

Predicción de la actividad neuromuscular durante la marcha en superficie plana y en la transición de sedestación a bipedestación mediante técnicas de aprendizaje automático.

Carlos Jové Blanco

Máster universitario en Bioinformática y Bioestadística UOC-UB
Bioinformática y Bioestadística Área 4

Nombre Consultor/a: Romina Astrid Rebrij

Nombre Profesor/a responsable de la asignatura: Antoni Pérez Navarro

Fecha Entrega: 8 de junio de 2021



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Predicción de la actividad neuromuscular durante la marcha en superficie plana y en la transición de sedestación a bipedestación mediante técnicas de aprendizaje automático.</i>
Nombre del autor:	<i>Carlos Jové Blanco</i>
Nombre del consultor/a:	<i>Romina Astrid Rebrij</i>
Nombre del PRA:	<i>Antoni Pérez Navarro</i>
Fecha de entrega (mm/aaaa):	06/2021
Titulación:	Máster universitario en Bioinformática y Bioestadística UOC-UB
Área del Trabajo Final:	<i>Bioinformática y Bioestadística Área 4</i>
Idioma del trabajo:	Castellano
Número de créditos:	15
Palabras clave	<i>Aprendizaje automático, electromiografía de superficie, unidad de medición inercial</i>
Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.</i>	
<p>La biomecánica es un área de conocimiento multidisciplinar que estudia los efectos de las fuerzas aplicadas sobre los organismos. Debido a su complejidad, requiere de la adquisición de señales de numerosos sensores, haciendo difícil su aplicación clínica.</p> <p>Las técnicas de aprendizaje automático permiten entrenar modelos para la predicción de valores numéricos, haciendo posible obtener resultados de parámetros que no han sido medidos. Su implementación en el área de la biomecánica permitirá reducir el número de sensores necesarios durante los registros de las actividades, con el impacto que ello tiene a nivel económico y de consumo de tiempo.</p> <p>Se ha realizado un proceso de cinco pasos: recogida de los datos, exploración y preparación de los datos, entrenamiento de los modelos, evaluación de los modelos y mejora de los modelos. Se han entrenado un total de 8 modelos para 21 combinaciones distintas de sensores, compuesta cada una por un sensor inercial y dos sensores de electromiografía, para predecir la actividad de cinco músculos.</p>	

Gracias a los avances en la creación de aplicaciones web resulta posible implementar estos modelos y hacerlos accesibles de forma gratuita, consiguiendo un impacto en la práctica clínica.

El resultado final de este trabajo es una aplicación web que implementa el modelo con mejor capacidad de **predicción de la actividad muscular** durante los siguientes movimientos: la marcha en superficie plana y la transición de sedestación a bipedestación. Además, la aplicación web también permitirá realizar de forma automática el **procesamiento de las señales** durante estos dos movimientos y la obtención de las variables seleccionadas.

Abstract (in English, 250 words or less):

Biomechanics is a multidisciplinary area of knowledge that studies the effects of forces on organisms. Due to its complexity, it requires the acquisition of signals from multiple sensors, making it difficult to implement in daily practice.

Machine learning techniques enable us to train models for the prediction of numeric values, making it possible to obtain parameters that have not been measured directly. Their implementation in biomechanics will allow to reduce the number of necessary sensors during measurements, reducing costs and time consumption.

The process consisted of five steps: accessing the data, exploring and preparing the data, training the models, evaluating the models and improving them. A total of 8 models for 21 different combinations, made up of one inertial sensor and two surface electromyography sensors, to predict five values of muscle activity.

Recent advancements in the creation of web application make it possible to implement these models and make them accessible for free, achieving an impact on daily practice.

The final product of this work is a web application that implements the model with the best results in predicting muscle activity during level walking and the sit to stand transition. Furthermore, it will also allow automatic processing of signals during these two movements and the obtention of the desired variables.

Índice

Contenido

1	Resumen	3
2	Introducción	4
2.1	Contexto y justificación del Trabajo	4
2.2	Objetivos del Trabajo.....	5
2.2.1	Objetivo general	5
2.2.2	Objetivos específicos	5
2.3	Enfoque y método seguido	5
2.3.1	Fase 1.	5
2.3.2	Fase 2.	9
2.4	Planificación del Trabajo.....	10
2.5	Breve resumen de contribuciones y productos obtenidos.....	12
2.5.1	Plan de trabajo	12
2.5.2	Memoria	12
2.5.3	Código fuente	12
2.5.4	Aplicación Web.....	12
2.5.5	Presentación Virtual	12
2.6	Breve descripción de los otros capítulos de la memoria	13
3	Estado del arte.....	14
4	Definición de aprendizaje automático	17
5	Metodología	18
5.1	Paso 1. Recogida de datos.....	18
5.1.1	Procesamiento de las señales del conjunto de datos.....	19
5.1.2	Elaboración del algoritmo de detección de ciclos de la marcha ..	21
5.1.3	Elaboración del algoritmo de detección de ciclos de la transición de sedestación a bipedestación.....	22
5.1.4	Extracción de las variables de las señales procesadas	23
5.2	Entrenamiento de los modelos	25
5.2.1	Modelos seleccionados	25
5.2.2	Entrenamiento de los modelos	26
5.2.3	Evaluación de los modelos.....	27
5.2.4	Mejora de los modelos	27
6	Resultados.....	30
6.1	Predicción de la actividad muscular en la transición de sedestación a bipedestación.	30

6.1.1	Evaluación de los modelos sobre los datos completos en la transición de sedestación a bipedestación	30
6.1.2	Evaluación de los modelos sobre los datos sin valores extremos en la transición de sedestación a bipedestación.....	32
6.1.3	Evaluación de los modelos sobre los datos seleccionados tras la eliminación recursiva de variables en la transición de sedestación a bipedestación.....	33
6.1.4	Evaluación de los modelos sobre los datos tras eliminar todas las variables de IMU no utilizadas para la detección del ciclo de transición de sedestación a bipedestación.....	36
6.2	Predicción de la actividad muscular en la marcha en superficie plana.	38
6.2.1	Evaluación de los modelos sobre los datos completos en la marcha en superficie plana.....	38
6.2.2	Evaluación de los modelos sobre los datos sin valores extremos en la marcha en superficie plana	40
6.2.3	Evaluación de los modelos sobre los datos seleccionados tras la eliminación recursiva de variables en la marcha en superficie plana.....	42
6.2.4	Evaluación de los modelos sobre los datos tras eliminar todas las variables de IMU no utilizadas para la detección del ciclo en la marcha en superficie plana.....	44
7	Discusión	47
7.1	Predicción de la actividad neuromuscular en la transición de sedestación a bipedestación	47
7.2	Predicción de la actividad neuromuscular en la marcha sobre superficie plana	49
7.3	Metodología del trabajo	51
8	Conclusiones	53
Los resultados obtenidos en este trabajo ponen de relieve tres conclusiones principales:		53
8.1	Líneas de futuro.....	53
8.2	Seguimiento de la planificación	54
9	Glosario	55
10	Bibliografía.....	56
11	Anexos.....	60
11.1	Anexo 1. Contenido del repositorio del trabajo.	60
11.2	Anexo 2. Contenido del repositorio de la aplicación web.	62
11.3	Anexo 3. Diagrama de Gantt de la planificación definitiva de la elaboración del trabajo	64

Lista de figuras

Ilustración 1. Imagen del artículo de Hu et al.(5) en la que se aprecia la complejidad de la instrumentación.	16
Ilustración 2. Características de los sujetos registrados en el estudio	19
Ilustración 3. Representación gráfica de un filtro paso banda. $f_1 = 20\text{Hz}$, $f_2 = 450\text{ Hz}$	20
Ilustración 4.Efecto del filtro en una señal electromiográfica	20
Ilustración 5. Resultado de la aplicación del algoritmo de detección de pasos	22
Ilustración 6. Resultado de la aplicación del algoritmo de detección de la transición de sedestación a bipedestación	23
Ilustración 7. Ejemplo de entrenamiento sobre una combinación de datos de marcha sobre superficie plana: IMU tibia, Tibial Anterior y Gemelo interno	27

Lista de tablas

Tabla 1. Combinaciones evaluadas en todos los modelos. IMU en marcha en superficie plana se encuentra en la tibia. IMU en la transición de sedestación a bipedestación se encuentra en la cintura.	7
Tabla 2. Número de registros que permanecen en el conjunto de datos tras la eliminación de valores extremos en la marcha en superficie plana.....	28
Tabla 3. Número de registros que permanecen en el conjunto de datos tras la eliminación de valores extremos en la transición de sedestación a bipedestación.	28
Tabla 4. RECM obtenido tras el entrenamiento de los modelos con los datos completos sobre el conjunto de datos de prueba. (Transición)	30
Tabla 5. EAM obtenido tras el entrenamiento de los modelos con los datos completos sobre el conjunto de datos de prueba. (Transición)	31
Tabla 6. RECM medio obtenido tras el entrenamiento de los modelos con los datos completos a través de validación cruzada. (Transición)	31
Tabla 7. RECM obtenido tras eliminar valores extremos sobre el conjunto de datos de prueba. (Transición).....	32
Tabla 8. EAM obtenido tras eliminar valores extremos de los datos sobre el conjunto de datos de prueba. (Transición)	32
Tabla 9. RECM obtenido tras eliminar valores extremos de los datos a través de validación cruzada. (Transición).....	33
Tabla 10. Número de variables seleccionadas para cada modelo a través de selección recursiva de variables. (Transición).....	34
Tabla 11. RECM obtenido tras la selección recursiva de variables sobre el conjunto de datos de prueba. (Transición)	34
Tabla 12. EAM obtenido tras la selección recursiva de variables sobre el conjunto de datos de prueba. (Transición)	35
Tabla 13. RECM obtenido tras la selección recursiva de variables por validación cruzada. (Transición).....	35
Tabla 14. RECM obtenido tras la selección por conveniencia de variables sobre el conjunto de datos de prueba. (Transición)	36
Tabla 15. EAM obtenido tras la selección por conveniencia de variables sobre el conjunto de datos de prueba. (Transición)	37
Tabla 16. RECM obtenido tras la selección por conveniencia de variables por validación cruzada. (Transición).....	37
Tabla 17. RECM obtenido tras el entrenamiento de los modelos con los datos completos sobre el conjunto de datos de prueba. (Marcha).....	38
Tabla 18. EAM obtenido tras el entrenamiento de los modelos con los datos completos sobre el conjunto de datos de prueba. (Marcha).....	39
Tabla 19. RECM medio obtenido tras el entrenamiento de los modelos con los datos completos a través de validación cruzada. (Marcha).....	39

Tabla 20. RECM obtenido tras eliminar valores extremos sobre el conjunto de datos de prueba. (Marcha)	40
Tabla 21. EAM obtenido tras eliminar valores extremos sobre el conjunto de datos de prueba. (Marcha)	41
Tabla 22. RECM obtenido tras eliminar valores extremos de los datos a través de validación cruzada. (Marcha)	41
Tabla 23. Número de variables seleccionadas para cada modelo a través de selección recursiva de variables. (Marcha)	42
Tabla 24. RECM obtenido tras la selección recursiva de variables sobre el conjunto de datos de prueba. (Marcha).....	43
Tabla 25. EAM obtenido tras la selección recursiva de variables sobre el conjunto de datos de prueba. (Marcha).....	43
Tabla 26. RECM obtenido tras la selección recursiva de variables por validación cruzada. (Marcha)	44
Tabla 27. RECM obtenido tras la selección por conveniencia de variables sobre el conjunto de datos de prueba. (Marcha)	45
Tabla 28. EAM obtenido tras la selección por conveniencia de variables sobre el conjunto de datos de prueba. (Marcha).....	45
Tabla 29. RECM obtenido tras la selección por conveniencia de variables por validación cruzada. (Marcha)	46
Tabla 30. Estadística descriptiva de las variables a predecir por el algoritmo.	49
Tabla 31. Estadística descriptiva de las variables a predecir por el algoritmo.	51
Tabla 32. Combinaciones musculares que obtienen los mejores resultados.	53

1 Resumen

La biomecánica es un área de conocimiento multidisciplinar que estudia los efectos de las fuerzas aplicadas sobre los organismos. Debido a su complejidad, requiere de la adquisición de señales de numerosos sensores, haciendo difícil su aplicación clínica.

Las técnicas de aprendizaje automático permiten entrenar modelos para la predicción de valores numéricos, haciendo posible obtener resultados de parámetros que no han sido medidos. Su implementación en el área de la biomecánica permitirá reducir el número de sensores necesarios durante los registros de las actividades, con el impacto que ello tiene a nivel económico y de consumo de tiempo.

Se ha realizado un proceso de cinco pasos: recogida de los datos, exploración y preparación de los datos, entrenamiento de los modelos, evaluación de los modelos y mejora de los modelos. Se han entrenado un total de ocho modelos para veintiuna combinaciones distintas de sensores, compuesta cada una por un sensor inercial y dos sensores de electromiografía, para predecir la actividad de cinco músculos.

Gracias a los avances en la creación de aplicaciones web resulta posible implementar estos modelos y hacerlos accesibles de forma gratuita, consiguiendo un impacto en la práctica clínica.

El resultado final de este trabajo es una aplicación web que implementa el modelo con mejor capacidad de **predicción de la actividad muscular** durante los siguientes movimientos: la marcha en superficie plana y la transición de sedestación a bipedestación. Además, la aplicación web también permitirá realizar de forma automática el **procesamiento de las señales** durante estos dos movimientos y la obtención de las variables seleccionadas.

2 Introducción

2.1 Contexto y justificación del Trabajo

La biomecánica es un área de conocimiento multidisciplinar que estudia los fenómenos naturales que ocurren en el cuerpo humano y otros organismos como consecuencia de sufrir la aplicación de fuerzas.(1)

Para ello se utilizan una gran variedad de tecnologías entre las que se incluyen: plataformas de presiones, electromiografía de superficie (EMGs), unidades de medición inercial (IMU) y fotogrametría. Con ellas se analizan de forma cuantitativa las distintas variables que caracterizan los movimientos.

En concreto, la electromiografía de superficie permite medir la actividad neuromuscular a través del uso de electrodos de superficie. Esta actividad eléctrica es la base de todos los movimientos que puede producir el cuerpo humano y es de especial interés debido a sus implicaciones en numerosas enfermedades neurológicas y patologías musculoesqueléticas como son el dolor lumbar, el dolor de hombro o el dolor de cuello. (2)

El movimiento humano tiene una gran complejidad y por ello es necesario el uso de numerosos dispositivos para medir todos sus componentes. Esto hace que para la mayoría de los clínicos (médicos, fisioterapeutas, licenciados en actividad física) sea prácticamente imposible implementar los últimos avances en su día a día.

Las dos barreras fundamentales que se encuentran son la inversión monetaria y el tiempo que consume la preparación del paciente para la realización de los registros.

Con este trabajo se buscará identificar el mejor modelo para la predicción de la actividad neuromuscular de cinco músculos durante los siguientes movimientos: la marcha en una superficie plana y en la transición de sedestación a bipedestación. El input de los modelos será la información obtenida por un IMU y dos sensores de EMGs.

Se desarrollará también una aplicación web amigable para el usuario que permitirá:

- Implementar el modelo desarrollado para la predicción de la actividad muscular durante dos movimientos: la marcha en superficie plana y la transición de sedestación a bipedestación. **Objetivo:** facilitar la implementación clínica de estas mediciones. Esto se logrará ya que será posible predecir la actividad muscular de cinco músculos a partir de la información obtenida por tres sensores exclusivamente, reduciendo el consumo de tiempo y reduciendo costes.
- Realizar el procesamiento de las señales derivadas de IMU y EMGs de forma automática durante estas dos actividades y el cálculo de las variables deseadas.

Todos los resultados obtenidos tras el uso de la aplicación web podrán ser descargados en formato CSV.

2.2 Objetivos del Trabajo

2.2.1 Objetivo general

Identificar el mejor modelo para la predicción de la actividad neuromuscular de cinco músculos durante la marcha en una superficie plana y en la transición de sedestación a bipedestación a partir de la información obtenida por un IMU y dos sensores de EMGs.

2.2.2 Objetivos específicos

- Desarrollar una aplicación amigable que permita extraer las principales características de una señal durante la marcha en superficie plana y la transición de sedestación a bipedestación.
- Identificar el algoritmo de aprendizaje automático más preciso en la predicción de la actividad neuromuscular durante la marcha en superficie plana y la transición de sedestación a bipedestación.
- Implementar en dicha aplicación el algoritmo para predecir la actividad muscular durante la marcha en superficie plana y la transición de sedestación a bipedestación.
- Presentar las predicciones en formato numérico y descargable.

2.3 Enfoque y método seguido

El proyecto está formado por dos fases:

1. Fase 1: Identificación del algoritmo de aprendizaje automático más preciso en la predicción de la actividad neuromuscular durante la marcha en superficie plana y en la transición de sedestación a bipedestación a partir de la información de un IMU y dos sensores de EMGs.
2. Fase 2: Desarrollo de una aplicación web para implementar el algoritmo y predecir la actividad neuromuscular durante la marcha en superficie plana y en la transición de sedestación a bipedestación. También permitirá extraer las principales características de una señal durante la marcha en superficie plana y la transición de sedestación a bipedestación

2.3.1 Fase 1.

Para la identificación del algoritmo de aprendizaje automático se van a seguir los pasos definidos por Brett Lantz (3). Existen otras estrategias como la desarrollada por Aurélien Gerón (4) que divide el proceso en etapas más

pequeñas, pero se ha escogido la metodología desarrollada por Brett Lantz (3) ya que es más genérica. A continuación, se enumeran los pasos:

Paso 1. Recogida de datos.

Para este caso de estudio se van a utilizar los datos del artículo titulado “Benchmark Datasets for Bilateral Lower-Limb Neuromechanical Signals from Wearable Sensors during Unassisted Locomotion in Able-Bodied Individuals” (5).

El conjunto de datos contiene:

- Registros electromiográficos de siete músculos por pierna: vasto lateral del cuádriceps (VL), recto femoral (RF), bíceps femoral (BF), semitendinoso (ST), tibial anterior (TA), gemelo interno (MG) y sóleo (SOL)
- Registros de dos sensores inerciales (tibia y fémur) por extremidad y uno a nivel de la cintura.
- 25 repeticiones por sujeto de un circuito que constaba de siete actividades (sedestación, bipedestación, caminar en plano, ascenso y descenso de escaleras, ascenso y descenso de una rampa).
- Registro del rango articular de la rodilla y el tobillo de ambas extremidades inferiores a través de un goniómetro.

De estos datos se han utilizado para este estudio aquellos derivados de los sensores inerciales de la tibia y la cintura y de los sensores de electromiografía de superficie exclusivamente.

Paso 2. Explorar y preparar los datos

Al tratarse de señales en crudo, es necesario procesarlas. El procesamiento consiste en la aplicación de un filtrado siguiendo los estándares del área de estudio (5,6). Una vez procesada la señal se identifican y extraen los ciclos de cada actividad.

A continuación, se detalla el concepto de ciclo dependiendo de la actividad:

- Marcha sobre superficie plana: la marcha humana es una actividad cíclica. En el estudio del movimiento humano se entiende como un ciclo de la marcha el movimiento que realiza la extremidad al dar un paso. Dicho de otra manera, es todo el movimiento realizado desde un apoyo de talón hasta el siguiente con la misma extremidad.
- Transición de sedestación a bipedestación: esta transición se produce cuando uno se levanta de un asiento. En concreto, el comienzo del ciclo lo marca la inclinación anterior del tronco y el final la posición completamente erguida del cuerpo.

De cada ciclo se extraen las variables a introducir en el modelo. Tras la extracción de todas las variables, se exploran los datos para conocer su

distribución, evaluar la existencia de datos anómalos y gestionar los datos perdidos.

Paso 3. Entrenar los modelos

Se llevará a cabo el entrenamiento de los siguientes ocho modelos:

- Regresión Ridge
- Regresión Lasso
- Regresión Elastic Net
- Árbol de decisión
- Árboles de decisión múltiples
- Árboles extremadamente aleatorios
- K-vecinos más próximos
- Perceptrón Multicapa

Ya que el objetivo del trabajo es encontrar el mejor modelo para la predicción de la actividad muscular de cinco músculos durante la marcha en superficie plana y la transición de sedestación a bipedestación a partir de la información obtenida por **un IMU y dos sensores de EMG**, es necesario definir todas las combinaciones posibles entre ellos.

Los elementos a combinar son:

- Marcha en superficie plana:
 - IMU ubicado en la tibia
 - Siete músculos registrados en el conjunto de datos
- Transición de sedestación a bipedestación:
 - IMU ubicado en la cintura
 - Siete músculos registrados en el conjunto de datos

En la tabla 1, se muestran las veintiuna combinaciones de valores de entrada sobre las que han sido entrenados los ocho modelos en cada movimiento.

IMU, TA, MG	IMU, TA, SOL	IMU, TA, BF	IMU, TA, ST	IMU, TA, VL	IMU, TA, RF
IMU, MG, SOL	IMU, MG, BF	IMU, MG, ST	IMU, MG, VL	IMU, MG, RF	IMU, SOL, BF
IMU, SOL, ST	IMU, SOL, VL	IMU, SOL, RF	IMU, BF, ST	IMU, BF, VL	IMU, BF, RF
IMU, ST, VL	IMU, ST, RF	IMU, VL, RF			

Tabla 1. Combinaciones evaluadas en todos los modelos. IMU en marcha en superficie plana se encuentra en la tibia. IMU en la transición de sedestación a bipedestación se encuentra en la cintura.

El resultado del entrenamiento de estos modelos será la predicción de la actividad muscular de los cinco músculos que no han sido introducidos como valores de entrada en el modelo, cuantificada por la media cuadrática (RMS), durante cada ciclo. A continuación, se expone un ejemplo ilustrativo del entrenamiento de los modelos:

- Valores de entrada: IMU, TA, MG
- Valores de salida: RMS de los cinco músculos no introducidos como input, en este ejemplo: VL, RF, BF, ST, SOL.

Como valor de referencia en este trabajo se utilizan los resultados obtenidos de los modelos sin técnicas de eliminación de valores extremos e incluyendo todas las variables extraídas de cada señal.

Durante el entrenamiento de los modelos se ha utilizado la función “RandomizedSearchCV” (7) para seleccionar los hiperparámetros óptimos para cada modelo en cada conjunto de datos. La contrapartida de la utilización de este método es que es computacionalmente más costoso.

Paso 4. Evaluar el rendimiento de los modelos

Las métricas de precisión que se utilizan son el error absoluto medio (EAM) y la raíz del error cuadrático medio (RECM) (8). A través de ellas se evalúa la precisión de los modelos en la predicción de la actividad muscular de los cinco músculos no introducidos como valores de entrada en los modelos.

A continuación, se detallan las características de cada una de las métricas de precisión:

- Error absoluto medio (EAM):(9) se trata de la media de los valores absolutos de cada error de predicción individual sobre todas las instancias en el conjunto de datos de prueba. Cada error de predicción es la diferencia entre el valor real y el predicho para cada instancia. Su fórmula es la siguiente:

$$EAM = \frac{\sum_{i=1}^n \text{abs}(y_i - \lambda(x_i))}{n}$$

- Raíz del error cuadrático medio (RECM):(10) es la raíz cuadrada de la media de todos los errores al cuadrado.

$$RECM = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$$

Se han seleccionado estas dos métricas ya que son las más utilizadas para evaluar los problemas de regresión.

Paso 5. Mejora de los modelos

En este paso se realizan las modificaciones necesarias en la implementación de los modelos para obtener mejores resultados.

Se lleva a cabo en tres pasos:

1. Eliminación de valores extremos: se eliminan del conjunto de datos aquellos valores que se identifiquen como valores extremos utilizando técnicas basadas en modelos.

2. Eliminación recursiva de variables: se utilizan técnicas de eliminación recursiva de variables en cada combinación de datos para explorar si los modelos mejoran disminuyendo el número de variables.
3. Eliminación de variables: se eliminarán las variables derivadas del IMU que no se utilizan para la detección de los ciclos con el objetivo de simplificar la interacción con la aplicación web.

El mejor modelo, seleccionado de acuerdo con las métricas del paso 4, se implementa en la aplicación web.

2.3.2 Fase 2.

La aplicación web se implementará a través del paquete 'Streamlit'. Los motivos principales para la elección de 'Streamlit' como método para la creación de la aplicación web son su fácil implementación, la posibilidad de desarrollar toda la aplicación en lenguaje Python y alojar la aplicación en su servidor de forma gratuita.

Existen un gran número de alternativas para implementar una aplicación web. Entre ellas se encuentran Django (<https://www.djangoproject.com>) y Flask (<https://flask.palletsprojects.com>), frameworks de Python, pero requieren de mayor elaboración y de la búsqueda de un servidor para alojar la aplicación. El resto de las posibilidades para la creación de la aplicación web han sido descartadas debido a que requieren del aprendizaje y utilización de otros lenguajes de programación (JavaScript, HTML).

El resultado final de la aplicación está compuesto por:

- Página principal: en ella se introduce la utilidad de la aplicación, su origen y se presenta un menú desplegable que permite acceder a la página de cada funcionalidad.
- Página 1. "Extracción de variables durante la transición de sedestación a bipedestación": en esta página el usuario puede introducir su matriz de datos, indicar a qué tipo de señal corresponde cada columna y seleccionar las variables que desea obtener. Los resultados de las operaciones se muestran en una tabla en pantalla y existe la posibilidad de descargar los datos en formato CSV.
- Página 2: "Predicción de la actividad neuromuscular durante la transición de sedestación a bipedestación": en esta página el usuario puede introducir su matriz de datos, indicar qué columna pertenece a las características con las que ha sido entrenado el modelo y obtener la predicción de la actividad muscular de cinco músculos.
- Página 3: "Extracción de variables durante la marcha en superficie plana": en esta página el usuario puede introducir su matriz de datos, indicar a qué tipo de señal corresponde cada columna y seleccionar las variables que desea obtener. Los resultados de las operaciones se muestran en una tabla en pantalla y existe la posibilidad de descargar los datos en formato CSV.

- Página 4: “Predicción de la actividad neuromuscular durante la marcha en superficie plana”: en esta página el usuario puede introducir su matriz de datos, indicar qué columna pertenece a las características con las que ha sido entrenado el modelo y obtener la predicción de la actividad muscular de cinco músculos.

2.4 Planificación del Trabajo

Descripción de tareas, fechas de inicio y fin y entregable en cada etapa:

Nombre	Fecha de inicio	Fecha de fin	Hito
Definición de los contenidos del trabajo	17-febrero	01-marzo	Entrega de PEC0
Plan de trabajo	02-marzo	16-marzo	Entrega de PEC1
Fase 1 del desarrollo de trabajo	17-marzo	19-abril	Entrega de PEC2
Revisión bibliográfica sobre predicción de EMGs (algoritmos y variables)	17-marzo	21-marzo	
Exploración de librerías de Python para aprendizaje automático	22-marzo	26-marzo	
Procesamiento de las señales de EMGs	26-marzo	28-marzo	
Procesamiento de las señales IMU.	26-marzo	28-marzo	
Desarrollo del algoritmo de detección de ciclos en la marcha.	29-marzo	4-abril	
Desarrollo del algoritmo de detección de la	29-marzo	4-abril	

fase de transición sedestación- bipedestación.			
Extracción de variables	5-abril	11-abril	
Selección de modelos	12-abril	19-abril	
Elaboración de la memoria de la primera fase del trabajo	12-abril	19-abril	
Fase 2 del desarrollo del trabajo	20-abril	17-mayo	Entrega de PEC3
Entrenamiento de los modelos	20-abril	25-abril	
Evaluación de los modelos	26-abril	30-abril	
Mejora de los modelos	30-abril	08-mayo	
Aprendizaje de desarrollo de aplicaciones empleando “Streamlit”	05-mayo	08-mayo	
Definir esquema de la aplicación web	09-mayo	10-mayo	
Elaboración de la aplicación web	11-mayo	15-mayo	
Deploy web app utilizando “Streamlit sharing”	16-mayo	17-mayo	
Pruebas de uso de la aplicación	16-mayo	17-mayo	
Elaboración de la memoria de la segunda fase del trabajo	10-mayo	17-mayo	
Cierre de la memoria	18-mayo	08-junio	Entrega de PEC4

Escribir capítulos extra de la memoria	18-mayo	08-junio	
Elaboración de la presentación	10-junio	13-junio	Entrega de PEC5a
Defensa Pública	16-junio	23-junio	Entrega de PEC5b

2.5 Breve resumen de contribuciones y productos obtenidos

2.5.1 Plan de trabajo

Documento en el que se detallan los objetivos, tareas y tiempos de elaboración del trabajo de fin de máster. Se incluye también en el documento un análisis de riesgos y sus posibles soluciones.

2.5.2 Memoria

Se entregará un documento en formato PDF con la descripción del desarrollo del trabajo de fin de máster.

2.5.3 Código fuente

Todo el código utilizado para la realización del trabajo estará disponible en el siguiente repositorio público de Github:

https://github.com/cjove/TFM_CarlosJove_UOC.

En el Anexo 1 se detalla el contenido del repositorio.

2.5.4 Aplicación Web

La aplicación web interactiva creada a través de Streamlit que permite realizar la extracción de variables de señales en crudo y su introducción en el modelo para realizar la predicción de la actividad neuromuscular.

El enlace para acceder a la aplicación web es:

https://share.streamlit.io/cjove/tfm_web_app/main/app.py.

El repositorio de datos en el que se encuentra todo el código fuente de la aplicación es: https://github.com/cjove/tfm_web_app.

En el Anexo 2 se detalla el contenido del repositorio.

2.5.5 Presentación Virtual

Será realizada a través de los medios facilitados por la Universitat Oberta de Catalunya. Se realizará una presentación de PowerPoint como herramienta de soporte a la comunicación oral.

2.6 Breve descripción de los otros capítulos de la memoria

- Estado del arte: en este capítulo se explica el conocimiento existente en el área de la biomecánica y el aprendizaje automático, los problemas que existen en la actualidad a la hora de recoger datos objetivos en el área clínica y la propuesta de este trabajo para mejorar dichas problemáticas.
- Metodología: se desarrollan en detalle todos los pasos llevados a cabo para conseguir los objetivos propuestos.
- Resultados: se describen los resultados obtenidos tras la aplicación de la metodología descrita en el paso anterior.
- Discusión: en ella se desarrolla el análisis de los resultados obtenidos en este trabajo.
- Conclusiones: último capítulo de la memoria en el que se describen los puntos más importantes de este trabajo se reflexiona a cerca de la consecución de los objetivos planteados inicialmente y se plantean líneas de futuro.

3 Estado del arte

El número de publicaciones en los últimos años que aplican técnicas de aprendizaje automático al área de la biomecánica pone de relieve el interés que despierta esta temática. En concreto, si se realiza una búsqueda en Google Scholar con los términos “biomechanics AND machine learning” se obtienen un total de aproximadamente 16.100 resultados desde 2017 en la fecha de entrega de este trabajo.

Como ponen de relieve Camargo et al. (11) las variables relacionadas con la biomecánica tienen dos aplicaciones principales:

1. Aportar información adicional a nivel clínico
2. Mejorar el control de dispositivos de rehabilitación tales como prótesis, órtesis y exoesqueletos.

Es el último punto el que mayor interés ha despertado en la comunidad científica debido a su potencial para mejorar la calidad de vida de aquellos individuos que requieren del uso de estos elementos. Estos mismos autores (11) han utilizado estrategias de aprendizaje automático para estimar variables del entorno críticas para conseguir un comportamiento fluido de una prótesis como es la altura de distintos escalones. Khairuddin et al. (12) utilizaron señales de electromiografía para clasificar la intención de movimiento del miembro superior en el movimiento de la flexión de codo. Utilizaron señales electromiográficas ya que muestran cambios antes de la producción del movimiento, haciendo de esta señal una herramienta muy interesante para lograr que elementos como las prótesis tengan mayor fluidez y facilidad de uso. En la misma línea de desarrollo Little et al. (13) utilizaron la información proveniente de distintos sensores portátiles para predecir la trayectoria del codo.

A nivel clínico, ya en 2004, Schollhorn et al. (14) ponían de relieve todas las aplicaciones prácticas que las redes neuronales tenían en la biomecánica clínica y citando al autor “el cuerpo de literatura de las redes neuronales ha crecido de una forma intratablemente vasta en los últimos quince años”. En ella se recogen la aplicación de redes neuronales a la clasificación de la marcha, la carrera, diferentes tareas motoras y de la actividad neuromuscular para, entre otros, el diagnóstico de patologías.

Desde entonces la aplicación de estos métodos estadísticos ha seguido avanzando en sus aplicaciones. Recientemente, Raval et al. (15) han estudiado la capacidad de predecir el ritmo de progresión individual de la enfermedad de Parkinson utilizando distintas medidas clínicas y biomecánicas de la marcha y de la estabilidad postural con unos resultados muy positivos. Por otro lado, en la aplicación de la biomecánica al área de la traumatología, Brisson et al. (16) han desarrollado una ecuación para predecir la fuerza de contacto de la porción medial de la tibia con el fémur (MCF) durante la marcha y poder con ella

determinar la relación entre la MCF y la pérdida de cartílago en personas con osteoartritis. Entre otros autores interesados en la predicción de las fuerzas se encuentran Sharma et al (17) que utilizan sensores portátiles para medir la fuerza de reacción del suelo de forma indirecta.

A nivel deportivo también se han utilizado técnicas de aprendizaje automático sobre distintas variables biomecánicas con dos objetivos principales: la mejora del rendimiento y la prevención de lesiones. Por ejemplo, Sarvestan et al. (18) utilizaron técnicas de aprendizaje automático para analizar el patrón de coordinación en los remates de voleibol y concluyeron que los patrones de los miembros superiores juegan un papel fundamental en la realización del gesto con éxito. En cuanto a su aplicación en la prevención de lesiones, Seshadri et al.(19) en su revisión narrativa desarrollan la aplicación de sensores portátiles y técnicas de aprendizaje automático para monitorizar la carga interna y externa del entrenamiento, el nivel de hidratación, descanso y estado para la vuelta al juego del deportista. Todos ellos, junto con la fatiga(20), factores cruciales para disminuir el riesgo de lesión, punto crítico tanto para deportistas profesionales como amateurs.

Muchos de estos estudios han sido posibles gracias a los avances tecnológicos que han permitido el desarrollo de sensores portátiles, más conocidos por su nombre en inglés 'wearables'. (19) Estos sensores son mucho más económicos y fáciles de usar que los instrumentos de medición convencionales.

Sin embargo, los avances en la aplicación de técnicas de aprendizaje automático y en los desarrollos tecnológicos, no se han visto trasladados a la práctica clínica habitual debido a que se necesita de mucho tiempo para una instrumentación completa y la recolección de los datos. Como ejemplo, Hu et al.(5) reflejan en su artículo que la instrumentación de los dos miembros inferiores con electromiografía de superficie, sensores inerciales y goniometría junto con el registro de 25 repeticiones de un circuito les llevó hasta dos horas. En la Ilustración 1, imagen del artículo que publicó el conjunto de datos utilizados en este trabajo, se puede apreciar la complejidad del sistema que es necesario utilizar para registrar todas las variables cinemáticas y electromiográficas. Si a esa duración se le añade el posterior tratamiento de los datos y la falta de formación en las ciencias de la salud y el ejercicio en esa área, se vuelve un esfuerzo inasumible en la mayoría de los casos.

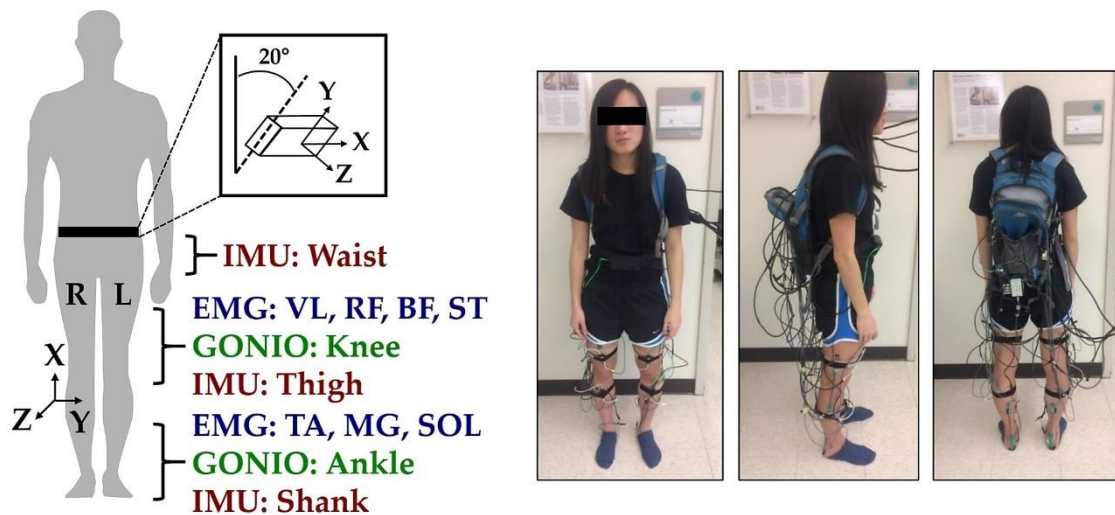


Ilustración 1. Imagen del artículo de Hu et al.(5) en la que se aprecia la complejidad de la instrumentación.

Para facilitar el tratamiento de los datos, se han desarrollado multitud de algoritmos de detección de las fases de la marcha (11,21), de la transición de sedestación a bipedestación (22–24), la carrera (25) y de distintas actividades (11,26,27). El objetivo de estos es identificar correctamente el movimiento que se está realizando y en que fragmento de la señal se produce. Su creación e implementación permite dar sentido espacial y temporal a los datos y realizar análisis mucho más precisos de manera más rápida, por lo que se vuelven una herramienta fundamental para conseguir generalizar el uso de dispositivos de medición entre las ciencias de la salud y el ejercicio.

Tomando como ejemplo los trabajos realizados por Rittenhouse et al.(28) y Heller et al. (29) en los que se utilizan datos cinemáticos para la predicción de la actividad neuromuscular se busca en este trabajo identificar el mejor modelo para la predicción de la actividad neuromuscular de cinco músculos durante la marcha en una superficie plana y en la transición de sedestación a bipedestación a partir de la información obtenida por un IMU y dos EMGs.

Una vez desarrollada esa fase del trabajo, para facilitar la implementación en la práctica clínica, tomando como modelo de referencia el trabajo realizado por mDurance Solutions (6) se creará una aplicación amigable que permita automatizar el procesamiento de las señales y la predicción de la actividad muscular.

La doble funcionalidad de la aplicación web busca cubrir las siguientes necesidades:

- Funcionalidad 1. Extracción de las características de una señal: la mayoría de los profesionales de la salud y el ejercicio carecen de nociones de programación. Sin embargo, toda investigación e interpretación requiere de la adquisición de características de la señal, por lo que se considera de utilidad que una plataforma amigable les permita extraer de

forma sencilla dicha información en cada actividad analizada en este trabajo.

- Funcionalidad 2. Implementación del algoritmo de aprendizaje automático: la reducción de los elementos necesarios para obtener la actividad neuromuscular tiene como finalidad posibilitar su uso al reducir costes y disminuir el tiempo necesario para llevarlo a cabo.

Con ello se busca acercar los últimos desarrollos en la aplicación de técnicas de aprendizaje automático en el área de la biomecánica a la práctica clínica y romper algunas de las barreras que impiden su implementación.

4 Definición de aprendizaje automático

“El aprendizaje automático es el área de estudio que proporciona a los ordenadores la habilidad de aprender sin ser programados de forma explícita.” Arthur Samuel, 1959.

Para ello el aprendizaje automático utiliza un número de muestras elevado, llamado conjunto de entrenamiento para afinar los parámetros de un modelo adaptativo. El resultado de la aplicación de algoritmos de aprendizaje automático puede ser expresado como una función $y(x)$ que toma nuevas muestras como valores de entrada y que genera un vector de salida codificado de la misma manera que el vector diana utilizado para el aprendizaje. Esta función se determina durante la fase de aprendizaje o entrenamiento y posteriormente se utiliza para clasificar/predecir correctamente nuevos valores de entrada. Esto se conoce como generalización. (30)

Existen distintos tipos de sistemas de aprendizaje automático que pueden dividirse en tres grandes grupos siguiendo los siguientes criterios:(4)

- Modo de entrenamiento: según reciban supervisión o no de un humano, encontramos los siguientes grupos:
 - Aprendizaje supervisado
 - Aprendizaje semisupervisado
 - Aprendizaje no supervisado
 - Aprendizaje por refuerzo
- Capacidad de aprendizaje: según sean o no capaces de aprender de nuevos datos se encuentran los siguientes grupos:
 - Aprendizaje online
 - Aprendizaje por lotes
- Modo de trabajo: según trabajen por comparación o detectando patrones se encuentran los siguientes grupos:
 - Aprendizaje basado en instancias
 - Aprendizaje basado en modelos

En la elaboración de este trabajo se han utilizado algoritmos que utilizan aprendizaje supervisado. La característica principal de este método de aprendizaje es que el conjunto de datos que se utiliza para entrenar el algoritmo

incluye las soluciones deseadas. Este tipo de algoritmos pueden ser utilizados para abordar tanto problemas de clasificación como de regresión, siendo este último el planteado en este trabajo.

Dentro de los problemas de regresión nos encontramos dos categorías:

- Regresión única
- Regresión múltiple.

Los problemas de regresión única tienen como objetivo predecir un único valor numérico y los problemas de regresión múltiple tienen como objetivo predecir de forma simultánea varios valores numéricos. En el caso de este trabajo se aborda un problema de **regresión múltiple**.

5 Metodología

Siguiendo la implementación descrita en el apartado 2.3 de esta memoria se desarrollan en detalle los puntos llevados a cabo en cada uno de ellos:

5.1 Paso 1. Recogida de datos.

Los datos utilizados para este estudio provienen del conjunto de datos publicado por Hu et al. (5) disponible en el siguiente enlace: https://figshare.com/articles/dataset/Benchmark_datasets_for_bilateral_lower_limb_neuromechanical_signals_from_wearable_sensors_during_unassisted_locomotion_in_able-bodied_individuals/5362627?file=9226057.

En el conjunto de datos se recogen también valores de goniometría del tobillo y de la rodilla, sin embargo, se decidió no incluir dichos elementos en la elaboración de este estudio por los siguientes motivos:

1. Los autores reflejan en el artículo que hubo algunos problemas en la adquisición de la señal de estos sensores.
2. La variable que registran estos sensores es el rango de movimiento articular, que puede ser obtenido o estimado a partir de las señales obtenidas por los IMU.

Paso 2. Explorar y preparar los datos

Los datos provienen de la medición de 10 sujetos sanos con las características descritas en la ilustración 2. Es de vital importancia conocer las características de los sujetos, ya que nos indican sobre qué tipo de sujetos se ha entrenado el modelo. En este caso supone una limitación ya que se ha registrado sólo a sujetos de edades comprendidas entre los 23 y 29 años y no sabemos si el modelo será capaz de predecir correctamente sobre otros rangos de edad.

Subject	Age (years)	Height (cm)	Weight (kg)
AB156	26	193	77
AB185	28	181	75
AB186	24	163	54
AB188	25	185	87
AB189	24	178	66
AB190	23	160	54
AB191	26	163	54
AB192	23	170	75
AB193	27	185	95
AB194	29	160	61

Ilustración 2. Características de los sujetos registrados en el estudio

Del conjunto de sensores utilizados en los datos aportados por Hu et al. (5) se utilizan en este trabajo el IMU de la tibia, el IMU de la cintura y la actividad muscular registrada con sensores de EMGs. Una vez seleccionados los sensores de interés se llevaron a cabo los siguientes pasos:

- Procesamiento de las señales del conjunto de datos inicial
- Elaboración del algoritmo de detección de ciclos de la marcha
- Elaboración del algoritmo de detección de la transición de sedestación a bipedestación
- Extracción de las variables de las señales procesadas

A continuación, se desarrolla cada punto de forma detallada.

5.1.1 Procesamiento de las señales del conjunto de datos

5.1.1.1 Electromiografía de superficie

La electromiografía de superficie es una técnica no invasiva que permite analizar la actividad eléctrica que producen los músculos. Al igual que cualquier otra señal, los datos recogidos por los sensores son una combinación de señal y ruido. Esta característica intrínseca a cualquier señal recogida en un entorno real hace necesario el procesamiento de ésta para poder eliminar lo máximo posible el ruido, ya que la presencia de éste podría llevar a la extracción de conclusiones incorrectas.(31)

En el caso de la señal de electromiografía, una de las técnicas más aplicadas para limpiar la señal es la aplicación de filtros de paso banda. Esto se debe a que se ha descrito que la actividad muscular produce señales con frecuencias

comprendidas entre los 20 y 450 Hz. La función de un filtro paso banda es atenuar aquellas frecuencias que se encuentran fuera del rango fijado en el filtro, reduciendo de forma muy efectiva la presencia de ruido.(6) En la Ilustración 3 se muestra la forma característica de un filtro paso banda.

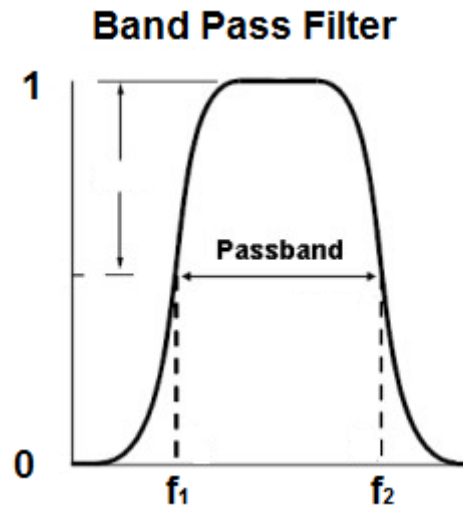


Ilustración 3. Representación gráfica de un filtro paso banda. $f_1 = 20\text{Hz}$, $f_2 = 450\text{ Hz}$.

Según el estudio que generó el conjunto de datos inicial(5) la frecuencia de muestreo del sistema de electromiografía de superficie fue de 1000 Hz, tuvo una amplificación de 1000x y fue filtrado a nivel de hardware con un filtro de paso banda entre 20 y 450 Hz. Sobre estas señales se aplicó durante el desarrollo de este trabajo un filtro de paso banda entre 20-450 Hz utilizando un filtro de Butterworth de orden 4. (6) En la Ilustración 4 se observa el efecto que tiene la aplicación de este filtro sobre la señal electromiográfica.

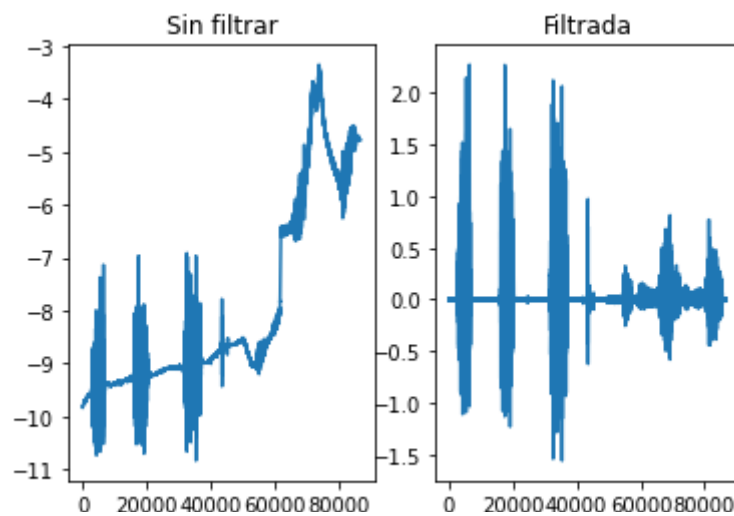


Ilustración 4. Efecto del filtro en una señal electromiográfica

5.1.1.2 Sensores inerciales

Los sensores inerciales son dispositivos electrónicos que miden y reportan las aceleraciones, velocidades angulares y, a veces, la orientación de los cuerpos utilizando una combinación de acelerómetros, giróscopos y, a veces, magnetómetros. Se han utilizado para multitud de finalidades, incluyendo la conducción de aeronaves, satélites y la biomecánica. (32)

En el conjunto de datos utilizados en este trabajo se utilizaron sensores inerciales con acelerómetros y giróscopos con 3 grados de libertad cada uno. Esto permite que se hayan recogido las aceleraciones y velocidades angulares en los tres planos del espacio, obteniendo así una descripción detallada del movimiento registrado. (5)

Igual que la señal de electromiografía, las señales recogidas por los sensores inerciales incluyen ruido. Este ruido suele estar compuesto por vibraciones y oscilaciones del dispositivo que no tienen relevancia física y que añaden una gran variabilidad a la señal. Por ello se aplican filtrados que reducen o eliminan esas perturbaciones, suavizando la señal y permitiendo una mejor detección de eventos significativos.

Sobre el conjunto de datos iniciales se aplicó un filtro de paso banda bajo de 25 Hz utilizando un filtro de Butterworth de orden 6 replicando el realizado por el estudio de origen (5)

5.1.2 Elaboración del algoritmo de detección de ciclos de la marcha

La elaboración de este algoritmo ha sido posible ya que las señales en crudo de cada registro realizado contienen una etiqueta identificando a que actividad de todas las realizadas corresponde cada dato de la señal (columna: Mode). Esto garantiza que los segmentos de la señal utilizados para posteriormente identificar los pasos pertenecen al tramo de caminar por una superficie plana.

La señal utilizada pertenece al componente Y del giróscopo colocado a nivel de la tibia (columnas: Right_Shank_Gy y Left_Shank_Gy). Siguiendo el procedimiento descrito en el artículo (5) , esta señal se filtra de nuevo, pero aplicando un filtro paso banda bajo de 6 Hz utilizando un filtro Butterworth de orden 1. Este filtro se aplica ya que es necesario un mayor suavizado de la señal. La identificación de los ciclos de marcha se llevó a cabo con un método muy similar al utilizado por Maqboor et al. (33) De forma general la marcha se puede dividir en dos fases, fase de apoyo y fase de balanceo, que se pueden subdividir en otras cuatro fases.(34) Siempre se toma como referencia el apoyo de talón como evento inicial y final de un paso. Al tratarse de un movimiento cíclico se pueden identificar distintos eventos de estas fases utilizando sensores inerciales. En este caso el algoritmo identifica el mayor pico de velocidad angular como la fase de balanceo y tomando ese punto como valor de referencia se selecciona el siguiente pico que representa el apoyo de talón y el segundo pico anterior que identifica el apoyo de talón anterior. A través de este método se identifica cada paso. Todos los parámetros utilizados en este algoritmo han sido obtenidos empíricamente y se verificó visualmente la correcta identificación.

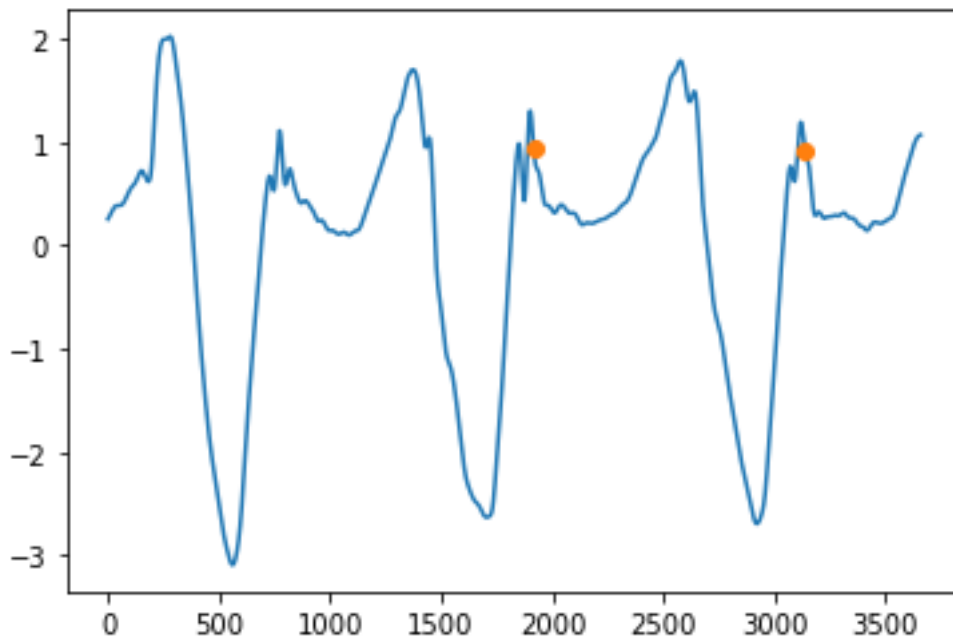


Ilustración 5. Resultado de la aplicación del algoritmo de detección de pasos

5.1.3 Elaboración del algoritmo de detección de ciclos de la transición de sedestación a bipedestación

Existen multitud de algoritmos en la literatura para detectar tanto la transición de sedestación a bipedestación como a la inversa en población sana(24) y patológica. (22,23)

Tomando dichos estudios como referencia, se seleccionó el componente X del giróscopo ubicado en la cintura, ya que la transición se inicia con un movimiento anterior del tronco y termina con la verticalidad de éste.

El filtrado que se realizó a esta señal es idéntico al realizado para el algoritmo desarrollado en el punto anterior.

El algoritmo se descompone en tres funciones:

- Identificación de los picos máximos y mínimos de la señal que son las velocidades máximas que se producen durante la transición.
- Identificación de los cambios de velocidad de la señal a través de un método de umbral (33) que permite encontrar el inicio y final del movimiento.
- Utilizando los datos anteriores, se seleccionan sólo el punto anterior al primer valor máximo encontrado y el punto posterior al primer valor mínimo encontrado.

La limitación principal de este abordaje es que sólo identificará correctamente la transición de sedestación a bipedestación si es el primer movimiento que

se produce y si no existen otros movimientos con valores más elevados en el componente X del giróscopo ubicado en la cintura.

Todos los parámetros utilizados en este algoritmo han sido obtenidos empíricamente y se verificó visualmente la correcta identificación.

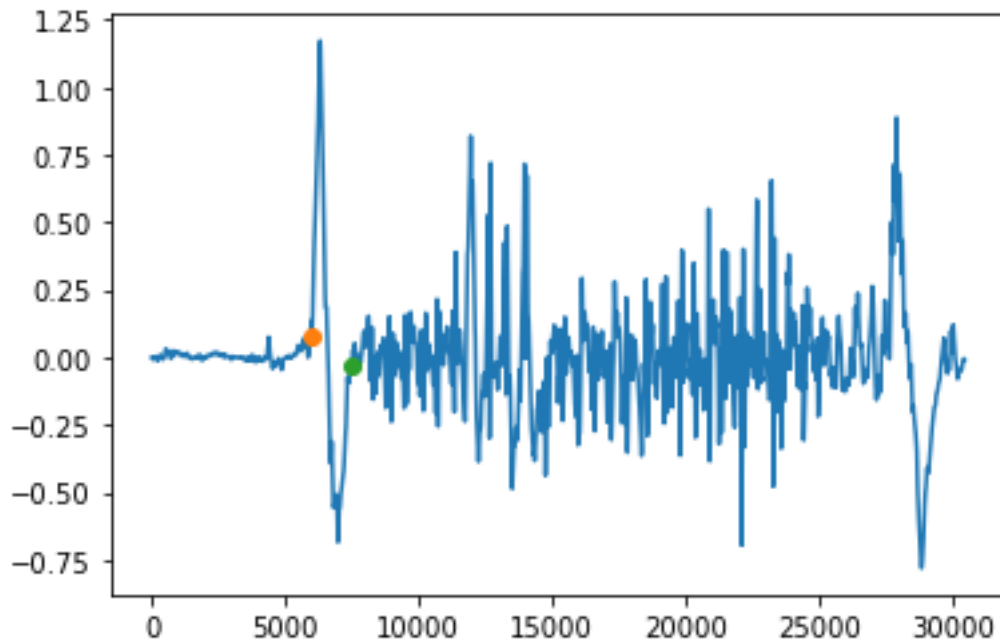


Ilustración 6. Resultado de la aplicación del algoritmo de detección de la transición de sedestación a bipedestación

5.1.4 Extracción de las variables de las señales procesadas

Al estar trabajando con señales provenientes de dos tipos de sensores (electromiografía de superficie y sensores inerciales) es necesario diferenciar las variables extraídas de cada una de ellas. Todas las variables descritas a continuación fueron extraídas de cada paso y de cada transición de sedestación a bipedestación respectivamente.

Las variables extraídas de todos los componentes del **sensor inercial** son las extraídas en el artículo de Pham et al (5):

- Valor mínimo (min): representa el valor mínimo de la señal de cada ciclo.
- Valor máximo (max): representa el valor máximo de la señal de cada ciclo.
- Desviación estándar (sd): representa la variabilidad de la señal de cada ciclo.
- Valor inicial (init): representa el valor del primer punto de la señal de cada ciclo.
- Valor final (final): representa el valor del último punto de la señal de cada ciclo.

Todas las señales eléctricas pueden ser analizadas desde dos prismas distintos, el **dominio de la frecuencia** y el **dominio del tiempo**. La utilización de variables de ambos dominios permite caracterizar de una forma más completa la señal de electromiografía:

- Espectro del tiempo (5): todas estas variables recogen la evolución de la señal en el tiempo.
 - Valor medio absoluto (MAV): representa el área debajo de la curva de la señal de electromiografía una vez que ha sido rectificada. Es un valor utilizado para representar la amplitud de una señal en un periodo de tiempo. (35)
 - Longitud de la forma de onda (WL): es una medida de la complejidad de la señal en un periodo de tiempo.(35)
 - Cruces por cero (ZC): representa el ritmo al que una señal cambia de signo en un periodo de tiempo.(36)
 - Cambio del signo de la pendiente (SSC): representa el número de veces que una señal cambia el signo de su pendiente en un periodo de tiempo.(36)
 - Integral cuadrada simple (SSI): expresa la energía de la señal en un periodo de tiempo. (35)
 - Varianza (VAR): expresa la potencia de la señal en un periodo de tiempo. (35)
 - Media cuadrática (RMS): es la media cuadrática de la señal y se utiliza para representar la amplitud de una señal en un periodo de tiempo. (35)
 - Amplitud de Willison (WA): describe el número de veces que la amplitud de la señal excede un determinado umbral. La literatura describe el uso de un umbral entre 50 y 100 mV. En este trabajo se ha utilizado un umbral de 50mV.(35)
- Espectro de frecuencia (37): todas estas variables muestran las componentes de la señal según la frecuencia en la que oscilan dentro de un rango determinado.
 - Frecuencia mediana (MDF): recoge la frecuencia que divide el espectro de frecuencia en dos partes iguales.(38)
 - Frecuencia media (MNF): recoge la frecuencia media que existe en el espectro de frecuencia.(38)
 - Entropía de Shannon (SE): recoge la incertidumbre de la señal.(39)
 - Entropía espectral (SPE): cuantifica la complejidad del espectro de una señal.(40)
 - Entropía de valor de descomposición único (SVDE): es un indicador del número de vectores propios necesarios para una correcta explicación de la señal. (39)

5.2 Entrenamiento de los modelos

La finalidad de este proyecto es predecir valores numéricos por lo que los modelos que se han tenido en consideración son aquellos capaces de resolver problemas de regresión.

No todos los algoritmos de regresión pueden abordar problemas de regresión múltiple por defecto. Este ha sido el criterio fundamental a la hora de determinar los algoritmos utilizados para la realización de este trabajo y que se detallan en el siguiente apartado. (13,28,41)

5.2.1 Modelos seleccionados

A continuación, se enumeran y describen los modelos utilizados en este trabajo:

1. Regresión Ridge: este algoritmo también es conocido como Regularización de Tikhonov y es una versión regularizada de la regresión lineal. Ello obliga al modelo no solo a aprender para predecir los datos sino a mantener los pesos del modelo lo más pequeños posibles. En concreto, en este trabajo se ha utilizado una solución de circuito cerrada, implementada en el paquete `sklearn.linear_model` con la función `Ridge` que usa la variante de André Louis Cholesky.(7)
2. Regresión Lasso: las siglas “lasso” en inglés significan “Least Absolute Shrinkage and Selection Operator Regression” y es otra versión regularizada de la regresión lineal. En este caso, utiliza la norma l_1 del vector de pesos en lugar de la mitad del cuadrado de la norma l_2 . Una de las características principales de esta regresión es que tiende a eliminar los pesos de las variables menos importantes. En este caso se utilizó en el trabajo la función `Lasso` del paquete `sklearn.linear_model`.(7)
3. Regresión Elastic net: se trata de un algoritmo que mezcla los términos de regularización de los algoritmos de Ridge y Lasso. Esta mezcla puede ser controlada a través de la ratio r . En este trabajo se utilizó la función `ElasticNet` del paquete `sklearn.linear_model`.(7)
4. Árbol de decisión: `scikit learn` utiliza el algoritmo “Classification and Regression Tree (CART)” para entrenar los árboles de decisión. El funcionamiento del algoritmo consiste en la división sucesiva en dos subgrupos utilizando una única variable k y un umbral t_k . Escoge estos datos buscando la pareja (λ) que produzca los subgrupos más puros. En este estudio se ha utilizado la función `DecisionTreeRegressor` del paquete `sklearn.tree`.(7)
5. Árbol de decisión múltiple: se trata de un método de ensamble de árboles de decisión. Se trata de un algoritmo que introduce mayor aleatoriedad cuando genera los árboles ya que en lugar de buscar la mejor característica cuando realiza la división del conjunto de características, la busca de un subconjunto aleatorio. En este trabajo se ha utilizado la función `RandomForestRegressor` del paquete `sklearn.ensemble`.(7)

6. Árboles extremadamente aleatorios: este algoritmo se diferencia del Árbol de decisión múltiple en que, en lugar de buscar el mejor umbral, utiliza umbrales aleatorios para cada característica. En este trabajo se ha utilizado la función `ExtraTreesRegressor` del paquete `sklearn.ensemble`.(7)
7. K-vecinos más próximos: se trata de un método no paramétrico que aproxima la asociación entre variables independientes y la variable dependiente a través de promediar las observaciones en la misma vecindad. En este trabajo se ha utilizado la función `KNeighborsRegressor` del paquete `sklearn.neighbors`.(7)
8. Perceptrón Multicapa: compuesto por una capa de entrada, una o más capas de unidades lógicas de umbral (TLU), llamadas capas ocultas, y una última capa de TLU llamada capa de salida. Por cada instancia de entrenamiento, el algoritmo de retropropagación hace una predicción, mide el error y recorre de nuevo todas las capas al revés para medir la contribución al error de cada conexión y finalmente modifica los pesos para reducir el error. En este trabajo se ha utilizado una función personalizada siguiendo las indicaciones de Aurélien Gerón.(4)

5.2.2 Entrenamiento de los modelos

El entrenamiento de todos los modelos se ha llevado a cabo para las veintiuna combinaciones disponibles (tabla 1), tanto en la marcha en superficie plana como en la transición de bipedestación a sedestación.

Para escoger los mejores hiper parámetros para cada modelo se ha utilizado la función “`RandomizedSearchCV`” del paquete `scikit-learn`, que realiza una búsqueda aleatoria entre las distribuciones de parámetros introducidas. Los parámetros del estimador utilizado para aplicar estos métodos están optimizados por una validación cruzada sobre los parámetros de ajuste. Los parámetros seleccionados por esta función son aquellos que maximizan el resultado de acuerdo con los parámetros de puntuación, en este caso el error medio cuadrado negativo. En la ilustración 4 se muestra el proceso completo del entrenamiento de los modelos sobre uno de los conjuntos de datos. Este proceso se ha repetido para todos los conjuntos de datos.

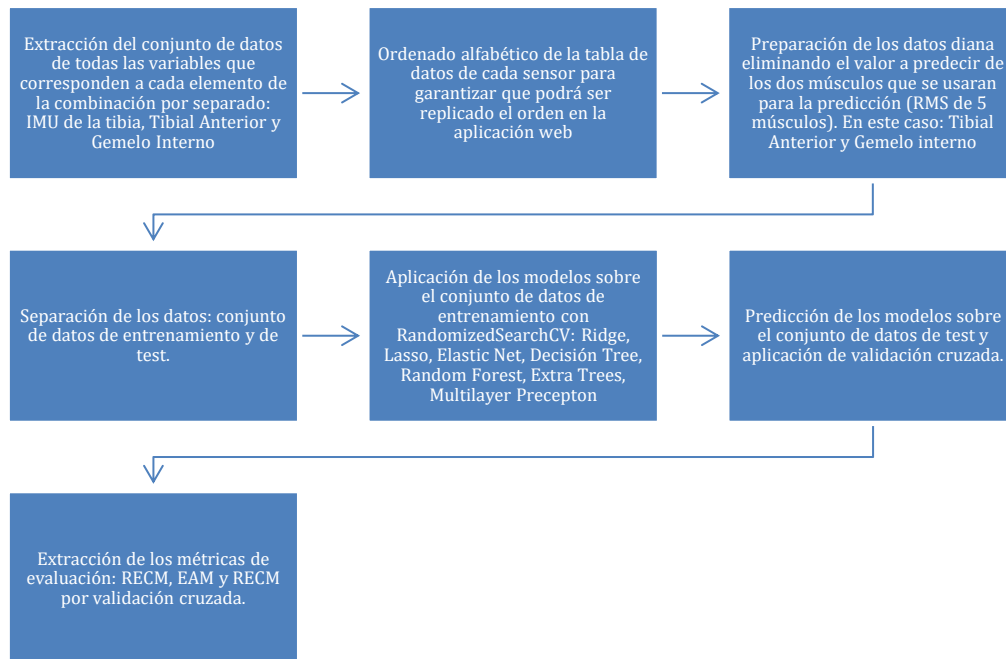


Ilustración 7. Ejemplo de entrenamiento sobre una combinación de datos de marcha sobre superficie plana: IMU tibia, Tibial Anterior y Gemelo interno

5.2.3 Evaluación de los modelos

Por cada una de las veintiuna combinaciones, se extrajo la mejor combinación de hiper parámetros para cada uno de los modelos. Sobre ellos se han utilizado los siguientes criterios de evaluación:

- Raíz del error cuadrático medio (RECM): este valor se obtuvo a través de la función “mean_squared_error” de scikit-learn (7) y posterior raíz cuadrada entre los datos de prueba y la predicción realizada por el modelo.
- Error absoluto medio (EAM): este valor se obtuvo a través de la función “mean_absolute_error” de scikit-learn (7) entre los datos de prueba y la predicción realizada por cada modelo.
- Raíz del error cuadrático medio por validación cruzada: se utilizó la función “cross_val_score” de scikit-learn (7) para obtener el valor medio de RECM obtenido con 10 repeticiones de validación cruzada. Como la métrica de puntuación utilizada fue “neg_mean_squared_error” se procedió a realizar la raíz cuadrada del ese valor multiplicado por (-1).

5.2.4 Mejora de los modelos

Teniendo en consideración que el primer entrenamiento de los modelos fue llevado a cabo sin eliminación de valores extremos, para intentar mejorar el resultado, el primer paso ha sido la eliminación de éstos a través de métodos de eliminación de valores extremos basados en modelos a través de las funciones “IsolationForest” y “EllipticEnvelope” y “OneClassSVM”. (7,42)

Se aplicaron sobre los datos escalados y para cada uno de ellos se validó empíricamente cual era el mejor valor para el parámetro “contaminación” a través del efecto de la eliminación sobre la distribución de las variables. Finalmente se escogió el valor 0.11 ya que los tres algoritmos aplicados daban un número de eliminaciones similares y homogeneizaba la distribución de las variables. Se aplicó sobre los datos el algoritmo “IsolationForest” para eliminar los registros y aplicar de nuevo los ocho modelos para cada una de las combinaciones. Las tablas 2 y 3 muestran el efecto de distintos valores del parámetro contaminación sobre el conjunto de datos de la marcha en superficie plana y en la transición de sedestación a bipedestación respectivamente en los tres métodos utilizados.

Index	Isolation Forest	Minimum Covariance Determinant	OneClassSVM
0.01	{'EMG': (2776,), 'IMU': (2776,,)}	{'EMG': (2776,), 'IMU': (2776,,)}	{'EMG': (2742,), 'IMU': (2709,,)}
0.11	{'EMG': (2496,), 'IMU': (2496,,)}	{'EMG': (2496,), 'IMU': (2496,,)}	{'EMG': (2498,), 'IMU': (2495,,)}
0.21000000000000002	{'EMG': (2216,), 'IMU': (2216,,)}	{'EMG': (2216,), 'IMU': (2215,,)}	{'EMG': (2214,), 'IMU': (2215,,)}
0.31000000000000005	{'EMG': (1935,), 'IMU': (1935,,)}	{'EMG': (1935,), 'IMU': (1935,,)}	{'EMG': (1934,), 'IMU': (1933,,)}
0.41000000000000003	{'EMG': (1655,), 'IMU': (1655,,)}	{'EMG': (1655,), 'IMU': (1655,,)}	{'EMG': (1654,), 'IMU': (1654,,)}
0.51	{'EMG': (1374,), 'IMU': (1375,,)}	{'EMG': (1374,), 'IMU': (1374,,)}	{'EMG': (1375,), 'IMU': (1374,,)}
0.61000000000000001	{'EMG': (1094,), 'IMU': (1094,,)}	{'EMG': (1094,), 'IMU': (1094,,)}	{'EMG': (1092,), 'IMU': (1096,,)}
0.71000000000000001	{'EMG': (814,), 'IMU': (814,,)}	{'EMG': (814,), 'IMU': (814,,)}	{'EMG': (813,), 'IMU': (813,,)}
0.81	{'EMG': (533,), 'IMU': (533,,)}	{'EMG': (533,), 'IMU': (533,,)}	{'EMG': (532,), 'IMU': (532,,)}
0.91	{'EMG': (253,), 'IMU': (253,,)}	{'EMG': (253,), 'IMU': (253,,)}	{'EMG': (252,), 'IMU': (253,,)}

Tabla 2. Número de registros que permanecen en el conjunto de datos tras la eliminación de valores extremos en la marcha en superficie plana.

Index	Isolation Forest	Minimum Covariance Determinant	OneClassSVM
0.01	{'EMG': (934,), 'IMU': (934,,)}	{'EMG': (934,), 'IMU': (934,,)}	{'EMG': (914,), 'IMU': (894,,)}
0.11	{'EMG': (840,), 'IMU': (840,,)}	{'EMG': (840,), 'IMU': (840,,)}	{'EMG': (836,), 'IMU': (838,,)}
0.21000000000000002	{'EMG': (745,), 'IMU': (745,,)}	{'EMG': (745,), 'IMU': (745,,)}	{'EMG': (746,), 'IMU': (748,,)}
0.31000000000000005	{'EMG': (651,), 'IMU': (651,,)}	{'EMG': (651,), 'IMU': (651,,)}	{'EMG': (652,), 'IMU': (650,,)}
0.41000000000000003	{'EMG': (557,), 'IMU': (557,,)}	{'EMG': (557,), 'IMU': (557,,)}	{'EMG': (557,), 'IMU': (556,,)}
0.51	{'EMG': (463,), 'IMU': (463,,)}	{'EMG': (463,), 'IMU': (463,,)}	{'EMG': (463,), 'IMU': (461,,)}
0.61000000000000001	{'EMG': (368,), 'IMU': (368,,)}	{'EMG': (368,), 'IMU': (368,,)}	{'EMG': (368,), 'IMU': (366,,)}
0.71000000000000001	{'EMG': (274,), 'IMU': (274,,)}	{'EMG': (274,), 'IMU': (274,,)}	{'EMG': (273,), 'IMU': (273,,)}
0.81	{'EMG': (180,), 'IMU': (180,,)}	{'EMG': (180,), 'IMU': (180,,)}	{'EMG': (180,), 'IMU': (178,,)}
0.91	{'EMG': (85,), 'IMU': (85,,)}	{'EMG': (85,), 'IMU': (85,,)}	{'EMG': (85,), 'IMU': (85,,)}

Tabla 3. Número de registros que permanecen en el conjunto de datos tras la eliminación de valores extremos en la transición de sedestación a bipedestación.

Posteriormente, con la intención de seguir mejorando los modelos se aplicó una eliminación recursiva de variables utilizando la función “RFECV” sobre los datos iniciales.(43) Se utilizó el algoritmo de Árbol de decisión (función: DecisionTreeRegression) sin valor mínimo de variables a mantener. Una vez seleccionadas las variables más importantes a través de este método para las veintiuna combinaciones se volvieron a aplicar todos los modelos sobre las variables escogidas de cada combinación.

El último paso para intentar mejorar el modelo y hacer más viable la implementación del algoritmo de predicción en la aplicación web, se eliminaron todas las variables derivadas del sensor inercial que no han sido utilizadas en los algoritmos de detección de las actividades. Se preservaron la variable del componente X del giróscopo de la cintura para la transición de sedestación a bipedestación y el componente Y del giróscopo de la tibia para la marcha en superficie plana.

6 Resultados

6.1 Predicción de la actividad muscular en la transición de sedestación a bipedestación.

6.1.1 Evaluación de los modelos sobre los datos completos en la transición de sedestación a bipedestación

Se muestran a continuación los resultados de las métricas de evaluación de los ocho modelos en las veintiuna combinaciones de datos estudiadas.

df_rmse - DataFrame

Index	Ridge	Lasso	ElasticNet	DecisionTree	randomFore	extraTree	kneighbors	MLP
('Waist', 'TA', 'MG')	0.0175247	0.0226455	0.0226455	0.0192221	0.0114492	0.0114424	0.0116363	0.0389437
('Waist', 'TA', 'SOL')	0.0166807	0.0163679	0.020388	0.0186759	0.0132253	0.0161021	0.0105614	0.0627372
('Waist', 'TA', 'BF')	0.0157992	0.0173348	0.0199621	0.0312507	0.010926	0.00963521	0.0129706	0.0790957
('Waist', 'TA', 'ST')	0.026407	0.0230856	0.0230856	0.0322561	0.0136699	0.0119874	0.0124115	0.111036
('Waist', 'TA', 'VL')	0.0161762	0.019716	0.0168441	0.0201653	0.0117618	0.0104265	0.0123152	0.0499129
('Waist', 'TA', 'RF')	0.0266244	0.0237089	0.0211452	0.0219412	0.0161196	0.0141882	0.017274	0.0681615
('Waist', 'MG', 'SOL')	0.0175134	0.0235252	0.0213021	0.0219334	0.0147548	0.01351	0.0123418	0.0913679
('Waist', 'MG', 'BF')	0.016563	0.024723	0.0192405	0.0238538	0.0130344	0.01171	0.0131258	0.0505574
('Waist', 'MG', 'ST')	0.0189628	0.025898	0.0255648	0.024814	0.0136866	0.0138828	0.0138903	0.0516055
('Waist', 'MG', 'VL')	0.0179776	0.0229452	0.0169311	0.0208314	0.0116055	0.012214	0.0144829	0.0651884
('Waist', 'MG', 'RF')	0.0286268	0.0264551	0.0192116	0.0242697	0.0127336	0.012816	0.0153156	0.0869235
('Waist', 'SOL', 'BF')	0.0164885	0.0189711	0.0228254	0.0194543	0.0134201	0.0108356	0.012215	0.0617229
('Waist', 'SOL', 'ST')	0.017304	0.0239491	0.0203114	0.023369	0.0199886	0.0168526	0.0126699	0.0648552
('Waist', 'SOL', 'VL')	0.0155265	0.0207204	0.0154908	0.0187764	0.0147056	0.0155546	0.0122023	0.0869221
('Waist', 'SOL', 'RF')	0.021964	0.0245505	0.0245505	0.0231472	0.0123442	0.0124942	0.0136296	0.0997945
('Waist', 'BF', 'ST')	0.017368	0.0254904	0.0254904	0.0242192	0.0139966	0.0116178	0.0134261	0.0590547
('Waist', 'BF', 'VL')	0.0148944	0.0224842	0.0195194	0.0260508	0.0132727	0.0108574	0.0144046	0.0574078
('Waist', 'BF', 'RF')	0.0227434	0.0260563	0.0197313	0.023597	0.013643	0.0134949	0.015947	0.0845563
('Waist', 'ST', 'VL')	0.0173139	0.0175562	0.0233797	0.0218425	0.0131189	0.0126098	0.014434	0.0923781
('Waist', 'ST', 'RF')	0.0257761	0.0268328	0.0268328	0.0276136	0.0126364	0.0151927	0.0156352	0.0880537
('Waist', 'VL', 'RF')	0.0417346	0.0239953	0.0228719	0.022995	0.0184637	0.0166858	0.0195504	0.0648769

Tabla 4. RECM obtenido tras el entrenamiento de los modelos con los datos completos sobre el conjunto de datos de prueba. (Transición)

Index	Ridge	Lasso	ElasticNet	decisionTree	andomFor	extraTree	kneighbors	MLP
('Waist', 'TA', 'MG')	0.0105028	0.0134106	0.0134106	0.0108069	0.00675138	0.00591514	0.00579847	0.0292538
('Waist', 'TA', 'SOL')	0.0099221	0.0112039	0.0128619	0.0109558	0.00670115	0.00638529	0.00569475	0.0442212
('Waist', 'TA', 'BF')	0.0100349	0.0113641	0.0122221	0.0172665	0.0066728	0.0058142	0.00611966	0.0594703
('Waist', 'TA', 'ST')	0.0169062	0.0142528	0.0142528	0.0153934	0.00755851	0.00641317	0.00660122	0.0836454
('Waist', 'TA', 'VL')	0.00956311	0.0106576	0.0101138	0.00910218	0.00615202	0.00535345	0.00537129	0.0377893
('Waist', 'TA', 'RF')	0.0176451	0.0140364	0.0133226	0.0119778	0.0071221	0.00740414	0.0066427	0.0496071
('Waist', 'MG', 'SOL')	0.011345	0.0157657	0.0146665	0.0132805	0.00787812	0.00742996	0.00710753	0.0723446
('Waist', 'MG', 'BF')	0.0112775	0.0155796	0.013353	0.0143199	0.00812766	0.0073008	0.00769819	0.0388342
('Waist', 'MG', 'ST')	0.0128224	0.0171565	0.017086	0.0165768	0.00835822	0.00801531	0.00840453	0.0404495
('Waist', 'MG', 'VL')	0.0102705	0.0135614	0.0113591	0.0122976	0.00681198	0.00709233	0.007098	0.0482619
('Waist', 'MG', 'RF')	0.0171103	0.0169402	0.0138243	0.0144813	0.0079621	0.00808185	0.00855534	0.0706262
('Waist', 'SOL', 'BF')	0.0107419	0.0130745	0.0150184	0.0118779	0.00758949	0.00666753	0.00707708	0.0432274
('Waist', 'SOL', 'ST')	0.0116071	0.0166079	0.0148889	0.0158326	0.00854574	0.00806972	0.00742081	0.0496355
('Waist', 'SOL', 'VL')	0.00901835	0.0130127	0.0106153	0.0116296	0.00718265	0.00668473	0.00637412	0.0561874
('Waist', 'SOL', 'RF')	0.0143882	0.0163915	0.0163915	0.0148772	0.00773955	0.00790513	0.00746783	0.0810316
('Waist', 'BF', 'ST')	0.0118083	0.0164616	0.0164616	0.0153408	0.00924751	0.00761867	0.00799482	0.0347468
('Waist', 'BF', 'VL')	0.00955374	0.0128665	0.0123761	0.0147967	0.00870575	0.00647017	0.00735714	0.0431867
('Waist', 'BF', 'RF')	0.0148243	0.0162452	0.0138912	0.0140569	0.00852645	0.00790371	0.00853638	0.0664838
('Waist', 'ST', 'VL')	0.0120702	0.0123435	0.0144036	0.0128716	0.00736197	0.00692912	0.00740943	0.0716673
('Waist', 'ST', 'RF')	0.0186365	0.0177823	0.0177823	0.0168321	0.00821943	0.00912323	0.00892568	0.0648798
('Waist', 'VL', 'RF')	0.0247115	0.0141872	0.0155857	0.0129377	0.00842228	0.00899631	0.0086392	0.0495719

Tabla 5. EAM obtenido tras el entrenamiento de los modelos con los datos completos sobre el conjunto de datos de prueba. (Transición)

Index	Ridge	Lasso	ElasticNet	decisionTree	andomFore	extraTree	kneighbors
('Waist', 'TA', 'MG')	0.0290327	0.024138	0.0242831	0.0253778	0.0193128	0.0188393	0.026247
('Waist', 'TA', 'SOL')	0.0222916	0.0221037	0.0225124	0.0233982	0.018378	0.0185904	0.0247901
('Waist', 'TA', 'BF')	0.0259543	0.0248973	0.0227627	0.0242649	0.0187598	0.0180668	0.0256552
('Waist', 'TA', 'ST')	0.0353861	0.0253461	0.025252	0.0250577	0.0205366	0.0208783	0.0288159
('Waist', 'TA', 'VL')	0.0272351	0.0200029	0.0214652	0.020654	0.0186462	0.0211795	0.0211317
('Waist', 'TA', 'RF')	0.0396852	0.0247311	0.0226609	0.0276744	0.0208781	0.01956	0.0280788
('Waist', 'MG', 'SOL')	0.0251347	0.0260526	0.0257383	0.0271754	0.0213888	0.021023	0.0290673
('Waist', 'MG', 'BF')	0.0300511	0.0263035	0.030575	0.0270659	0.0211343	0.0214397	0.0296712
('Waist', 'MG', 'ST')	0.0374641	0.0283366	0.0274374	0.0292526	0.0228464	0.0231689	0.0322065
('Waist', 'MG', 'VL')	0.0294669	0.0238214	0.0283548	0.0266345	0.0206564	0.0200136	0.026157
('Waist', 'MG', 'RF')	0.0382614	0.0281406	0.030689	0.0286203	0.0227647	0.0220493	0.0318847
('Waist', 'SOL', 'BF')	0.0217956	0.0254287	0.0249328	0.0268796	0.0209441	0.0203219	0.02854
('Waist', 'SOL', 'ST')	0.0249494	0.0270094	0.0274394	0.0303804	0.0228258	0.0232969	0.0310149
('Waist', 'SOL', 'VL')	0.0211355	0.0223094	0.0229775	0.0250392	0.0197467	0.0198798	0.0233967
('Waist', 'SOL', 'RF')	0.032043	0.026739	0.0266846	0.0273258	0.0213458	0.0214838	0.0295637
('Waist', 'BF', 'ST')	0.0254601	0.0271618	0.0267918	0.0284707	0.0226079	0.0226405	0.0309577
('Waist', 'BF', 'VL')	0.0268159	0.0224612	0.0232297	0.0252674	0.0204022	0.0193629	0.0248021
('Waist', 'BF', 'RF')	0.04159	0.0267525	0.0259694	0.0276258	0.0222086	0.0216906	0.0307201
('Waist', 'ST', 'VL')	0.0254954	0.0248428	0.0249052	0.029419	0.0216306	0.0236	0.0274019
('Waist', 'ST', 'RF')	0.0354227	0.029747	0.0295912	0.0311151	0.0229731	0.0241917	0.0325494
('Waist', 'VL', 'RF')	0.047015	0.024754	0.0268686	0.030957	0.0224769	0.0216598	0.0280778

Tabla 6. RECM medio obtenido tras el entrenamiento de los modelos con los datos completos a través de validación cruzada. (Transición)

6.1.2 Evaluación de los modelos sobre los datos sin valores extremos en la transición de sedestación a bipedestación.

Se muestran a continuación los resultados de las métricas de evaluación de los ocho modelos en las veintiuna combinaciones de datos estudiadas sobre el conjunto de datos una vez que se han eliminado aquellos valores que se han considerado extremos.

Index	Ridge	Lasso	ElasticNet	decisionTree	andomFores	extraTree	kneighbors	MLP
('Waist', 'TA', 'MG')	0.0120603	0.0178735	0.0178735	0.0198424	0.00878898	0.00797648	0.00785493	0.165017
('Waist', 'TA', 'SOL')	0.0114519	0.016972	0.013927	0.0153889	0.00888242	0.0077524	0.00745984	0.0677663
('Waist', 'TA', 'BF')	0.011328	0.0173425	0.0173425	0.0171971	0.0082909	0.00750594	0.0076278	0.106718
('Waist', 'TA', 'ST')	0.0127837	0.0150608	0.0173844	0.0174629	0.00947589	0.00846182	0.00879101	0.064719
('Waist', 'TA', 'VL')	0.0086213	0.0124117	0.0116919	0.0121623	0.00680612	0.00614178	0.00679375	0.0337813
('Waist', 'TA', 'RF')	0.0114045	0.0181407	0.0132513	0.0177376	0.00860296	0.00793117	0.00779394	0.0852372
('Waist', 'MG', 'SOL')	0.0135059	0.0207919	0.0193652	0.0188368	0.0113227	0.0101448	0.0105859	0.0478251
('Waist', 'MG', 'BF')	0.0138559	0.0210954	0.0202808	0.0206529	0.0110062	0.00982196	0.0100586	0.0820613
('Waist', 'MG', 'ST')	0.0149137	0.019051	0.0195009	0.0196154	0.0117208	0.0106039	0.0106783	0.0541861
('Waist', 'MG', 'VL')	0.0114814	0.0172714	0.0172714	0.0157054	0.00950555	0.00879505	0.00882974	0.0903975
('Waist', 'MG', 'RF')	0.0135721	0.0217563	0.0160121	0.0202482	0.0108661	0.0101452	0.011469	0.0471884
('Waist', 'SOL', 'BF')	0.0132977	0.0203372	0.0148177	0.0186492	0.0104749	0.00968042	0.0133616	0.151005
('Waist', 'SOL', 'ST')	0.0139955	0.0211022	0.0200539	0.0184823	0.011884	0.0103577	0.0101613	0.0821675
('Waist', 'SOL', 'VL')	0.0103615	0.0163368	0.0150507	0.0153942	0.00907106	0.00847297	0.00858029	0.130347
('Waist', 'SOL', 'RF')	0.0123367	0.021022	0.021022	0.0195299	0.0110103	0.00928356	0.0159581	0.0516321
('Waist', 'BF', 'ST')	0.0147023	0.0214013	0.0194842	0.0204192	0.0114023	0.010197	0.0109145	0.122713
('Waist', 'BF', 'VL')	0.0111058	0.0167213	0.01656	0.0145647	0.00964018	0.00868091	0.0091746	0.0396319
('Waist', 'BF', 'RF')	0.0133732	0.0213222	0.0210675	0.0195111	0.0110971	0.00989004	0.0108624	0.0671498
('Waist', 'ST', 'VL')	0.0119567	0.0163096	0.0176438	0.017177	0.00975828	0.00905676	0.00892487	0.0545365
('Waist', 'ST', 'RF')	0.014392	0.0220531	0.0201054	0.0210424	0.0114518	0.0107465	0.0109306	0.124104
('Waist', 'VL', 'RF')	0.0118717	0.0175478	0.0138674	0.0160571	0.0101665	0.00876617	0.0097108	0.0561512

Tabla 7. RECM obtenido tras eliminar valores extremos sobre el conjunto de datos de prueba. (Transición)

Index	Ridge	Lasso	ElasticNet	decisionTree	andomFores	extraTree	kneighbors	MLP
('Waist', 'TA', 'MG')	0.00829935	0.011809	0.011809	0.0122601	0.00583124	0.00525222	0.00492687	0.121429
('Waist', 'TA', 'SOL')	0.00790419	0.011566	0.010068	0.0101827	0.00592032	0.00511657	0.00472896	0.0533445
('Waist', 'TA', 'BF')	0.00769192	0.0115259	0.0115259	0.00993871	0.00562978	0.00513761	0.00478446	0.0795402
('Waist', 'TA', 'ST')	0.00895183	0.010564	0.0119933	0.0110266	0.00640624	0.00554717	0.00533791	0.0483107
('Waist', 'TA', 'VL')	0.00603919	0.00887056	0.00833246	0.00854959	0.00470588	0.00419473	0.00437139	0.0258827
('Waist', 'TA', 'RF')	0.00821549	0.012316	0.00949712	0.0109002	0.00565676	0.00520115	0.00449368	0.0567203
('Waist', 'MG', 'SOL')	0.00951549	0.0145402	0.0135904	0.0122414	0.00765634	0.0067525	0.00637236	0.039526
('Waist', 'MG', 'BF')	0.00967285	0.0145001	0.0139826	0.013054	0.00763871	0.0067376	0.00639535	0.0641372
('Waist', 'MG', 'ST')	0.0106416	0.0134988	0.0137487	0.0132164	0.00816017	0.00735746	0.0069356	0.0437984
('Waist', 'MG', 'VL')	0.00793244	0.0118448	0.0118448	0.0107123	0.00634565	0.00580687	0.00571164	0.0790337
('Waist', 'MG', 'RF')	0.00981124	0.0152903	0.011498	0.013284	0.00739646	0.0067126	0.00704379	0.0367182
('Waist', 'SOL', 'BF')	0.0090725	0.0142571	0.0103293	0.0122871	0.00721709	0.00642215	0.00883609	0.0978911
('Waist', 'SOL', 'ST')	0.00998953	0.0151805	0.0144916	0.0124139	0.00816876	0.00685903	0.00625476	0.0565416
('Waist', 'SOL', 'VL')	0.00709863	0.0116018	0.0107499	0.0103183	0.00589154	0.00541078	0.00511438	0.110267
('Waist', 'SOL', 'RF')	0.00885503	0.0150472	0.0150472	0.0126262	0.00756493	0.00614139	0.0114792	0.0382852
('Waist', 'BF', 'ST')	0.0104007	0.0151405	0.0138199	0.0132746	0.00793217	0.00703715	0.00696313	0.085708
('Waist', 'BF', 'VL')	0.00763995	0.0115617	0.0114605	0.00981898	0.00630881	0.00596709	0.00576124	0.0302551
('Waist', 'BF', 'RF')	0.00967073	0.0150072	0.0149018	0.0119549	0.00754549	0.00653518	0.00651319	0.0551158
('Waist', 'ST', 'VL')	0.00857129	0.0115034	0.0124851	0.0114441	0.00662155	0.005907	0.00570745	0.0423283
('Waist', 'ST', 'RF')	0.0105218	0.0159306	0.0147701	0.013711	0.0077073	0.00695033	0.00669609	0.0951055
('Waist', 'VL', 'RF')	0.0085644	0.0123518	0.00994066	0.010644	0.00683334	0.00581136	0.00605894	0.0411691

Tabla 8. EAM obtenido tras eliminar valores extremos de los datos sobre el conjunto de datos de prueba. (Transición)

Index	Ridge	Lasso	ElasticNet	DecisionTree	randomFore	extraTree	kneighbors
('Waist', 'TA', 'MG')	0.0171636	0.0194619	0.0190118	0.0177097	0.0146322	0.014382	0.0209949
('Waist', 'TA', 'SOL')	0.0199639	0.0179197	0.0207632	0.0182679	0.0137744	0.0139545	0.0199137
('Waist', 'TA', 'BF')	0.0158282	0.0174537	0.0180312	0.0190949	0.0139529	0.013747	0.0199906
('Waist', 'TA', 'ST')	0.0189437	0.0201651	0.0201846	0.0198105	0.0154195	0.0154247	0.0233967
('Waist', 'TA', 'VL')	0.0128948	0.0123956	0.012465	0.0130874	0.0104102	0.0104524	0.013572
('Waist', 'TA', 'RF')	0.0167039	0.0188449	0.0175022	0.0189919	0.0140124	0.0139812	0.0222149
('Waist', 'MG', 'SOL')	0.0217891	0.022015	0.0242404	0.0218333	0.0173321	0.0173023	0.0244411
('Waist', 'MG', 'BF')	0.0193603	0.0215125	0.0234232	0.0228335	0.0174135	0.017373	0.0244312
('Waist', 'MG', 'ST')	0.0217025	0.0241031	0.0241158	0.0238047	0.0190144	0.0190538	0.0261959
('Waist', 'MG', 'VL')	0.0167097	0.017813	0.0175001	0.0192259	0.0147953	0.0148976	0.0193305
('Waist', 'MG', 'RF')	0.0197181	0.0234049	0.0211189	0.0229371	0.0182044	0.0183996	0.0262001
('Waist', 'SOL', 'BF')	0.0194212	0.0207368	0.0232907	0.0218428	0.0160535	0.0161614	0.0221802
('Waist', 'SOL', 'ST')	0.0225356	0.0224546	0.0246783	0.0228102	0.0176858	0.0180753	0.0259654
('Waist', 'SOL', 'VL')	0.0158187	0.0166943	0.018101	0.0164886	0.0134852	0.0132328	0.0186096
('Waist', 'SOL', 'RF')	0.0198255	0.0221353	0.0220111	0.0224884	0.0170611	0.0169596	0.022627
('Waist', 'BF', 'ST')	0.019773	0.0221618	0.0235258	0.0231427	0.0179295	0.0177343	0.0255758
('Waist', 'BF', 'VL')	0.0161613	0.0168643	0.0174482	0.0179931	0.0144372	0.014223	0.0187202
('Waist', 'BF', 'RF')	0.0190084	0.0220263	0.0209182	0.0222977	0.0173623	0.0174047	0.0248
('Waist', 'ST', 'VL')	0.0181178	0.0192794	0.0183623	0.0198625	0.0153264	0.0156905	0.0207061
('Waist', 'ST', 'RF')	0.0209602	0.0232232	0.0226298	0.023351	0.0187937	0.0192382	0.0265134
('Waist', 'VL', 'RF')	0.0183219	0.0179423	0.0193032	0.0183302	0.0151893	0.0149559	0.0209511

Tabla 9. RECM obtenido tras eliminar valores extremos de los datos a través de validación cruzada. (Transición)

6.1.3 Evaluación de los modelos sobre los datos seleccionados tras la eliminación recursiva de variables en la transición de sedestación a bipedestación

La eliminación recursiva de variables ha sido realizada para las veintiuna combinaciones de datos, dando lugar a un grupo de variables distinto para cada combinación.

Key	Type	Size	
('Waist', 'TA', 'BF')	list	9	['Waist_Az_mean', 'Waist_Gx_std', 'TA_ssc', 'TA_svde', 'BF_mav', 'BF_r ...
('Waist', 'SOL', 'VL')	list	8	['Waist_Ay_initial', 'Waist_Az_initial', 'SOL_mav', 'SOL_spe', 'VL_mav ...
('Waist', 'TA', 'VL')	list	25	['Waist_Ay_initial', 'Waist_Ay_max', 'Waist_Ay_mean', 'Waist_Az_mean', ...
('Waist', 'VL', 'RF')	list	18	['Waist_Ax_mean', 'Waist_Ay_initial', 'Waist_Ay_max', 'Waist_Ay_mean', ...
('Waist', 'MG', 'SOL')	list	21	['Waist_Ax_mean', 'Waist_Ay_initial', 'Waist_Ay_max', 'Waist_Ay_mean', ...
('Waist', 'ST', 'RF')	list	21	['Waist_Ax_mean', 'Waist_Ay_initial', 'Waist_Ay_max', 'Waist_Ay_mean', ...
('Waist', 'MG', 'ST')	list	23	['Waist_Ax_mean', 'Waist_Ax_min', 'Waist_Ay_initial', 'Waist_Ay_max', ...
('Waist', 'MG', 'RF')	list	30	['Waist_Ax_mean', 'Waist_Ax_min', 'Waist_Ay_initial', 'Waist_Ay_max', ...
('Waist', 'TA', 'MG')	list	40	['Waist_Ax_mean', 'Waist_Ax_min', 'Waist_Ay_initial', 'Waist_Ay_max', ...
('Waist', 'TA', 'ST')	list	28	['Waist_Ax_max', 'Waist_Ax_std', 'Waist_Ay_initial', 'Waist_Ay_max', ' ...
('Waist', 'MG', 'BF')	list	18	['Waist_Ax_initial', 'Waist_Ax_mean', 'Waist_Az_initial', 'Waist_Az_me ...
('Waist', 'ST', 'VL')	list	15	['Waist_Ax_initial', 'Waist_Ax_mean', 'Waist_Ay_initial', 'Waist_Az_in ...
('Waist', 'MG', 'VL')	list	17	['Waist_Ax_initial', 'Waist_Ax_mean', 'Waist_Ay_final', 'Waist_Ay_max' ...
('Waist', 'BF', 'ST')	list	25	['Waist_Ax_initial', 'Waist_Ax_mean', 'Waist_Ax_std', 'Waist_Ay_final' ...
('Waist', 'SOL', 'RF')	list	23	['Waist_Ax_final', 'Waist_Ax_mean', 'Waist_Ay_initial', 'Waist_Ay_max' ...
('Waist', 'SOL', 'BF')	list	38	['Waist_Ax_final', 'Waist_Ax_mean', 'Waist_Ax_min', 'Waist_Ax_std', 'W ...
('Waist', 'BF', 'RF')	list	56	['Waist_Ax_final', 'Waist_Ax_max', 'Waist_Ax_mean', 'Waist_Ax_std', 'W ...
('Waist', 'TA', 'RF')	list	49	['Waist_Ax_final', 'Waist_Ax_initial', 'Waist_Ax_mean', 'Waist_Ax_std' ...
('Waist', 'SOL', 'ST')	list	61	['Waist_Ax_final', 'Waist_Ax_initial', 'Waist_Ax_max', 'Waist_Ax_mean' ...
('Waist', 'TA', 'SOL')	list	58	['Waist_Ax_final', 'Waist_Ax_initial', 'Waist_Ax_max', 'Waist_Ax_mean' ...
('Waist', 'BF', 'VL')	list	56	['Waist_Ax_final', 'Waist_Ax_initial', 'Waist_Ax_max', 'Waist_Ax_mean' ...

Tabla 10. Número de variables seleccionadas para cada modelo a través de selección recursiva de variables. (Transición)

Se muestran a continuación los resultados de las métricas de evaluación de los ocho modelos en las veintiuna combinaciones de datos estudiadas.

RMSE - DataFrame								
Index	Ridge	Lasso	ElasticNet	DecisionTree	andomFore	extraTree	kneighbors	MLP
('Waist', 'TA', 'MG')	0.0171023	0.0226455	0.0176868	0.0208545	0.0112021	0.0112471	0.0113842	0.0561549
('Waist', 'TA', 'SOL')	0.0167377	0.020388	0.0163669	0.017504	0.0152166	0.0145322	0.0102151	0.0887337
('Waist', 'TA', 'BF')	0.0168989	0.0221782	0.0198119	0.0215809	0.0120893	0.0105559	0.0106695	0.0602954
('Waist', 'TA', 'ST')	0.0229765	0.0238856	0.0168701	0.0185258	0.0156596	0.0119942	0.0123897	0.0427296
('Waist', 'TA', 'VL')	0.0163263	0.019716	0.019716	0.0198544	0.0114162	0.0141878	0.0130552	0.113101
('Waist', 'TA', 'RF')	0.0264937	0.0237089	0.0246388	0.0624723	0.0165018	0.0155482	0.0172867	0.12239
('Waist', 'MG', 'SOL')	0.0194348	0.0235252	0.0205755	0.0225635	0.0152322	0.0121946	0.0126318	0.0494378
('Waist', 'MG', 'BF')	0.0184756	0.0185095	0.0189037	0.0216243	0.0131533	0.0121432	0.012572	0.0529888
('Waist', 'MG', 'ST')	0.0195416	0.025898	0.0201653	0.0250511	0.0143702	0.0136934	0.0123813	0.0650003
('Waist', 'MG', 'VL')	0.0196563	0.0229452	0.0227856	0.0200429	0.0127522	0.0131883	0.0178572	0.0700527
('Waist', 'MG', 'RF')	0.0185801	0.0264551	0.0260765	0.0226725	0.0129424	0.0124055	0.0193017	0.0659078
('Waist', 'SOL', 'BF')	0.0170545	0.0230757	0.0230757	0.0239684	0.0138119	0.0115755	0.0163523	0.0487004
('Waist', 'SOL', 'ST')	0.017426	0.0219994	0.0226963	0.0230228	0.0200265	0.0168838	0.0127479	0.0531974
('Waist', 'SOL', 'VL')	0.0206367	0.0207204	0.0207204	0.020071	0.0128565	0.0127277	0.0149834	0.0577991
('Waist', 'SOL', 'RF')	0.0208573	0.0245505	0.0245505	0.0200621	0.0119841	0.0122995	0.0144939	0.0496534
('Waist', 'BF', 'ST')	0.0188957	0.0198943	0.0254904	0.0206292	0.015491	0.0119414	0.011881	0.0914045
('Waist', 'BF', 'VL')	0.0151785	0.0215293	0.0224842	0.0247514	0.0131934	0.0108204	0.0122841	0.0663341
('Waist', 'BF', 'RF')	0.0236318	0.0255793	0.0209816	0.0206691	0.0137256	0.0139454	0.0158458	0.0953898
('Waist', 'ST', 'VL')	0.016584	0.0225683	0.0233797	0.0220038	0.0120883	0.0121582	0.0134907	0.0635645
('Waist', 'ST', 'RF')	0.0197332	0.0268328	0.0259675	0.0263613	0.01347	0.0135564	0.0172474	0.056462
('Waist', 'VL', 'RF')	0.0296165	0.0239953	0.0224889	0.0374157	0.0191319	0.0191382	0.0187201	0.0574736

Tabla 11. RECM obtenido tras la selección recursiva de variables sobre el conjunto de datos de prueba. (Transición)

Index	Ridge	Lasso	ElasticNet	DecisionTree	RandomForest	extraTree	kneighbors	MLP
('Waist', 'TA', 'MG')	0.0105239	0.0134106	0.0118319	0.0111982	0.00649056	0.00584305	0.00560552	0.0411959
('Waist', 'TA', 'SOL')	0.00989584	0.0128619	0.0112833	0.010762	0.00666591	0.00634388	0.00553191	0.0598869
('Waist', 'TA', 'BF')	0.0113329	0.0127156	0.0121282	0.0126941	0.00764549	0.00645907	0.00597284	0.0457142
('Waist', 'TA', 'ST')	0.0151114	0.0142528	0.0119896	0.0113456	0.00761764	0.0063603	0.00636513	0.0298547
('Waist', 'TA', 'VL')	0.00960093	0.0106576	0.0106576	0.0108658	0.00560277	0.00577574	0.00502203	0.0822637
('Waist', 'TA', 'RF')	0.0175801	0.0140364	0.0164537	0.0287564	0.00724027	0.0084343	0.00651561	0.0846571
('Waist', 'MG', 'SOL')	0.0123812	0.0157657	0.0141567	0.013515	0.00825502	0.00706592	0.00702349	0.0381578
('Waist', 'MG', 'BF')	0.012785	0.0127135	0.0130812	0.0128718	0.008214	0.00725099	0.00713042	0.0388034
('Waist', 'MG', 'ST')	0.0136642	0.0171565	0.0146143	0.0162014	0.00965099	0.0078536	0.0074262	0.0504735
('Waist', 'MG', 'VL')	0.0113647	0.0135614	0.0135298	0.0122168	0.00722893	0.00732569	0.0086294	0.0491346
('Waist', 'MG', 'RF')	0.0129196	0.0169402	0.0168142	0.0144856	0.00777451	0.00734162	0.0103303	0.0513985
('Waist', 'SOL', 'BF')	0.0109096	0.0150707	0.0150707	0.0143791	0.00745863	0.00702828	0.009337	0.0370027
('Waist', 'SOL', 'ST')	0.0117065	0.0159794	0.0162646	0.0140568	0.00882933	0.00806867	0.00743695	0.039655
('Waist', 'SOL', 'VL')	0.0142398	0.0130127	0.0130127	0.0121764	0.00732226	0.00715206	0.00691504	0.0440553
('Waist', 'SOL', 'RF')	0.0148405	0.0163915	0.0163915	0.0123492	0.00752719	0.00745045	0.00682202	0.0375227
('Waist', 'BF', 'ST')	0.0125309	0.0143478	0.0164616	0.0131558	0.00996105	0.00766964	0.00704962	0.0680647
('Waist', 'BF', 'VL')	0.00977294	0.0127132	0.0128665	0.0129565	0.00817624	0.00632965	0.00695914	0.0522811
('Waist', 'BF', 'RF')	0.0152436	0.0161002	0.0140171	0.0127673	0.00826397	0.0078865	0.00825855	0.0733363
('Waist', 'ST', 'VL')	0.0115196	0.0142707	0.0144036	0.0131033	0.00725172	0.00693178	0.00697417	0.048088
('Waist', 'ST', 'RF')	0.014673	0.0177823	0.0173867	0.0172976	0.00844576	0.00804948	0.00827181	0.0433196
('Waist', 'VL', 'RF')	0.0195461	0.0141872	0.0150278	0.0233759	0.00855218	0.0106596	0.0074096	0.0465219

Tabla 12. EAM obtenido tras la selección recursiva de variables sobre el conjunto de datos de prueba. (Transición)

Index	Ridge	Lasso	ElasticNet	DecisionTree	RandomForest	extraTree	kneighbors
('Waist', 'TA', 'MG')	0.0289406	0.0241132	0.0280501	0.0262946	0.0198289	0.0185116	0.026622
('Waist', 'TA', 'SOL')	0.0223468	0.0225247	0.0221041	0.0216331	0.0181155	0.0187927	0.0247902
('Waist', 'TA', 'BF')	0.0234966	0.0223794	0.0218368	0.0224181	0.0196175	0.0185119	0.0236442
('Waist', 'TA', 'ST')	0.0238021	0.025342	0.024288	0.0295981	0.0214362	0.0207894	0.0288417
('Waist', 'TA', 'VL')	0.0217066	0.0198604	0.0197958	0.021181	0.0196943	0.0189364	0.0214486
('Waist', 'TA', 'RF')	0.0399524	0.0247827	0.0362153	0.0249611	0.0205799	0.0194387	0.028078
('Waist', 'MG', 'SOL')	0.0240004	0.0258884	0.026492	0.0277419	0.0211999	0.0209196	0.0294154
('Waist', 'MG', 'BF')	0.0268125	0.0274856	0.0282447	0.0293205	0.0209273	0.0211426	0.0291349
('Waist', 'MG', 'ST')	0.0267257	0.0285044	0.0277419	0.0303804	0.0235798	0.0228769	0.031681
('Waist', 'MG', 'VL')	0.0255515	0.0238631	0.0226911	0.0262427	0.0206631	0.0201656	0.0254373
('Waist', 'MG', 'RF')	0.0262053	0.0280457	0.0253584	0.0287887	0.0231511	0.0220308	0.0299288
('Waist', 'SOL', 'BF')	0.0243705	0.0245156	0.0247817	0.0262002	0.0218143	0.021162	0.0266514
('Waist', 'SOL', 'ST')	0.0249965	0.0272387	0.0272114	0.0285423	0.0227964	0.0232056	0.0310149
('Waist', 'SOL', 'VL')	0.0222027	0.02248	0.02248	0.0270193	0.0230945	0.0260429	0.0245938
('Waist', 'SOL', 'RF')	0.0254646	0.0265271	0.0268402	0.0287084	0.0219167	0.0224387	0.0309941
('Waist', 'BF', 'ST')	0.0271082	0.0289351	0.0268187	0.034713	0.02248	0.0224897	0.0307957
('Waist', 'BF', 'VL')	0.0269127	0.0219723	0.0223121	0.0260985	0.0199907	0.0187335	0.0248021
('Waist', 'BF', 'RF')	0.0415647	0.0239015	0.0371537	0.0266573	0.0223814	0.0217967	0.0309751
('Waist', 'ST', 'VL')	0.0264783	0.0263281	0.0248186	0.0249143	0.0212585	0.0210602	0.0280273
('Waist', 'ST', 'RF')	0.02835	0.029249	0.0282479	0.0301588	0.0234249	0.0238807	0.0329327
('Waist', 'VL', 'RF')	0.0248505	0.0246645	0.0250126	0.028618	0.021968	0.0214312	0.0284673

Tabla 13. RECM obtenido tras la selección recursiva de variables por validación cruzada. (Transición)

6.1.4 Evaluación de los modelos sobre los datos tras eliminar todas las variables de IMU no utilizadas para la detección del ciclo de transición de sedestación a bipedestación.

Se muestran a continuación los resultados de las métricas de evaluación de los ocho modelos en las veintiuna combinaciones de datos estudiadas tras eliminar del conjunto de datos todas las variables derivadas del IMU no utilizadas para la detección del ciclo de transición de sedestación a bipedestación.

Index	Ridge	Lasso	ElasticNet	DecisionTree	andomFore	extraTre	kneighbors	MLP
('Waist', 'SOL', 'ST')	0.019498	0.0239491	0.0196626	0.0221794	0.02067	0.0195492	0.0129595	0.0463448
('Waist', 'VL', 'RF')	0.03092	0.0228262	0.0229327	0.0300242	0.0192009	0.0180857	0.0190812	0.05916
('Waist', 'ST', 'RF')	0.0272795	0.0268328	0.0268328	0.0267389	0.0156632	0.0162233	0.0151253	0.0516873
('Waist', 'TA', 'SOL')	0.017831	0.0190821	0.020388	0.0183019	0.0143125	0.0160268	0.0112773	0.0674626
('Waist', 'TA', 'RF')	0.0276857	0.0237089	0.0237089	0.0188694	0.0155025	0.0159636	0.0186674	0.0392168
('Waist', 'MG', 'ST')	0.0211325	0.0213485	0.025898	0.0247693	0.0164834	0.0158259	0.0141676	0.0442932
('Waist', 'MG', 'SOL')	0.0197699	0.0235252	0.0211286	0.0215302	0.0163379	0.0146051	0.012991	0.0596247
('Waist', 'TA', 'ST')	0.0269927	0.0230856	0.0230856	0.0206486	0.0144708	0.0145616	0.0124488	0.0446935
('Waist', 'SOL', 'VL')	0.0176403	0.0207204	0.0166593	0.0199792	0.016089	0.0144575	0.0123863	0.0606116
('Waist', 'MG', 'RF')	0.0284951	0.0264551	0.0244722	0.0249114	0.0142833	0.014455	0.0151385	0.0656976
('Waist', 'BF', 'RF')	0.0242875	0.0260563	0.0196045	0.0298504	0.0147737	0.0142701	0.0146662	0.103992
('Waist', 'ST', 'VL')	0.0187671	0.0233797	0.0233797	0.018747	0.0134787	0.0141595	0.0144437	0.0629395
('Waist', 'SOL', 'RF')	0.0227578	0.0224617	0.0201372	0.0262711	0.0146131	0.0140858	0.0123803	0.0506672
('Waist', 'MG', 'VL')	0.0196489	0.0229452	0.0196845	0.0216765	0.013481	0.0134476	0.0133331	0.0474238
('Waist', 'SOL', 'BF')	0.0182326	0.0230757	0.0208887	0.0220311	0.0175226	0.0131451	0.0125379	0.049659
('Waist', 'MG', 'BF')	0.0187232	0.0250926	0.0249319	0.025582	0.0149752	0.0127083	0.01314	0.226408
('Waist', 'TA', 'MG')	0.0184522	0.0226455	0.0179214	0.0192835	0.0121749	0.0126265	0.0128158	0.0470705
('Waist', 'BF', 'ST')	0.0191109	0.0254904	0.0198729	0.026092	0.0155457	0.0125116	0.0126305	0.0522584
('Waist', 'TA', 'VL')	0.0161621	0.019716	0.019716	0.0215546	0.0130042	0.0122376	0.0138583	0.0336757
('Waist', 'BF', 'VL')	0.0162463	0.0224842	0.01725	0.0241522	0.0128373	0.0121235	0.0137848	0.121965
('Waist', 'TA', 'BF')	0.0172579	0.0221782	0.0162506	0.0203228	0.0114609	0.0100868	0.0119349	0.101827

Tabla 14. RECM obtenido tras la selección por conveniencia de variables sobre el conjunto de datos de prueba. (Transición)

Index	Ridge	Lasso	ElasticNet	DecisionTree	RandomForest	extraTree	kneighbors	MLP
('Waist', 'TA', 'MG')	0.0111097	0.0134106	0.0120286	0.0107453	0.00660021	0.0067156	0.00611773	0.032174
('Waist', 'TA', 'SOL')	0.0106381	0.0125725	0.0128619	0.0115735	0.00737199	0.00687766	0.00598964	0.0456224
('Waist', 'TA', 'BF')	0.0107234	0.0127156	0.0108762	0.0111354	0.00701851	0.00618473	0.00584836	0.0666105
('Waist', 'TA', 'ST')	0.0169174	0.0142528	0.0142528	0.0123159	0.0081987	0.00727587	0.0065015	0.0337645
('Waist', 'TA', 'VL')	0.0100184	0.0106576	0.0106576	0.0107987	0.00649387	0.00615464	0.00568289	0.0227412
('Waist', 'TA', 'RF')	0.0182857	0.0140364	0.0140364	0.0106288	0.00766121	0.00861005	0.00863277	0.0286187
('Waist', 'MG', 'SOL')	0.0131068	0.0157657	0.0145003	0.0137852	0.00928872	0.00876685	0.00756682	0.0402173
('Waist', 'MG', 'BF')	0.0126036	0.0156194	0.0155986	0.0158232	0.00964324	0.00781908	0.0076199	0.143374
('Waist', 'MG', 'ST')	0.0146057	0.0155775	0.0171565	0.0157951	0.00946034	0.00904058	0.00844372	0.0280516
('Waist', 'MG', 'VL')	0.0116773	0.0135614	0.0128538	0.0123204	0.00777481	0.0071267	0.00669542	0.0325852
('Waist', 'MG', 'RF')	0.017835	0.0169402	0.0162731	0.0153994	0.00908275	0.00929812	0.00794605	0.0472276
('Waist', 'SOL', 'BF')	0.0118171	0.0150707	0.014153	0.013917	0.00898856	0.00773861	0.00724673	0.0374404
('Waist', 'SOL', 'ST')	0.013173	0.0166079	0.0131001	0.0146685	0.00995013	0.00922744	0.00771659	0.0349353
('Waist', 'SOL', 'VL')	0.0102786	0.0130127	0.011283	0.0126264	0.00791749	0.00721143	0.00643525	0.0448905
('Waist', 'SOL', 'RF')	0.0154746	0.0156686	0.0145464	0.0160578	0.00903151	0.00907839	0.00673367	0.0364988
('Waist', 'BF', 'ST')	0.0127879	0.0164616	0.0138934	0.0154744	0.0101463	0.00839707	0.00758878	0.0403101
('Waist', 'BF', 'VL')	0.0104584	0.0128665	0.0114596	0.0144196	0.00819407	0.00713515	0.00683629	0.0816771
('Waist', 'BF', 'RF')	0.0158306	0.0162452	0.0138461	0.0188941	0.00918349	0.00896247	0.0076466	0.0782231
('Waist', 'ST', 'VL')	0.0130578	0.0144036	0.0144036	0.0114881	0.00812262	0.00900228	0.0073858	0.0452055
('Waist', 'ST', 'RF')	0.0198095	0.0177823	0.0177823	0.0176297	0.0103002	0.0107694	0.0085485	0.03949
('Waist', 'VL', 'RF')	0.0203786	0.0147786	0.0140354	0.0148217	0.00905168	0.010407	0.00815568	0.0432577

Tabla 15. EAM obtenido tras la selección por conveniencia de variables sobre el conjunto de datos de prueba. (Transición)

Index	Ridge	Lasso	ElasticNet	DecisionTree	RandomForest	extraTree	kneighbors
('Waist', 'TA', 'BF')	0.0262922	0.0224806	0.0266987	0.024766	0.0184924	0.0183399	0.0255013
('Waist', 'TA', 'MG')	0.0286627	0.0241138	0.0280444	0.0277252	0.0192441	0.0189162	0.0262448
('Waist', 'TA', 'SOL')	0.0229092	0.0214805	0.0226734	0.0243299	0.0194181	0.0192559	0.024974
('Waist', 'BF', 'VL')	0.0268598	0.0224541	0.0257029	0.0258226	0.0200754	0.0198512	0.0248024
('Waist', 'TA', 'RF')	0.0418276	0.0247332	0.024749	0.0258101	0.0198457	0.0199839	0.0265914
('Waist', 'MG', 'VL')	0.0299211	0.0239135	0.023595	0.0249617	0.0216513	0.0203796	0.0261329
('Waist', 'SOL', 'BF')	0.0236965	0.0249178	0.0250004	0.0265705	0.0220472	0.0204479	0.0285344
('Waist', 'SOL', 'VL')	0.0235946	0.0223108	0.0223962	0.0220377	0.0210549	0.0206286	0.0233994
('Waist', 'TA', 'VL')	0.0265791	0.0195265	0.0196386	0.0201482	0.0206197	0.0210281	0.0210374
('Waist', 'MG', 'BF')	0.0306893	0.0259552	0.0262068	0.030637	0.0224408	0.0213249	0.0296729
('Waist', 'TA', 'ST')	0.0361239	0.0252913	0.0252271	0.0277337	0.0212411	0.0215542	0.0288252
('Waist', 'BF', 'RF')	0.0436705	0.0267887	0.0270876	0.0283038	0.0225643	0.0216276	0.0307259
('Waist', 'SOL', 'RF')	0.0341466	0.0238904	0.0237447	0.0298616	0.0225918	0.0221723	0.0293536
('Waist', 'MG', 'SOL')	0.0268582	0.0260627	0.0257437	0.0261121	0.0226339	0.0223672	0.0290757
('Waist', 'BF', 'ST')	0.0261889	0.0271774	0.0277941	0.0296672	0.0233494	0.022412	0.0309491
('Waist', 'VL', 'RF')	0.044054	0.0259214	0.0258999	0.0266281	0.0234139	0.0226277	0.0280783
('Waist', 'MG', 'RF')	0.0425239	0.0287546	0.0253595	0.0297969	0.0234901	0.0229833	0.0319659
('Waist', 'ST', 'VL')	0.0262609	0.0249044	0.0248856	0.0305423	0.0230236	0.0231538	0.0274086
('Waist', 'MG', 'ST')	0.036732	0.0284136	0.0284529	0.0319215	0.0232995	0.0236374	0.0322068
('Waist', 'SOL', 'ST')	0.0266054	0.0270025	0.0270179	0.0274238	0.0239828	0.0236509	0.031013
('Waist', 'ST', 'RF')	0.0391591	0.0290576	0.0295268	0.0302113	0.0252064	0.025196	0.0325412

Tabla 16. RECM obtenido tras la selección por conveniencia de variables por validación cruzada. (Transición)

6.2 Predicción de la actividad muscular en la marcha en superficie plana.

6.2.1 Evaluación de los modelos sobre los datos completos en la marcha en superficie plana

Se muestran a continuación los resultados de las métricas de evaluación de los ocho modelos en las veintiuna combinaciones de datos estudiadas.

Index	Ridge	Lasso	ElasticNet	DecisionTree	randomForest	extraTree	kneighbors	MLP
('Shank', 'TA', 'MG')	0.0154031	0.0158219	0.0277767	0.0243605	0.00847968	0.00770563	0.0135367	0.0568872
('Shank', 'TA', 'SOL')	0.0143227	0.0269455	0.0184976	0.0250925	0.011032	0.0108562	0.021224	0.0352168
('Shank', 'TA', 'BF')	0.0148648	0.027979	0.027979	0.0248357	0.0120152	0.00766226	0.014288	0.0368983
('Shank', 'TA', 'ST')	0.0159559	0.0302187	0.0191861	0.0262563	0.00879843	0.0159333	0.0162412	0.0634416
('Shank', 'TA', 'VL')	0.0137098	0.0264456	0.0224946	0.0256013	0.00802225	0.00890295	0.0140716	0.0940569
('Shank', 'TA', 'RF')	0.0158501	0.0298145	0.0188675	0.0268063	0.00819337	0.00947829	0.0147663	0.0445742
('Shank', 'MG', 'SOL')	0.0195383	0.0306383	0.0306383	0.0251876	0.0117506	0.0126187	0.015739	0.0784551
('Shank', 'MG', 'BF')	0.018394	0.0315511	0.0265497	0.0275286	0.0135681	0.0123338	0.0148529	0.0444328
('Shank', 'MG', 'ST')	0.0194481	0.0335532	0.0272347	0.0288997	0.0143571	0.0137047	0.0160103	0.0878438
('Shank', 'MG', 'VL')	0.0184094	0.0301996	0.0275513	0.0263382	0.011915	0.00951157	0.015215	0.123087
('Shank', 'MG', 'RF')	0.0175059	0.0331896	0.0331896	0.0259376	0.0106204	0.0124403	0.0162225	0.0886961
('Shank', 'SOL', 'BF')	0.0178888	0.0308219	0.0308219	0.0259231	0.0132349	0.0125482	0.0155803	0.0625621
('Shank', 'SOL', 'ST')	0.0193678	0.0328684	0.0281833	0.0287047	0.0143518	0.0154464	0.0169293	0.118303
('Shank', 'SOL', 'VL')	0.0174006	0.0294369	0.0294369	0.0228907	0.00861751	0.0123499	0.0236965	0.0548568
('Shank', 'SOL', 'RF')	0.0167116	0.0324972	0.0324972	0.0274776	0.0130865	0.0144121	0.0164364	0.055941
('Shank', 'BF', 'ST')	0.0190833	0.0330681	0.0260613	0.0293216	0.0144252	0.00983389	0.0163841	0.0698957
('Shank', 'BF', 'VL')	0.0171739	0.0245542	0.0170889	0.0294904	0.00927779	0.00894336	0.0153161	0.0901885
('Shank', 'BF', 'RF')	0.0172513	0.0333592	0.0235669	0.0305062	0.0100989	0.012567	0.0160025	0.0454801
('Shank', 'ST', 'VL')	0.0177078	0.027951	0.0323797	0.0298976	0.0141362	0.00865868	0.0146469	0.0858535
('Shank', 'ST', 'RF')	0.0176388	0.0352587	0.0216923	0.0286761	0.0102539	0.00983257	0.0179487	0.105635
('Shank', 'VL', 'RF')	0.0182432	0.0320839	0.0246378	0.0302646	0.0145228	0.00915293	0.0202975	0.0476726

Tabla 17. RECM obtenido tras el entrenamiento de los modelos con los datos completos sobre el conjunto de datos de prueba. (Marcha)

Index	Ridge	Lasso	ElasticNet	decisionTree	andomFore	extraTree	kneighbors	MLP
('Shank', 'TA', 'MG')	0.00953247	0.00958895	0.0135716	0.0101224	0.00518608	0.00459848	0.00452082	0.0395192
('Shank', 'TA', 'SOL')	0.00860225	0.0125991	0.0104951	0.0109308	0.00538923	0.00491648	0.00689783	0.0211531
('Shank', 'TA', 'BF')	0.009536	0.0132945	0.0132945	0.0106603	0.00540815	0.00447756	0.00468384	0.0226881
('Shank', 'TA', 'ST')	0.00968205	0.0142023	0.0110922	0.0116488	0.00547473	0.00544582	0.00518559	0.0425783
('Shank', 'TA', 'VL')	0.00760507	0.0105763	0.010177	0.00952251	0.00419996	0.00382686	0.00414342	0.06442
('Shank', 'TA', 'RF')	0.0102805	0.0137931	0.0102229	0.0108191	0.00504855	0.00497355	0.00451138	0.0283392
('Shank', 'MG', 'SOL')	0.0113382	0.0164747	0.0164747	0.0124453	0.00697355	0.00662214	0.00571628	0.056467
('Shank', 'MG', 'BF')	0.0115591	0.0171702	0.0155147	0.0138365	0.00727072	0.00631625	0.00536311	0.0289178
('Shank', 'MG', 'ST')	0.0122295	0.018078	0.0164493	0.014069	0.0076281	0.00714	0.00588118	0.0627201
('Shank', 'MG', 'VL')	0.00976986	0.014452	0.0137601	0.0112935	0.00679903	0.00563333	0.00499597	0.0818689
('Shank', 'MG', 'RF')	0.010888	0.0176687	0.0176687	0.0117927	0.00688428	0.00661776	0.00581652	0.0526057
('Shank', 'SOL', 'BF')	0.0104788	0.0161977	0.0161977	0.0123535	0.00709941	0.00639509	0.00544567	0.0413364
('Shank', 'SOL', 'ST')	0.0116838	0.0171055	0.0160686	0.0133397	0.0071869	0.00704633	0.00585152	0.0746899
('Shank', 'SOL', 'VL')	0.00873804	0.0134795	0.0134795	0.00991271	0.00575999	0.00565172	0.00736539	0.0326341
('Shank', 'SOL', 'RF')	0.0101219	0.0166963	0.0166963	0.0125634	0.0070138	0.0061341	0.00541477	0.035954
('Shank', 'BF', 'ST')	0.012027	0.0177131	0.0156058	0.0136929	0.00786529	0.00622354	0.00564504	0.0437734
('Shank', 'BF', 'VL')	0.00944375	0.0128747	0.00957967	0.0128309	0.00612627	0.00520929	0.00475051	0.0714443
('Shank', 'BF', 'RF')	0.0107009	0.0173917	0.013813	0.0145927	0.00633793	0.00601625	0.00520107	0.0282746
('Shank', 'ST', 'VL')	0.00994334	0.0141309	0.0150391	0.0121821	0.00679425	0.00549153	0.00496619	0.0586361
('Shank', 'ST', 'RF')	0.0111453	0.0182995	0.0131845	0.0141633	0.00660453	0.00608303	0.00602113	0.0773056
('Shank', 'VL', 'RF')	0.0100486	0.0146735	0.0129577	0.0128691	0.00704395	0.00557854	0.00540231	0.0312514

Tabla 18. EAM obtenido tras el entrenamiento de los modelos con los datos completos sobre el conjunto de datos de prueba. (Marcha)

Index	Ridge	Lasso	ElasticNet	decisionTree	andomFore	extraTree	kneighbors
('Shank', 'TA', 'VL')	0.0244798	0.017831	0.0165987	0.0205892	0.0151164	0.0150981	0.0189283
('Shank', 'TA', 'SOL')	0.0206947	0.0202794	0.0190438	0.0220645	0.0163921	0.0161659	0.0208051
('Shank', 'BF', 'VL')	0.0271713	0.0209058	0.0251253	0.0258101	0.0195198	0.019966	0.0249174
('Shank', 'TA', 'BF')	0.0258395	0.021061	0.0210403	0.0240769	0.017982	0.0169818	0.0226801
('Shank', 'ST', 'VL')	0.0235948	0.0213029	0.0221618	0.024801	0.0197361	0.0188782	0.0257201
('Shank', 'SOL', 'VL')	0.0197702	0.0218467	0.0218086	0.0233089	0.0182601	0.0181689	0.0221631
('Shank', 'TA', 'RF')	0.038055	0.0221723	0.0235682	0.0247775	0.0173225	0.0171198	0.0221497
('Shank', 'TA', 'ST')	0.0238152	0.0226646	0.0197564	0.0231124	0.0178104	0.0182336	0.0242281
('Shank', 'MG', 'VL')	0.0265359	0.0232322	0.0218699	0.0246523	0.0199503	0.0188943	0.0259626
('Shank', 'VL', 'RF')	0.0442448	0.024049	0.0234667	0.0247981	0.0203862	0.0194349	0.0265996
('Shank', 'MG', 'SOL')	0.0256664	0.0251294	0.0251237	0.0246959	0.0201705	0.0203018	0.0270683
('Shank', 'SOL', 'BF')	0.0277301	0.0253083	0.0253104	0.0257968	0.0198916	0.0199056	0.0270151
('Shank', 'TA', 'MG')	0.0259548	0.0254032	0.0216957	0.0229321	0.0176863	0.0172468	0.0224429
('Shank', 'SOL', 'RF')	0.0325761	0.02568	0.0257017	0.0256693	0.0203578	0.0204968	0.0267036
('Shank', 'BF', 'ST')	0.0346782	0.0260959	0.0249292	0.0299522	0.0219481	0.021529	0.0284392
('Shank', 'SOL', 'ST')	0.0249493	0.0262417	0.0242094	0.0290179	0.020748	0.0212781	0.0280693
('Shank', 'MG', 'BF')	0.0283191	0.0265235	0.0261205	0.0271381	0.0205745	0.0201355	0.0276026
('Shank', 'BF', 'RF')	0.0355148	0.0267062	0.0240408	0.0269925	0.0202938	0.0199036	0.0274739
('Shank', 'MG', 'RF')	0.0367673	0.0269147	0.0270263	0.0292045	0.0207665	0.0201561	0.0288134
('Shank', 'MG', 'ST')	0.0286942	0.027328	0.0278459	0.0298091	0.021238	0.0212354	0.0285638
('Shank', 'ST', 'RF')	0.0348406	0.027724	0.0231389	0.0280461	0.0209641	0.0204807	0.0289593

Tabla 19. RECM medio obtenido tras el entrenamiento de los modelos con los datos completos a través de validación cruzada. (Marcha)

6.2.2 Evaluación de los modelos sobre los datos sin valores extremos en la marcha en superficie plana

Se muestran a continuación los resultados de las métricas de evaluación de los ocho modelos en las veintiuna combinaciones de datos estudiadas sobre el conjunto de datos una vez que se han eliminado aquellos valores que se han considerado extremos

Index	Ridge	Lasso	ElasticNet	decisionTree	randomForest	extraTree	kneihbors	MLP
('Shank', 'TA', 'MG')	0.0124045	0.0186851	0.0186851	0.0142136	0.00785238	0.00651373	0.00747163	0.0426467
('Shank', 'TA', 'SOL')	0.0115316	0.0175602	0.0147465	0.0137927	0.00754329	0.00661775	0.0070776	0.102737
('Shank', 'TA', 'BF')	0.0120928	0.0184876	0.0176573	0.0159343	0.00778089	0.00686319	0.00667133	0.0389448
('Shank', 'TA', 'ST')	0.0125558	0.0188028	0.0188028	0.0151764	0.00809977	0.00717815	0.00763795	0.0292832
('Shank', 'TA', 'VL')	0.00827561	0.0117681	0.0117681	0.0108511	0.00572333	0.00488578	0.00554809	0.124567
('Shank', 'TA', 'RF')	0.010725	0.0120815	0.015196	0.0143393	0.00701955	0.00613681	0.00870185	0.0284433
('Shank', 'MG', 'SOL')	0.0158626	0.0224728	0.0224728	0.0185162	0.010077	0.00881508	0.0099874	0.0661988
('Shank', 'MG', 'BF')	0.0154878	0.0232047	0.022437	0.0195331	0.00991428	0.00873684	0.00963105	0.0421025
('Shank', 'MG', 'ST')	0.0158709	0.0234566	0.0159291	0.0192297	0.0100341	0.00951005	0.00980192	0.0704443
('Shank', 'MG', 'VL')	0.0129132	0.0183072	0.0181081	0.0145472	0.0083467	0.00768581	0.00873998	0.0338664
('Shank', 'MG', 'RF')	0.0144724	0.0232694	0.0207184	0.018891	0.00947671	0.0084562	0.00963051	0.0374531
('Shank', 'SOL', 'BF')	0.0152094	0.0223089	0.0223089	0.017178	0.0108108	0.00939298	0.00961072	0.0973331
('Shank', 'SOL', 'ST')	0.0155652	0.0225708	0.0188151	0.0179632	0.00991572	0.00875425	0.0101785	0.0667187
('Shank', 'SOL', 'VL')	0.012146	0.0171576	0.0171576	0.0149995	0.00805684	0.00743698	0.00802345	0.122595
('Shank', 'SOL', 'RF')	0.0142046	0.0161784	0.0214406	0.0174401	0.00905584	0.00851206	0.00884306	0.049106
('Shank', 'BF', 'ST')	0.0163764	0.0232996	0.0232996	0.0189153	0.0112263	0.00951754	0.0103896	0.0405913
('Shank', 'BF', 'VL')	0.0128906	0.0140507	0.0181056	0.0140991	0.00872742	0.00758147	0.00832476	0.0885644
('Shank', 'BF', 'RF')	0.0146367	0.0231111	0.0231111	0.0169056	0.0095274	0.00852657	0.00910429	0.0373214
('Shank', 'ST', 'VL')	0.0130856	0.0184274	0.0184274	0.0148998	0.00862995	0.00759311	0.00880595	0.03662
('Shank', 'ST', 'RF')	0.014982	0.021638	0.0213037	0.0172348	0.00967439	0.00952051	0.00961188	0.0833947
('Shank', 'VL', 'RF')	0.0131427	0.0181885	0.0142151	0.0153461	0.00846183	0.0074815	0.00834532	0.179992

Tabla 20. RECM obtenido tras eliminar valores extremos sobre el conjunto de datos de prueba. (Marcha)

Index	Ridge	Lasso	ElasticNet	decisionTree	andomFores	extraTree	kneighbors	MLP
('Shank', 'TA', 'MG')	0.00828716	0.0118476	0.0118476	0.00840448	0.00463844	0.00382118	0.00406467	0.0308773
('Shank', 'TA', 'SOL')	0.00747776	0.0111714	0.00961066	0.00802492	0.00441891	0.00386511	0.00388123	0.0826033
('Shank', 'TA', 'BF')	0.00822812	0.0118002	0.0114003	0.00939876	0.00461765	0.0040597	0.00380483	0.0283369
('Shank', 'TA', 'ST')	0.00841242	0.0121523	0.0121523	0.00934722	0.00482095	0.00415171	0.00412207	0.0209675
('Shank', 'TA', 'VL')	0.00585512	0.00848814	0.00848814	0.00744862	0.00368524	0.00315102	0.00333062	0.077999
('Shank', 'TA', 'RF')	0.00740105	0.00828166	0.0101336	0.00904708	0.00418816	0.00352654	0.00556844	0.0198519
('Shank', 'MG', 'SOL')	0.0105424	0.0151225	0.0151225	0.0117939	0.00618012	0.00538324	0.00544954	0.0489224
('Shank', 'MG', 'BF')	0.0104187	0.0157513	0.0152858	0.0126156	0.00592979	0.0051151	0.00519296	0.029706
('Shank', 'MG', 'ST')	0.010902	0.0161034	0.0109195	0.0122379	0.00616207	0.00590902	0.00536511	0.0527119
('Shank', 'MG', 'VL')	0.00855843	0.0124392	0.0123304	0.00925897	0.00524448	0.00473457	0.00471919	0.0234366
('Shank', 'MG', 'RF')	0.00996621	0.0156979	0.0141821	0.0120359	0.00585994	0.00497336	0.00534003	0.0263898
('Shank', 'SOL', 'BF')	0.00977317	0.0150751	0.0150751	0.0107937	0.0064131	0.00550031	0.0052036	0.0727003
('Shank', 'SOL', 'ST')	0.0102619	0.0154272	0.0128195	0.0115708	0.0062145	0.00530851	0.00544137	0.0518963
('Shank', 'SOL', 'VL')	0.00767199	0.011763	0.011763	0.00973705	0.00509114	0.00456833	0.00435544	0.0892633
('Shank', 'SOL', 'RF')	0.00932691	0.0110646	0.0144977	0.0110441	0.00550482	0.00492288	0.00475581	0.0375
('Shank', 'BF', 'ST')	0.0111612	0.016056	0.016056	0.0121415	0.00682347	0.0058148	0.00548786	0.0264432
('Shank', 'BF', 'VL')	0.00840458	0.00927337	0.0123918	0.00891898	0.00559901	0.00462078	0.00447898	0.057403
('Shank', 'BF', 'RF')	0.00989949	0.0156505	0.0156505	0.0108984	0.00577022	0.00485606	0.0048115	0.0262978
('Shank', 'ST', 'VL')	0.00862038	0.0127439	0.0127439	0.00961673	0.00537453	0.00464201	0.00469095	0.0274079
('Shank', 'ST', 'RF')	0.0102066	0.0149897	0.0147868	0.0112362	0.00601488	0.00578131	0.00517602	0.0620979
('Shank', 'VL', 'RF')	0.00864843	0.0123384	0.00953071	0.00971473	0.00545046	0.00467518	0.00459827	0.151389

Tabla 21. EAM obtenido tras eliminar valores extremos sobre el conjunto de datos de prueba. (Marcha)

Index	Ridge	Lasso	ElasticNet	decisionTree	randomForest	extraTree	kneighbors
('Shank', 'TA', 'MG')	0.0162763	0.0170257	0.017127	0.0177158	0.0135535	0.0133221	0.0176919
('Shank', 'TA', 'SOL')	0.014528	0.0160274	0.0163606	0.0153085	0.0121594	0.011964	0.0181087
('Shank', 'TA', 'BF')	0.0158606	0.0169735	0.0169517	0.0179351	0.0129698	0.012823	0.0176713
('Shank', 'TA', 'ST')	0.0163392	0.0175025	0.017546	0.0174195	0.0138389	0.0135771	0.0193616
('Shank', 'TA', 'VL')	0.0117142	0.0114994	0.0113882	0.0124887	0.0103016	0.0103413	0.013884
('Shank', 'TA', 'RF')	0.0144147	0.0151277	0.015108	0.0174404	0.012338	0.0121495	0.0165078
('Shank', 'MG', 'SOL')	0.019853	0.0214964	0.0216207	0.0207527	0.0169748	0.0169986	0.0235767
('Shank', 'MG', 'BF')	0.0197474	0.0221686	0.0213486	0.021321	0.01694	0.0166983	0.0236303
('Shank', 'MG', 'ST')	0.0206216	0.0226356	0.0204201	0.021954	0.0177952	0.0178096	0.0238754
('Shank', 'MG', 'VL')	0.0173816	0.0181824	0.0176734	0.0179556	0.0146744	0.0146987	0.0214757
('Shank', 'MG', 'RF')	0.0187611	0.0222962	0.0201828	0.0220434	0.0164283	0.0162359	0.024108
('Shank', 'SOL', 'BF')	0.0191937	0.0216659	0.0217826	0.0193324	0.0163393	0.0160997	0.0227496
('Shank', 'SOL', 'ST')	0.0201602	0.0223251	0.0217552	0.021151	0.0167825	0.0167475	0.0238665
('Shank', 'SOL', 'VL')	0.0162816	0.0171454	0.0171493	0.0164376	0.0135808	0.013653	0.0193012
('Shank', 'SOL', 'RF')	0.0180157	0.0189307	0.0198677	0.0197166	0.0156214	0.0154513	0.0233064
('Shank', 'BF', 'ST')	0.0210891	0.0232506	0.0231394	0.0217614	0.0182356	0.017587	0.0236676
('Shank', 'BF', 'VL')	0.0170638	0.0168134	0.0180665	0.0184921	0.0145289	0.0147541	0.019898
('Shank', 'BF', 'RF')	0.0188751	0.022359	0.0221219	0.0211835	0.0158689	0.0155971	0.0224609
('Shank', 'ST', 'VL')	0.0175492	0.0181254	0.0181448	0.018658	0.0147753	0.0148641	0.0210068
('Shank', 'ST', 'RF')	0.0194526	0.0209292	0.0208443	0.0218532	0.016835	0.0168086	0.0241428
('Shank', 'VL', 'RF')	0.0173505	0.0183	0.0173115	0.017658	0.0143745	0.0142502	0.0207396

Tabla 22. RECM obtenido tras eliminar valores extremos de los datos a través de validación cruzada. (Marcha)

6.2.3 Evaluación de los modelos sobre los datos seleccionados tras la eliminación recursiva de variables en la marcha en superficie plana.

La eliminación recursiva de variables ha sido realizada para las veintiuna combinaciones de datos, dando lugar a un grupo de variables distinto para cada combinación.

Key	Type	Size	
('Shank', 'BF', 'RF')	list	9	['Shank_Ay_mean', 'Shank_Az_max', 'Shank_Az_std', 'BF_rms', 'BF_spe', ...
('Shank', 'BF', 'ST')	list	10	['Shank_Ax_mean', 'Shank_Ay_initial', 'Shank_Ay_mean', 'Shank_Az_max', ...
('Shank', 'BF', 'VL')	list	34	['Shank_Ax_max', 'Shank_Ax_mean', 'Shank_Ax_min', 'Shank_Ax_std', 'Sha ...
('Shank', 'MG', 'BF')	list	9	['Shank_Ax_mean', 'Shank_Ay_mean', 'Shank_Az_max', 'Shank_Az_std', 'MG ...
('Shank', 'MG', 'RF')	list	21	['Shank_Ax_mean', 'Shank_Ax_std', 'Shank_Ay_initial', 'Shank_Ay_mean', ...
('Shank', 'MG', 'SOL')	list	10	['Shank_Ax_mean', 'Shank_Ay_initial', 'Shank_Ay_mean', 'Shank_Az_mean' ...
('Shank', 'MG', 'ST')	list	35	['Shank_Ax_mean', 'Shank_Ax_min', 'Shank_Ax_std', 'Shank_Ay_initial', ...
('Shank', 'MG', 'VL')	list	19	['Shank_Ax_min', 'Shank_Ax_std', 'Shank_Ay_mean', 'Shank_Az_initial', ...
('Shank', 'SOL', 'BF')	list	21	['Shank_Ax_mean', 'Shank_Ax_min', 'Shank_Ay_initial', 'Shank_Ay_mean', ...
('Shank', 'SOL', 'RF')	list	41	['Shank_Ax_initial', 'Shank_Ax_mean', 'Shank_Ax_min', 'Shank_Ax_std', ...
('Shank', 'SOL', 'ST')	list	19	['Shank_Ax_mean', 'Shank_Ax_std', 'Shank_Ay_initial', 'Shank_Ay_mean', ...
('Shank', 'SOL', 'VL')	list	10	['Shank_Ax_min', 'Shank_Ay_mean', 'Shank_Az_max', 'Shank_Az_std', 'Sha ...
('Shank', 'ST', 'RF')	list	12	['Shank_Ay_mean', 'Shank_Az_max', 'Shank_Az_mean', 'Shank_Az_min', 'Sh ...
('Shank', 'ST', 'VL')	list	11	['Shank_Ax_min', 'Shank_Ay_mean', 'Shank_Az_mean', 'Shank_Az_std', 'Sh ...
('Shank', 'TA', 'BF')	list	15	['Shank_Ay_initial', 'Shank_Ay_mean', 'Shank_Az_max', 'Shank_Az_mean', ...
('Shank', 'TA', 'MG')	list	15	['Shank_Ay_initial', 'Shank_Ay_mean', 'Shank_Az_max', 'Shank_Az_mean', ...
('Shank', 'TA', 'RF')	list	15	['Shank_Ay_mean', 'Shank_Az_initial', 'Shank_Az_max', 'Shank_Az_std', ...
('Shank', 'TA', 'SOL')	list	15	['Shank_Ax_mean', 'Shank_Ay_initial', 'Shank_Ay_mean', 'Shank_Az_initi ...
('Shank', 'TA', 'ST')	list	9	['Shank_Ax_mean', 'Shank_Ay_mean', 'Shank_Az_initial', 'Shank_Az_std', ...
('Shank', 'TA', 'VL')	list	44	['Shank_Ax_initial', 'Shank_Ax_mean', 'Shank_Ay_final', 'Shank_Ay_init ...
('Shank', 'VL', 'RF')	list	7	['Shank_Ay_mean', 'Shank_Az_std', 'VL_rms', 'VL_svde', 'RF_svde', 'RF_ ...

Tabla 23. Número de variables seleccionadas para cada modelo a través de selección recursiva de variables. (Marcha)

Se muestran a continuación los resultados de las métricas de evaluación de los ocho modelos en las veintiuna combinaciones de datos estudiadas.

Index	Ridge	Lasso	ElasticNet	DecisionTree	andomFores	extraTree	kneighbors	MLP
('Shank', 'TA', 'MG')	0.0176679	0.0277767	0.0277767	0.0244974	0.0118677	0.00950136	0.0141305	0.111041
('Shank', 'TA', 'SOL')	0.0157135	0.0269455	0.0269455	0.0217931	0.00857761	0.0112249	0.0143362	0.0390061
('Shank', 'TA', 'BF')	0.0168959	0.027979	0.027979	0.0217972	0.0126427	0.0115225	0.0140729	0.0361437
('Shank', 'TA', 'ST')	0.0186188	0.0302187	0.022323	0.0277279	0.0139725	0.0168393	0.01666	0.0518867
('Shank', 'TA', 'VL')	0.0146375	0.0264456	0.0264456	0.0262121	0.00946788	0.00709984	0.0150492	0.0329057
('Shank', 'TA', 'RF')	0.0166015	0.0298145	0.0222243	0.0262489	0.0107745	0.0100423	0.0150425	0.0324472
('Shank', 'MG', 'SOL')	0.0214446	0.0306383	0.0225246	0.0283272	0.0176969	0.0141079	0.0169019	0.0477301
('Shank', 'MG', 'BF')	0.0204614	0.0207509	0.0233971	0.0269041	0.014031	0.0133789	0.0149969	0.0747807
('Shank', 'MG', 'ST')	0.0200009	0.0335532	0.0335532	0.0313028	0.0106349	0.0133551	0.0161296	0.0577221
('Shank', 'MG', 'VL')	0.0179783	0.0301996	0.0301996	0.0265206	0.0120927	0.0118639	0.015691	0.0618495
('Shank', 'MG', 'RF')	0.0186341	0.0331896	0.0329262	0.0278172	0.0135853	0.0127774	0.0154955	0.0572387
('Shank', 'SOL', 'BF')	0.0189959	0.0308219	0.0260596	0.0263541	0.0115794	0.012905	0.0144376	0.0587125
('Shank', 'SOL', 'ST')	0.0209655	0.0328684	0.0328684	0.0302366	0.0135293	0.0139292	0.0261091	0.121915
('Shank', 'SOL', 'VL')	0.0171151	0.0294369	0.0294369	0.026357	0.0120934	0.00992672	0.0235886	0.0638221
('Shank', 'SOL', 'RF')	0.0167606	0.0324972	0.0225302	0.025489	0.0124608	0.0103029	0.0171629	0.199986
('Shank', 'BF', 'ST')	0.0214835	0.0234516	0.0337209	0.0299404	0.0151231	0.0149459	0.0162308	0.065862
('Shank', 'BF', 'VL')	0.017808	0.0303858	0.0202672	0.0283061	0.0158762	0.0104448	0.0241574	0.374409
('Shank', 'BF', 'RF')	0.0197413	0.0333592	0.0201877	0.0296341	0.0139249	0.0142929	0.0240426	0.0704467
('Shank', 'ST', 'VL')	0.0195505	0.0324599	0.0324599	0.0261169	0.00908369	0.0129815	0.0121848	0.0479548
('Shank', 'ST', 'RF')	0.0210127	0.0277893	0.0214315	0.030371	0.0144805	0.0114865	0.0174075	0.0641125
('Shank', 'VL', 'RF')	0.0207175	0.0320839	0.0207897	0.0272007	0.0130719	0.0106912	0.0180353	0.0367825

Tabla 24. RECM obtenido tras la selección recursiva de variables sobre el conjunto de datos de prueba. (Marcha)

Index	Ridge	Lasso	ElasticNet	decisionTree	andomFores	extraTree	kneighbors	MLP
('Shank', 'TA', 'MG')	0.010453	0.0135716	0.0135716	0.010231	0.00588678	0.00520521	0.00477086	0.0700354
('Shank', 'TA', 'SOL')	0.00949947	0.0125991	0.0125991	0.00986148	0.00525738	0.00542761	0.004527	0.0266415
('Shank', 'TA', 'BF')	0.0101322	0.0132945	0.0132945	0.0102014	0.0059583	0.00539877	0.00449982	0.0219931
('Shank', 'TA', 'ST')	0.0112108	0.0142023	0.0126245	0.0119971	0.00706338	0.00653962	0.0064064	0.0355555
('Shank', 'TA', 'VL')	0.00729573	0.0105763	0.0105763	0.0097953	0.00457009	0.0036492	0.00423543	0.0191288
('Shank', 'TA', 'RF')	0.00955853	0.0137931	0.0113232	0.0109716	0.0055571	0.00470517	0.00445927	0.0189428
('Shank', 'MG', 'SOL')	0.0132446	0.0164747	0.0138355	0.0146568	0.00778376	0.0068375	0.00661891	0.0316685
('Shank', 'MG', 'BF')	0.0127952	0.016363	0.0143763	0.0116889	0.00764499	0.00700867	0.00586777	0.0517557
('Shank', 'MG', 'ST')	0.0123669	0.018078	0.018078	0.0156524	0.00686821	0.00684275	0.00565369	0.0405606
('Shank', 'MG', 'VL')	0.01054	0.014452	0.014452	0.0105108	0.00623244	0.00534662	0.00486458	0.0444406
('Shank', 'MG', 'RF')	0.0117455	0.0176687	0.0175416	0.0133577	0.00730597	0.00649968	0.00522812	0.036406
('Shank', 'SOL', 'BF')	0.0118842	0.0161977	0.0142063	0.0122751	0.00619695	0.00587323	0.00544989	0.0411199
('Shank', 'SOL', 'ST')	0.0126791	0.0171055	0.0171055	0.014303	0.00729053	0.00643234	0.00972139	0.062673
('Shank', 'SOL', 'VL')	0.00986051	0.0134795	0.0134795	0.0107588	0.00578553	0.00502187	0.00721677	0.0419698
('Shank', 'SOL', 'RF')	0.010427	0.0166963	0.0129042	0.01163	0.00646342	0.0059564	0.00556654	0.156911
('Shank', 'BF', 'ST')	0.0136877	0.0145309	0.0178009	0.0129491	0.00791966	0.00742351	0.00649183	0.0424666
('Shank', 'BF', 'VL')	0.00941397	0.0141749	0.0110036	0.0117083	0.00706179	0.00572958	0.00735881	0.187193
('Shank', 'BF', 'RF')	0.0126741	0.0173917	0.0126784	0.0132692	0.00662832	0.00663787	0.00809711	0.0486953
('Shank', 'ST', 'VL')	0.0113281	0.0150827	0.0150827	0.00997953	0.00501903	0.00499191	0.00482523	0.030529
('Shank', 'ST', 'RF')	0.013022	0.0155226	0.0130485	0.0135449	0.00757197	0.00616178	0.0054754	0.0408373
('Shank', 'VL', 'RF')	0.0123172	0.0146735	0.012315	0.0121486	0.00667583	0.0058773	0.00634616	0.0215094

Tabla 25. EAM obtenido tras la selección recursiva de variables sobre el conjunto de datos de prueba. (Marcha)

Index	Ridge	Lasso	ElasticNet	DecisionTree	RandomForest	extraTree	kneighbors
('Shank', 'TA', 'MG')	0.0257954	0.0217148	0.0218825	0.0219594	0.0179315	0.017343	0.0225032
('Shank', 'TA', 'SOL')	0.0187451	0.0208366	0.0205269	0.0221618	0.0161846	0.0161752	0.020453
('Shank', 'TA', 'BF')	0.0237714	0.0215246	0.0213289	0.0220386	0.018258	0.0173171	0.021911
('Shank', 'TA', 'ST')	0.0203336	0.0227966	0.0202589	0.0244141	0.0181487	0.0181837	0.0227351
('Shank', 'TA', 'VL')	0.0243772	0.0179095	0.0176957	0.0179272	0.0155871	0.0151115	0.018975
('Shank', 'TA', 'RF')	0.0247805	0.0215585	0.0206222	0.0227536	0.017227	0.0175973	0.0223382
('Shank', 'MG', 'SOL')	0.0258893	0.0253697	0.0259284	0.0255126	0.0205663	0.0202718	0.0279482
('Shank', 'MG', 'BF')	0.023358	0.0262619	0.0262626	0.0253589	0.021111	0.0202394	0.0290564
('Shank', 'MG', 'ST')	0.0289781	0.0272581	0.0273308	0.0290569	0.0213514	0.0210854	0.0285641
('Shank', 'MG', 'VL')	0.0265702	0.024123	0.0212242	0.0241636	0.0198794	0.0195988	0.0229689
('Shank', 'MG', 'RF')	0.0345217	0.027249	0.0248832	0.0277371	0.0209971	0.0208213	0.028219
('Shank', 'SOL', 'BF')	0.0222851	0.0256868	0.026699	0.0246602	0.0212387	0.0197189	0.0278978
('Shank', 'SOL', 'ST')	0.023131	0.0263984	0.0264233	0.0287169	0.021289	0.0208231	0.0269022
('Shank', 'SOL', 'VL')	0.0171136	0.0207335	0.0224001	0.0223093	0.0181128	0.0179284	0.0223149
('Shank', 'SOL', 'RF')	0.0299044	0.0256419	0.0279705	0.0246636	0.0197305	0.019498	0.0275191
('Shank', 'BF', 'ST')	0.0234893	0.0255587	0.0279561	0.0320956	0.0229848	0.022154	0.0256632
('Shank', 'BF', 'VL')	0.0229083	0.0229744	0.023651	0.0234412	0.0206565	0.0193205	0.0231333
('Shank', 'BF', 'RF')	0.0234656	0.0279323	0.0257916	0.0257954	0.0207964	0.0201723	0.0255493
('Shank', 'ST', 'VL')	0.0227745	0.0249706	0.0249706	0.0270298	0.0197701	0.0196152	0.0248631
('Shank', 'ST', 'RF')	0.0249559	0.025997	0.0247316	0.0271665	0.0221786	0.0224543	0.0274556
('Shank', 'VL', 'RF')	0.0240011	0.0249407	0.0242814	0.0257084	0.0201999	0.0196619	0.0263877

Tabla 26. RECM obtenido tras la selección recursiva de variables por validación cruzada. (Marcha)

6.2.4 Evaluación de los modelos sobre los datos tras eliminar todas las variables de IMU no utilizadas para la detección del ciclo en la marcha en superficie plana.

Se muestran a continuación los resultados de las métricas de evaluación de los ocho modelos en las veintiuna combinaciones de datos estudiadas tras eliminar del conjunto de datos todas las variables derivadas del IMU no utilizadas para la detección del ciclo en la marcha sobre superficie plana.

Index	Ridge	Lasso	ElasticNet	DecisionTree	andomFores	extraTree	kneighbors	MLP
('Shank', 'TA', 'MG')	0.0171203	0.0277767	0.0243483	0.0207066	0.0107121	0.0123951	0.0221655	0.0835955
('Shank', 'TA', 'SOL')	0.0158037	0.0269455	0.0269455	0.0265355	0.00958634	0.0148945	0.0148559	0.064859
('Shank', 'TA', 'BF')	0.016554	0.027979	0.0222686	0.0227757	0.0114727	0.00957749	0.0153376	0.0783184
('Shank', 'TA', 'ST')	0.0178935	0.0290735	0.0302187	0.0280727	0.0109991	0.0105131	0.0162071	0.0484638
('Shank', 'TA', 'VL')	0.0148199	0.0186975	0.0264456	0.0232995	0.0123548	0.00838165	0.014405	0.045501
('Shank', 'TA', 'RF')	0.0171178	0.0298145	0.0286704	0.0243641	0.0123917	0.0114922	0.0153149	0.0827479
('Shank', 'MG', 'SOL')	0.0212445	0.0306383	0.0247464	0.0269356	0.0172716	0.0151194	0.0176909	0.0420636
('Shank', 'MG', 'BF')	0.0196119	0.0315511	0.0240436	0.0261602	0.0159381	0.012346	0.0237229	0.103214
('Shank', 'MG', 'ST')	0.0210289	0.0335532	0.0330624	0.0308952	0.0164568	0.015689	0.0184857	0.0481074
('Shank', 'MG', 'VL')	0.0194057	0.0301996	0.0273949	0.0260564	0.015053	0.012184	0.0159603	0.0930361
('Shank', 'MG', 'RF')	0.0185532	0.0280008	0.0331896	0.0299194	0.0160676	0.0121623	0.0165176	0.0502441
('Shank', 'SOL', 'BF')	0.0198872	0.0308219	0.0224479	0.0300513	0.0151659	0.0150911	0.0238864	0.0844362
('Shank', 'SOL', 'ST')	0.0208789	0.0328684	0.0328684	0.0302821	0.0165304	0.0160603	0.0275546	0.0628961
('Shank', 'SOL', 'VL')	0.0180711	0.0290258	0.0294369	0.0263737	0.00973773	0.0170287	0.0175816	0.0538608
('Shank', 'SOL', 'RF')	0.0175114	0.0324972	0.0236743	0.0283149	0.014389	0.0112993	0.0172455	0.0731522
('Shank', 'BF', 'ST')	0.0206308	0.0337209	0.0337209	0.0288463	0.0165629	0.0171978	0.0263041	0.0693964
('Shank', 'BF', 'VL')	0.0180677	0.0303858	0.0303858	0.0276498	0.0154232	0.0129471	0.0161353	0.0566916
('Shank', 'BF', 'RF')	0.0181182	0.0333592	0.0333592	0.0280193	0.0118842	0.0140517	0.01659	0.0656479
('Shank', 'ST', 'VL')	0.0184755	0.0324599	0.0324599	0.0259085	0.0153304	0.0137776	0.0158622	0.0903847
('Shank', 'ST', 'RF')	0.0186672	0.0352587	0.0310132	0.0317097	0.0155878	0.0178856	0.0175696	0.1914
('Shank', 'VL', 'RF')	0.0192355	0.0320839	0.0320839	0.0301211	0.0109245	0.010939	0.0192325	0.0421272

Tabla 27. RECM obtenido tras la selección por conveniencia de variables sobre el conjunto de datos de prueba. (Marcha)

Index	Ridge	Lasso	ElasticNet	decisionTree	andomFores	extraTree	kneighbors	MLP
('Shank', 'TA', 'MG')	0.0104773	0.0135716	0.0127982	0.0116111	0.00683552	0.00662611	0.00849869	0.0574263
('Shank', 'TA', 'SOL')	0.0095974	0.0125991	0.0125991	0.0117534	0.00632708	0.00636417	0.00542634	0.0446158
('Shank', 'TA', 'BF')	0.0104856	0.0132945	0.0119949	0.0117006	0.00646727	0.00610457	0.00578266	0.050117
('Shank', 'TA', 'ST')	0.0105597	0.0139373	0.0142023	0.0130042	0.00714979	0.00649917	0.00598371	0.0289056
('Shank', 'TA', 'VL')	0.00790556	0.0091847	0.0105763	0.00859584	0.00555792	0.00453125	0.00451703	0.0281639
('Shank', 'TA', 'RF')	0.0108224	0.0137931	0.0131708	0.0104712	0.00652592	0.00591626	0.00518275	0.0587996
('Shank', 'MG', 'SOL')	0.0126903	0.0164747	0.0151857	0.0135698	0.00946433	0.00845208	0.00739697	0.0274963
('Shank', 'MG', 'BF')	0.0123292	0.0171702	0.0146996	0.0144558	0.00915803	0.00800523	0.00900527	0.0615738
('Shank', 'MG', 'ST')	0.0132329	0.018078	0.0180031	0.0154613	0.00947234	0.00877057	0.00768463	0.0300359
('Shank', 'MG', 'VL')	0.0102126	0.014452	0.013712	0.0110388	0.00821792	0.00706854	0.00570412	0.055937
('Shank', 'MG', 'RF')	0.0115539	0.0157781	0.0176687	0.0151395	0.00861901	0.00765515	0.00651324	0.0330523
('Shank', 'SOL', 'BF')	0.0113747	0.0161977	0.0135473	0.0160391	0.00839516	0.00838476	0.00954932	0.060065
('Shank', 'SOL', 'ST')	0.012852	0.0171055	0.0171055	0.01427	0.00892727	0.00850851	0.0118496	0.0466708
('Shank', 'SOL', 'VL')	0.0092204	0.0132804	0.0134795	0.0114687	0.00653726	0.00649635	0.00588376	0.0377929
('Shank', 'SOL', 'RF')	0.0107948	0.0166963	0.0136403	0.0130854	0.00798737	0.00703496	0.00620332	0.0479201
('Shank', 'BF', 'ST')	0.0130435	0.0178009	0.0178009	0.0154529	0.00967188	0.00872592	0.0105696	0.0482129
('Shank', 'BF', 'VL')	0.00995857	0.0141749	0.0141749	0.0112007	0.00784902	0.00654643	0.00551736	0.038865
('Shank', 'BF', 'RF')	0.0114372	0.0173917	0.0173917	0.0134648	0.00778093	0.0073887	0.00604927	0.0445034
('Shank', 'ST', 'VL')	0.0104889	0.0150827	0.0150827	0.0118217	0.00808468	0.00749788	0.00606696	0.0499948
('Shank', 'ST', 'RF')	0.0120222	0.0182995	0.0167339	0.0152812	0.00871477	0.00819408	0.00662395	0.13493
('Shank', 'VL', 'RF')	0.0106651	0.0146735	0.0146735	0.0128701	0.00742167	0.00653874	0.00604542	0.0244455

Tabla 28. EAM obtenido tras la selección por conveniencia de variables sobre el conjunto de datos de prueba. (Marcha)

Index	Ridge	Lasso	ElasticNet	DecisionTree	RandomForest	extraTree	kneighbors
('Shank', 'TA', 'MG')	0.0258705	0.0218367	0.0221751	0.0214131	0.0183949	0.0179466	0.0215742
('Shank', 'TA', 'SOL')	0.022331	0.0203391	0.0202614	0.0243256	0.0171208	0.01693	0.0219236
('Shank', 'TA', 'BF')	0.0260757	0.0215889	0.0222783	0.02138	0.0180521	0.0171982	0.0226887
('Shank', 'TA', 'ST')	0.023873	0.0209574	0.0226188	0.0232452	0.0181987	0.0184145	0.0241913
('Shank', 'TA', 'VL')	0.0238502	0.0202885	0.0175827	0.0189754	0.0152085	0.0147869	0.018928
('Shank', 'TA', 'RF')	0.0374317	0.0221873	0.0193388	0.0227202	0.0174072	0.0171652	0.0221501
('Shank', 'MG', 'SOL')	0.0281929	0.0253981	0.0239365	0.0251963	0.0210093	0.0207318	0.027065
('Shank', 'MG', 'BF')	0.027953	0.0265425	0.0271264	0.0262419	0.0208484	0.0201752	0.0261969
('Shank', 'MG', 'ST')	0.0294261	0.0273278	0.0260464	0.0269725	0.0212907	0.0210329	0.0286177
('Shank', 'MG', 'VL')	0.0260587	0.0233712	0.0219164	0.0231026	0.0191729	0.0183052	0.0259626
('Shank', 'MG', 'RF')	0.0369171	0.024657	0.0268961	0.0257886	0.0210786	0.0201264	0.0288177
('Shank', 'SOL', 'BF')	0.0263169	0.0255251	0.022616	0.0254755	0.0212299	0.020279	0.0253935
('Shank', 'SOL', 'ST')	0.0265334	0.0262929	0.0262231	0.0260607	0.021424	0.0216924	0.0266321
('Shank', 'SOL', 'VL')	0.0197695	0.0186158	0.0214598	0.0236834	0.0180481	0.0177272	0.0229784
('Shank', 'SOL', 'RF')	0.0312282	0.0256613	0.0208473	0.0247581	0.0203835	0.020208	0.0267024
('Shank', 'BF', 'ST')	0.0340594	0.0278321	0.0278096	0.0271371	0.0220032	0.021651	0.0272995
('Shank', 'BF', 'VL')	0.0264556	0.023087	0.0227972	0.0235932	0.0192742	0.0185565	0.0249206
('Shank', 'BF', 'RF')	0.03595	0.0269215	0.0264344	0.026472	0.0201254	0.0198481	0.0274757
('Shank', 'ST', 'VL')	0.0230015	0.0239341	0.0238506	0.0239601	0.0193349	0.0188321	0.0257218
('Shank', 'ST', 'RF')	0.0344855	0.027823	0.024692	0.0299385	0.0211808	0.0212269	0.0289155
('Shank', 'VL', 'RF')	0.0429361	0.0235231	0.0241225	0.0260677	0.0195844	0.0195628	0.0265978

Tabla 29. RECM obtenido tras la selección por conveniencia de variables por validación cruzada.
(Marcha)

7 Discusión

7.1 Predicción de la actividad neuromuscular en la transición de sedestación a bipedestación

Desde una perspectiva global, se aprecia como existen diferencias en la precisión de la predicción según las combinaciones utilizadas dentro de cada modelo.

Centrando la atención sobre los algoritmos con mayor precisión, Árbol de decisión múltiple y Árboles extremadamente aleatorios, se aprecia que, tanto en el entrenamiento inicial como en los siguientes pasos de mejora de los modelos, de las veintiuna combinaciones disponibles, estos dos algoritmos logran los mejores resultados en un subgrupo de ellas, independientemente de la métrica de evaluación utilizada. Este subgrupo de combinaciones son las formadas por los músculos: TA/BF, TA/VL, SOL/BF, BF/VL, TA/MG y TA/SOL.

A partir de los datos mostrados en el apartado 6.1 de esta memoria se puede apreciar que, en el primer entrenamiento de los modelos, con el conjunto de datos completos, son los algoritmos de Árbol de decisión múltiple y Árboles extremadamente aleatorios los que obtienen mejores resultados, tanto para los valores de RECM obtenidos en el conjunto de datos de prueba como por validación cruzada. Teniendo en consideración el valor de la métrica de evaluación EAM obtenido sobre el conjunto de prueba se aprecia también que el algoritmo de K-vecinos próximos tiene una precisión en la predicción similar al Árbol de decisión múltiple y Árboles extremadamente aleatorios.

Tras la eliminación de los valores extremos mejoran los resultados de todos los modelos, pero especialmente para el algoritmo de K-vecinos próximos en la métrica de evaluación de RECM tanto sobre el conjunto de prueba como por validación cruzada.

Esto pone de relieve el importante efecto que tienen las muestras eliminadas del conjunto de datos sobre la precisión en la predicción de los algoritmos. Sin embargo, hay que tener cautela en la interpretación de estos datos como valores extremos o incorrectos ya que estos pueden en realidad representar la variabilidad existente en los patrones de activación entre individuos y repeticiones, característica de la electromiografía de superficie y los sensores inerciales. También podrían provenir dichos valores de las posibles variaciones en la ejecución del gesto, como pueden ser la velocidad o el patrón de movimiento. (44)

Debido a que esta variabilidad está presente en la práctica clínica es necesario optar por otras mejoras de los modelos que los hagan más resistentes a la presencia de valores extremos.

En la tabla 8 se puede apreciar como la selección recursiva aplicada a cada una de las veintiuna combinaciones da lugar a un número de variables seleccionadas muy dispar, en concreto, entre 8 y 61 variables. Esta selección de variables se aplica sobre el conjunto de datos al completo y se observa de nuevo que los algoritmos con mayor precisión según el RECM obtenido por validación cruzada son el Árbol de decisión múltiple y Árboles extremadamente aleatorios.

Sin embargo, comparando los mejores resultados obtenidos en esta mejora con los mejores resultados obtenidos en el primer paso del entrenamiento de los modelos sólo existe una mejoría en los resultados del Árbol de decisión múltiple aplicado sobre la combinación que incluye al TA y SOL.

Cabe destacar también de los datos obtenidos en esta mejora de los modelos la capacidad de predicción de las combinaciones de TA/BF y SOL/VL en los modelos de Árbol de decisión múltiple y Árboles extremadamente aleatorios. Utilizando 9 y 8 variables respectivamente tienen un rendimiento muy comparable a los mejores modelos con un número más elevado de variables con los dos algoritmos mencionados anteriormente. En concreto el algoritmo de Árboles extremadamente aleatorios aplicado sobre 9 variables de la combinación que incluye a los músculos TA y BF obtiene un valor muy similar al modelo de Árbol de decisión múltiple aplicado sobre la combinación que incluye a los músculos TA y SOL con 58 variables (Árbol de decisión múltiple TA/SOL RECM:0.0181155, Árboles extremadamente aleatorios TA/BF RECM:0.0185118).

Estos resultados ponen de relieve que existen un pequeño grupo de variables en algunas combinaciones de músculos capaces de condensar la información del resto de músculos de una forma más precisa y sobre los que podrían centrarse futuras mejoras de estos modelos.

De nuevo, tras la aplicación de la eliminación recursiva de variables, podemos observar como el grupo de combinaciones que obtuvo los mejores resultados en el primer entrenamiento de los modelos utilizando los algoritmos de Árbol de decisión múltiple y Árboles extremadamente aleatorios, siguen obteniendo los mejores resultados tras la aplicación de esta mejora.

El último paso llevado a cabo para la mejora de los modelos fue una eliminación de variables por conveniencia, en la que se eliminaron todas las variables extraídas del IMU excepto la utilizada para la detección del ciclo, en este caso, el componente X del giróscopo de la cintura. Los resultados reflejan que tomando como referencia el valor de RECM obtenido por validación cruzada, son los algoritmos de Árbol de decisión múltiple y Árboles extremadamente aleatorios los que consiguen mejores resultados de nuevo, sin embargo, no existe una mejoría en la precisión de su predicción respecto al entrenamiento inicial. Al haber eliminado todas las variables del IMU excepto una, podemos intuir que el peso de las variables cinemáticas para la predicción de la actividad muscular no es muy grande ya que no suponen una clara mejora de los modelos.

Analizando los datos de la métrica EAM se observa cómo no sufre grandes variaciones entre el entrenamiento inicial y los pasos de mejora de los modelos que incluyen el conjunto de datos completo, sin embargo, si tiene una mejoría importante al eliminar los valores extremos. Esta métrica otorga el mismo peso a todos los errores, en lugar de elevarlos al cuadrado como la métrica RECM. La métrica RECM da mayor peso a los errores de mayor magnitud. Partiendo de esta premisa y aunque no se ha calculado el valor de EAM por validación cruzada, podemos intuir que los modelos deben estar cometiendo errores de gran magnitud en algunas predicciones lo que aumenta su valor de RECM pero que el error medio es de menor magnitud.

En la tabla 28, que se presenta a continuación, se muestra la distribución de las variables a predecir:

Index	TA rms	MG rms	SOL rms	BF rms	ST rms	VL rms	RF rms
count	944	944	944	944	944	944	944
mean	0.0719411	0.0204979	0.0284612	0.0260305	0.0172221	0.0562884	0.0247382
std	0.0349745	0.0254997	0.030437	0.0306815	0.0125	0.0346428	0.014911
min	0.0142959	0.00303844	0.00331875	0.00509226	0.00505004	0.0106712	0.00523901
25%	0.0461494	0.00915593	0.015196	0.0126902	0.0109754	0.0327361	0.0171673
50%	0.0737035	0.0164314	0.0230709	0.0172222	0.0149125	0.0430499	0.0221897
75%	0.0911024	0.0275197	0.0308264	0.0314565	0.0208145	0.0703858	0.030435
max	0.602682	0.555127	0.526307	0.539054	0.245382	0.302767	0.300957

Tabla 30. Estadística descriptiva de las variables a predecir por el algoritmo.

Tomando como referencia los valores obtenidos de RECM por validación cruzada en este trabajo, parece que el nivel de precisión en la predicción no es suficiente, ya que, en algunos músculos como el gemelo interno (MG) y el semitendinoso (ST), el valor del error registrado es casi de la misma magnitud que el valor medio de los datos y es casi de la misma magnitud que el valor entre el primer y tercer cuartil. Al tratarse de datos cuyo uso es clínico, errores de esta magnitud no son asumibles para implementar ninguno de estos modelos en un producto final.

Pese a ello, ya que uno de los objetivos de este trabajo es implementar un modelo de predicción de la actividad muscular en la transición de sedestación a bipedestación en una aplicación web, se implementará el modelo obtenido por la selección recursiva de variables que contiene como input al algoritmo de Árboles extremadamente aleatorios las siguientes variables: 'Waist_Az_mean', 'Waist_Gx_std', 'TA_ssc', 'TA_svde', 'BF_mav', 'BF_rms', 'BF_ssi', 'BF_svde', 'BF_wfl' y tiene como salida la predicción del valor de RMS del gemelo interno, sóleo, semitendinoso, vasto lateral y recto femoral. Los parámetros utilizados en este modelo son: {bootstrap = False, criterion = mae, max_depth = 10, min_samples_leaf = 1, min_samples_split = 5, n_estimators = 300}. Se selecciona este modelo ya que es el modelo de menor complejidad capaz de obtener unos resultados muy parejos a modelos más complejos.

7.2 Predicción de la actividad neuromuscular en la marcha sobre superficie plana

Igual que en el caso de la transición de sedestación a bipedestación, desde una perspectiva global, se aprecia como existen diferencias en la precisión de la predicción según las combinaciones utilizadas dentro de cada modelo.

Centrando la atención sobre los algoritmos con mayor precisión, Árbol de decisión múltiple y Árboles extremadamente aleatorios, se aprecia que, tanto en el entrenamiento inicial como en los siguientes pasos de mejora de los modelos, de las veintiuna combinaciones disponibles, estos dos algoritmos logran los mejores resultados en un subgrupo de ellas, independientemente de la métrica de evaluación utilizada, formadas por los músculos: TA/VL, TA/RF, TA/MG, TA/BF, ST/VL, TA/SOL.

A partir de los datos mostrados en el apartado 6.2 de esta memoria se puede apreciar que, en el primer entrenamiento de los modelos, con el conjunto de datos completos, son los algoritmos de Árbol de decisión múltiple, Árboles extremadamente aleatorios y K-vecinos próximos los que obtienen mejores resultados, tanto para los valores obtenidos en el conjunto de datos de prueba como por validación cruzada en ambas métricas de evaluación.

Tras la eliminación de los valores extremos se puede observar como la precisión en la predicción de los modelos aumenta, tanto en los valores de EAM como RECM. En total se han eliminado 309 pasos, que representan un 11% del total de registros. Igual que en el caso de la transición de sedestación a bipedestación, hay que tener cautela con la eliminación de registros debido a la variabilidad existente en datos de estas características.(44)

Al realizar la eliminación recursiva de variables sobre los datos completos podemos apreciar cómo no existe una clara mejoría en los resultados obtenidos por los modelos que mejor rendimiento tenían en el entrenamiento inicial. A nivel de la métrica de evaluación EAM observamos que no existe mejoría en la precisión de la predicción. Igual que en el caso de la transición de sedestación a bipedestación, parece que la eliminación recursiva de variables no logra una mejoría similar a la eliminación de los datos extremos. Sin embargo, sí que se aprecia como los modelos Árbol de decisión múltiple y Árboles extremadamente aleatorios logran mantener su precisión con un número mucho más reducido de variables con las siguientes combinaciones de músculos: BF/RF, MG/BF y MG/SOL, aunque no son capaces de alcanzar los niveles más altos de precisión del conjunto.

Estos resultados muestran que es posible obtener valores de precisión similares con un número de variables menor, pero que el efecto de los valores extremos sigue influyendo en la precisión de forma importante.

El último paso llevado a cabo para la mejora de los modelos fue una eliminación de variables por conveniencia, en la que se eliminaban todas las variables extraídas del IMU excepto la utilizada para la detección del ciclo, en este caso, el componente Y del giróscopo de la tibia. Si comparamos los resultados obtenidos en el entrenamiento inicial, solo se observa mejoría en el modelo Árboles extremadamente aleatorios sobre la combinación TA y VL. Si lo comparamos con los resultados obtenidos con la eliminación recursiva de variables, observamos que existe una mejoría en los modelos de Árbol de decisión múltiple y Árboles extremadamente aleatorios para el conjunto de datos TA y VL.

En la tabla que se presenta a continuación se muestra la distribución de las variables a predecir:

Index	TA rms	MG rms	SOL rms	BF rms	ST rms	VL rms	RF rms
count	2805	2805	2805	2805	2805	2805	2805
mean	0.0490768	0.0159555	0.0243713	0.017506	0.0141464	0.036311	0.0189279
std	0.0399469	0.0226427	0.0288516	0.023241	0.0134042	0.0342073	0.0143794
min	0.00211425	0.00203276	0.00218715	0.00143076	0.00161219	0.00193882	0.00183257
25%	0.0219686	0.00632715	0.0111313	0.00765083	0.00714161	0.0119014	0.0106829
50%	0.0380318	0.0111824	0.0181347	0.0126131	0.0121535	0.0266057	0.0167826
75%	0.0732717	0.0216319	0.0287529	0.0204572	0.0185904	0.0480832	0.0249211
max	0.75631	0.697166	0.660153	0.630016	0.307741	0.37754	0.377634

Tabla 31. Estadística descriptiva de las variables a predecir por el algoritmo.

Tomando como referencia los valores obtenidos de RECM por validación cruzado en este trabajo, parece que el nivel de precisión en la predicción no es suficiente, ya que, en algunos músculos como el gemelo medial (MG) y el semitendinoso (ST), el valor del error registrado es casi de la misma magnitud que el valor medio de los datos y es casi de la misma magnitud que el valor entre el primer y tercer cuartil. Al tratarse de datos cuyo uso es clínico, errores de esta magnitud no se consideran asumibles.

A la luz de los resultados obtenidos y teniendo en consideración que ante la misma precisión es más conveniente escoger el modelo más sencillo, se aplicará en la aplicación web el modelo Árboles extremadamente aleatorios con los siguientes datos como input: componente Y del giróscopo de la tibia, variables derivadas de la señal de EMGs del TA y del VL en este orden. El modelo contiene los siguientes parámetros: {bootstrap = False, criterion = mse, max_depth = 90, min_samples_leaf = 1, min_samples_split = 5, n_estimators = 250}

7.3 Metodología del trabajo

Cabe destacar en la discusión de este trabajo que de todas las combinaciones de hiperparámetros disponibles para los modelos se han probado un número pequeño de las mismas, en concreto 10 combinaciones aleatorias de las distribuciones de parámetros definidas para cada modelo y que es muy probable que existan ajustes que no se han realizado que mejoren la capacidad de predicción de los modelos.

También es importante tener en cuenta que la selección recursiva de variables se ha llevado a cabo a través de un árbol de decisión aplicado sobre cada conjunto de datos, por lo que combinaciones obtenidas por otros algoritmos como la regresión logística o el Árbol de decisión múltiple no han sido aplicados. Otra alternativa que podría haber sido llevada a cabo para escoger las variables a incluir en el entrenamiento de los modelos es la utilización de técnicas de reducción de la dimensionalidad como son el Análisis de componentes principales o el Análisis Discriminante Lineal.

Durante el desarrollo de este trabajo no se ha extraído el valor de la métrica EAM por validación cruzada para comparar con el obtenido sobre el conjunto de datos de prueba, por lo que, en futuros desarrollos derivados de este trabajo será conveniente introducir el método de valoración cruzada para la métrica de evaluación EAM.

8 Conclusiones

Los resultados obtenidos en este trabajo ponen de relieve tres conclusiones principales:

1. De todos los algoritmos utilizados, para ambos movimientos, son el Árbol de decisión múltiple, los Árboles extremadamente aleatorios y K- vecinos próximos aquellos que logran mayor precisión.
2. Existen un grupo de combinaciones que consiguen en ambos movimientos la mayor precisión al aplicar los algoritmos mencionados en el punto 1. Esto resulta de especial interés ya que se trata de movimientos bien diferenciados y pese a ello tienen dichas combinaciones en común. Este resultado puede ayudar a guiar las combinaciones a tener en consideración a la hora de replicar este trabajo para movimientos distintos de la extremidad inferior.

Combinaciones musculares con mejores resultados en la transición de sedestación a bipedestación	Combinaciones musculares con mejores resultados en la marcha sobre superficie plana
TA/BF, TA/VL, SOL/BF, BF/VL, TA/MG y TA/SOL.	TA/VL, TA/RF, TA/MG, TA/BF, ST/VL, TA/SOL.

Tabla 32. Combinaciones musculares que obtienen los mejores resultados.

3. La información derivada de tres sensores, estos algoritmos de aprendizaje automático y los parámetros utilizados no son suficientes para realizar una predicción precisa de la actividad muscular de cinco músculos durante la marcha en superficie plana y en la transición de sedestación a bipedestación. Esta conclusión implica que a futuro es necesario analizar el efecto de aumentar el número de sensores, nuevas combinaciones de parámetros y/o otros algoritmos de aprendizaje automático.

Siendo el objetivo principal de este trabajo la identificación del mejor modelo para la predicción de la actividad muscular de cinco músculos sin fijar un valor mínimo de precisión se puede considerar que el objetivo se ha visto satisfecho. Sin embargo, las conclusiones de este trabajo ponen de relieve que no se ha logrado la precisión necesaria como para ser de utilidad clínica, por lo que desde esa perspectiva puede considerarse que no se ha visto satisfecho.

En cuanto a los objetivos específicos todos se han visto satisfechos con éxito, ya que la aplicación web realiza todos los puntos propuestos en ellos.

8.1 Líneas de futuro

En la planificación del trabajo se planteó realizar una exploración de todas las posibles combinaciones de sensores y explorar cual era el número mínimo de sensores que sería necesario incluir para obtener buenas predicciones. Debido al tiempo limitado y la problemática surgida durante la realización del trabajo se decidió optar por todas las combinaciones posibles entre un IMU y dos sensores

de EMGs, pero a futuro, sería interesante retomar la propuesta inicial ya que con los resultados obtenidos en este trabajo no son suficientemente buenos como para tener implicaciones clínicas.

En el trabajo actual para las dos actividades se han llevado a cabo predicciones unilaterales, sin embargo, sería interesante explorar también la posibilidad de predecir la actividad de una extremidad registrando la contralateral, pudiendo reducir aún más los costes y los tiempos de aplicación.

En la elaboración de este trabajo se ha implementado cada modelo de forma independiente, por lo que el siguiente paso en el futuro desarrollo de este trabajo podría ser agrupar predictores creando un conjunto de predictores capaces de realizar una predicción mejor.

8.2 Seguimiento de la planificación

La planificación del trabajo se siguió correctamente hasta la PEC2, pero en la PEC3 aparecieron inconvenientes no previstos en la planificación que obligaron a la reestructuración de la aplicación de los modelos. La metodología prevista ha sido la adecuada ya que ha recogido todos los elementos necesarios para una ejecución completa de un proceso de análisis de datos: desde el preprocesamiento de los datos hasta la implementación de los modelos entrenados en un servidor web. Sin embargo, ha habido dos elementos clave que han hecho necesaria la introducción de cambios para garantizar el éxito del trabajo:

- Mala previsión en la ejecución del preprocesamiento de las señales que ha obligado a realizar duplicidad de código para la implementación de ese apartado en el servidor web
- Mala previsión en el número de posibles combinaciones entre sensores que han provocado una reestructuración de los datos para hacer el trabajo viable en el tiempo que restaba disponible.
- Tras los problemas surgidos en la elaboración de la PEC3, el autor de este trabajo no se dio cuenta de que a través de la aplicación de técnicas de reducción de la dimensionalidad podría haber abordado el objetivo inicial del trabajo que era: Identificar los elementos imprescindibles para la predicción de la actividad neuromuscular durante la marcha en superficie plana y en la transición de sedestación a bipedestación y realizar la predicción de la actividad muscular a partir de dichos elementos.

Tras la entrega de la PEC3 se llevó a cabo la reestructuración necesaria para hacer viable el trabajo, siendo el resultado el presente trabajo.

En el Anexo 1 se adjunta el diagrama de Gantt que representa la planificación reestructurada tras la elaboración de la PEC3.

9 Glosario

CART: Algoritmo de Árbol de Regresión y Clasificación

EMA: Error medio absoluto

FINAL: Valor final

INIT: Valor inicial

MAV: Valor medio promediado

MAX: Máximo

MCF: Porción medial de la tibia con el fémur

MDF: Frecuencia mediana

MIN: Mínimo

MNF: Frecuencia media

RECM: Raíz del error cuadrático medio

RMS: Media cuadrática

SD: Desviación estándar

SE: Entropía de Shannon

SPE: Entropía espectral

SSC: Cambio de signo de la pendiente

SSI: Integral cuadrada simple

SVDE: Entropía de valor de descomposición único

TLU: Unidades lógicas de umbral

VAR: Varianza

WA: Amplitud de Willison

WL: Longitud de la onda

ZC: Cruce de cero

10 Bibliografía

1. Wikipedia contributors. Biomecánica. [Internet]. Wikipedia, The Free Encyclopedia. [cited 2021 Jun 5]. Available from: <https://es.wikipedia.org/wiki/Biomec%C3%A1nica>
2. Seeley MK, Son SJ, Kim H, Ty Hopkins J. Biomechanics Differ for Individuals With Similar Self-Reported Characteristics of Patellofemoral Pain During a High-Demand Multiplanar Movement Task. *Journal of Sport Rehabilitation*. 2020;
3. Lantz B. *Machine Learning with R*. . 3. Packt Publishing.; 2019.
4. Géron A. *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. 2nd ed. O'Reilly; 2019.
5. Hu B, Rouse E, Hargrove L. Benchmark datasets for Bilateral Lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals. *Frontiers in Robotics and AI*. 2018;5(FEB).
6. Molina-Molina A, Ruiz-Malagón EJ, Carrillo-Pérez F, Roche-Seruendo LE, Damas M, Banos O, et al. Validation of mDurance, A Wearable Surface Electromyography System for Muscle Activity Assessment. *Frontiers in Physiology*. 2020 Nov 27;11.
7. Pedregosa F. VG, GA, MV, TB, GOBM, PP, WR, DV, VJ, PA, CD, BM, PM, DE. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12(85):2825–30.
8. Acheron Analytics. How to measure the accuracy of predictive models. <http://www.acheronanalytics.com/acheron-blog/how-to-measure-the-accuracy-of-predictive-models>. 2018.
9. Sammut Claude and Webb GI, editor. Mean Absolute Error. In: *Encyclopedia of Machine Learning* [Internet]. Boston, MA: Springer US; 2010. p. 652. Available from: https://doi.org/10.1007/978-0-387-30164-8_525
10. Neill SP, Hashemi MR. Chapter 8 - Ocean Modelling for Resource Characterization. In: Neill SP, Hashemi MR, editors. *Fundamentals of Ocean Renewable Energy* [Internet]. Academic Press; 2018. p. 193–235. (E-Business Solutions). Available from: <https://www.sciencedirect.com/science/article/pii/B9780128104484000082>
11. Camargo J, Flanagan W, Csomay-Shanklin N, Kanwar B, Young A. A Machine Learning Strategy for Locomotion Classification and Parameter Estimation using Fusion of Wearable Sensors. *IEEE Transactions on Biomedical Engineering*. 2021;
12. Mohd Khairuddin I, Sidek SN, P.P. Abdul Majeed A, Mohd Razman MA, Ahmad Puzi A, Md Yusof H. The classification of movement intention through machine learning models: the identification of significant time-domain EMG features. *PeerJ Computer Science*. 2021 Feb 25;7:e379.
13. Little K, Pappachan BK, Yang S, Noronha B, Campolo D, Accoto D. Elbow motion trajectory prediction using a multi-modal wearable system: A

- comparative analysis of machine learning techniques. *Sensors* (Switzerland). 2021 Jan 2;21(2):1–16.
14. Schöllhorn WI. Applications of artificial neural nets in clinical biomechanics. Vol. 19, *Clinical Biomechanics*. Elsevier Ltd; 2004. p. 876–98.
 15. Raval V, Nguyen KP, Gerald A, Dewey RB, Montillo A. Prediction of individual progression rate in parkinson's disease using clinical measures and biomechanical measures of gait and postural stability. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. Institute of Electrical and Electronics Engineers Inc.; 2020. p. 1319–23.
 16. Brisson NM, Gatti AA, Damm P, Duda GN, Maly MR. Association of machine learning based predictions of medial knee contact force with cartilage loss over 2.5 years in knee osteoarthritis. *Arthritis & Rheumatology* [Internet]. 2021 Mar 24;art.41735. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/art.41735>
 17. Sharma D, Davidson P, Müller P, Piché R. Indirect estimation of vertical ground reaction force from a body-mounted ins/gps using machine learning. *Sensors*. 2021 Feb 2;21(4):1–19.
 18. Sarvestan J, Svoboda Z, Alaei F, Mulloy F. Analysis of whole-body coordination patterning in successful and faulty spikes using self-organising map-based cluster analysis: A secondary analysis. *Sensors* (Switzerland). 2021 Feb 2;21(4):1–14.
 19. Seshadri DR, Thom ML, Harlow ER, Gabbett TJ, Geletka BJ, Hsu JJ, et al. Wearable Technology and Analytics as a Complementary Toolkit to Optimize Workload and to Reduce Injury Burden. *Frontiers in Sports and Active Living*. 2021 Jan 21;2.
 20. Jiang Y, Hernandez V, Venture G, Kulić D, Chen BK. A data-driven approach to predict fatigue in exercise based on motion data from wearable sensors or force plate. *Sensors*. 2021 Feb 2;21(4):1–16.
 21. Zhou B, Wang H, Hu F, Feng N, Xi H, Zhang Z, et al. Accurate recognition of lower limb ambulation mode based on surface electromyography and motion data using machine learning. *Computer Methods and Programs in Biomedicine*. 2020 Sep 1;193.
 22. Pham MH, Warmerdam E, Elshehabi M, Schlenstedt C, Bergeest LM, Heller M, et al. Validation of a lower back “wearable”-based sit-to-stand and stand-to-sit algorithm for patients with Parkinson's disease and older adults in a home-like environment. *Frontiers in Neurology*. 2018 Aug 10;9(AUG).
 23. Neřuková S, Klempíř O, Krupička R, Dušek P, Kutílek P, Szabó Z, et al. The timed up & go test sit-to-stand transition: Which signals measured by inertial sensors are a viable route for continuous analysis? *Gait and Posture*. 2021 Feb 1;84:8–10.
 24. Rodríguez-Martín D, Samà A, Pérez-López C, Català A. Identification of Sit-to-Stand and Stand-to-Sit transitions using a single inertial sensor. In: *Studies in Health Technology and Informatics*. IOS Press; 2012. p. 113–7.
 25. Gonzalez S, Stegall P, Edwards H, Stirling L, Siu HC. Ablation analysis to selectwearable sensors for classifying standing, walking, and running. *Sensors* (Switzerland). 2021 Jan 1;21(1):1–21.

26. Olivares A, Ramírez J, Górriz JM, Olivares G, Damas M. Detection of (In)activity periods in human body motion using inertial sensors: A comparative study. *Sensors (Switzerland)*. 2012 May;12(5):5791–814.
27. Demir F, Bajaj V, Ince MC, Taran S, Şengür A. Surface EMG signals and deep transfer learning-based physical action classification. *Neural Computing and Applications*. 2019 Dec 1;31(12):8455–62.
28. Matheson Rittenhouse D, Abdullah HA, John Runciman R, Basir O. A neural network model for reconstructing EMG signals from eight shoulder muscles: Consequences for rehabilitation robotics and biofeedback. *Journal of Biomechanics*. 2006;39(10):1924–32.
29. Heller BW, Veltink PH, Rijkhoff NJM, Rutten WLC, Andrews L' BJ, Veltink PH. Reconstructing muscle activation during normal walking: a comparison of symbolic and connectionist machine learning techniques. Vol. 69, *Biol. Cybern.* 1993.
30. Bishop CM. Pattern recognition and machine learning. Jordan M., Kleinberg J., B.Schölkopf, editors. Springer; 2006. 1–4.
31. del Olmo M, Domingo R. EMG characterization and processing in production engineering. Vol. 13, *Materials*. MDPI AG; 2020. p. 1–28.
32. Wikipedia contributors. Inertial measurement unit [Internet]. Wikipedia, The Free Encyclopedia. [cited 2021 May 29]. Available from: https://en.wikipedia.org/w/index.php?title=Inertial_measurement_unit&oldid=1025660823
33. Maqbool HF, Husman MAB, Awad MI, Abouhossein A, Mehryar P, Iqbal N, et al. Real-time gait event detection for lower limb amputees using a single wearable sensor. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Institute of Electrical and Electronics Engineers Inc.; 2016. p. 5067–70.
34. Martín Nogueras A, Calvo Arenillas JL, Orejuela Rodríguez J, Barbero Iglesias FJ, Sánchez Sánchez C. Revista iberoamericana de fisioterapia y kinesiología. [Internet]. Vol. 2, *Revista Iberoamericana de Fisioterapia y Kinesiología*. Elsevier; 1999 [cited 2021 Apr 18]. 44–49. Available from: <http://www.elsevier.es/es-revista-revista-iberoamericana-fisioterapia-kinesiologia-176-articulo-fases-marcha-humana-13012714>
35. Spiewak C. A Comprehensive Study on EMG Feature Extraction and Classifiers. *Open Access Journal of Biomedical Engineering and Biosciences*. 2018 Feb 7;1(1).
36. Tkach D, Huang H, Kuiken TA. Study of stability of time-domain features for electromyographic pattern recognition. *Journal of NeuroEngineering and Rehabilitation*. 2010;7(1).
37. Phinyomark A, Limsakul C, Phukpattaranont P. A Novel Feature Extraction for Robust EMG Pattern Recognition. Vol. 1, *JOURNAL OF COMPUTING*. 2009.
38. Phinyomark A, Limsakul C, Phukpattaranont P. A Novel Feature Extraction for Robust EMG Pattern Recognition. Vol. 1, *JOURNAL OF COMPUTING*. 2009.
39. Galar D, Kumar U. Preprocessing and Features. In: *eMaintenance*. Elsevier; 2017.

40. Acharya UR, Fujita H, Sudarshan VK, Bhat S, Koh JEW. Application of entropies for automated diagnosis of epilepsy using EEG signals: A review. *Knowledge-Based Systems*. 2015 Nov;88.
41. Heller BW, Veltink PH, Rijkhoff NJM, Rutten WLC, Andrews L' BJ, Veltink PH. Reconstructing muscle activation during normal walking: a comparison of symbolic and connectionist machine learning techniques. Vol. 69, *Biol. Cybern.* 1993.
42. Jason Brownlee. 4 Automatic Outlier Detection Algorithms in Python [Internet]. *Machine Learning Mastery*. 2020 [cited 2021 Jun 6]. Available from: <https://machinelearningmastery.com/model-based-outlier-detection-and-removal-in-python/>
43. Jason Brownlee. Recursive Feature Elimination (RFE) for Feature Selection in Python [Internet]. *Machine Learning Mastery*. 202AD [cited 2021 Jun 6]. Available from: <https://machinelearningmastery.com/rfe-feature-selection-in-python/#:~:text=RFE%20is%20a%20wrapper%2Dtype%20feature%20selection%20algorithm.&text=This%20is%20achieved%20by%20fitting,specified%20number%20of%20features%20remains.>
44. Agostini V, Ghislieri M, Rosati S, Balestra G, Knaflitz M. Surface Electromyography Applied to Gait Analysis: How to Improve Its Impact in Clinics? *Frontiers in Neurology* [Internet]. 2020;11:994. Available from: <https://www.frontiersin.org/article/10.3389/fneur.2020.00994>

11 Anexos

11.1 Anexo 1. Contenido del repositorio del trabajo.

- `data_extraction.py`: contiene el código necesario para extraer la información de los archivos utilizados en este trabajo.
- `dataset_generation_s2s.py`: contiene el código necesario para generar la matriz de datos de las variables extraídas de las señales utilizadas en este trabajo de la transición de sedestación a bipedestación.
- `dataset_generation_step.py`: contiene el código necesario para generar la matriz de datos de las variables extraídas de las señales utilizadas en este trabajo de la marcha en superficie plana.
- `signal_processing.py`: contiene el código necesario para realizar el filtrado de las señales.
- `sit2stand_functions.py`: contiene las funciones necesarias para la detección de la transición de sedestación a bipedestación.
- `step_functions.py`: contiene las funciones necesarias para la detección de los ciclos de la marcha en superficie plana.
- `variable_extraction_s2s.py`: contiene el código para realizar la extracción de las variables de la transición de sedestación a bipedestación.
- `variable_extraction_step.py`: contiene el código para realizar la extracción de las variables de la marcha en superficie plana.
- `implementation.py`: contiene el código para la implementación de los archivos descritos en los puntos anteriores.
- `plotting.py`: contiene el código utilizado para la validación de los algoritmos de este trabajo a través de visualización.
- `model_application_s2s_2.py`: contiene el código del proceso completo de entrenamiento y evaluación sobre los datos de transición de sedestación a bipedestación.
- `model_application_step.py`: contiene el código del proceso completo de entrenamiento y evaluación sobre los datos de la marcha en superficie plana.
- `model_creation_functions.py`: contiene las funciones utilizadas para el entrenamiento de los modelos.
- `model_improvement.py`: contiene el código de los tres pasos llevados a cabo para la mejorar los modelos y la adquisición de sus métricas de evaluación.
- `multioutput_regression_step_def.pkl`: archivo que contiene el modelo entrenado para la predicción de la actividad muscular durante la marcha en superficie plana.
- `multioutput_regression_s2s_def.pkl`: archivo que contiene el modelo entrenado para la predicción de la actividad muscular durante la transición de sedestación a bipedestación.
- `data_exploration_s2s.py`: incluye fragmentos del código utilizados para la exploración de los datos y el estudio y aplicación de la eliminación de valores extremos en la transición de sedestación a bipedestación.

- `data_exploration_step.py`: incluye fragmentos del código utilizados para la exploración de los datos y el estudio y aplicación de la eliminación de valores extremos en la transición de sedestación a bipedestación.
- Memoria TFM – Carlos Jové -PDF.pdf: contiene la memoria del trabajo final del máster de Carlos Jové.
- README.md: contiene el contenido de este anexo.

11.2 Anexo 2. Contenido del repositorio de la aplicación web.

- Imagen.jpg: contiene la imagen que se muestra en la página de inicio de la aplicación web.
- app.py: archivo necesario para la creación de una aplicación web con aplicaciones internas.
- multiapp.py: archivo necesario para la creación de una aplicación web con aplicaciones internas.
- home.py: contiene la información que se presenta en la página de inicio de la aplicación web.
- app_lw.py: contiene el código de la ventana “Extracción de variables durante la marcha en superficie plana”.
- app_s2s_2.py: contiene el código de la ventana “Extracción de variables durante la transición de sedestación a bipedestación”.
- model_s2s.py: contiene el código de la ventana “Predicción de la actividad neuromuscular durante la transición de sedestación a bipedestación”.
- model_step.py: contiene el código de la ventana “Predicción de la actividad neuromuscular durante la marcha en superficie plana”.
- multioutput_regression_step_def.pkl: archivo que contiene el modelo entrenado para la predicción de la actividad muscular durante la marcha en superficie plana.
- multioutput_regression_s2s_def.pkl: archivo que contiene el modelo entrenado para la predicción de la actividad muscular durante la transición de sedestación a bipedestación.
- requirements.txt: contiene los requerimientos que necesita la aplicación web para funcionar.
- data_extraction_web.py: contiene el código necesario para extraer la información de la matriz de datos facilitada por el usuario.
- dataset_generation_s2s_web.py: contiene el código necesario para generar la matriz de datos de las variables extraídas de las señales introducidas por el usuario de la transición de sedestación a bipedestación.
- dataset_generation_step_web.py: contiene el código necesario para generar la matriz de datos de las variables extraídas de las señales introducidas por el usuario de la marcha en superficie plana.
- signal_processing.py: contiene el código necesario para realizar el filtrado de las señales introducidas por el usuario.
- sit2stand_functions_web.py: contiene las funciones necesarias para la detección de la transición de sedestación a bipedestación de los datos introducidos por el usuario.
- step_functions_web.py: contiene las funciones necesarias para la detección de los ciclos de la marcha en superficie plana de los datos introducidos por el usuario.
- variable_extraction_s2s_web.py: contiene el código para realizar la extracción de las variables de la transición de sedestación a bipedestación de los datos introducidos por el usuario.

- `variable_extraction_step_web.py`: contiene el código para realizar la extracción de las variables de la marcha en superficie plana de los datos introducidos por el usuario.
- `README.md`: contiene el contenido de este anexo.

11.3 Anexo 3. Diagrama de Gantt de la planificación definitiva de la elaboración del trabajo

