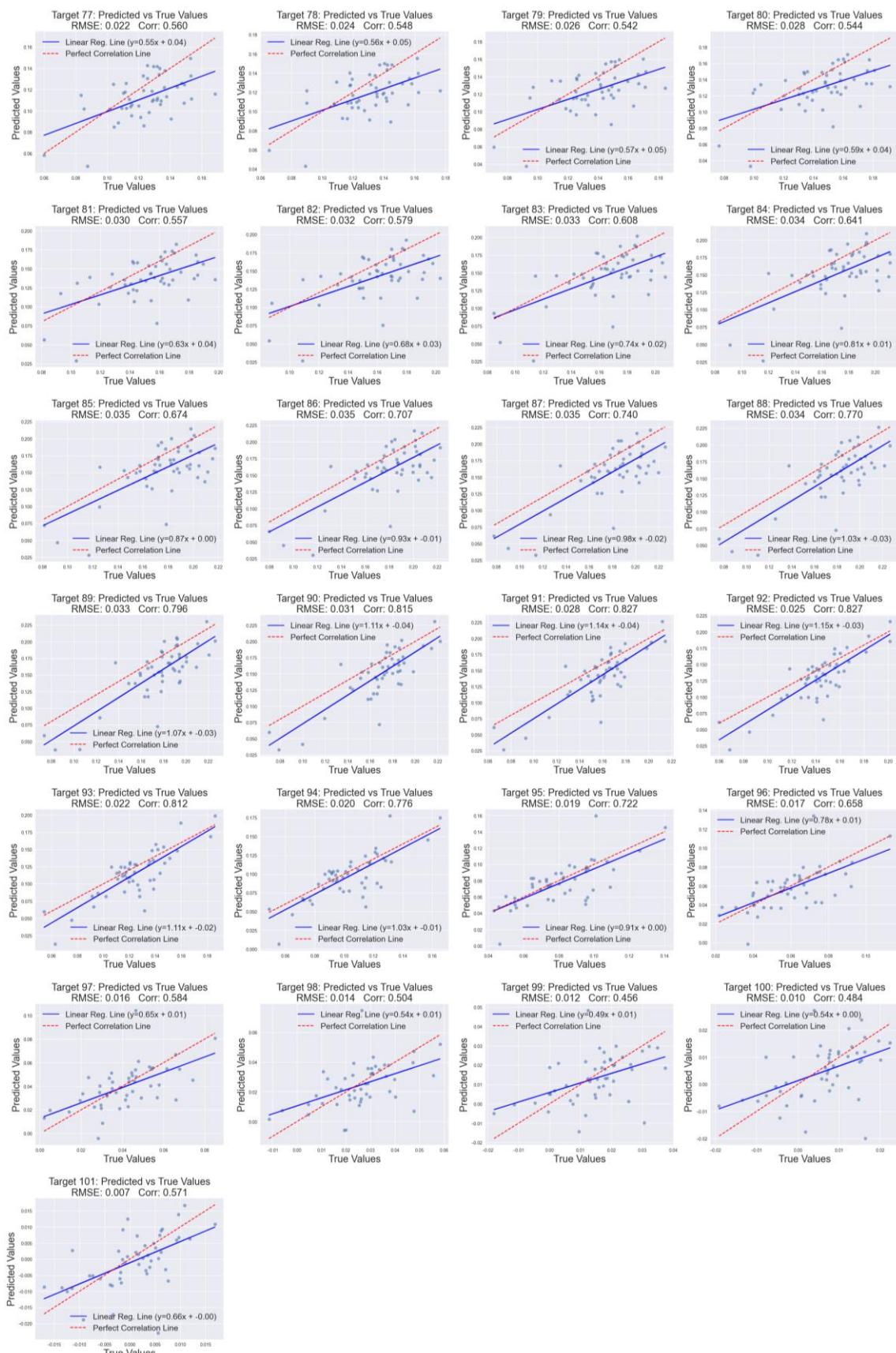
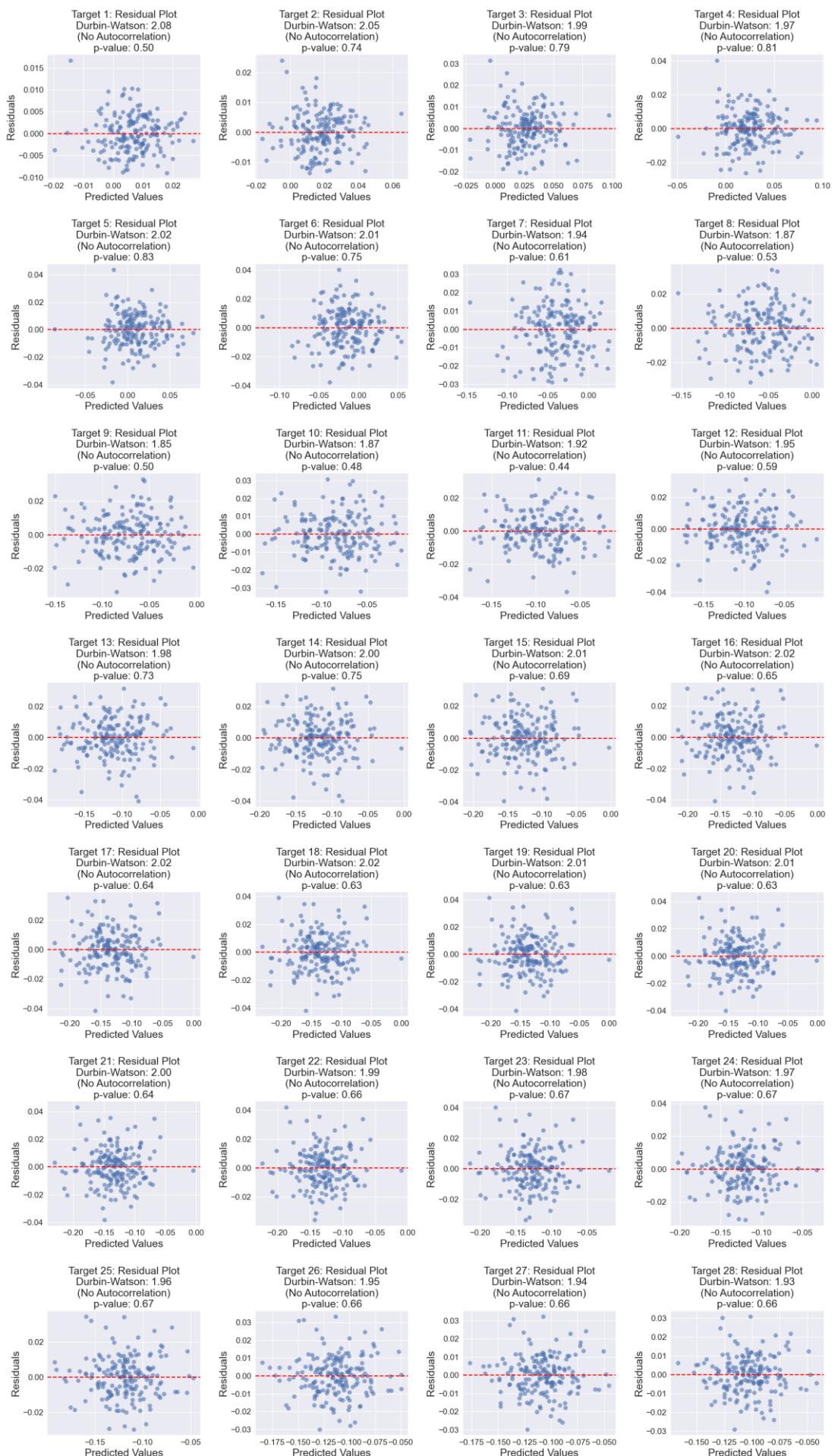
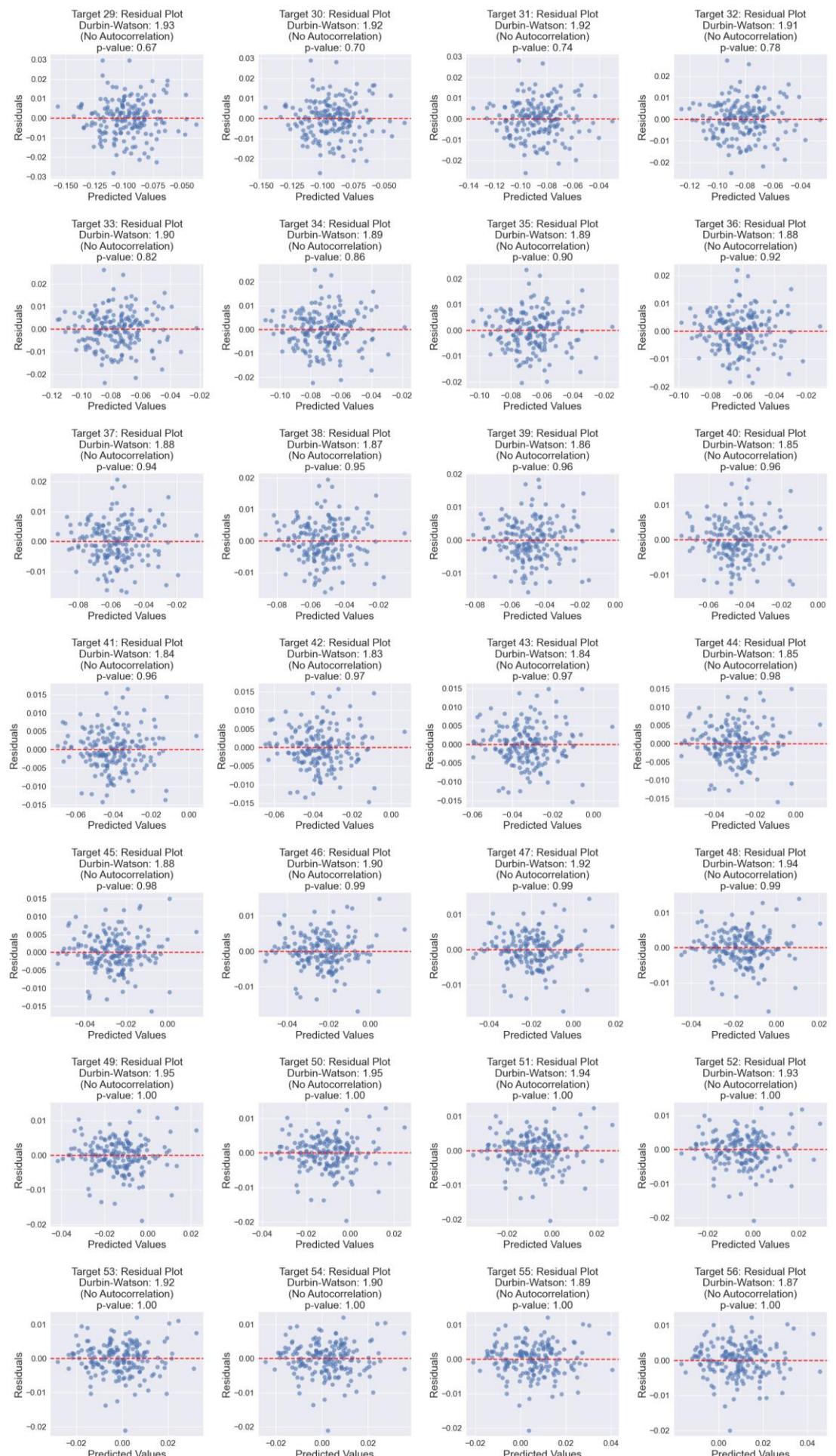
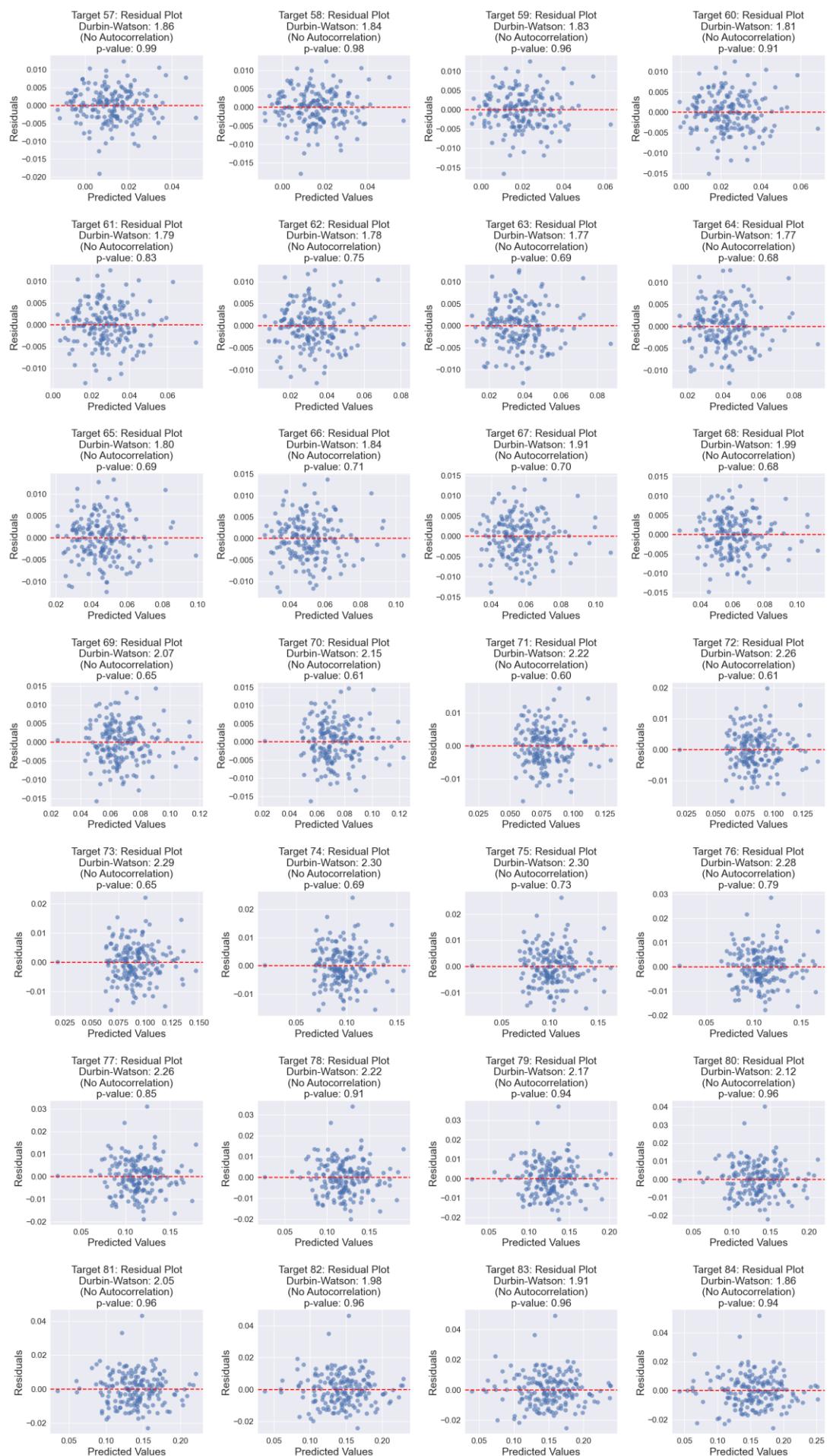


Figura A3. Valores predichos por el primer modelo de regresión lineal múltiple en el conjunto de prueba frente a los valores reales para cada una de las 101 variable objetivo. Se muestra el RMSE, el coeficiente de correlación de Pearson, la línea de correlación perfecta como referencia, la línea de la regresión lineal, y la ecuación de la recta de regresión.







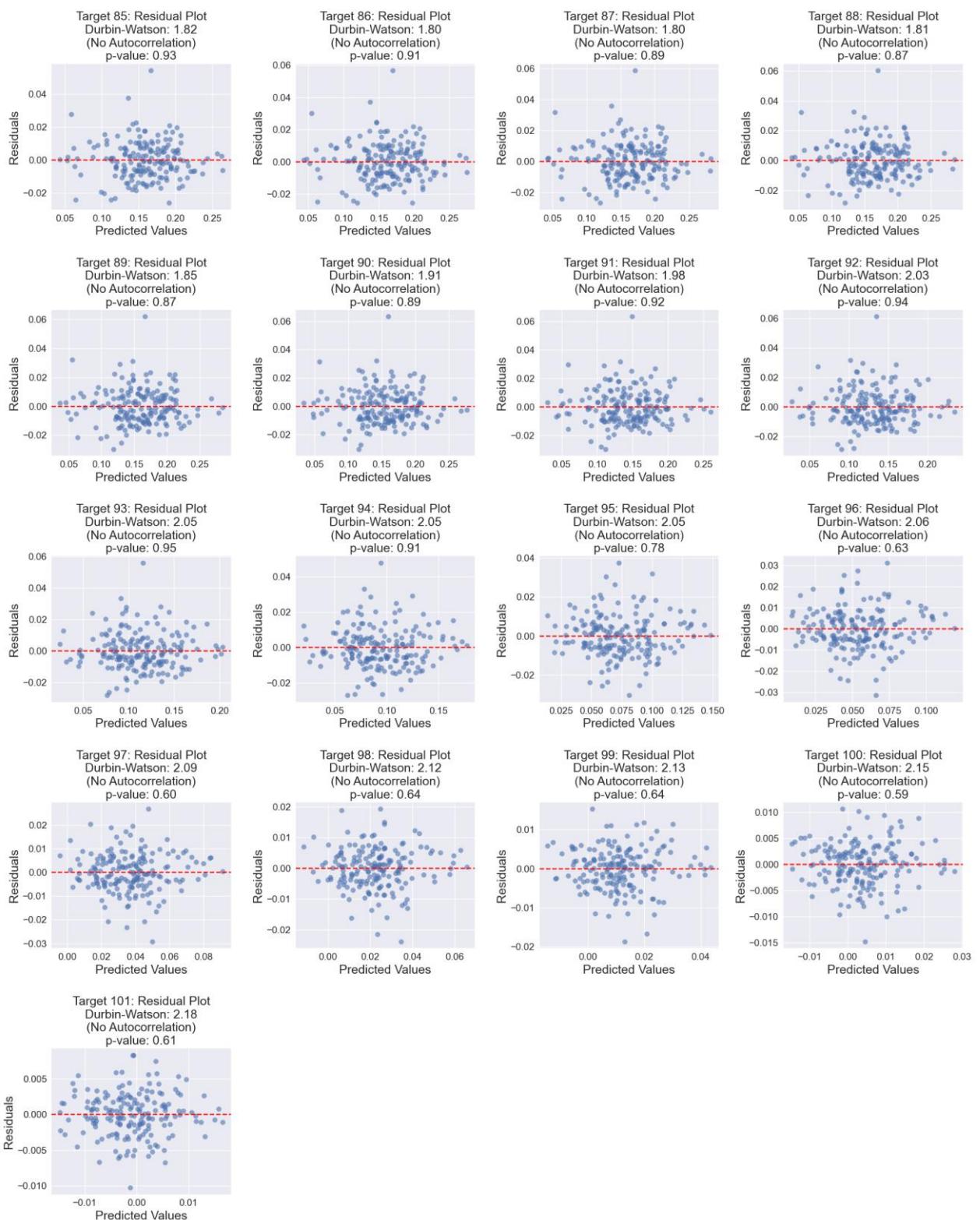


Figura A4. Gráficos de dispersión de residuos frente a valores predichos para los 101 modelos de regresión lineal múltiple (MLR) utilizados en la predicción de las 101 variables objetivo. Cada gráfico incluye los resultados de las pruebas estadísticas de Durbin-Watson y Breusch-Pagan. Estos gráficos nos permitieron evaluar los supuestos de linealidad, independencia y normalidad de los residuos."