



**UNIVERSIDAD NACIONAL DE EDUCACIÓN A
DISTANCIA
ESCUELA TÉCNICA SUPERIOR DE INFORMÁTICA
MÁSTER UNIVERSITARIO EN INGENIERÍA Y CIENCIA
DE DATOS
TRABAJO FIN DE MÁSTER**

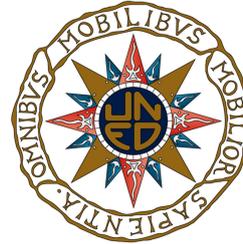
**Valoración cuantitativa de la marcha mediante
un sistema óptico de captura y análisis de
movimiento y técnicas de Aprendizaje
Automático**

Sergio González Velázquez

Director: Rafael Pastor Vargas

Co-Director: Iván González Díaz

**UNIVERSIDAD NACIONAL DE EDUCACIÓN A
DISTANCIA
ESCUELA TÉCNICA SUPERIOR DE INFORMÁTICA**



**MÁSTER UNIVERSITARIO EN INGENIERÍA Y CIENCIA
DE DATOS**

TRABAJO FIN DE MÁSTER

**Valoración cuantitativa de la marcha mediante
un sistema óptico de captura y análisis de
movimiento y técnicas de Aprendizaje
Automático**

Autor: Sergio González Velázquez

Director: Rafael Pastor Vargas

Co-Director: Iván González Díaz

septiembre, 2021

Trabajo Fin de Máster

© Sergio González Velázquez, 2021

Este documento se distribuye con licencia CC BY-NC-SA 4.0. El texto completo de la licencia puede obtenerse en <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

La copia y distribución de esta obra está permitida en todo el mundo, sin regalías y por cualquier medio, siempre que esta nota sea preservada. Se concede permiso para copiar y distribuir traducciones de este libro desde el español original a otro idioma, siempre que la traducción sea aprobada por el autor del libro y tanto el aviso de copyright como esta nota de permiso, sean preservados en todas las copias.

Este texto ha sido preparado con la plantilla \LaTeX de TFG para la UCLM publicada por [Jesús Salido](#) en GitHub¹ y Overleaf² como parte del curso « *\LaTeX esencial para preparación de TFG, Tesis y otros documentos académicos*» impartido en la Escuela Superior de Informática de la Universidad de Castilla-La Mancha.



¹https://github.com/JesusSalido/TFG_ESI_UCLM

²<https://www.overleaf.com/latex/templates/plantilla-de-tfg-escuela-superior-de-informatica-uclm/phjgscmfqtsw>

TRIBUNAL:

Presidente: _____

Vocal: _____

Secretario: _____

FECHA DE DEFENSA: _____

CALIFICACIÓN: _____

PRESIDENTE

VOCAL

SECRETARIO

Fdo.:

Fdo.:

Fdo.:

*A mis seres queridos
Por apoyarme y ayudarme a seguir creciendo*

Resumen

El estudio y valoración de la marcha humana es una importante herramienta diagnóstica que permite identificar anomalías y trastornos del aparato locomotor, en especial, en el adulto mayor. Tradicionalmente, el análisis de la marcha se ha realizado siguiendo un enfoque cualitativo, caracterizado por la aplicación de distintas pruebas y escalas basadas en observaciones visuales por parte de los profesionales sanitarios.

En este sentido, una de las escalas más completas y validadas es el test de Tinetti, que estudia de forma sistemática, tanto la marcha como el equilibrio, siendo una de sus principales aplicaciones la evaluación del riesgo de caídas en personas de avanzada edad. Además, a pesar de su fiabilidad y validez, el marcado carácter observacional de este método provoca que su aplicación conlleve procesos rutinarios que requieren de una evaluación subjetiva por parte de uno o varios expertos.

Dentro de este contexto, el presente Trabajo Fin de Máster pretende automatizar y objetivar los tradicionales análisis observacionales de la marcha, y en particular, el uso de la escala de Tinetti. Para ello, se analizan, aplican y evalúan diferentes técnicas de Aprendizaje Automático sobre los datos recopilados por un sistema óptico profesional para la captura y análisis de movimiento. Sin embargo, teniendo en cuenta que estos datos en crudo tienen una alta dimensionalidad y son bastante ruidosos, como paso previo es necesario realizar una importante fase de preprocesamiento y limpieza de los datos que posteriormente alimentan a los algoritmos de *Machine Learning*.

Finalmente, se obtienen varios modelos predictivos que son integrados y desplegados en una herramienta que, a partir de datos sensoriales de movimiento, es capaz de generar automáticamente las respuestas a las distintas categorías que constituyen el criterio de marcha en el test Tinetti.

Abstract

Human gait assessment is an important diagnostic tool for identifying abnormalities and disorders of the locomotor system. Typically, gait analysis has been conducted in a qualitative way, using different tests and scales based on visual observations carried out by healthcare professionals.

In this sense, one of the most complete and validated scales is the Tinetti test, which studies both gait and balance in order to, for example, measure the risk of falls in elderly people. Although this method is robust and valid, its observational nature makes its application a routine task that involves a subjective assessment from one or more experts.

Against this background, this Master's thesis aims to automate and objectify the traditional observational gait analysis, and particularly the use of the Tinetti scale. For this purpose, different Machine Learning techniques are analysed, applied and evaluated on the data recorded using an optical motion capture system. However, since this raw data has a high dimensionality and is very noisy, it is necessary to preprocess and clean the data that will serve as input for the Machine Learning algorithms.

Finally, various predictive models are integrated and deployed in a tool which, using motion data, is able to automatically generate the responses to the different categories of the Tinetti test gait criterion.

Agradecimientos

En primer lugar, quiero expresar mi infinito agradecimiento a José Bravo, Iván González y Jesús Fontecha, por darme su confianza y dejarme aprender de ellos, ya que sin su conocimiento y consejos no hubiera sido posible el planteamiento y desarrollo de este trabajo. Gracias Iván y Jesús, por todas las horas dedicadas a dirigir, revisar y mejorar el trabajo realizado. Son un ejemplo a seguir del que me gustaría seguir aprendiendo. También he de agradecer los comentarios, recomendaciones y críticas constructiva de Rafael Pastor.

Cómo no, tengo que dar las gracias a mi padres, por su enorme esfuerzo diario para que hoy pueda estar escribiendo estas líneas. Gracias por guiarme en todas las etapas de mi vida. Gracias también a mi entorno familiar más cercano: mi hermano, y mi abuela; y en general, a toda mi familia, por apoyarme y motivarme tanto en los momentos buenos como en los más complicados.

Gracias a Sandra, por quererme tanto y hacerme feliz. Gracias por creer en mí más que nadie y animarme a conseguir todo lo que me propongo. Tu apoyo constante durante todos estos años ha sido esencial para que pueda llegar hasta aquí. Este nuevo logro es en gran parte gracias a ti.

También me gustaría dar las gracias a mis antiguos compañeros de carrera y a todas las personas que me han acompañado durante esta etapa universitaria. En especial, gracias a David, con quien he pasado tantísimas horas de trabajo durante esta nueva experiencia. Gracias por compartir conmigo tu curiosidad, ganas de aprender e interminable lluvia de ideas.

Por último, y no por eso menos importante, también estoy muy agradecido a todas las personas que con un gran esfuerzo sacrificaron un valioso tiempo de sus ajetreadas vidas para participar voluntariamente en la fase de adquisición de datos. Sin su colaboración, este trabajo no habría sido posible. En particular, me gustaría dar las gracias a Cosmin, técnico del laboratorio, ya que su ayuda y disposición ha sido fundamental para manejar los sistemas de captura y análisis de movimiento.

Una vez más, me siento orgulloso de haber comprobado que con esfuerzo y perseverancia se puede conseguir todo lo que te propongas, y es que, *'el esfuerzo constante, y no la fuerza o inteligencia, es la clave para liberar todo nuestro potencial'* (Winston Churchill).

Sergio González Velázquez
Ciudad Real, 2021

Índice general

Resumen	IX
Abstract	XI
Agradecimientos	XIII
Índice de figuras	XIX
Índice de tablas	XXIII
Glosario de acrónimos	XXV
1. Introducción	1
1.1. Descripción del proyecto	2
1.2. Contexto del trabajo	3
1.3. Estructura del documento	3
2. Objetivos	5
2.1. Objetivo principal	5
2.2. Objetivos específicos	5
3. Antecedentes y Fundamentos Teóricos	7
3.1. Análisis de la marcha humana	7
3.1.1. Ciclo de la marcha	7
3.1.2. Parámetros espaciotemporales	9
3.1.3. Parámetros cinemáticos	12
3.1.4. Parámetros cinéticos	16
3.1.5. Normalización	17
3.2. Valoración de la marcha	18
3.2.1. Análisis observacional de la marcha	18
3.2.2. Análisis cuantitativo de la marcha	22
3.3. Marcha y Aprendizaje Automático	27

4. Herramientas y Metodología de Trabajo	29
4.1. Sistema óptico de captura y análisis de movimiento	29
4.1.1. Cámaras <i>OptiTrack</i>	29
4.1.2. Clinical 3DMA	31
4.1.3. Otras herramientas utilizadas durante la captura de datos	32
4.2. Herramientas <i>software</i>	33
4.2.1. Entorno de programación y librerías	33
4.2.2. Control de versiones	34
4.2.3. Documentación y gráficos	34
4.3. Metodología de trabajo	34
4.3.1. Metodología Fundamental para la Ciencia de Datos	35
4.4. Planificación	36
5. Evolución y Resultados	39
5.1. Recopilación de datos	39
5.1.1. Configuración del laboratorio y calibración del sistema	39
5.1.2. Diseño del experimento y participantes	41
5.1.3. Exportación de datos y construcción del <i>dataset</i>	46
5.1.4. Análisis exploratorio de datos sobre el <i>dataset</i> inicial	54
5.2. Preparación de datos	59
5.2.1. Eventos de marcha y trayectoria de marcadores	60
5.2.2. Creación de variables explicativas	64
5.2.3. Limpieza de datos: filtrado de zancadas	70
5.2.4. Escalar parámetros espaciotemporales	75
5.2.5. <i>Script</i> de preprocesamiento y <i>dataset</i> resultante	78
5.3. Modelado y evaluación	84
5.3.1. Análisis exploratorio de datos sobre el <i>dataset</i> preprocesado	85
5.3.2. Descripción del proceso de modelado	90
5.3.3. LAP1. El pie derecho no sobrepasa al izquierdo en la fase de balanceo	96
5.3.4. LAP2. El pie izquierdo no sobrepasa al derecho en la fase de balanceo	100
5.3.5. LAP3. El pie derecho no se levanta completamente	103
5.3.6. LAP4. El pie izquierdo no se levanta completamente	107
5.3.7. DC. Discontinuidad entre pasos	110
5.3.8. PM. Postura en la marcha	114
5.3.9. Resumen y discusión de resultados	117
5.4. Implementación	119
5.4.1. <i>Script</i> para la generación de informes con valoración de la marcha	119
5.4.2. Informe biomecánico de la marcha humana	123

6. Conclusiones y Trabajo Futuro	127
6.1. Objetivos alcanzados	127
6.2. Aplicación de competencias adquiridas en el máster	129
6.3. Trabajo Futuro	129
6.4. Dedicación y costes	130
Bibliografía	131
A. Repositorios del Proyecto	145
B. Librerías Python	147
C. Hoja de información y consentimiento informado	149
D. Resumen del <i>dataset</i>	153
E. Informe biomecánico de marcha humana	159
F. Dedicación y costes	169
F.1. Dedicación horaria	169
F.2. Presupuesto	170

Índice de figuras

1.1. Esquema general del proyecto	2
3.1. División funcional del ciclo de la marcha	8
3.2. Parámetros espaciotemporales, cinemáticos y cinéticos obtenidos generalmente en estudios de marcha humana	9
3.3. Parámetros espaciales de la marcha	10
3.4. Planos y ejes de referencia en análisis del movimiento	12
3.5. Movimientos sobre la articulación de la cadera y rodilla	13
3.6. Cinemática normal de la marcha	14
3.7. Desplazamiento del centro de gravedad durante la marcha en sentido vertical	15
3.8. Fuerza de reacción al suelo durante el contacto del pie opuesto	16
3.9. Fuerzas de reacción al suelo en las direcciones medio-lateral, anteroposterior y vertical	17
3.10. Escala de Tinetti	19
3.11. Clasificación de sistemas para la captura de movimiento	22
3.12. Análisis de la marcha con un sistema de captura óptico basado en marcadores reflectantes	23
3.13. Análisis de la marcha con sensores inerciales	24
3.14. Registro de la actividad muscular durante la marcha con electromiografía .	25
3.15. Estudio de la marcha humana mediante plantillas instrumentadas	26
3.16. Plataformas de fuerza colocadas de manera consecutiva para el análisis de la marcha	26
3.17. Ejemplos de siluetas al caminar	28
4.1. Cámara <i>OptiTrack Primex 13W</i>	30
4.2. Productos 3DMA	31
4.3. Interfaz de usuario de <i>Clinical 3DMA</i>	32
4.4. Materiales adicionales utilizados durante la captura de movimiento	33
4.5. Metodología Fundamental para la Ciencia de Datos	35
4.6. Diagrama Gantt del proyecto	38

5.1.	Configuración del laboratorio de captura de movimiento: vista esquemática	40
5.2.	Configuración del laboratorio de captura de movimiento: vista real	40
5.3.	Calibración del sistema de captura	41
5.4.	Puntos anatómicos y medidas antropométricas	43
5.5.	Modelo de ubicación de marcadores pasivos según el protocolo Helen Hayes	45
5.6.	Marcadores reflectantes colocados en uno de los participantes siguiendo el protocolo Helen Hayes	46
5.7.	Ejemplo de formatos exportados desde <i>Clinical 3DMA</i>	48
5.8.	Organización en directorios y archivos del conjunto de datos	49
5.9.	Sistema de coordenadas en una pantalla con resolución 1920x1080.	51
5.10.	Pantalla de base de datos en <i>Clinical 3DMA</i>	53
5.11.	Pantalla principal en <i>Clinical 3DMA</i>	53
5.12.	Visualización de las relaciones entre variables antropométricas del <i>dataset</i> diferenciadas por género	55
5.13.	Distribución de los marcadores en diferentes instantes de la captura	56
5.14.	Desplazamiento vertical del centro de gravedad	57
5.15.	Curvas cinemáticas a partir de los datos exportados	58
5.16.	Esquema de las tareas abordadas en la fase de preprocesamiento	59
5.17.	Ciclo de la marcha a partir del fichero de eventos	60
5.18.	Extracción de zancadas en el fichero de eventos	61
5.19.	Fichero con las trayectoria de marcadores en formato TXT	62
5.20.	Datos faltantes en la trayectoria los marcadores colocados en la espina ilíaca y en el talón de la extremidad derecha	62
5.21.	Interpolación de los datos faltantes en trayectoria de los marcadores colocados en la espina ilíaca y en talón de la extremidad derecha	63
5.22.	Suavizado mediante Savitzky–Golay de la trayectoria de los marcadores colocados en la espina ilíaca y en el talón	64
5.23.	Cálculo de parámetros temporales de la marcha	65
5.24.	Cálculo del ángulo de paso	67
5.25.	Cálculo de la altura de pie y talón durante la marcha	67
5.26.	Cálculo del ángulo de contacto y despegue	68
5.27.	Desplazamiento vertical y horizontal del centro de gravedad	69
5.28.	Posición del centro de gravedad en el eje de avance durante la marcha	71
5.29.	Trayectoria filtrada y sin filtrar del centro de gravedad, en la dirección de avance, para una marcha patológica	72
5.30.	Relación entre la posición del centro de gravedad en el eje de avance durante la marcha y la longitud de zancada	73
5.31.	Relación entre la posición del centro de gravedad en el eje de avance durante la marcha y el ángulo de paso	74

5.32. Valores atípicos en los datos cinemáticos	75
5.33. Esquema del flujo de operaciones llevadas a cabo por el <i>script</i> de preprocesamiento	80
5.34. Conjunto de datos preprocesado	81
5.35. Longitud de altura y paso y altura máxima de talón y pie en cada tipo de grabación	86
5.36. Duración promedio de las fases de balanceo y soporte en cada tipo de grabación	86
5.37. Mapa de calor con la correlación entre las variables espaciotemporales . .	87
5.38. Mapa de calor con la correlación entre en los parámetros cinemáticos relacionados con la amplitud angular	89
5.39. Muestreo estratificado en LAP1	96
5.40. Características más importantes en LAP1	97
5.41. Características más importantes en LAP2	101
5.42. Características más importantes en LAP3	104
5.43. Árbol de decisión para el problema LAP3 con parámetros espaciotemporales	106
5.44. Características más importantes en LAP4	108
5.45. Arquitectura de PMC para el problema de clasificación LAP4	110
5.46. Características más importantes en DC	112
5.47. Características más importantes en PM	115
5.48. Árbol de decisión para el problema PM con parámetros espaciotemporales	116
5.49. Evaluación en el conjunto de prueba (exactitud) de cada familia de algoritmos utilizando los cinco predictores espaciotemporales y cinemáticos más relevantes	118
5.50. Esquema de la generación de informes con valoración de la marcha	119
5.51. Entrada de datos en la herramienta de generación de informes	121
5.52. Funcionamiento de la herramienta para la generación de informes con valoración de la marcha	122
5.53. Información general incluida en los informes de marcha	123
5.54. Análisis detallado del ciclo de la marcha en los informes generados	124
5.55. Esquema para representar visualmente la longitud de paso y zancada en los informes generados	124
5.56. Esquema para representar visualmente la altura y ángulos de tobillo y pie en los informes generados	125
5.57. Esquema y tabla para representar el desplazamiento del centro de gravedad durante la marcha en los informes generados	125
5.58. Generación automática de las respuestas a la escala de Tinetti	126
F.1. Distribución del trabajo por tipo de tarea	170
F.2. Dedicación horaria y distribución de tareas por meses	170

Índice de tablas

4.1. Comparativa entre productos <i>OptiTrack</i>	30
5.1. Ubicación de los marcadores según el protocolo Helen Hayes modificado por STT	44
5.2. Etiquetado del conjunto de datos	48
5.3. Columnas del archivo <i>anthropometry.csv</i>	50
5.4. Resumen de características demográficas y antropométricas	54
5.5. Variables informativas calculadas para resumir los parámetros cinemáticos	70
5.6. Umbrales para el filtrado de zancadas en los datos cinemáticos	75
5.7. Contraste de normalidad para los parámetros espaciales de la marcha y medidas antropométricas	76
5.8. Coeficiente de correlación entre los parámetros espaciotemporales de la marcha y variables antropométricas	77
5.9. Normalización adimensional de los parámetros espaciotemporales	78
5.10. Variables predictoras basadas en los parámetros espaciotemporales	83
5.11. Variables predictoras basadas en los parámetros cinemáticos	84
5.12. Media y desviación estándar de las variables espaciotemporales para cada tipo de marcha en el conjunto de datos	85
5.13. Media y desviación estándar del rango articular (variables cinemáticas) para cada tipo de marcha en el conjunto de datos	88
5.14. Selección de características espaciotemporales y cinemáticas para el problema LAP1	97
5.15. Rendimiento y tiempo de entrenamiento para el problema LAP1	98
5.16. Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC) para el problema LAP1	98
5.17. Modelos seleccionados para determinar si el pie derecho supera al izquierdo durante la fase de balanceo (LAP1)	99
5.18. Selección de características espaciotemporales y cinemáticas para el problema LAP2	100
5.19. Rendimiento y tiempo de entrenamiento para el problema LAP2	101
5.20. Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC) para el problema LAP2	102

5.21. Modelos seleccionados para determinar si el pie izquierdo supera al derecho durante la fase de balanceo (LAP2)	103
5.22. Selección de características espaciotemporales y cinemáticas para el problema LAP3	103
5.23. Rendimiento y tiempo de entrenamiento para el problema LAP3	105
5.24. Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC) para el problema LAP3	105
5.25. Modelos seleccionados para determinar si el pie derecho se levanta completamente del suelo (LAP3)	106
5.26. Selección de características espaciotemporales y cinemáticas para el problema LAP4	107
5.27. Rendimiento y tiempo de entrenamiento para el problema LAP4	108
5.28. Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC) para el problema LAP4	109
5.29. Modelos seleccionados para determinar si el pie izquierdo se levanta completamente (LAP4)	110
5.30. Selección de características espaciotemporales y cinemáticas para el problema DC	111
5.31. Rendimiento y tiempo de entrenamiento para el problema DC	112
5.32. Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC) para el problema DC	113
5.33. Modelos seleccionados para determinar si se trata de una marcha discontinua (DC)	113
5.34. Selección de características espaciotemporales y cinemáticas para el problema PM	114
5.35. Rendimiento y tiempo de entrenamiento para el problema PM	115
5.36. Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC) para el problema PM	116
5.37. Modelos seleccionados para valora la postura durante la marcha (PM) . . .	117
6.1. Aplicación de competencias adquiridas en el Máster	129
B.1. Librerías <i>Python</i> utilizadas en el proyecto	147
D.1. Resumen del conjunto de datos	153
F.1. Horas de dedicación al TFM	169
F.2. Coste de personal	171
F.3. Coste total	171

Glosario de acrónimos

- AA** Aprendizaje Automático. 27
- AEMPS** Agencia Española de Medicamentos y Productos Sanitarios. 149
- ANOVA** *Analysis of Variance*. 91, 96, 97, 100, 101, 103, 104, 107, 108, 111, 112, 114, 115, 128
- API** *Application Programming Interfaces*. 34, 94, 147
- ASCII** *American Standard Code for Information Interchange*. 32
- AUC** *area under the curve*. 95, 99, 102, 105, 109, 112, 113, 116–118, 128, 129
- CE** Competencias Específicas. 129
- CG** Competencias Generales. 129
- CLI** *command-line interface*. 34, 119
- CNB** *Complement Naive Bayes*. 93, 99, 102, 109, 112, 113, 117, 118, 128, 129
- COG** *center of gravity*. 15, 16, 47, 57, 68, 71–74, 77, 79, 88, 120, 125
- CRISP-DM** *Cross Industry Standard Process for Data Mining*. 35
- CSV** *Comma Separated Values*. 32, 47, 48, 50, 51, 53, 57, 64, 68, 69, 71–73, 80, 81, 120, 121
- DGI** *Dynamic Gait Index*. 21
- DT** *Decision Tree*. 27, 93, 105, 109, 112, 116, 117
- ECTS** *European Credit Transfer and Accumulation System*. 36, 169
- EDA** *Exploratory Data Analysis*. 54, 85
- EM** Esclerosis Múltiple. 1, 21
- EMG** electromiografía. 25
- EVGS** *Edinburgh Visual Gait Score*. 21
- FMDS** *Foundational Methodology for Data Science*. 35
- FOV** *Field of View*. 30
- FPR** *false positive rate*. 95
- GAIT** *Gait Assessment and Intervention Tool*. 20
- GARS** *Gait Abnormality Rating Scale*. 20
- GARS-M** *Modified Gait Abnormality Rating Scale*. 20
- GRF** *ground reaction force*. 16, 17
- GUI** *Graphical User Interface*. 51, 52, 130
- HMM** *Hidden Markov Model*. 27
- HS** *heel strike*. 8, 11, 56, 60, 61, 65, 66, 68
- iEMG** electromiografía intramuscular. 25
- IMUs** Unidades de Medición Inercial. 24, 70
- ITSI** Instituto de Tecnologías y Sistemas de Información. 3, 29

- JSON** *JavaScript Object Notation*. 52, 79
- KDD** *Knowledge Discovery in Databases*. 35
- KNN** *K-Neighbors Classifier*. 27, 92, 99, 102, 105, 109, 112, 117, 128, 129
- LED** *Light Emitting Diode*. 23
- MAmI** *Modelling Ambient Intelligence*. 3, 25
- MEMS** sistemas microelectromecánicos. 24
- ML** *Machine Learning*. 27, 129
- MLP** *Multi-layer Perceptron*. 94, 109, 118, 129
- MSWS-12** *Multiple Sclerosis Walking Scale*. 21
- OGA** *Observacional Gait Analysis*. 18
- OGS** *Observational Gait Scale*. 21
- PCA** *Principal Component Analysis*. 28
- PDF** *Portable Document Format*. 119, 120, 122, 128, 129
- POMA** *Performance-Oriented Mobility Assessment*. 1, 18, 39, 42, 84, 100, 119, 130
- QGA** *Quantitative Gait Analysis*. 18
- RBF** *Radial Basis Function*. 93, 94, 117
- RF** *Random Forest*. 27
- RGPD** Reglamento General de Protección de Datos. 43, 149
- RNA** Redes Neuronales Artificiales. 94
- ROC** *Receiver Operating Characteristic*. 94, 95, 98, 104, 115, 118, 128, 129
- RVGA** *Rivermead Visual Gait Analysis*. 21
- sEMG** electromiografía de superficie. 25
- SF-GT** *Salford Gait Tool*. 21
- SVC** *Support Vector Classification*. 99, 102, 105, 106, 109, 110, 112, 113, 117
- SVM** *Support Vector Machine*. 27, 93, 118, 128, 129
- T25-FW** *Timed 25-Foot Walk*. 21
- TFM** Trabajo Fin de Máster. 3–5, 7, 22, 25, 28, 29, 31, 33–36, 78, 127, 129, 130, 169, 170
- TO** *toe-off*. 8, 11, 56, 60, 65, 66, 68
- TRC** *Track Row Colum*. 47, 52
- TUG** *Timed Up and Go*. 20
- UCLM** Universidad de Castilla-La Mancha. 3, 29
- UNED** Universidad Nacional de Educación a Distancia. 129
- WGS** *Wisconsin Gait Scale*. 21
- XLSX** *Microsoft Excel Open XML Spreadsheet*. 82

CAPÍTULO 1

Introducción

La marcha humana es un complejo proceso de locomoción que requiere la interacción de varios subsistemas anatómicos que trabajan coordinadamente para permitir el desplazamiento bípedo del cuerpo [1]. Ya sea en personas sanas o con alguna patología, el estudio de la marcha proporciona una gran información sobre la actividad neuromuscular y esqueleto-muscular [2], pues la forma de caminar constituye un claro indicador del estado de salud, discapacidad y necesidad de ayuda sociosanitaria en muchas enfermedades crónicas neurológicas como el Alzheimer, Parkinson o *Esclerosis Múltiple (EM)* [3]. Por tanto, su caracterización y diagnóstico es fundamental para identificar precozmente patologías y trastornos locomotores, así como prevenir posibles anomalías antes de que se desarrollen.

Actualmente, existen numerosas herramientas estandarizadas (pruebas y escalas) que se utilizan en el contexto de la práctica clínica neurológica para valorar la marcha humana y la capacidad funcional de las personas, con el objetivo de detectar alteraciones en componentes de la capacidad física como la movilidad, equilibrio o fuerza muscular [3]. A pesar de su viabilidad y validez, este tipo de evaluaciones se caracterizan, en su mayoría, por estar realizadas mediante procesos rutinarios y un marcado carácter observacional que conduce a resultados relativamente subjetivos, por lo que pueden depender del punto de vista del evaluador.

Una de las pruebas más completas y validadas es la Escala de Tinetti (*POMA, Performance-Oriented Mobility Assessment*) [4], que estudia de forma sistemática, a través de dos subescalas evaluadas independientemente, la marcha y el equilibrio. Debido a su carácter observacional, este método requiere una evaluación subjetiva en la que uno o varios profesionales atienden cómo el paciente realiza una serie de tareas básicas (levantarse de una silla, permanecer de pie, caminar a paso normal y rápido) para determinar, entre otras cuestiones, la base de sustentación, la postura durante la marcha, la simetría, fluidez o la longitud y altura del paso. De esta forma, se obtienen unas puntuaciones que se pueden utilizar para, entre otras aplicaciones, determinar precozmente el riesgo de caída de una persona anciana.

En contraposición a las características de estos métodos observacionales, la constante innovación en los sistemas de captura y análisis de movimiento permite obtener datos objetivos que describen, con una alta precisión, la cinemática, dinámica y otros aspectos de la marcha humana. Consecuentemente, en los últimos años han surgido numerosas investigaciones, apoyadas en el uso de este tipo de sistemas sensorizados, para ayudar a los profesionales de la salud a determinar una valoración más precisa de la marcha. En este contexto, resultan especialmente interesantes las técnicas de Aprendizaje Automático como potentes herramientas capaces de analizar grandes y complejos volúmenes de datos para detectar automáticamente patrones y relaciones inmersos en ellos.

Centrándose en la subescala de Tinetti para la valoración de la marcha, este proyecto persigue un enfoque híbrido entre ambas metodologías, bajo el cual se utilicen los datos sensoriales obtenidos con un sistema óptico de captura de movimiento para elaborar y evaluar uno o varios modelos predictivos orientados a la valoración automática y objetiva de la marcha humana. En este sentido, dichos modelos serán capaces de generar automáticamente las respuestas a las distintas categorías que constituyen el criterio de marcha en la escala de Tinetti, de forma que mediante técnicas de *Machine Learning* se podrá construir una herramienta de valoración precisa y objetiva de un test hasta ahora observacional y subjetivo.

1.1. DESCRIPCIÓN DEL PROYECTO

La motivación del presente proyecto consiste en combinar la captura y análisis de datos sensoriales con la aplicación de técnicas de Ciencia de Datos, a fin de construir una herramienta que permita automatizar, agilizar y objetivar la valoración de la marcha humana, y en particular, el empleo de una de las escalas funcionales más extendidas en el contexto de la práctica clínica: la escala de Tinetti. El esquema general de la Figura 1.1 representa gráficamente las principales actividades y tecnologías a utilizar en el proyecto que se pretende abordar.

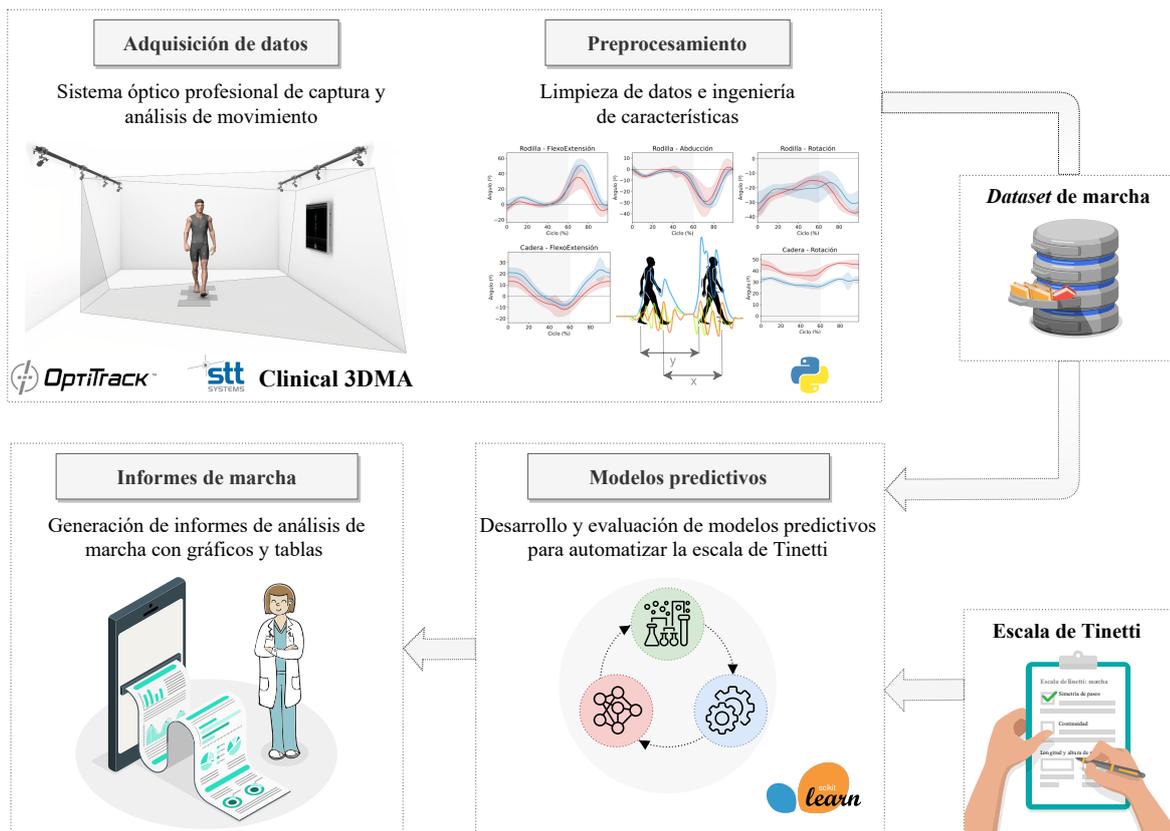


Figura 1.1: Esquema general del proyecto (elaboración propia).

El trabajo comenzará con la construcción de un *dataset* que sirva de línea base para la valoración automática de la marcha, utilizando como fuente de adquisición de datos un sistema de *tracking* óptico profesional para la captura y análisis de movimiento, basado en cámaras y marcadores reflectantes (*OptiTrack + Clinical 3DMA*). Después de la recopilación de datos inicial, se llevarán a cabo una serie de actividades para preparar el conjunto de datos

resultante, incluyendo tareas de limpieza de datos y procesos de ingeniería de características para crear variables explicativas adicionales, tales como parámetros espaciotemporales o curvas cinemáticas de la marcha a partir de los datos en crudo procedentes del sistema de sensorización.

Sobre los datos preprocesados, se aplicarán y evaluarán diversos algoritmos de *Machine Learning* con la intención de proponer uno o varios modelos predictivos que permitan generar automáticamente las respuestas a las distintas categorías que constituyen el criterio de marcha en la mencionada prueba de Tinetti. De esta forma, la implementación, combinación y despliegue de los modelos entrenados permitirá generar informes detallados con análisis cuantitativos de la marcha humana.

1.2. CONTEXTO DEL TRABAJO

Este Trabajo Fin de Máster (TFM) se ha realizado en el marco del grupo de investigación MAmI¹ de la Universidad de Castilla-La Mancha (UCLM), que centra una parte relevante de sus esfuerzos de investigación en la monitorización de ambientes inteligentes, con especial atención en aquellos contextos de aplicación relacionados con la salud y la atención sanitaria.

De esta forma, durante la realización del trabajo se han podido utilizar los recursos disponibles en el laboratorio técnico *WeCareLab*, situado en el Instituto de Tecnologías y Sistemas de Información (ITSI) de la UCLM. Este laboratorio cuenta con una potente infraestructura para la adquisición, seguimiento y procesamiento de datos de usuarios en contextos relacionados con la atención a la salud y al bienestar, y que ha sido financiada por el Ministerio de Ciencia e Innovación con el proyecto **EQC2019-006053-P**. Entre los recursos disponibles en el *WeCareLab* se encuentran sistemas profesionales de captura y análisis de movimiento (*Optitrack Motion Capture* y *Xsens 3D Motion Tracking*), electromiografía, electroencefalogramas y plantillas instrumentadas que han sido íntegramente diseñadas por miembros del grupo MAmI [5].

1.3. ESTRUCTURA DEL DOCUMENTO

El resto de la memoria de este TFM se organiza en una serie de capítulos con los contenidos que se describen a continuación:

Capítulo 2: Objetivos

En este capítulo se describe el objetivo principal, así como una serie de objetivos específicos, que se pretenden alcanzar al realizar este TFM.

Capítulo 3: Antecedentes y Fundamentos Teóricos

Este capítulo permite hacer una introducción al contexto teórico en el que se desarrolla este trabajo, es decir, se presentan y describen los procedimientos habituales en el análisis biomecánico de la marcha humana. Además, se incluye una revisión de estudios previos en los que se hayan aplicado técnicas de Aprendizaje Automático para el análisis de la marcha.

¹<https://mamilab.eu/>

Capítulo 4: Herramientas y Metodología de Trabajo

Este capítulo detalla las metodologías empleadas para la planificación y desarrollo del proyecto, así como las principales herramientas *software* y sistemas *hardware* utilizados durante dicho proceso de desarrollo.

Capítulo 5: Resultados

En este capítulo se describe cómo la aplicación de las herramientas y métodos de trabajo descritos en el Capítulo 4, han permitido abordar tanto el objetivo principal como los objetivos específicos planteados en este TFM.

Capítulo 6: Conclusiones y Trabajo Futuro

En este capítulo se realizará una valoración crítica de los resultados obtenidos, y se apuntarán algunas líneas de continuación a abordar a partir del trabajo desarrollado.

De manera adicional, se han incluido una serie de anexos que proporcionan información complementaria sobre el trabajo realizado.

CAPÍTULO 2

Objetivos

En este capítulo se describe el objetivo principal que se pretende alcanzar con el desarrollo del presente TFM, destacando el problema específico que trata de resolver. Además, se enumeran una serie de subobjetivos que se corresponden con las diferentes fases de un proyecto de Ciencia de Datos.

2.1. OBJETIVO PRINCIPAL

El **objetivo principal** de este proyecto consiste en analizar, aplicar y evaluar técnicas de Aprendizaje Automático para la valoración cuantitativa de la marcha humana, partiendo de una fuente de adquisición de datos basada en un sistema de captura de movimiento óptico y siguiendo el criterio de marcha en la prueba de Tinetti.

2.2. OBJETIVOS ESPECÍFICOS

De manera adicional al objetivo principal, enunciado en la sección anterior, se establecen varios **objetivos específicos**,

- **O1. Construir un *dataset* que sirva de línea base para la valoración automática de la marcha.** Utilizando como fuente de adquisición de datos un sistema óptico de captura de movimiento con marcadores reflectivos, se tendrá que construir un *dataset* básico a partir de las trayectorias de dichos marcadores colocados sobre los participantes mientras realizan algunos de los movimientos definidos en la escala de Tinetti para valoración de la marcha. En este contexto, como paso previo a la recogida de datos, el autor del TFM participará como sujeto piloto para familiarizarse con el sistema de *tracking* (*hardware* y *software*) y poder entender mejor los datos recogidos.

Además, con el objetivo de asegurar la homogeneidad y validez de los datos obtenidos, se deberá determinar y adoptar un protocolo estandarizado para la colocación de los marcadores en los sujetos de ensayo, así como valorar qué variables (antropométricas, demográficas, clínico-funcionales) se incluyen en el estudio. Finalmente, los datos se etiquetarán según las distintas categorías consideradas de la escala de Tinetti para valoración de la marcha (simetría, continuidad y altura del paso, trayectoria con o sin desviación, postura en la marcha, etc.).

- **O2. Preprocesar y limpiar los datos**, siguiendo una metodología adecuada para la depuración, descubrimiento y corrección de datos erróneos (*data cleansing*), ga-

rantizando así la calidad de los datos que servirán de entrada a los algoritmos de Aprendizaje Automático construidos en **O3**.

La preparación de datos implicará además un proceso de ingeniería de características para crear variables explicativas adicionales, a través de la combinación del conocimiento en el dominio y los datos en crudo proporcionados por el sistema de sensorización. De esta forma, se incluirán en el conjunto de datos características de más alto nivel como parámetros espaciotemporales de la marcha o valores angulares de las principales articulaciones durante el movimiento (cinemáticos).

- **O3. Aplicar diversas técnicas de *Machine Learning*** para desarrollar varios modelos predictivos que permitan automatizar la valoración cuantitativa de la marcha humana, siguiendo los criterios considerados en la mencionada prueba de Tinetti.

En este sentido, se aplicarán distintos métodos supervisados de Aprendizaje Automático sobre el conjunto de datos adquirido y etiquetado en **O1**, y preprocesado en **O2**. Por tanto, será necesario utilizar alguna técnica de reducción de la dimensionalidad o selección de características para identificar y eliminar las variables irrelevantes; entrenar varios modelos basados en diferentes algoritmos supervisados de *Machine Learning* y, ajustar sus hiperparámetros para perfeccionar los modelos, reduciendo el riesgo de sobreajuste y tratando de que sean más robustos en el conjunto de datos.

- **O4. Evaluar la capacidad predictiva de los modelos de Aprendizaje Automático** desarrollados en **O3**, con la finalidad de determinar su capacidad de generalización ante observaciones nuevas no vistas durante el entrenamiento y escoger los estimadores que mejor se adapten al conjunto de datos. Para ello, se deberán seleccionar las medidas de rendimiento más apropiadas para el problema en concreto.
- **O5. Desplegar e integrar los modelos de *Machine Learning* elaborados para la valoración automática de la marcha.** Se deberá desarrollar una herramienta que tome como entrada datos nuevos recogidos con el mismo sistema de sensorización para, aplicando los modelos predictivos que han sido desarrollados en **O3** y evaluados en **O4**, generar automáticamente detallados informes sobre análisis de la marcha.

Con la intención de facilitar el diagnóstico y caracterización de la marcha, estos informes contendrán una respuesta automática a la escala de Tinetti, resultado de integrar los modelos construidos y evaluados durante el trabajo, así como gráficos y tablas con otros parámetros y variables específicas del dominio.

Antecedentes y Fundamentos Teóricos

Este capítulo tiene como objetivo exponer los fundamentos teóricos del contexto en el que se enmarca el presente TFM y hacer una revisión de trabajos relacionados. Para ello, primero se describe el análisis biomecánico de la marcha humana, presentando los tipos de variables comúnmente estudiadas (espaciotemporales, cinemáticas y cinéticas), así como la instrumentación y técnicas disponibles para la adquisición de datos de movimiento. Se explican con especial detalle los sistemas basados en cámaras por ser los utilizados en este trabajo. Finalmente, se hace una revisión de estudios previos en los que se hayan aplicado técnicas de Aprendizaje Automático para el análisis de la marcha.

3.1. ANÁLISIS DE LA MARCHA HUMANA

La marcha humana se ha descrito tradicionalmente utilizando diferentes definiciones que van desde “sucesión de pasos”, “movimiento por sucesión alternante del apoyo de un pie o de los dos” hasta “locomoción bipodal con actividad alternante de los miembros inferiores y mantenimiento del equilibrio” [6]. En cualquier caso, esta forma tan característica de desplazarnos es una tarea funcional compleja que implica la interacción simultánea de múltiples factores, y requiere la coordinación de un elevado número de articulaciones del cuerpo, especialmente, de las extremidades inferiores [7, 8].

Como consecuencia de la alta complejidad que conlleva el análisis de la marcha humana y su naturaleza periódica, para su estudio se suele descomponer en una secuencia cíclica de movimientos (fases y periodos) a los que comúnmente se refiere como **ciclo de la marcha**. Generalmente, se considera como inicio del ciclo de la marcha el instante en el que el talón de uno de los pies golpea el suelo, extendiéndose hasta que el talón del mismo pie vuelve a tocar el piso. No obstante, el ciclo de la marcha puede definirse y segmentarse a partir de cualquier otro evento relevante [9].

3.1.1. Ciclo de la marcha

Durante un ciclo de la marcha cada extremidad atraviesa dos fases principales, conocidas como fase de soporte, o de apoyo (*stance*), y fase de balanceo, o de oscilación (*swing*). El apoyo comienza con el contacto inicial del pie con el suelo, mientras que el balanceo se asocia al tiempo en el que el pie está en el aire para el avance de la pierna [10]. A su vez, el ciclo de la marcha se subdivide en ocho periodos, determinados por la posición del miembro inferior: cinco dentro de la fase de soporte y tres en la fase de balanceo. Los periodos que conforman el ciclo de la marcha, presentados en la Figura 3.1, son [9-11]:

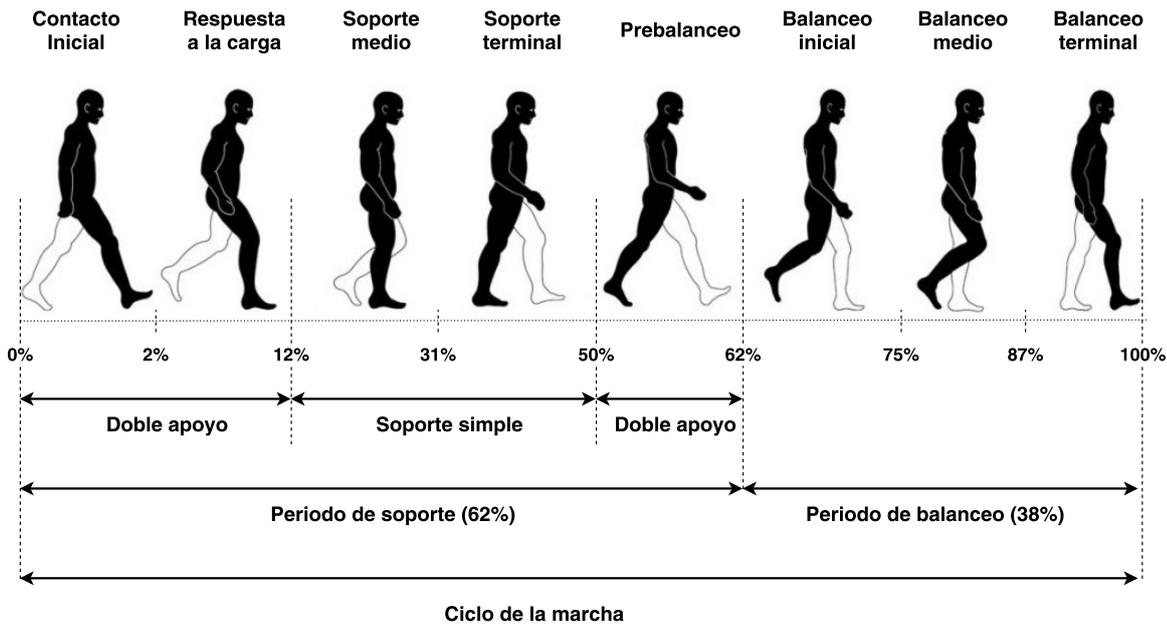


Figura 3.1: División funcional del ciclo de la marcha. Imagen elaborada a partir de [12]

- **Contacto Inicial.** Es el instante en el cual el talón del pie que se adelanta (sombreado en la Figura 3.1) hace contacto sobre el suelo (*heel strike*, HS). En esta fase, la cadera se encuentra flexionada y la rodilla extendida. Comprende del 0 % al 2 % del ciclo de la marcha.
- **Respuesta a la carga.** Es el intervalo en el que el pie hace contacto total con el suelo y el peso corporal es transferido a la pierna adelantada. Esta fase, que abarca en torno al 2-12 % del ciclo de la marcha, está incluida en el periodo de doble apoyo inicial.
- **Soporte medio.** Se inicia con la elevación del pie del miembro opuesto (claro en la Figura 3.1) y continua hasta que el peso corporal se transfiere al antepié (cabeza de los metatarsianos). Coincide con la primera mitad del soporte simple (monopodal) e incluye del 12 % al 31 % del ciclo de la marcha.
- **Soporte terminal.** El talón se empieza a elevar y la extremidad (sombreada) avanza sobre el antepié mientras la otra extremidad (clara) está completando su balanceo. Esta fase completa el intervalo de soporte simple y engloba del 31 % al 50 % del ciclo de la marcha.
- **Prebalanceo.** Esta última fase del soporte se corresponde con el segundo intervalo de apoyo doble en el ciclo de la marcha, y representa la transición hacia la fase de balanceo. Se inicia cuando el talón del pie opuesto entra en contacto con el suelo y termina cuando la extremidad sombreada despegar del piso. Cubre en torno al 50 %-62 % del ciclo de la marcha.
- **Balanceo inicial.** Se inicia en el instante en que los dedos del pie dejan de contactar el suelo (*toe-off*, TO) y termina cuando la rodilla alcanza la flexión máxima durante la marcha. Abarca del 62 % al 75 % del ciclo de la marcha.
- **Balanceo medio.** Comienza cuando la rodilla de la extremidad oscilante alcanza la flexión máxima y finaliza cuando la tibia se dispone en posición perpendicular con el suelo, es decir, se orienta verticalmente. Comprende del 75 % al 87 % del ciclo de la marcha.
- **Balanceo terminal.** Se inicia en la posición vertical de la tibia, y termina cuando el talón hace contacto con el suelo. Representa el último 13 % del ciclo de la marcha (del

87 % al 100 %).

De esta forma, cada ciclo completo de la marcha se normaliza en un intervalo de 0-100 %, haciendo el análisis independiente de cada individuo y generando así un patrón de marcha fácilmente comparable [13]. En este sentido, la distribución genérica normal de las fases del ciclo de la marcha se aproxima a un 60 % para el apoyo y un 40 % para el balanceo, pero hay que tener en cuenta que estas duraciones son relativas y varían a medida que aumenta o disminuye la velocidad de deambulación de la persona [10].

Por otro lado, el ciclo de la marcha humana se puede caracterizar empleando diferentes tipos de variables que evalúan distintos aspectos de la función neuromuscular [1]. De la forma más básica, la marcha se describe mediante parámetros espaciotemporales relacionados con la biomecánica de la marcha. Sin embargo, se obtiene una descripción aún más detallada del patrón de este movimiento al utilizar variables cinemáticas y cinéticas [14]. Las primeras, estudian el desplazamiento angular de las articulaciones del cuerpo, así como la inclinación, torsión y oscilación de los segmentos corporales. Por su parte, los parámetros cinéticos se ocupan de las fuerzas que producen o modifican dicho movimiento. En las siguientes secciones se presentan los principales parámetros espaciotemporales, cinemáticos y cinéticos que usualmente se analizan en estudios marcha, listados en la Figura 3.2.

PARÁMETROS DE LA MARCHA		
ESPACIOTEMPORALES	CINEMÁTICOS	CINÉTICOS
Temporales		
Tiempo en fase de soporte (ms)	Valores de los ángulos articulares con el paciente de pie (valores de <i>offset</i> para la cinemática)	Gráficas de momentos en las articulaciones (generalmente en el plano sagital)
Tiempo en fase de balanceo (ms)		
Tiempo de soporte (% del ciclo)	Plano frontal	
Tiempo de balanceo (% del ciclo)	Oblicuidad de la pelvis	Gráficas de potencias en las articulaciones (generalmente en el plano sagital)
Duración del ciclo o zancada (ms)	Aducción-abducción de cadera	
Cadencia (pasos/min)	Plano sagital	
Duración de doble soporte (ms)	Basculación de pelvis	Fuerza de reacción antero-posterior
Porcentaje de doble soporte (% ciclo)	Flexo-extensión de cadera	Fuerza de reacción medio-lateral
Espaciales	Flexo-extensión de rodilla	Fuerza de reacción vertical
Largo del paso (mm)	Dorsiflexión-plantiflexión de tobillo	
Longitud del ciclo o zancada (mm)	Plano transversal	Gráficas del centro de presión antero-posterior
Ancho del paso (mm)	Rotación pelvis	
Espaciotemporales	Rotación interna-externa de cadera	
Velocidad (m/s)	Rotación interna-externa de rodilla	
Velocidad de balanceo (m/s)	Ángulo de progresión del pie	
Velocidad media (m/s)		

*Las gráficas de rodilla y tobillo en el plano frontal no son tan significativas para el análisis, pero pueden obtenerse.

Figura 3.2: Parámetros espaciotemporales, cinemáticos y cinéticos obtenidos generalmente en estudios de marcha humana. Imagen tomada de [14]

3.1.2. Parámetros espaciotemporales

A partir de los eventos temporales y espaciales de las fases del ciclo de la marcha, se deducen una serie de variables espaciotemporales que los expertos utilizan para el análisis detallado de este modo de locomoción. En este sentido, los principales parámetros de la marcha relacionados con el tiempo son la duración de paso y zancada, el tiempo de apoyo, tiempo de balanceo, cadencia y velocidad. Por su parte, entre los parámetros espaciales que se deben explorar bilateralmente cuando se estudia la marcha humana, se encuentran la longitud de zancada, longitud de paso, ancho de paso y ángulo de paso. A continuación, se describen estas variables utilizando como referencia [9, 11, 15, 16].

Tiempo y longitud de zancada y de paso

Una zancada es el equivalente a un ciclo de la marcha. En este sentido, la duración o **tiempo de zancada** se corresponde con el intervalo entre dos contactos iniciales con el pie consecutivos, por parte de una misma extremidad inferior [10].

Asimismo, en cada zancada hay dos pasos que hacen referencia al cronometraje entre el contacto con el suelo de las dos extremidades. Por tanto, el **tiempo de paso** es el intervalo entre los contactos iniciales con el suelo de cada pie. A velocidad normal, se pueden observar diferencias en el tiempo de paso que no están relacionadas ni con la edad ni con la altura de los sujetos [17]. Por tanto, para poder realizar comparaciones sin que afecten las distintas velocidades de marcha de cada registro, el tiempo de paso se puede normalizar por el tiempo de zancada [15].

En este sentido, la **longitud de paso**, o paso corto, es la distancia lineal, en la dirección de progresión, entre los contactos pie-suelo sucesivos con la extremidad opuesta. Tradicionalmente, para registrar la longitud de paso se mide la distancia comprendida entre las dos líneas horizontales trazadas por el borde posterior de ambos talones (véase Figura 3.3).

En cuanto a la **longitud de zancada**, o paso largo, es la distancia lineal, entre los contactos pie-suelo sucesivos con la misma extremidad. Aunque una zancada está compuesta por dos pasos cortos (derecho e izquierdo), esto no significa que el valor de la longitud de zancada sea siempre el doble del valor de un paso corto, pues pueden existir asimetrías entre los pasos izquierdo y derecho [11]. Generalmente, para la valoración de la marcha se mide la longitud de zancada, registrando la distancia existente entre los bordes posteriores de los talones de dos huellas consecutivas de una misma extremidad, y se compara la longitud de zancada izquierda y derecha [9].

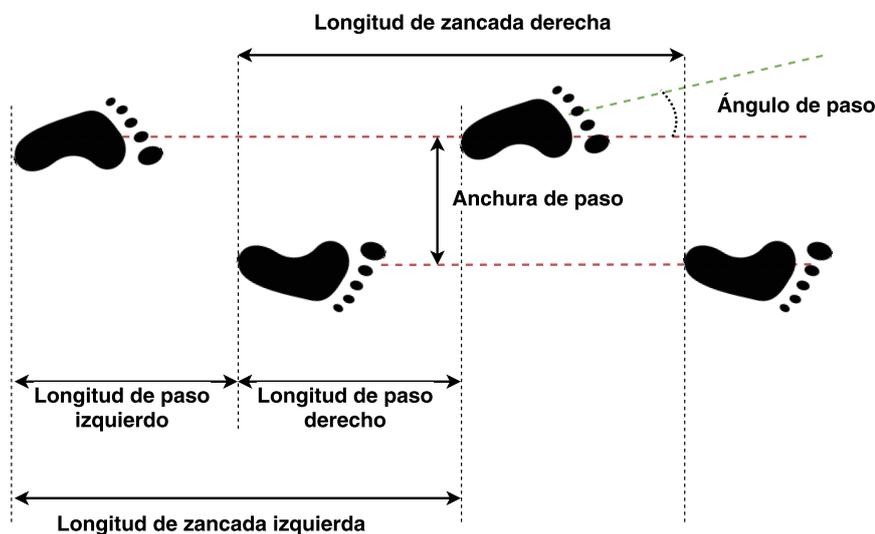


Figura 3.3: Parámetros espaciales de la marcha: longitud de paso, longitud de zancada, anchura de paso y ángulo de paso. (Elaboración propia).

Ancho de paso

La **anchura de paso** viene dada por la distancia lineal entre los apoyos de ambos pies. Generalmente, se mide como la distancia entre las dos líneas paralelas trazadas sobre el punto medio de ambos talones [9] (ver Figura 3.3). El ancho de paso es una variable directamente

relacionada con la estabilidad y el equilibrio, que se puede ver aumentada en pacientes con alteraciones neurológicas debido a que así se obtiene una base de apoyo mayor [11].

Ángulo de paso

El **ángulo de paso** se corresponde con el ángulo que se forma entre la línea que une los talones de dos pasos consecutivos de la misma extremidad y la línea central que divide al pie en dos mitades [9] (ver Figura 3.3). Normalmente, en personas no patológicas este ángulo oscila entre los 5° y 8°, pero algunas deficiencias en las extremidades inferiores pueden provocar el incremento o su inversión (hacia dentro), en cuyo caso se registra como un valor negativo [11].

Tiempo de apoyo y balanceo

El **tiempo de apoyo** es el cronometraje en el que un pie está en contacto con el suelo. Por tanto, es la duración del intervalo temporal que comienza con el contacto inicial del talón un pie con el suelo (HS) y termina con el contacto final del antepié de la misma extremidad con el piso (TO). Se puede segmentar en tiempo de apoyo bipodal (ambos pies están en contacto con el suelo) y monopodal (solo un pie está en contacto con el piso). Por otro lado, el **tiempo de balanceo** es el tiempo en el que un pie no está en contacto con el suelo, por lo que coincide en magnitud con el tiempo de apoyo monopodal del pie contrario [15].

Velocidad de marcha

La **velocidad de la marcha** define la relación entre la distancia recorrida por un sujeto en un tiempo determinado. Las personas modifican la velocidad mediante la variación conjunta de la longitud y tiempo de zancada, por lo que esta variable también se puede calcular como la relación entre la longitud y el tiempo de zancada. Por tanto, las fórmulas que definen la velocidad de la marcha son las siguientes [15]:

$$\text{Velocidad}(m/s) = \frac{\text{distancia recorrida}(m)}{\text{tiempo}(s)} \quad (3.1)$$

$$\text{Velocidad}(m/s) = \frac{\text{longitud de zancada}(m)}{\text{tiempo de zancada}(s)} \quad (3.2)$$

Cadencia de pasos

La **cadencia** es el número de pasos por unidad de tiempo que, generalmente, se fija en un minuto. Por tanto, es una variable que está directamente relacionada con la velocidad de la marcha. Puede variar de acuerdo con la longitud de las extremidades, el peso corporal y disminuye con el proceso de envejecimiento [9, 11]. En sujetos no patológicos caminando a una velocidad de entre 1 y 3 m/s, la cadencia oscila en torno a 100 y 155 pasos/minuto [16]. La cadencia de pasos se puede determinar mediante las siguientes fórmulas [15]:

$$\text{Cadencia de pasos}(pasos/s) = \frac{\text{número de pasos}}{\text{tiempo}(s)} \quad (3.3)$$

$$\text{Cadencia de pasos}(\text{pasos}/\text{min}) = \frac{\text{velocidad}(\text{m}/\text{s})}{\text{longitud de zancada}(\text{m})} * 120 \quad (3.4)$$

Como valores orientativos de los parámetros espaciotemporales de la marcha [16], en sujetos no patológicos caminando a velocidad espontánea la longitud de paso media es de aproximadamente 0,6 a 0,8 m; el ancho de paso no supera los 10 cm; el ángulo del pie oscila alrededor de los 7° y la cadencia puede variar entre 100 y 155 pasos/minuto. Además, a una velocidad de paso normal, el tiempo de apoyo es de 600 ms (60-62 % del ciclo) mientras que el tiempo de balanceo suele durar en torno a 125 ms (38-40 % del ciclo). En cualquier caso, es importante tener en cuenta que los parámetros espaciotemporales son variables que pueden variar en función de numerosos factores como la estatura, la edad, el sexo, la indumentaria, el calzado o de las estructuras neuromusculares y osteoarticulares [11].

3.1.3. Parámetros cinemáticos

El estudio cinemático de la marcha describe el movimiento articular sin considerar las fuerzas que lo causan o modifican [9]. Desde el punto de vista de la cinemática, la descripción de la locomoción se explica por medio de los tres ejes y planos anatómicos de movimiento, presentados en la Figura 3.4. Los ejes anatómicos sirven como referencia para ver cómo se realizan los movimientos: **eje longitudinal** o **vertical**, perpendicular al suelo; **eje transversal** u **horizontal**, paralelo al suelo; y **eje sagital** o **anteroposterior**, oblicuo al suelo y dirigido de atrás hacia adelante [18]. En relación con estos ejes se definen tres planos que sirven como referencia espacial para describir y ubicar los movimientos corporales [19]: **plano sagital**, que pasa por el eje longitudinal del cuerpo para dividirlo en una mitad derecha y otra izquierda; **plano coronal** o **frontal**, que divide el cuerpo en una parte anterior y otra posterior; y **plano transversal** u **horizontal**, que segmenta el cuerpo humano en una parte inferior y otra superior.

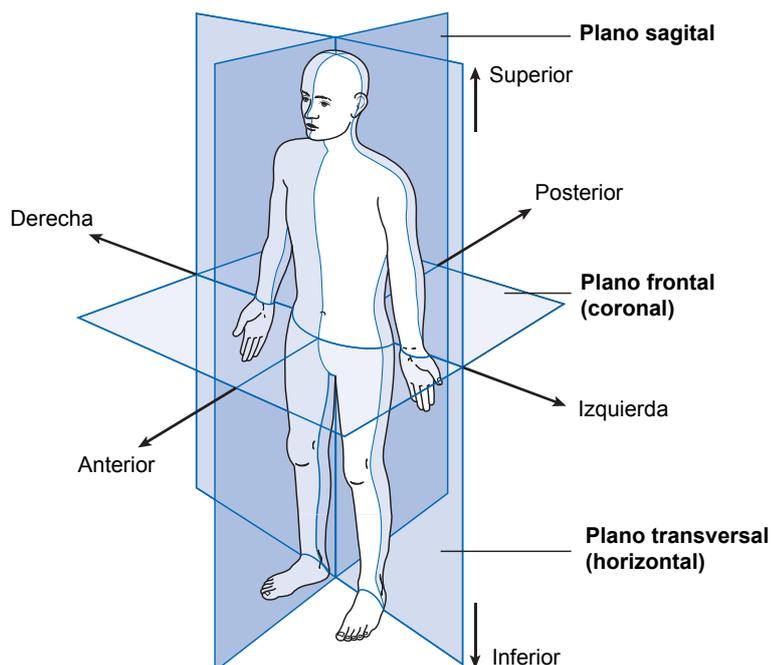


Figura 3.4: Planos y ejes de referencia en análisis del movimiento. Imagen tomada de [20].

En este sentido, entre las variables cinemáticas usualmente referidas en este tipo de análisis se incluyen los desplazamientos, velocidades y/o aceleraciones lineales y angulares de las articulaciones de la cadera, rodilla y tobillo, así como los segmentos del tronco, pelvis, muslo y pie [21] en cada uno de los tres planos.

Más concretamente, para las articulaciones de cadera, rodilla y tobillo, se estudian los movimientos en los planos sagital, frontal y horizontal, conocidos respectivamente como **flexión/extensión**, **abducción/aducción** y **rotación interna/externa**. La dirección de estos desplazamientos para la cadera y rodilla se muestra esquematizada en la Figura 3.5: la flexión implica una disminución del ángulo formado entre los huesos de la articulación (fémur y tibia para la rodilla) en dirección anteroposterior, paralela al plano sagital, mientras que el desplazamiento opuesto, la extensión, supone un aumento de dicho ángulo, y generalmente, se produce para enderezar una articulación tras una flexión previa; los movimientos de abducción y aducción se entienden como la separación y acercamiento del segmento analizado a la línea media del cuerpo trazada a lo largo del plano frontal (eje transversal); los movimientos de rotación son circulares e implican el movimiento del hueso alrededor de su eje longitudinal [18, 22].

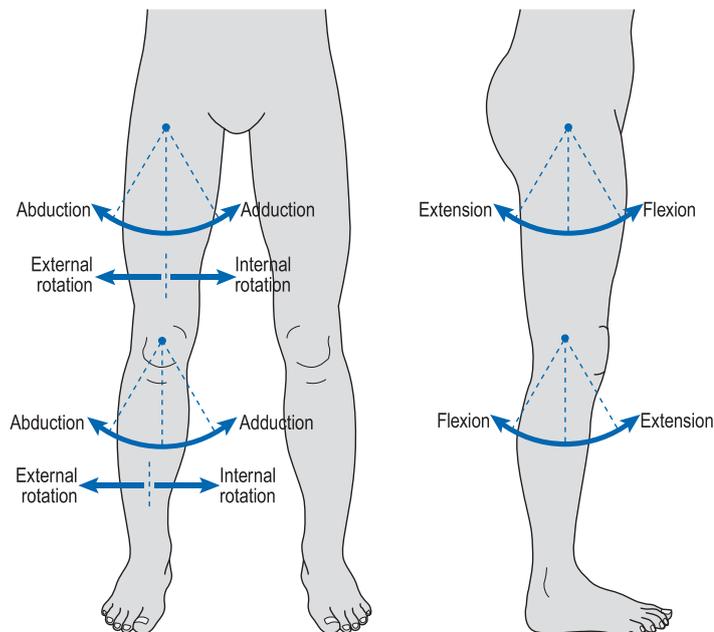


Figura 3.5: Movimientos sobre la articulación de la cadera y rodilla en los planos frontal y horizontal (izquierda), y sagital (derecha). Imagen tomada de [20].

Por otro lado, los desplazamientos analizados para la pelvis se refieren como **inclinación** (plano sagital), **oblicuidad** (plano frontal) y **rotación interna** y **externa** (plano horizontal). Por su parte, el desplazamiento del tobillo está limitado por la morfología de su articulación y solo permite los movimientos de **dorsi/plantiflexión** [8], que consiste en doblar el tobillo en la dirección del dorso o cara (dorsiflexión) o girar el tobillo en la dirección de la planta (flexión plantar) [18].

Generalmente, la cinemática articular de la marcha suele analizarse mediante una serie de diagramas que siguen una misma convención [21]: el eje horizontal indica la línea del tiempo y se representa como un porcentaje del ciclo de la marcha (0-100 %, considerando que el ciclo comienza con la fase de contacto inicial), mientras que el eje vertical representa los valores que toman las variables cinemáticas a lo largo del ciclo de la marcha. Como ejemplo, en la

Figura 3.6 se muestran los gráficos de la cinemática articular promedio en los planos sagital (primera columna), coronal (segunda columna) y transversal (tercera columna) para los casos de la pelvis, la cadera, la rodilla, el tobillo y el pie. En las gráficas de flexión/extensión, una curva ascendente indica un movimiento de flexión, mientras que una forma descendente señala que hay un movimiento de extensión [13]. De igual forma, una curva ascendente en la gráfica de abducción-aducción representa un desplazamiento de aducción (acercamiento) y una tendencia descendente indica un movimiento de abducción (alejamiento).

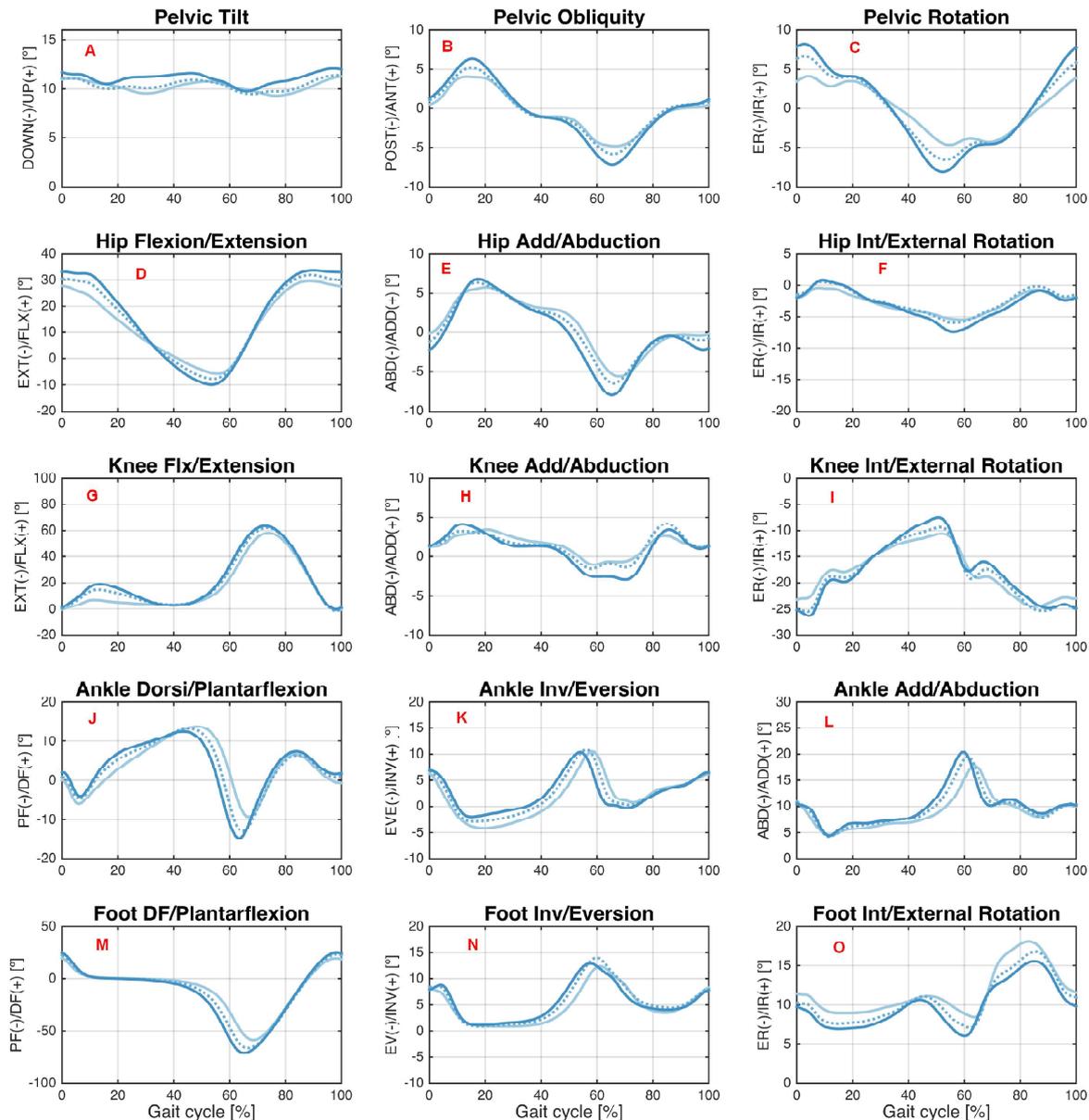


Figura 3.6: Cinemática normal de la marcha. Se muestran los diagramas de la cinemática articular promedio en los planos sagital (primera columna), coronal (segunda columna) y transversal (tercera columna) en el caso de la pelvis, la cadera, la rodilla, el tobillo y el pie. Imagen tomada de [23].

Desplazamiento del centro de gravedad durante la marcha

En relación con los movimientos que se producen en el cuerpo durante la marcha humana, es importante resaltar los desplazamientos horizontales y verticales del centro de gravedad corporal (COG, por sus siglas en inglés), ya que se relacionan directamente con el consumo energético de la persona [11] y puede ser un importante indicador de alteraciones en la marcha. Por ejemplo, caminar con un amplio ancho de paso o con movimientos exagerados genera mayores desplazamientos del centro de gravedad y, consecuentemente, un incremento proporcional del gasto energético [11].

El centro de gravedad se define como el punto en el que se aplica la fuerza peso de todo el cuerpo, por lo que es el punto alrededor del cual se equilibran las diferentes partes del cuerpo [13]. Aproximadamente, se calcula que en adultos sin patologías el COG se localiza a la altura de la primera o segunda vértebra sacra, unos 2 o 3 cm debajo del ombligo visto desde el plano frontal, aunque su ubicación varía en cada persona en función de sus características antropométricas [11], edad, sexo y postura adoptada [13]. Durante la marcha, el recorrido del centro de gravedad describe una leve curva sinusoidal, tanto en sentido vertical como horizontal [11], con excursiones de aproximadamente 3 cm en dirección vertical y 4 cm en dirección lateral [8].

La Figura 3.7 representa la trayectoria normal del centro de gravedad durante la marcha en sentido vertical (A) y horizontal (B). En sentido vertical, desde el contacto inicial con el suelo, el centro de gravedad se mueve de manera ascendente y lateral hasta la etapa de apoyo medio, en la que comienza a descender hasta un punto central. Con el contacto inicial de la extremidad opuesta, el COG se eleva de nuevo, alcanzando el pico de altura. Finalmente, con el segundo contacto inicial de la primera extremidad vuelve a descender una vez más [10]. Además, este rítmico desplazamiento vertical del COG está estrechamente relacionado con los apoyos bipodales, de forma que en los períodos de doble apoyo el centro de gravedad se encuentra en el punto más bajo [6].

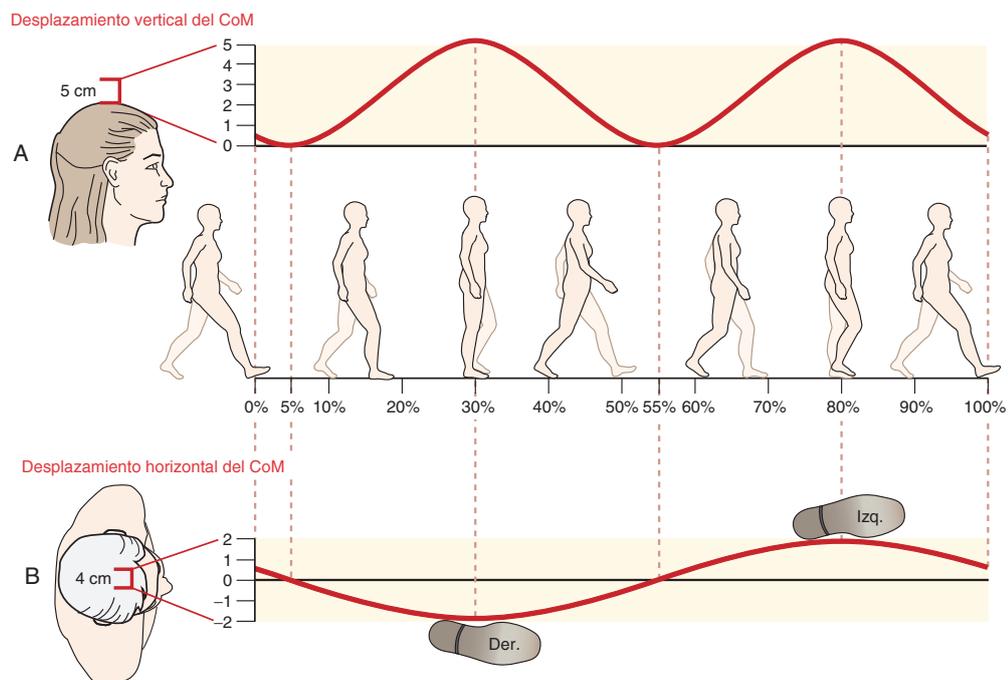


Figura 3.7: Desplazamiento del centro de gravedad durante la marcha en sentido vertical. Imagen tomada de [24]

Asimismo, durante la marcha, el COG se desplaza lateralmente desde la extremidad derecha a la izquierda, dando lugar a un sencillo patrón sinusoidal que oscila de lado a lado en cada ciclo de marcha (gráfico B de la Figura 3.7) [24].

3.1.4. Parámetros cinéticos

El patrón de movimiento de las extremidades que caracteriza el ciclo de la marcha incluye un conjunto de interacciones de fuerza generadas con la finalidad de sujetar, levantar y propulsar el cuerpo durante la trayectoria de progresión [10]. En este sentido, la cinética de la marcha estudia el movimiento teniendo en cuenta las fuerzas que lo producen o que lo modifican mediante el cálculo de momentos de fuerza, fuerzas de reacción y potencia [7]. El hecho de estudiar las fuerzas que generan el movimiento permite comprender y profundizar en las causas que provocan alternaciones en la marcha [25].

La fuerza más estudiada en el análisis de la marcha es la **fuerza de reacción al suelo** (*ground reaction force, GRF*), que es la ejercida por el suelo sobre el pie [26] según la ley de acción y reacción de Newton: durante la marcha se ejerce una fuerza contra el suelo, y éste a su vez lo hace en sentido contrario y con igual valor sobre la persona [13]. Puesto que la fuerza es una magnitud vectorial con dirección y sentido, la GRF se describe en dos dimensiones a partir de una componente vertical y una componente anteroposterior, tal como muestra la Figura 3.8. En tres dimensiones, aparece una componente medio-lateral adicional, de muy baja intensidad, que traduce los desplazamientos laterales del centro de gravedad [6].

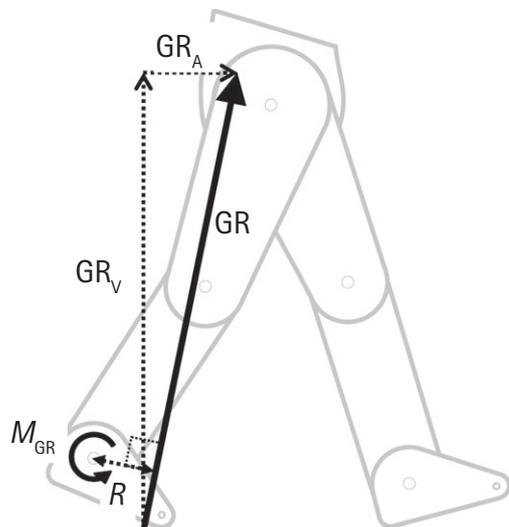


Figura 3.8: Fuerza de reacción al suelo durante el contacto del pie opuesto (GR) en sus componentes vertical (GR_V) y anteroposterior (GR_A). Imagen tomada de [26]

La Figura 3.9 representa las fuerzas de reacción al suelo durante la marcha en las direcciones medio lateral (A), anteroposterior (B) y vertical (C). En este tipo de gráficos el eje de ordenadas suele expresar la fuerza por unidad de masa corporal (N/kg), para comparar entre diferentes individuos sin que influya su masa [13]. En la dirección vertical, el gráfico de la GRF suele denominarse “diagrama de mariposa” por el patrón visual similar con las alas de estos insectos [13]. Con el contacto del talón, el cuerpo es impulsado hacia arriba para pasar por encima de la pierna que apoya, aumentando así la fuerza de reacción de suelo; cuando el cuerpo desciende para apoyarse en la pierna contraria disminuye su intensidad (valle);

su valor aumenta y aparece el segundo pico de la curva durante el impulso de despegue sobre los metatarsianos, para finalmente hacerse cero al despegar los dedos, ya que no hay contacto del pie con el piso [6, 13].

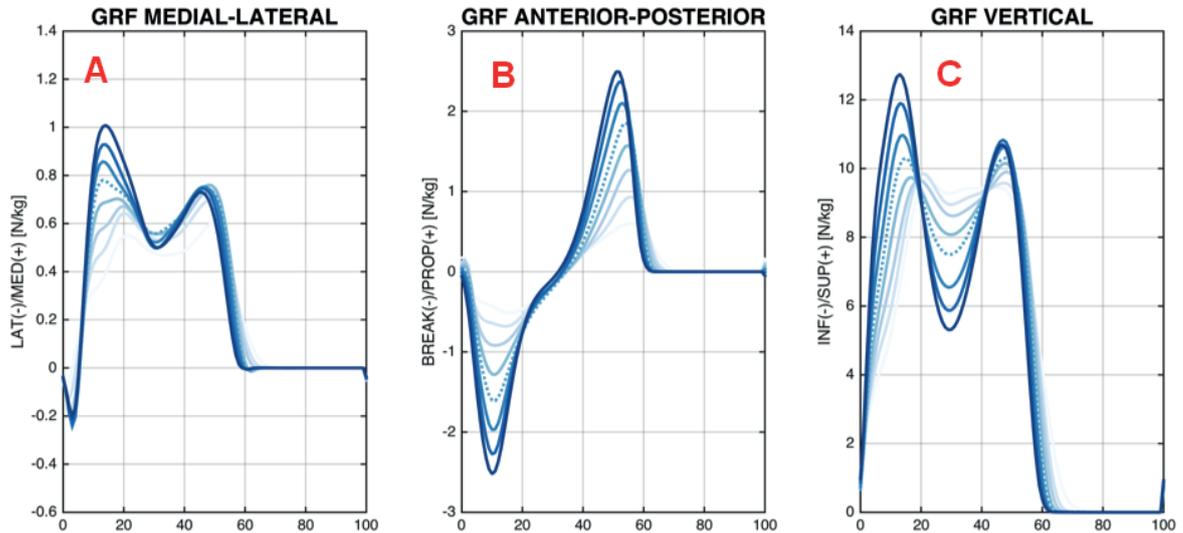


Figura 3.9: Fuerzas de reacción al suelo en las direcciones medio-lateral (A), anteroposterior (B) y vertical (C). Imagen tomada de [23]

Además de la **GRF**, el pie ejerce sobre el suelo una **fuerza de torsión**, de baja intensidad, relacionada con el movimiento de rotación de la extremidad inferior durante la marcha.

3.1.5. Normalización

La variabilidad en algunos parámetros espaciotemporales, como la longitud de zancada o la cadencia, es debida, entre otros factores, a las diferencias de altura entre sujetos. Igualmente, las variaciones de masa y momentos de inercia en los segmentos corporales pueden afectar a los datos cinéticos de fuerza o potencia. Por ello, los parámetros utilizados para evaluar la marcha humana son habitualmente normalizados por variables antropométricas, tales como el peso, la altura o la longitud de las extremidades inferiores. De esta forma, se anula la influencia de dichos factores y es posible comparar datos de individuos con diferentes medidas y masa corporal [26, 27].

Una forma aceptada de abordar la normalización de los parámetros de la marcha consiste en presentarlos de forma adimensional, es decir, como cantidades sin una dimensión física asociada. Teniendo en cuenta que una cantidad q puede expresarse como una magnitud sin dimensión \hat{q} si se divide por otra cantidad q_0 con las mismas unidades, en el contexto de la marcha humana se pueden dividir los distintos parámetros por la masa corporal m_0 , longitud de pierna l_0 y/o aceleración gravitatoria g , para conseguir variables adimensionales menos susceptibles de variabilidad según factores antropométricos [28].

En este sentido, para anular la influencia de la altura de las personas sobre la longitud de paso y de zancada, algunos investigadores las normalizan por la longitud del miembro inferior [29], o directamente por la altura de los sujetos [30]. La velocidad de marcha también suele normalizarse por la altura de los individuos y/o por la longitud de su extremidad inferior [31], mientras que variables temporales como el tiempo de apoyo [17, 29, 32, 33], balanceo [17, 29, 34] o tiempo de doble apoyo [17, 29, 34] se dividen por el tiempo de zancada.

3.2. VALORACIÓN DE LA MARCHA

El examen de la marcha consta de dos tipos de análisis o valoraciones: cuantitativo y subjetivo u observacional. Por un lado, el **análisis cuantitativo de la marcha** (*Quantitative Gait Analysis, QGA*) consiste en el estudio de los parámetros espaciotemporales, cinemáticos y/o cinéticos presentados en la sección anterior, así como cualquier otra información objetiva y cuantitativa que proporcione una descripción básica de la marcha de una persona [35]. Este tipo de exámenes puede ser tan sencillo como medir la longitud de los pasos con una cinta métrica o determinar la cadencia usando un cronómetro. Sin embargo, los avances tecnológicos han permitido el desarrollo de complejos laboratorios de análisis de movimiento dotados con distintos tipos de sistemas sensorizados y programas computarizados que ofrecen una gran precisión en la obtención y estudio de los parámetros de interés [11].

Por otro lado, el **examen observacional de la marcha** (*Observational Gait Analysis, OGA*) se refiere al enfoque cualitativo que identifica posibles alteraciones de la marcha mediante observaciones visuales, por lo que depende, en gran medida, del juicio de los especialistas observadores [36]. Para llevar a cabo este tipo de estudios han surgido pruebas clínicas y escalas que permiten conocer con gran detalle las alteraciones presentes en la marcha.

Este trabajo persigue un enfoque híbrido que permita objetivar los tradicionales análisis observacionales de la marcha mediante el uso de técnicas de Aprendizaje Automático alimentadas con los datos de los parámetros típicamente estudiados en el análisis cuantitativo de la marcha. En particular, se utiliza un sistema basado en cámaras para la adquisición de datos de movimiento, y se siguen las directrices de la extendida escala de Tinetti para la valoración observacional de la marcha. No obstante, en los siguientes apartados se hace una revisión de las distintas herramientas disponibles para la captura y análisis de la marcha humana (Sección 3.2.2), así como de las principales escalas utilizadas para la valoración funcional de la marcha (Sección 3.2.1).

3.2.1. Análisis observacional de la marcha

Existe una amplia bibliografía que evidencia que las pruebas observacionales de marcha son indicadores útiles de movilidad funcional, especialmente en personas mayores [37]. Además, algunas de estas pruebas evalúan otras capacidades funcionales como el equilibrio o la fuerza muscular, de forma que sus resultados se relacionan con la probabilidad de inestabilidad postural, discapacidad o riesgo de caídas. Sin embargo, no existe una escala ideal ni admitida en todas las culturas, sino que se han utilizado muchas dependiendo del tipo de estudio y de sus objetivos [3]. A continuación, se describen algunas de las pruebas de valoración de marcha más utilizadas.

Escala de Tinetti

La escala de Tinetti [4], también conocida como *Performance-Oriented Mobility Assessment (POMA)*, es una de las pruebas más extendidas y validadas para evaluar la capacidad funcional de personas mayores mediante la observación de movimientos habituales en situaciones de la vida cotidiana [38]. El test de Tinetti estudia la marcha y el equilibrio de forma sistemática, a través de dos subescalas evaluadas independientemente. La escala del equilibrio se evalúa sobre 16 puntos e incluye tareas como levantarse de una silla, permanecer de pie, sentarse

y equilibrarse tras un empujón. Por su parte, la subescala de la marcha se valora sobre 12 puntos y consiste en caminar, aproximadamente 8 metros a paso normal y volver a paso ligero, evaluando el inicio de la marcha, longitud y altura de paso, simetría y fluidez, trayectoria, desviación del tronco y postura [39]. La prueba de Tinetti completa se muestra en la Figura 3.10.

Test de Tinetti (Performance Oriented Mobility Assessment [POMA])	
Marcha	Equilibrio
Instrucciones	Instrucciones
El paciente permanece de pie con el examinador, camina por el pasillo o por la habitación (unos 8 m) a «paso normal» y luego regresa a «paso ligero pero seguro».	El paciente está sentado en una silla dura sin apoyabrazos. Se realizan las siguientes maniobras:
1. Iniciación de la marcha (inmediatamente después de pedirle que ande):	1. Equilibrio sentado:
a) Algunas vacilaciones o múltiples para empezar: 0.	a) Se inclina o se desliza en la silla: 0.
b) No vacila: 1.	b) Se mantiene seguro: 1.
2. Longitud y altura de paso:	2. Levantarse:
a) El movimiento del pie derecho no sobrepasa el izquierdo con el paso 0.	a) Imposible sin ayuda: 0.
b) El movimiento del pie derecho sobrepasa al izquierdo 1.	b) Capaz, pero usa los brazos para ayudarse: 1.
c) El pie derecho no se separa completamente del suelo con el paso 0.	c) Capaz de levantarse de un solo intento: 2.
d) El pie derecho se separa completamente del suelo 1.	3. Intentos para levantarse:
e) El movimiento del pie izquierdo no sobrepasa el derecho con el paso 0.	a) Incapaz sin ayuda: 0.
f) El movimiento del pie izquierdo sobrepasa al derecho 1.	b) Capaz pero necesita más de un intento: 1.
g) El pie izquierdo no se separa completamente del suelo con el paso 0.	c) Capaz de levantarse de un solo intento: 2.
h) El pie izquierdo se separa completamente del suelo con el paso 1.	4. Equilibrio en bipedestación inmediata (los primeros 5 s):
3. Simetría del paso:	a) Inestable (se tambalea, mueve los pies), marcado balanceo del tronco: 0.
a) La longitud de los pasos con los pies derecho e izquierdo no es igual 0.	b) Estable pero usa el andador, bastón o se agarra a otro objeto para mantenerse: 1.
b) La longitud parece igual 1.	c) Estable sin andador, bastón u otros soportes 2.
4. Fluidez del paso:	5. Equilibrio en bipedestación
a) Paradas entre los pasos: 0.	a) Inestable 0.
b) Los pasos parecen continuos: 1.	b) Estable pero usa el andador, bastón o se agarra a otro objeto para mantenerse: 1.
5. Trayectoria (observar el trazado que realiza uno de los pies durante unos 3m):	c) Estable sin andador, bastón u otros soportes 2.
a) Desviación grave de la trayectoria 0.	6. Empujar (el paciente en bipedestación con el tronco erecto y los pies)
b) Leve/moderada desviación o uso de ayudas para mantener trayectoria 1.	a) Empieza a caerse 0.
c) Sin desviación o ayudas 2.	b) Se tambalea y se agarra pero se mantiene 1.
6. Tronco:	c) Estable 2.
a) Balanceo marcado o uso de ayudas: 0.	7. Ojos cerrados (en la posición 6):
b) No se balancea, pero flexiona las rodillas o la espalda o separa los brazos al caminar 1.	a) Inestable 0.
c) No se balancea, no se reflexiona ni utiliza otras ayudas 2.	b) Estable 1.
7. Postura al caminar:	8. Vuelta de 360°:
a) Talones separados: 0.	a) Pasos discontinuos 0.
b) Talones casi juntos al caminar: 1.	b) Continuos 1.
	c) Inestable (se tambalea y agarra) 0.
	d) Estable 1.
	9. Sentarse:
	a) Inseguro, calcula mal la distancia, cae en la silla. 0.
	b) Usa los brazos o el movimiento es brusco. 1.
	c) Seguro, movimiento suave. 2.
Puntuación marcha: /12	Puntuación equilibrio: /16

Figura 3.10: Escala de Tinetti para la evaluación de equilibrio y marcha. Imagen tomada de [39]

Una de las principales aplicaciones de la escala de Tinetti es la evaluación del riesgo de caídas en personas de avanzada edad, pues diferentes estudios han demostrado que los pacientes que presentan anomalías en la marcha y el equilibrio, y consecuentemente, alcanzan

una puntuación baja en el test de Tinetti, tienen un riesgo más elevado de sufrir caídas. En particular, sobre los 28 puntos totales, una puntuación inferior a 19 es un indicativo de alto riesgo de caídas; entre 19 y 23 puntos sugieren un riesgo moderado y valores superiores a 24 indican un bajo riesgo de caídas [38, 40].

Timed Up and Go

La prueba de *Timed Up and Go* (TUG) es una herramienta para valorar la movilidad y la función del miembro inferior [39]. Inicialmente fue propuesta por (Mathias et al., 1986) [41], sin embargo, la versión más empleada es la cronometrada, descrita por (Podsiadlo et al., 1991) [42]. La prueba consiste en observar al paciente mientras se levanta de una silla, camina una distancia de 3 metros, gira y vuelve caminando a paso normal para sentarse de nuevo en la silla, cronometrando el tiempo que tarda en realizar la actividad completa.

El TUG también es un predictor de caídas. En este caso, se considera que el participante tiene un alto riesgo de caídas si tarda más de 20 segundos en completar la prueba, está en el límite de normalidad (riesgo moderado) si la actividad le conlleva entre 10 y 20 segundos, y se califica como normal si tarda menos de 10 segundos [40].

Gait Assessment and Intervention Tool

La escala *Gait Assessment and Intervention Tool* (GAIT) [43] es una prueba observacional de la marcha que valora la calidad del patrón locomotor, especialmente aspectos cinemáticos como la extensión de cadera o el movimiento de la rodilla y del tobillo en diferentes fases del ciclo. Está formada por 31 ítems organizados en tres categorías: 4 cuestiones evalúan la extremidad superior y el tronco; 14 valoran la extremidad inferior durante la fase de apoyo y los 13 restantes evalúan la extremidad inferior durante la fase de balanceo. Cada uno de estos ítems se puntúan en una escala numérica de 0 (normal) a 1,2 o 3 (máxima asimetría) dependiendo de la cuestión, de forma que la valoración máxima es de 62 puntos [44].

Algunas investigaciones han demostrado que esta herramienta es más sensible a las alteraciones en el patrón de la marcha que la escala de Tinetti, siendo apropiada para valorar la capacidad de marcha en pacientes que han sufrido un ictus [45].

Gait Abnormality Rating Scale

Gait Abnormality Rating Scale (GARS) [46] es una sencilla prueba basada en vídeos para evaluar posibles anomalías de la marcha. Fue específicamente diseñada para determinar el riesgo de caídas en personas de avanzada edad dentro de residencias de ancianos donde hay instrumentos y recursos limitados. No obstante, se puede utilizar para valorar diferentes aspectos de la marcha humana en cualquier otro entorno [47]. La escala está formada por 16 ítems organizados en tres categorías: 5 cuestiones generales, 4 ítems sobre extremidades inferiores, y 7 sobre tronco, cabeza y miembros superiores. Cada ítem se valora de 0 (sin anomalías) a 3 (deterioro funcional). En este sentido, la puntuación total de GARS se calcula como la suma de las 16 facetas individuales, de forma que una mayor puntuación implica un mayor riesgo de caídas.

Existe una versión modificada, *Modified Gait Abnormality Rating Scale* (GARS-M) [48] que reduce la escala original a 7 ítems.

Wisconsin Gait Scale

La escala *Wisconsin Gait Scale* (WGS) [49] es una prueba observacional que fue propuesta para evaluar los trastornos de la marcha experimentados por sujetos con hemiparesia¹ después de haber sufrido un ictus.

La WGS consta de 14 ítems que evalúan aspectos de la marcha humana: (1) uso de asistencia de apoyo durante la marcha, (2) tiempo de apoyo, (3) longitud de paso, (4) traslado del peso, (5) anchura del paso, (6) seguridad o vacilación en la trayectoria, (7) extensión de cadera, (8) rotación externa durante la oscilación inicial, (9) circunducción durante la oscilación, (10) elevación de la pelvis durante la oscilación, (11) flexión de rodilla, (12) separación del pie del suelo, (13) rotación de la pelvis en la oscilación final, (14) contacto inicial del pie. Cada uno de estos ítems se evalúa de 1 (normal) a 3 (atípico), excepto la primera cuestión (uso de asistencia de apoyo) que se puntúa de 1 a 5, y la undécima (flexión de rodilla) que se valora de 1 a 4 [50].

La puntuación mínima de la WGS es de 13.35 puntos y la máxima de 42 puntos, de forma que resultados más altos son indicadores de mayores alteraciones en la marcha.

Otras escalas observacionales

La importancia de analizar la marcha humana se traduce en la existencia de numerosas herramientas que tradicionalmente se han utilizado para evaluar alteraciones en la forma de caminar. Además de las descritas anteriormente, entre la bibliografía especializada se encuentran los siguientes test/escalas:

- *Rivermead Visual Gait Analysis* (RVGA). Evalúa las desviaciones cinemáticas de la marcha, siendo una medida habitualmente utilizada en personas que han sufrido accidentes cerebrovasculares [51].
- *Timed 25-Foot Walk* (T25-FW). Es un test cronometrado que permite cuantificar las alteraciones en la marcha sufridas por pacientes que padecen *Esclerosis Múltiple* [52].
- *Multiple Sclerosis Walking Scale* (MSWS-12). Es una herramienta diseñada con el mismo propósito que T25-FW, pero en este caso, en lugar de ser una prueba cronometrada, se trata de un cuestionario formado 12 ítems que determinan en qué grado ha afectado la EM a diferentes aspectos relacionados con la movilidad [53].
- *Observational Gait Scale* (OGS). Es una escala para valorar de manera estructurada el patrón de marcha a partir de grabaciones de vídeo. Según [54], esta prueba es de gran utilidad con niños y adultos jóvenes con parálisis cerebral, para los que es complicado llevar a cabo un análisis instrumentado de la marcha.
- *Dynamic Gait Index* (DGI). Se propuso como una herramienta clínica para evaluar la marcha, el equilibrio y el riesgo de caídas en personas con trastornos de vértigo, independientemente de la edad [55, 56].
- *Edinburgh Visual Gait Score* (EVGS). Se utiliza habitualmente como una puntuación validada de la calidad de la marcha en personas con parálisis cerebral [57, 58].
- *Salford Gait Tool* (SF-GT). Escala orientada al tratamiento de anomalías en la marcha en niños con parálisis cerebral [59].

De las diferentes escalas observacionales disponibles para valorar la marcha, este trabajo se centrará en la prueba de Tinetti, ya que actualmente es una de las más validadas,

¹La hemiparesia es una condición neurológica que dificulta el movimiento de una mitad del cuerpo como consecuencia de una lesión cerebral.

confiables y ampliamente utilizadas en la práctica clínica. Además, esta prueba ofrece una ventaja sobre otras evaluaciones, ya que también valora el equilibrio y permite descubrir la presencia de trastornos neurológicos o musculoesqueléticos [60]. En este TFM se pretende automatizar esta técnica observacional gracias a la potencia y precisión que ofrecen los sistemas tecnológicos para la captura y análisis de la marcha humana, descritos en el siguiente apartado.

3.2.2. Análisis cuantitativo de la marcha

Los laboratorios tradicionales y más sencillos para el estudio de la marcha humana simplemente obtienen parámetros espaciotemporales mediante la utilización de cintas métricas, cronómetros y un análisis visual sistemático por parte de los profesionales [14]. Sin embargo, en los últimos años han surgido diferentes sistemas y equipos tecnológicos que permiten obtener datos objetivos para evaluar de forma cuantitativa y eficaz la marcha a través de la cinemática, dinámica y otros aspectos como la actividad eléctrica de los músculos durante el movimiento.

En este sentido, cuando se utiliza esta tecnología altamente especializada para el análisis de la condición clínica de marcha se habla de **análisis instrumentado** de la marcha [61], siendo los métodos más usados los basados en cámaras, sensorización inercial, electromiografía, plantillas con sensores de presión y plataformas de fuerza. Esta clasificación se esquematiza en la Figura 3.11. A continuación, se describen y comparan las principales técnicas disponibles para el estudio de la marcha, comenzando por los sistemas ópticos de captura de movimiento por ser los utilizados en el presente proyecto.

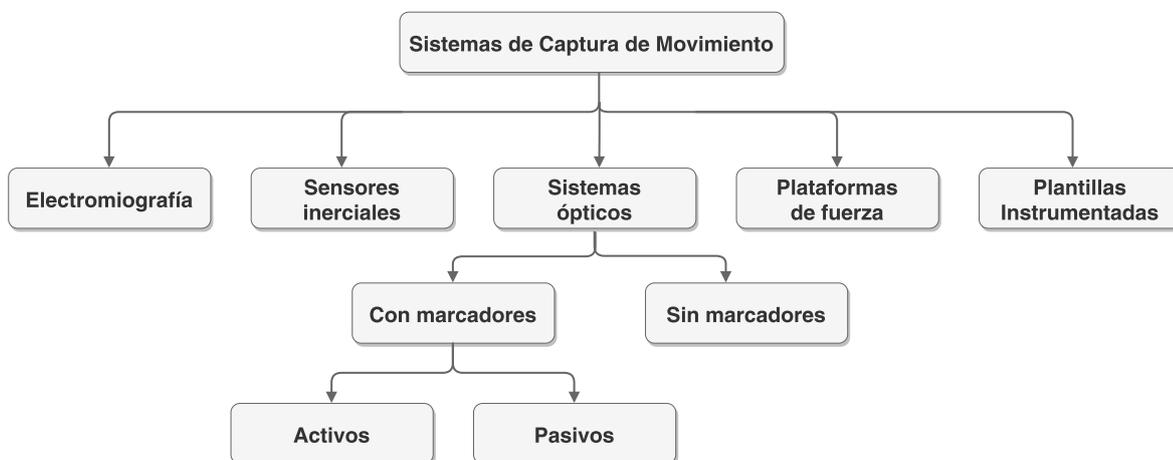


Figura 3.11: Clasificación de sistemas para la captura de movimiento. Imagen basada en [62]

Sistemas de análisis basados en cámaras

Los sistemas ópticos para el análisis de movimiento utilizan los datos recogidos por sensores de imagen para inferir la posición de los segmentos corporales en el espacio y adquirir así las características del movimiento de una persona con un grado de precisión relativamente alto [62]. Los primeros sistemas ópticos simplemente utilizaban la fotografía y el vídeo como métodos básicos para la captura y análisis del movimiento. Sin embargo, la mayoría de los sistemas ópticos actuales están basados en marcadores reflectantes (pasivos) o activos, ubicados en puntos anatómicos concretos. Independiente del tipo de marcador utilizado, los

sujetos de ensayo caminan sobre un área delimitada por la zona captura de las cámaras, y el sistema devuelve las coordenadas tridimensionales de cada uno de los marcadores [16].

En este sentido, el procedimiento por el que se determina la posición de una serie de marcadores en el espacio en tiempo real se conoce como **estereofotogrametría**: conociendo la ubicación y orientación de varias cámaras con respecto a un centro de coordenadas teórico, los pares de coordenadas (2D) con que cada cámara registra un punto determinado (marcador), se pueden integrar para obtener la posición 3D de ese punto con respecto a dicho sistema cartesiano [63]. De esta forma, a partir de la posición tridimensional de varios marcadores colocados en puntos anatómicos concretos, los sistemas ópticos de captura de movimiento estiman la posición de las articulaciones y la orientación de los segmentos corporales (Figura 3.12). Por tanto, el número de cámaras favorece a la precisión del movimiento capturado y la facilidad para procesar los datos obtenidos, siendo necesarios entre 6 y 8 dispositivos para capturar de manera precisa el movimiento que se produce a las velocidades típicas de la marcha humana [10].

En lo que se refiere a tipos de marcadores, los **sistemas ópticos pasivos** utilizan marcas reflectantes, de forma que las lentes de cada cámara están rodeadas por diodos emisores de luz infrarroja que emiten luz pulsátil. Esta luz infrarroja se refleja en los marcadores y regresa a la lente de la cámara, y las coordenadas de los puntos, que describen las trayectorias de los marcadores se transmiten digitalmente a un ordenador para su procesamiento [10]. En la Figura 3.12 se muestra el uso de un sistema óptico pasivo para la captura de movimiento. Por otro lado, en los **sistemas ópticos activos** (optoeléctricos) cada marcador es un diodo emisor de luz LED que está activado por una fuente eléctrica. Aunque los principios generales de colocación de marcadores y el tratamiento de los datos es similar en ambos tipos de sistemas, la dependencia de un aporte de energía y los cables transmisores necesarios son limitaciones de los sistemas ópticos activos respecto a los pasivos.

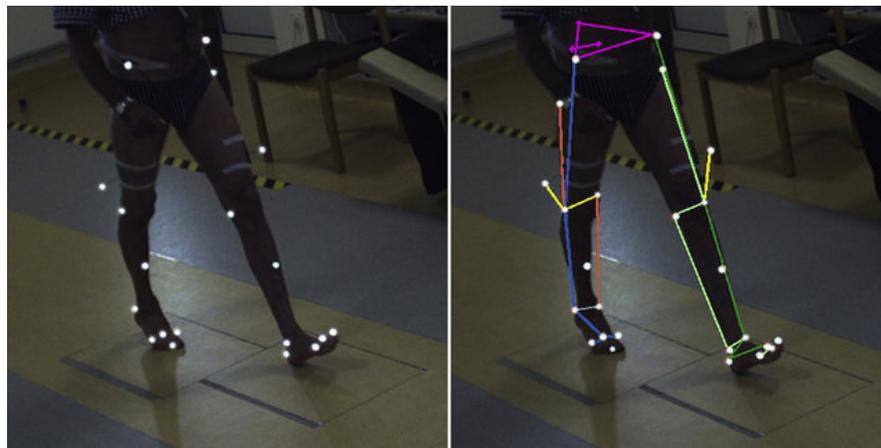


Figura 3.12: Análisis de la marcha con un sistema de captura óptico basado en marcadores reflectantes. Imagen tomada de [64]

A pesar de la alta precisión de los datos obtenidos con estas técnicas, una de las principales desventajas del análisis de movimiento óptico es que su uso está limitado al lugar de trabajo en el que estén fijadas las cámaras. Además, aparece el problema de obstrucción y sombras que pueden alterar la imagen capturada, así como la incertidumbre que puede surgir cuando se superponen segmentos corporales de interés durante la grabación del movimiento [2]. Por tanto, es fundamental ubicar adecuadamente los marcadores, en puntos corporales que representen con precisión las acciones de las articulaciones durante la marcha [10].

Equipos de sensorización inercial

Los sensores inerciales, también conocidos como **Unidades de Medición Inercial (IMUs)**, son unos dispositivos basados en **sistemas microelectromecánicos (MEMS)** capaces de medir la aceleración lineal (acelerómetro), velocidad angular (giroscopio) y, en algunos casos, fuerza de señales magnéticas (magnetómetro). De esta forma, los sensores se ubican en distintas partes del cuerpo, especialmente en las extremidades inferiores, y a partir de los datos registrados se pueden estimar parámetros espaciotemporales y cinemáticos [65].

En ese sentido, los sensores inerciales ofrecen dos ventajas principales respecto a otras herramientas de captura de movimiento [66]. Por un lado, el reducido tamaño, peso y consumo energético de las **IMUs** permite fabricar sistemas portátiles pequeños que reducen las posibles alteraciones al movimiento natural. Además, pueden registrar el movimiento sin necesidad de puntos de referencia, a diferencia de, por ejemplo, los sistemas ópticos, donde se requieren laboratorios de captura bien definidos. Sin embargo, el gran inconveniente de estos dispositivos para la captura de movimiento es que la necesidad de calibración periódica y el significativo ruido de la señal externa hacen que sean menos precisos y fiables que los sistemas ópticos [67, 68].

No obstante, actualmente existen numerosas empresas dedicadas a la comercialización de sistemas de alta precisión para la captura y análisis de movimiento basados en sensores inerciales. Es el caso de *Xsens*, *InterSense*, *Technaid*, *IMeasureU* o *Noraxon* [67]. Uno de los productos más extendidos es *Xsens MVN Awinda*², mostrado en la Figura 3.13, un sistema formado por 17 dispositivos inerciales capaz de operar durante más de 6 horas de forma inalámbrica a una distancia de hasta 50 metros [69].

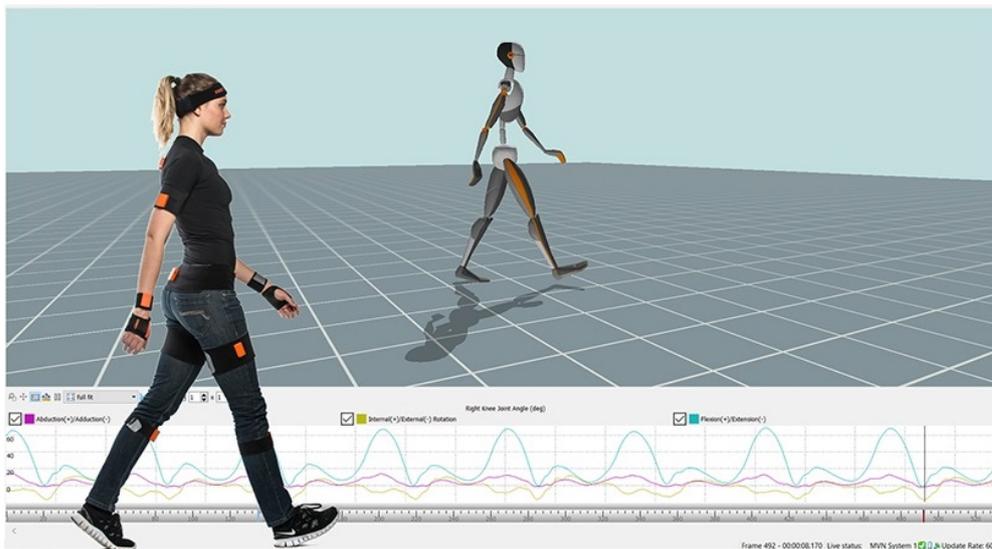


Figura 3.13: Análisis de la marcha con sensores inerciales: sistema Xsens MVN Awinda. Imagen extraída de www.xsens.com/.

Electromiografía

Además del análisis de los parámetros espaciotemporales, cinemáticos y cinéticos, en algunas investigaciones se realizan estudios complementarios que implican el registro de la actividad

²<https://www.xsens.com/products/mtw-awinda>

muscular mediante técnicas de **electromiografía (EMG)**. La **EMG** permite detectar, amplificar y mostrar gráficamente la actividad eléctrica de la contracción muscular, es decir, cuándo un músculo está o no activo durante el ciclo de la marcha [70].

Actualmente, para el análisis de la marcha se suele utilizar **electromiografía de superficie (sEMG)** [71], una técnica no invasiva que se basa en el uso de electrodos colocados en la superficie de la piel con adhesivos y gel para facilitar la transmisión de señales [72]. Además, a diferencia de la tradicional **electromiografía intramuscular (iEMG)**, la **sEMG** permite la libre ejecución de movimientos porque no se pincha en la piel ni en el músculo [73]. La Figura 3.14 muestra un ejemplo del uso de electromiografía superficial durante el análisis de la marcha humana.

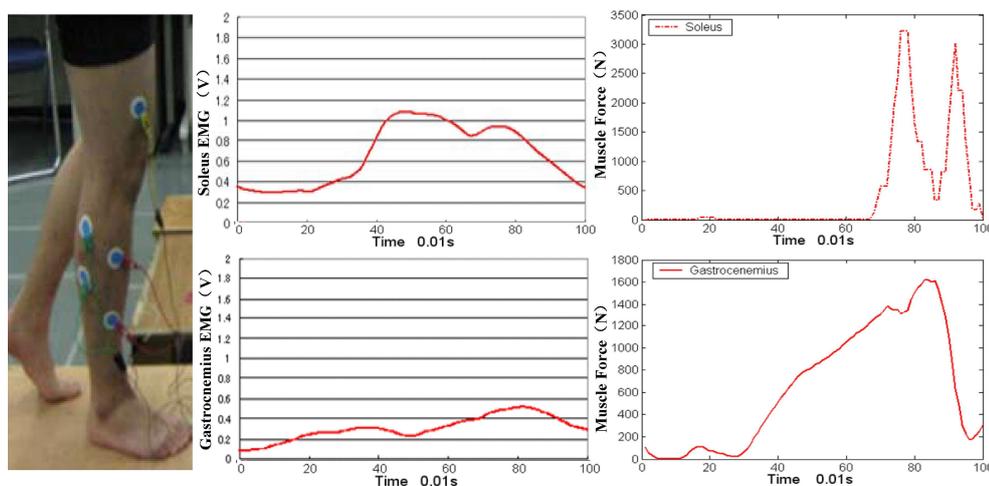
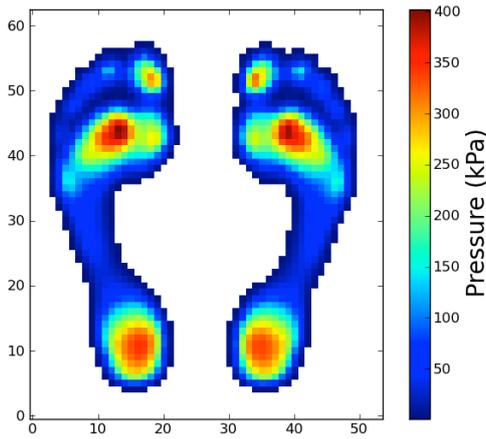


Figura 3.14: Registro de la actividad muscular durante la marcha con electromiografía. Imagen tomada de [74]

Plantillas instrumentadas

Esta aproximación consiste en el uso de zapatos instrumentados o plantillas sensorizadas que integran una serie de sensores de presión, dispositivos inerciales o combinaciones de ambos, útiles para realizar estimaciones de las fuerzas de reacción del suelo (análisis cinético) o para delimitar las distintas fases del ciclo de la marcha [75]. En este sentido, el uso de sensores de presión colocados sobre la planta del pie permite captar los momentos en que cada parte de la planta contacta con el suelo [6]. Cada sensor registra una fuerza vertical independiente que define una localización separada de presión (fuerza/área del sensor) dentro del pie [10], de forma que se pueden construir diagramas de presión plantar, como el mostrado en la Figura 3.15a.

Como ejemplo de estos sistemas, en la Figura 3.15b se muestran unas plantillas instrumentadas (*InGAIT*) formadas por 32 sensores de presión y un sensor inercial con acelerómetro y giroscopio, que han sido íntegramente diseñadas y desarrolladas por investigadores de **MAMl**, el grupo de investigación de la Universidad de Castilla-La Mancha [5] en el que se enmarca el presente **TFM**.



(a) Diagrama de presión plantar. Imagen tomada de [76].



(b) Plantillas instrumentadas para el análisis de la marcha

Figura 3.15: Estudio de la marcha humana mediante plantillas instrumentadas

Plataformas de fuerza

Las plataformas de fuerza, o dinamométricas, son unos dispositivos típicamente rectangulares con unas dimensiones que oscilan entre 40 y 90 cm [16], que se instalan a ras del piso y están equipadas con unos sensores para registrar las fuerzas verticales, laterales, anteroposteriores y/o de torsión ejercidas sobre el suelo durante la ejecución de movimientos [6]. En este sentido, proporcionan la fuerza total aplicada por el pie sobre el suelo pero, a diferencia de las plantillas instrumentadas no muestran la distribución de las diferentes partes de la fuerza plantar (por ejemplo, el talón o el antepié) durante la marcha [20].

En cualquier caso, utilizando los registros de la fuerza de reacción del suelo, se pueden calcular las fuerzas y los momentos que actúan en las articulaciones de las extremidades inferiores del sujeto (tobillos, rodillas y caderas) [77]. Sin embargo, su principal inconveniente es la limitación en la medida de movimientos posturales [66], y aunque se utilizan para estudios de la marcha, se requieren de varias plataformas de fuerza colocadas de manera consecutiva, tal y como se muestra en la Figura 3.16, por lo que es complicado obtener pasos limpios sobre ellas [14].



Figura 3.16: Cuatro plataformas de fuerza colocadas de manera consecutiva para el análisis de la marcha. Imagen tomada de [78]

En este trabajo se ha utilizado un sistema óptico basado en marcadores reflectantes (pasivos) para la captura del movimiento durante la marcha, con la intención de aplicar

técnicas de *Machine Learning* sobre los datos recogidos para valorar automáticamente la forma de caminar. Por tanto, resulta conveniente realizar un estudio de los trabajos previos en los que se han utilizado métodos de Aprendizaje Automático sobre este tipo de datos.

3.3. MARCHA Y APRENDIZAJE AUTOMÁTICO

Las técnicas de *Aprendizaje Automático (AA)*, o *Machine Learning (ML)*, son unas potentes herramientas capaces de analizar grandes y complejos volúmenes de datos para detectar automáticamente patrones y relaciones inmersos en ellos. Su capacidad para analizar y pronosticar datos temporales de alta dimensionalidad, identificar patrones en imágenes o agrupar los datos en clases, hace que sean técnicas idóneas para llevar a cabo diferentes tipos de análisis relacionados con la marcha humana. En esta sección se hace una revisión de trabajos en los que se aplican técnicas de *ML* sobre este tipo de conjuntos de datos.

Al igual que en cualquier otro dominio, modelar un problema relacionado con la marcha humana mediante algoritmos de *ML* consiste en construir un sistema biomecánico $T(x)$ identificando la relación existente entre los datos de entrada, $f(x)$, y las salidas $y(x)$, teniendo en cuenta que los datos de entrada pueden estar distorsionados por un ruido $n(t)$ que hace necesario un tratamiento previo. En este contexto, los datos de entrada suelen ser un vector multidimensional, u_i, v_i donde u se corresponde con el número de sujetos o ensayos y v son las características cinemáticas, cinéticas o antropométricas consideradas [79, 80].

Las principales aplicaciones de aprendizaje supervisado y no supervisado sobre datos de marcha humana son las siguientes:

- **Aprendizaje supervisado.** Para los casos en los que se dispone de un conjunto de datos etiquetado, es decir se conoce la solución deseada para una serie de ejemplos, el Aprendizaje Automático ha sido ampliamente utilizado en el desarrollo de modelos que detecten los eventos y fases de la marcha [81-85]; reconozcan entre distintos tipos de actividades [86, 87] (como caminar, correr o subir y bajar escaleras); distinguan las superficies por las que se camina [88, 89]; actúen como herramienta biométrica para reconocer personas a través del análisis de la marcha [90-93]; y, clasifiquen distintas alteraciones y asimetrías en la marcha [94-99].

En cuanto a los algoritmos de *Machine Learning* empleados, conviene señalar que en una gran cantidad de trabajos de clasificación de marcha se utiliza el algoritmo *KNN* por su simplicidad y capacidad para manejar grandes conjuntos de datos [93, 99-101]. No obstante, las diferentes variantes de *SVM* han demostrado ofrecer una alta capacidad de generalización en estudios de marcha [102-105], incluso para conjuntos de datos no demasiado grandes. Otros algoritmos utilizados en tareas de este tipo son las redes neuronales [106-108], bosques aleatorios (*RF*) [109-111], árboles de decisión (*DT*) [112-114] o modelos ocultos de Márkov (*HMM*) [115-118].

- **Aprendizaje no supervisado.** Las técnicas utilizadas cuando solo se conocen los datos de entrada, sin una etiqueta o clase asignada a cada observación, han sido menos estudiadas en trabajos sobre marcha, ya que es complicado definir con precisión los objetivos del aprendizaje [79]. No obstante, se han utilizado técnicas de *clustering* para agrupar patrones de marcha e investigar, por ejemplo, si las instancias dentro de un mismo grupo están relacionadas con atributos físicos como el tamaño corporal, el género o la edad [119]; organizar las posturas adoptadas en el instante de contacto

inicial por sujetos que padecen parálisis cerebral [120] o, agrupar los valores cinemáticas de la rodilla para extraer una curva representativa y característica de la marcha normal [121]. Asimismo, PCA es la técnica más frecuentemente utilizada para reducir la dimensionalidad de datos de marcha [122, 123].

Por otro lado, cuando la fuente de adquisición de los datos de marcha es un sistema basado en cámaras, como el utilizado en este TFM, los trabajos previos que han sido consultados se pueden organizar en dos categorías, dependiendo del enfoque seguido para procesar los datos:

- **Basados en la forma de la silueta.** Partiendo de una secuencia temporal de imágenes tomadas durante la marcha, este enfoque analiza las siluetas del cuerpo en 2D [101]. Para ello, generalmente se separan los píxeles de la figura en movimiento de los píxeles del fondo, con el objetivo de conseguir una silueta binaria del sujeto mientras camina, tal y como se muestra en el ejemplo de la Figura 3.17. Esta aproximación se utiliza principalmente en aplicaciones biométricas de la marcha [124, 125].
- **Basados en parámetros estáticos.** Consiste en utilizar los parámetros que caracterizan la forma de caminar, como la longitud de zancada, la cadencia o la velocidad, junto con medidas corporales estáticas como la relación entre las dimensiones de distintas partes del cuerpo [126].

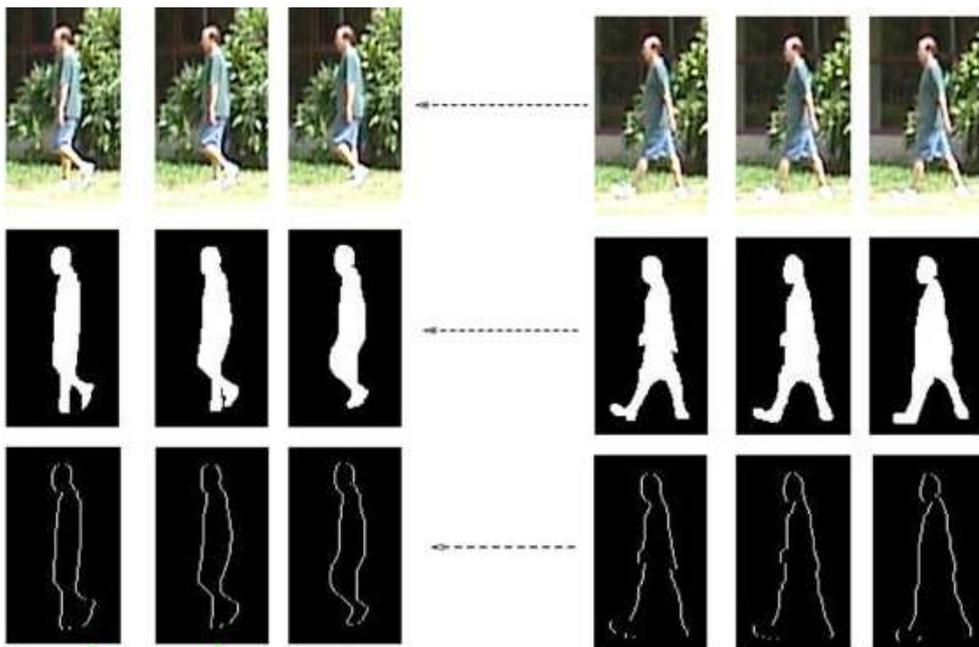


Figura 3.17: Ejemplos de siluetas al caminar. Imagen tomada de [124]

En relación con lo expuesto en esta sección, en el presente TFM se sigue un enfoque supervisado de clasificación, basado en parámetros estáticos, para reconocer las diferentes alteraciones y anomalías de la marcha consideradas en el test de Tinetti.

Herramientas y Metodología de Trabajo

Este capítulo tiene el objetivo de describir las diferentes tecnologías *software*, así como herramientas *hardware*, utilizadas en el proceso de desarrollo del presente TFM. Además, se presenta la metodología de trabajo empleada para la planificación y desarrollo del proyecto.

4.1. SISTEMA ÓPTICO DE CAPTURA Y ANÁLISIS DE MOVIMIENTO

De los distintos tipos de equipos tecnológicos utilizados actualmente para adquirir datos objetivos de la marcha humana (Sección 3.2.2), en este trabajo se ha usado un sistema basado en cámaras con marcadores reflectantes (pasivos). En particular, se ha contado con el sistema profesional *Clinical 3DMA* instalado en el laboratorio técnico *WeCareLab*, situado en el Instituto de Tecnologías y Sistemas de Información (ITSI) de la Universidad de Castilla-La Mancha (UCLM).

*Clinical 3DMA*¹ es una herramienta *software* para el análisis de movimiento 3D distribuida por *STT Systems*², una empresa con sede en San Sebastián especializada en el desarrollo de productos con diversas aplicaciones de análisis de movimiento, tanto ámbitos clínicos como deportivos. Para su funcionamiento, *Clinical 3DMA* se sincroniza con un conjunto de cámaras *OptiTrack*³ y aplica una transformación lineal directa para conseguir una reconstrucción tridimensional del movimiento a partir de las coordenadas 2D de un conjunto de marcadores reflectantes. De esta forma, el *software* calcula las ubicaciones de las articulaciones y huesos para proporcionar diferentes tipos de análisis de movimiento del cuerpo completo.

A continuación, se presentan detalladamente las cámaras *OptiTrack* y la plataforma *Clinical 3DMA*, así como algunas de las herramientas y materiales adicionalmente utilizados durante el presente trabajo para la construcción de un *dataset* con datos de marcha humana.

4.1.1. Cámaras *OptiTrack*

Las cámaras *OptiTrack*, desarrolladas por la empresa estadounidense *NaturalPoint*, se utilizan en un amplia variedad de sistemas de captura de movimiento, con aplicaciones orientadas tanto a estudios biomecánicos como robótica, realidad virtual o animaciones 3D. El significativo número de trabajos de investigación, publicados en los últimos años, y en los que

¹<https://www.stt-systems.com/motion-analysis/3d-optical-motion-capture/clinical-3dma/>

²<https://www.stt-systems.com/>

³<https://optitrack.com/cameras/>

se utiliza este tipo de cámaras, demuestra que *OptiTrack* se ha convertido en un sistema validado, aceptado y ampliamente extendido dentro la comunidad científica, constituyendo una alternativa más barata que los tradicionales sistemas de *Vicon*⁴ [127].

De los distintos sistemas de cámaras que ofrece *OptiTrack*, en este trabajo se han utilizado ocho cámaras **Primex 13W**⁵ (Figura 4.1) fijadas al techo del laboratorio, las cuales cuentan con una resolución de 1280 x 1024 píxeles (1.3 MP), 240 FPS, una precisión de aproximadamente 0.30 mm y una latencia de 420 ms. El campo de visión (FOV), medido en grados, es de 70° horizontalmente y 58° en sentido vertical. En cuanto a las dimensiones del producto, cada cámara tiene un alto y ancho de 6,86 cm, con una profundidad de 5,6 cm y un peso de 320 gramos. Según la documentación oficial, cada cámara tiene un coste aproximado de \$2,499, por lo que el coste del sistema instalado en el *WeCareLab*, formado por ocho dispositivos, asciende hasta \$19.992.



Figura 4.1: Cámara *OptiTrack Primex 13W*: vista frontal (izquierda), lateral (centro) y trasera (derecha). Imagen tomada de <https://optitrack.com/>

Para complementar lo explicado anteriormente, la Tabla 4.1 incluye un resumen de las características técnicas de los sistemas ofrecidos por *OptiTrack*.

Tabla 4.1: Comparativa entre productos *OptiTrack*

	Precio	Resolución	Frame Rate	FOV Horizontal	Conectividad	Latencia
Prime 41	\$5,999	4.1 MP (2048 × 2048)	180 FPS	51°	GigE/PoE+	5,5 ms
Prime 17W	\$3,499	1.7 MP (1664 × 1088)	360 FPS	70°	GigE/PoE+	2,8 ms
Prime 13	\$1,999	1.3 MP (1280 × 1024)	240 FPS	42°, 56°	GigE/PoE+	4,2 ms
Prime 13W	\$2,499	1.3 MP (1280 × 1024)	240 FPS	82°	GigE/PoE+	4,2 ms
Flex 13	\$999	0.3 MP (640 × 480)	120 FPS	42°, 56°	USB 2.0	8,3 ms
Flex 3	\$599	1.3 MP (1280 × 1024)	100 FPS	38°, 46°, 58°	USB 2.0	10 ms
Slim 13E	\$1,499	0.3 MP (640 × 480)	240 FPS	42°, 56°, 82°	GigE/PoE+	4,2 ms

⁴<https://www.vicon.com/>

⁵<https://www.optitrack.com/cameras/prime-13w/specs.html>

4.1.2. Clinical 3DMA

Clinical 3DMA es una herramienta *software* especializada en ofrecer precisos exámenes de movimiento biomecánico en 3D de todo el cuerpo, basándose en los datos recopilados por un sistema de sensorización óptico. Está especialmente orientada a estudios de la marcha humana, aplicable en diferentes entornos, desde un punto de vista clínico hasta laboratorios de rendimiento deportivo o centros de fisioterapia. En el presente TFM se ha utilizado la **versión OT 2020.1** de *Clinical 3DMA*.

En general, *Clinical 3DMA* pertenece a un conjunto de paquetes desarrollados por *STT Sytems* (*Sports 3DMA*, *Clinical 3DMA*, *Cycling 3DMA*, *Running 3DMA*, *Golf 3DMA*, *Human 3DMA*, *EDDO Biomechanics*), los cuales comparten la misma tecnología y capacidades, pero cada uno ha sido diseñado para proporcionar al usuario un conjunto específico y práctico de herramientas y protocolos de análisis. El funcionamiento común de estos paquetes, así como las peculiaridades de cada uno de ellos, se ilustra gráficamente en la Figura 4.2.

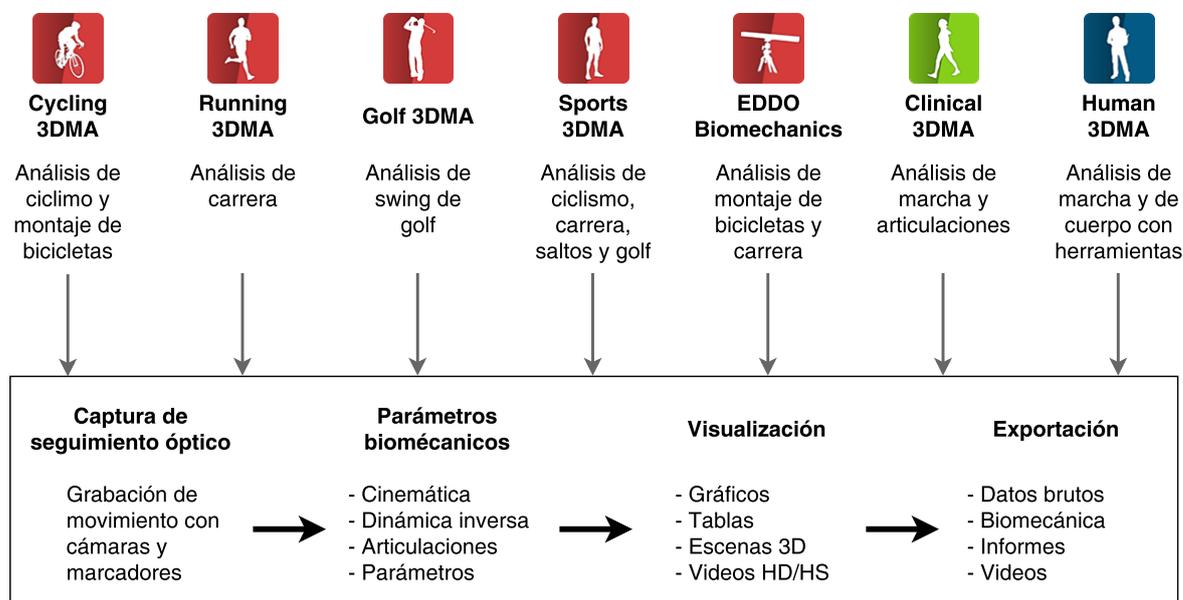


Figura 4.2: Características de los productos 3DMA

Entre las funcionalidades que ofrece *Clinical 3DMA* se encuentran las siguientes:

- **Análisis de marcha y otros protocolos.** Aunque *Clinical 3DMA* está especializado en el análisis de la marcha, cuenta con 25 protocolos de análisis (por ejemplo, análisis cervical o lumbar). Además, cabe señalar que los protocolos de análisis utilizados por *STT* están respaldados por la literatura científica, y analizan la marcha humana con procedimientos ampliamente aceptados como Helen-Hayes o Davis.
- **Detección de eventos.** Incluye un algoritmo para la detección automática de eventos durante la marcha: contacto inicial, y despegue de talón.
- **Representación de parámetros biomecánicos** (cinemática, dinámica inversa, articulaciones y parámetros espaciotemporales) a través de gráficas temporales en tiempo real y tablas numéricas.
- **Biofeedback en tiempo real.** Permite definir condiciones simples que se verifican durante el movimiento que las cámaras están captando, de forma que siempre que se cumpla alguna de las reglas especificadas el usuario recibe un aviso inmediato.
- **Generación automática de informes** a partir de los datos integrados de todos los

dispositivos, facilitando significativamente la tarea del analista.

- **Exportación de datos de captura** (trayectorias en bruto) en formato **ASCII**, y de curvas biomecánicas en formato **CSV**, para permitir el procesamiento en paquetes de software de terceros.
- **Integración con dispositivos de terceros**, ofreciendo una interfaz única para trabajar simultáneamente con plataformas de fuerza **AMTI**⁶ (para registrar la fuerza de reacción del suelo), cámaras de vídeo de alta velocidad (para grabar la imagen en color), el sistema de electromiografía superficial **Trigno Lab** de **Delsys**⁷ (para analizar la actividad eléctrica de los músculos durante el movimiento) y con el sistema **iSen** de sensores inerciales⁸.

Como ejemplo, la Figura 4.3 muestra una pantalla de la interfaz de usuario de *Clinical 3DMA*.

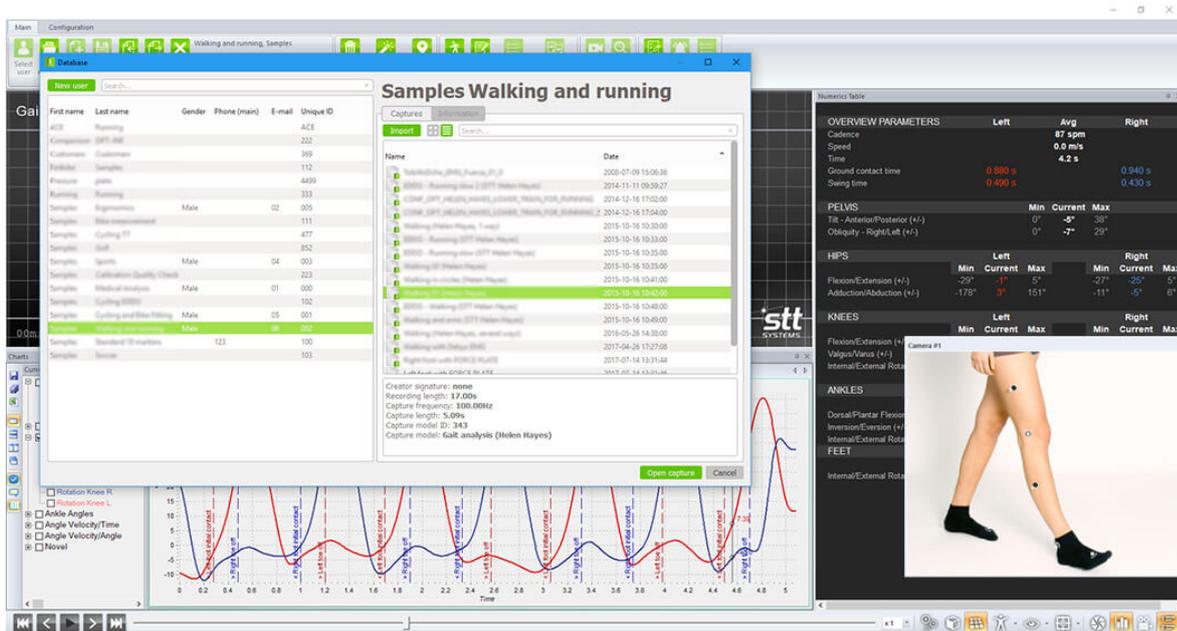


Figura 4.3: Captura de la interfaz de usuario de *Clinical 3DMA*. Imagen tomada de <https://www.stt-systems.com/>

4.1.3. Otras herramientas utilizadas durante la captura de datos

Además de las cámaras *OptiTrack* y el sistema *Clinical 3DMA*, durante la fase de captura de datos se han utilizado los siguientes materiales y herramientas:

- Un segmómetro flexible (Figura 4.4a) para medir longitudes y segmentos corporales, con el fin de obtener una estimación más precisa de la ubicación de los centros de rotación articulares. El dispositivo utilizado presenta una longitud de 5 metros; y la escala de medida está tanto en centímetros como en milímetros, por lo que tiene una precisión de un milímetro.
- Una cámara de vídeo *Anvasck Full HD 1080P 30 FPS* (Figura 4.4b) para grabar la imagen en color durante el movimiento.

⁶https://www.amti.biz/index_es.aspx

⁷<https://delsys.com/trigno/>

⁸<https://www.stt-systems.com/motion-analysis/inertial-motion-capture/isen/>

- Un juego de marcadores reflectantes de 11 mm y un rollo de adhesivos de velcro para fijarlos al cuerpo durante la captura del movimiento (Figura 4.4c).
- Una barra de calibración formada por 7 marcadores reflectantes uniformemente distribuidos a intervalos fijos de 30 cm.



Figura 4.4: Materiales adicionales utilizados durante la captura de movimiento

4.2. HERRAMIENTAS SOFTWARE

Además del sistema óptico de captura y análisis de movimiento (*software + hardware*), expuesto en la sección anterior, para el desarrollo de las diferentes etapas del trabajo, mostradas en la Sección 1.1 y esquematizados en la Figura 1.1, se hace necesario el uso de una variedad de herramientas y tecnologías *software* que se describen a continuación.

4.2.1. Entorno de programación y librerías

Durante el desarrollo del trabajo se han implementado varios *scripts*/servicios orientados al preprocesamiento y analítica de datos utilizando *Python*. Aunque se trata de un lenguaje de programación de propósito general, en los últimos años *Python* se ha consolidado como una de las herramientas más utilizadas en el ámbito de la Ciencia de Datos, gracias a la abundancia de potentes librerías de *software* libre que permiten desarrollar, de manera sencilla y rápida, desde los análisis de datos más básicos hasta los modelos de *Machine Learning* más complejos. Además, la elección de *Python* como lenguaje de programación para el desarrollo del proyecto se ha apoyado en que es una tecnología multiplataforma y lo suficientemente versátil como para resolver cualquier otro tipo de tarea que surgiera durante la realización del trabajo.

Más concretamente, se ha utilizado la versión 3.8.5 de *Python* sobre una distribución de *Anaconda*⁹ (conda 4.9.2). Como entornos de trabajo se han empleado tanto *Visual Studio Code* como *Jupyter Notebooks*. Además, en los diferentes *scripts* desarrollados en este TFM se han importado, principalmente, las siguientes librerías de código abierto:

- **Librerías especializadas en Ciencia de Datos:** *pandas*¹⁰, para el manejo y análisis de estructuras de datos; *NumPy*¹¹, una biblioteca orientada al cálculo numérico y computación científica; *SciPy*¹², un paquete que implementa algoritmos matemáticos

⁹<https://www.anaconda.com/>

¹⁰<https://pandas.pydata.org/>

¹¹<https://numpy.org/>

¹²<https://www.scipy.org/>

para la manipulación y filtrado de datos y, *Matplotlib*¹³ y *seaborn*¹⁴ para la visualización de datos y generación de gráficos.

- **Herramientas de Aprendizaje Automático:** *Scikit-learn*¹⁵, una librería que implementa de forma eficiente los principales algoritmos de *Machine Learning*; *TensorFlow*¹⁶, para entrenar y ejecutar redes neuronales a gran escala; y, *Keras*¹⁷, una API de *Deep Learning* de alto nivel para compilar y entrenar redes neuronales.
- **Librerías de propósito general:** *pyautogui*¹⁸, un módulo que permite la automatización de la interfaz usuario administrando las operaciones del ratón y teclado; *docxtpl*¹⁹, una herramienta para generar documentos de *Microsoft Word*; y, *PyInquirer*²⁰, para la construcción de interfaces de línea de comandos (CLI, por sus siglas en inglés).

Para más detalles, en el Anexo B se proporciona el listado completo de todas las bibliotecas importadas, especificando la versión exacta junto con un enlace al repositorio del que han sido descargadas.

4.2.2. Control de versiones

Con el objetivo de mantener y llevar un control de los cambios realizados durante el desarrollo del *software* asociado al proyecto, se ha utilizado *Git* como sistema de control de versiones, y *Github* como servicio web para almacenar, en un repositorio remoto, el código del proyecto. El enlace a dicho repositorio en la nube puede consultarse en el Anexo A.

4.2.3. Documentación y gráficos

- *Draw.io*. Herramienta de creación y edición de diagramas, que ha sido empleada en la construcción de algunas figuras incluidas en esta memoria.
- *Adobe Illustrator*. Editor de gráficos vectoriales, que se ha usado para crear y editar iconos e ilustraciones mostrados en el trabajo.
- \LaTeX . Herramienta orientada a la creación de documentos escritos, utilizada para desarrollar la memoria de este TFM.
- *OpenProj*. *Software* libre de gestión de proyectos mediante el cual se ha construido el diagrama de Gantt mostrado en la Figura 4.6.

4.3. METODOLOGÍA DE TRABAJO

De nada sirve disponer de buenas herramientas *hardware* y tecnologías *software* si no se cuenta con las directivas apropiadas para su aplicación. Al igual que en cualquier otra disciplina, cuando se trata de un trabajo en el dominio de Ciencia de Datos, estas directivas deben ser proporcionadas por un marco de referencia que estructure, planifique y guíe los procesos y actividades del proyecto.

¹³<https://matplotlib.org/>

¹⁴<https://seaborn.pydata.org/>

¹⁵<https://scikit-learn.org/stable/>

¹⁶<https://www.tensorflow.org/>

¹⁷<https://keras.io/>

¹⁸<https://pyautogui.readthedocs.io/en/latest/>

¹⁹<https://docxtpl.readthedocs.io/>

²⁰<https://github.com/CITGuru/PyInquirer>

Las metodologías de trabajo tradicionales están basadas en un estricto control del proyecto mediante una rigurosa definición de roles y actividades, así como una documentación extensa y detallada. Este tipo de metodologías han demostrado ser efectivas y dar buenos resultados cuando se aplican a proyectos estables, en cuanto a tiempo y recursos, desarrollados por grupos de trabajo grandes [128]. Sin embargo, en este TFM se va a llevar a cabo un proyecto estrictamente acotado a un período corto de tiempo, con una cierta naturaleza experimental, un importante trasfondo teórico y un número muy reducido de participantes.

Para un proyecto con estas características es más apropiada una metodología ágil, pues están especialmente orientadas a proyectos pequeños en los que se exige reducir tiempos de desarrollo sin renunciar a las prácticas fundamentales que garantizan la calidad de la solución alcanzada. Además, hay que tener en cuenta que el presente trabajo se enmarca en el dominio de la Ciencia de Datos, y se pretende seguir un enfoque “de arriba abajo” por el que partiendo de un problema bien definido se analizarán los datos para obtener una solución (en contraposición a las estrategias ascendentes en las que el científico de datos primero analiza grandes volúmenes de datos para determinar cuál es el objetivo).

En este sentido, el proyecto abordado se ha basado en la **Metodología Fundamental para la Ciencia de Datos** [129], propuesta por John B. Rollins, ya que se adapta perfectamente a las características del trabajo, y proporciona una estrategia de orientación independiente a las tecnologías y herramientas específicas, los volúmenes de datos y los enfoques involucrados [130].

4.3.1. Metodología Fundamental para la Ciencia de Datos

La Metodología Fundamental para la Ciencia de Datos (FMDS por sus siglas en inglés), comparte algunas similitudes con otros marcos reconocidos para la minería de datos como KDD o CRISP-DM, pero además, proporciona una serie de nuevas prácticas en la Ciencia de Datos [131]. FMDS se basa en un proceso iterativo de 10 etapas, ilustradas gráficamente en la Figura 4.5, y descritos a continuación [129]:

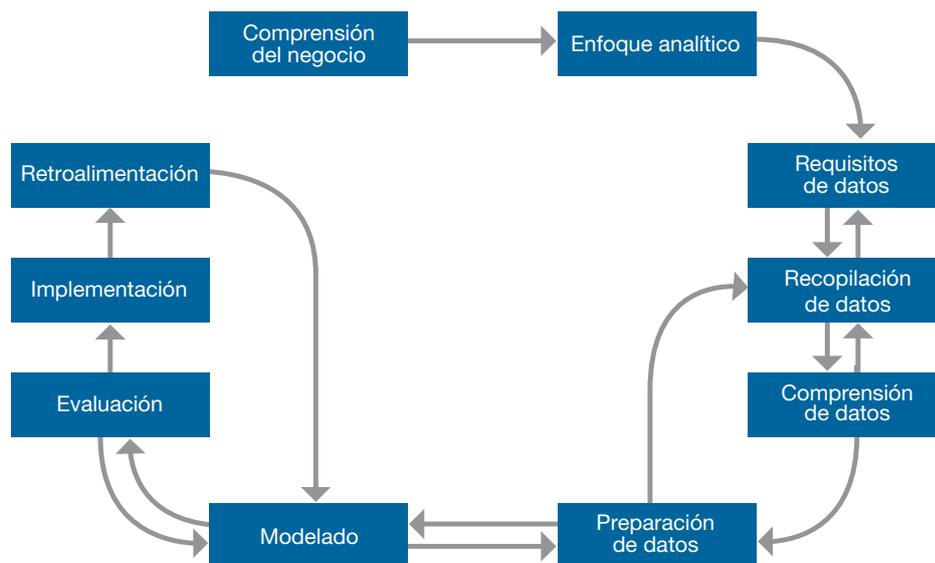


Figura 4.5: Metodología Fundamental para la Ciencia de Datos. Imagen tomada de [129]

1. **Comprensión del negocio.** Todo proyecto comienza con la comprensión del negocio para definir el problema, los objetivos del proyecto y los requisitos de la solución.

2. **Enfoque analítico.** Una vez establecido el problema, el científico de datos deberá definir el enfoque analítico para resolverlo. Por ejemplo, si el objetivo es predecir una respuesta como "sí" o "no", el enfoque analítico podría definirse como la construcción, pruebas e implementación de un modelo de clasificación binaria.
3. **Requisitos de datos.** El enfoque analítico elegido determina los requisitos de datos: contenido, formatos y representaciones.
4. **Recopilación de datos.** En la primera fase de adquisición de datos, se identifican los recursos de datos disponibles (estructurados, no estructurados y semiestructurados) y relevantes para el dominio del problema.
5. **Comprensión de datos.** Después de la adquisición de datos inicial, se suelen utilizar estadísticas descriptivas y técnicas de visualización para analizar el contenido de los datos y valorar su calidad.
6. **Preparación de datos.** Esta fase incluye todas las tareas destinadas a construir el conjunto de datos que se utilizará en la siguiente etapa de modelado.
7. **Modelado.** Consiste en desarrollar modelos predictivos según el enfoque analítico previamente definido, utilizando el conjunto de datos preparado en la etapa anterior.
8. **Evaluación.** Antes de su implementación, es necesario evaluar el modelo desarrollado para determinar su calidad y garantizar que aborda el problema de manera válida y completa.
9. **Implementación.** Cuando el modelo ha sido desarrollado, se implementa en un entorno de producción. Generalmente, de forma limitada, hasta que su rendimiento se evalúa completamente.
10. **Retroalimentación.** Al recopilar los resultados del modelo implementado, se obtiene una retroalimentación sobre su rendimiento e impacto en el entorno en el que se implementó. Con esta información, es posible ajustar el modelo para mejorar su precisión y calidad.

Con esta metodología, los científicos de datos vuelven frecuentemente a etapas previas para realizar ajustes a medida que van aprendiendo más sobre los datos y el modelado. Así pues, los modelos se mejoran y adaptan constantemente a través de la retroalimentación, ajustes y reimplementaciones para que el trabajo pueda proporcionar un valor continuo mientras la solución sea necesaria [129].

4.4. PLANIFICACIÓN

De acuerdo con el carácter iterativo de la metodología de desarrollo utilizada, y en relación con los objetivos definidos en el Capítulo 2, el trabajo del TFM (12 ECTS = 300 horas) se ha distribuido de forma secuencial en las siguientes fases de trabajo:

- **Estudio del dominio.** Teniendo en cuenta la naturaleza experimental y el importante trasfondo teórico del trabajo, como punto de partida se ha realizado una exhaustiva búsqueda y lectura de contribuciones científicas sobre los siguientes aspectos:
 - Estudio biomecánico de la marcha humana.
 - Escalas y pruebas de valoración observacional de la marcha, con especial atención en el test de Tinetti.
 - Revisión de sistemas de captura y análisis de movimiento, orientando el estudio a aquellos basados en cámaras y marcadores reflectantes.
 - Aplicación de técnicas de *Machine Learning* sobre datos de movimiento, y en especial, de la marcha.

El resultado de esta fase de trabajo se ha presentado en el Capítulo 3.

- **Recopilación de datos.** El trabajo se centra la construcción de un *dataset* básico para valoración de la marcha, siguiendo los criterios de la escala de Tinetti, y utilizando como fuente de adquisición de datos un sistema basado en cámaras y marcadores reflectantes. El esfuerzo se organiza en varias actividades:
 - Pruebas iniciales para familiarizarse con el sistema de *tracking* (*hardware* y *software*) y poder entender mejor los datos recogidos.
 - Grabaciones de movimiento, diseñando previamente un protocolo de captura que considere la ubicación estandarizada de marcadores, y elaborando una hoja de información y consentimiento informado para ser firmada por los participantes.
 - Exportación de datos desde el *software* de captura y construcción del *dataset*, etiquetando los datos según las categorías consideradas en la escala de Tinetti para valoración de la marcha.

- **Preparación de datos.** Con el objetivo de asegurar la calidad y validez de los datos adquiridos, y que servirán de entrada a los algoritmos de Aprendizaje Automático, se tendrá que normalizar, preprocesar y limpiar el conjunto de datos generado. En este contexto, entre las tareas de preparación de datos abordadas tendrán que considerarse las siguientes:
 - Limpiar los datos adquiridos depurando, descubriendo y corrigiendo datos erróneos (*data cleansing*).
 - Crear variables explicativas relacionadas con diferentes parámetros que caracterizan la marcha humana (espaciotemporales y cinemáticos), de forma que puedan ser utilizadas como variables predictoras en la fase de modelado.
 - Normalizar los datos de la marcha para reducir o eliminar la influencia de las características antropométricas de cada sujeto.

- **Modelado y evaluación.** Sobre el conjunto de datos preprocesado en la fase de trabajo anterior, desarrollar y evaluar uno o varios modelos predictivos que permitan generar automáticamente las respuestas a la prueba de Tinetti para valoración de la marcha. Para ello, será necesario llevar a cabo los siguientes procesos típicos de Aprendizaje Automático:
 - Realizar un análisis exploratorio de los datos para tener un mayor entendimiento antes de aplicar los algoritmos de *Machine Learning*.
 - Seleccionar las características más relevantes para predecir la variable objetivo, y escalar adecuadamente los datos de entrada.
 - Entrenar y perfeccionar múltiples modelos, ajustando sus respectivos parámetros para encontrar el estimador que mejor rendimiento ofrezca con las variables disponibles.
 - Evaluar los modelos desarrollados con datos nuevos no vistos durante el entrenamiento, utilizando las métricas más apropiadas para el problema en concreto.

- **Implementación.** Los modelos desarrollados se integrarán y desplegarán en una herramienta final que, tomando como entrada nuevos datos grabados con el sistema de captura, generará informes detallados con gráficos y tablas que faciliten a los profesionales de la salud la valoración de la marcha siguiendo el criterio de la prueba de Tinetti.

Para completar esta planificación de tareas, en la Figura 4.6 se incluye un diagrama de

Gantt que representa gráficamente el tiempo de dedicación previsto para las diferentes actividades del proyecto, considerando que se trabajará una media de 2-3 horas diarias de lunes a viernes, y que la duración total del proyecto se extiende desde la última semana de febrero de 2021 hasta la primera semana de septiembre de 2021.

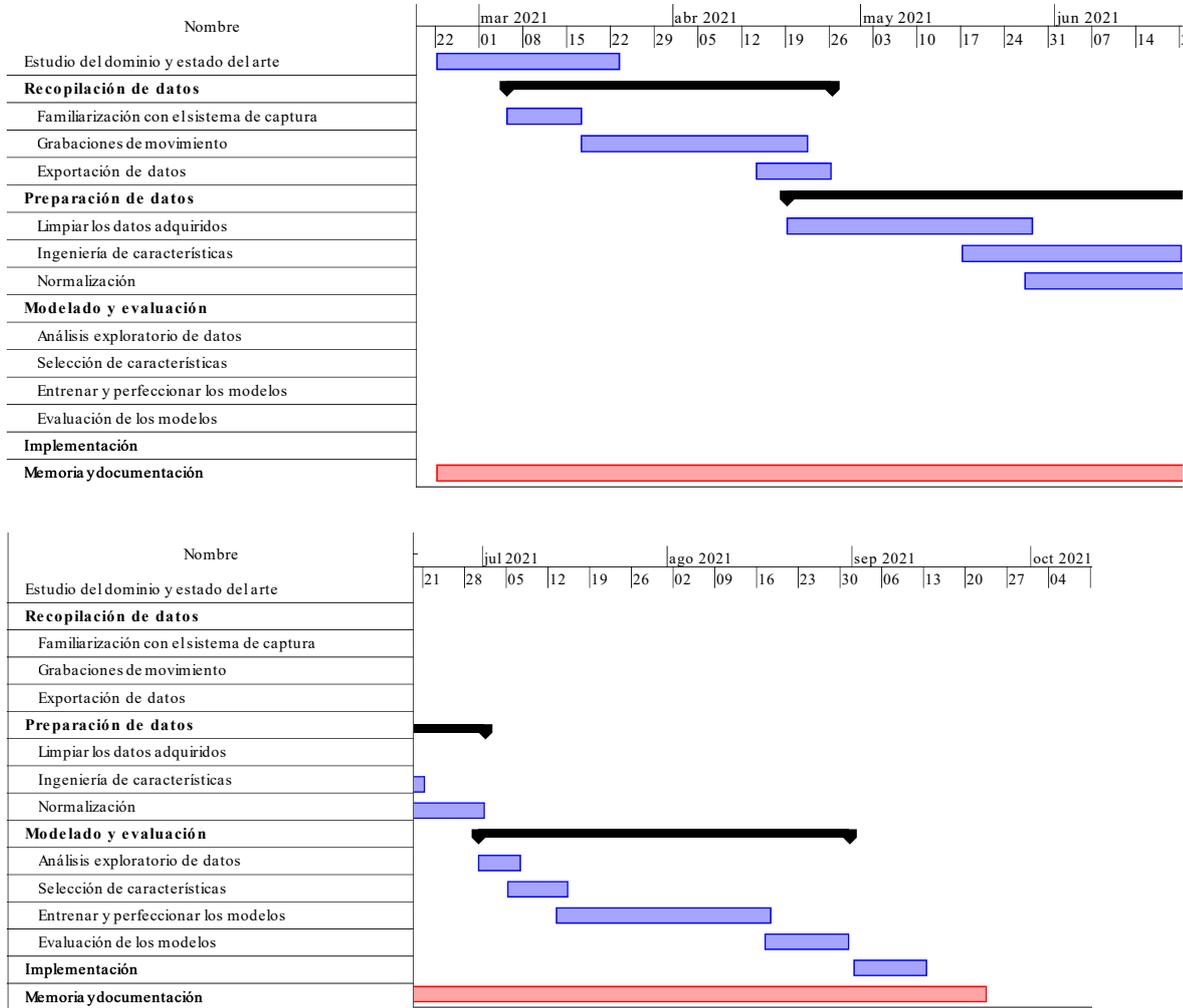


Figura 4.6: Diagrama Gantt del proyecto

CAPÍTULO 5

Evolución y Resultados

En este capítulo se describen las diferentes iteraciones/fases por las que ha ido evolucionando el desarrollo del proyecto, atendiendo a la metodología de trabajo presentada en la Sección 4.4. En este sentido, se comienza presentando el trabajo realizado para la construcción de un *dataset* de marcha basado en sensorización óptica, se describen las tareas de preprocesamiento previo y limpieza de datos llevadas a cabo, para, posteriormente, desarrollar y evaluar modelos de Aprendizaje Automático que permitan valorar la marcha según algunas de las cuestionarios consideradas en la prueba de Tinetti. Finalmente, se expone el desarrollo de un módulo que combina toda la información anterior para generar informes de análisis cuantitativos de marcha.

En cada etapa se presentan una serie de hitos y dificultades, así como las decisiones tomadas para solventarlas y cumplir con los objetivos propuestos en el Capítulo 2.

5.1. RECOPIACIÓN DE DATOS

La primera fase de trabajo ha consistido en la construcción de un conjunto de datos (*dataset*) que sirve como línea base para el desarrollo de modelos predictivos orientados a la valoración automática de la marcha humana según algunas de las cuestiones descritas en POMA. Para ello, se han seguido las consideraciones que se detallan en los siguientes apartados.

5.1.1. Configuración del laboratorio y calibración del sistema

Para la adquisición de datos durante la marcha se ha utilizado el sistema profesional de captura y análisis de movimiento *Optitrack Motion Capture + Clinical 3DMA*, descrito en la Sección 4.1. Más concretamente, se ha contado con **8 cámaras** *OptiTrack Primex 13W* con una frecuencia de muestreo de **100 Hz** (frames/sec). Estas cámaras están instaladas sobre soportes fijados al techo del laboratorio *WeCareLab*, distribuidas según se ilustra esquemáticamente en la Figura 5.1. Cada cámara está colocada a una altura de 3 metros sobre el suelo, y separada entre 2.80 y 4 metros de las cámaras adyacentes, de forma que con esta configuración se alcanza un área de captura óptima de 6 x 4 metros, ubicada en la parte central de la habitación. En la Figura 5.2 se incluye una fotografía real del laboratorio descrito.

Adicionalmente, en una de las esquinas del laboratorio se ha colocado una cámara de vídeo *Anvasik Full HD 30 FPS* para grabar y sincronizar la imagen en color con los datos de movimiento generados por el *software Clinical 3DMA*. Esta cámara se ha colocado de manera perpendicular al suelo utilizando un pequeño trípode fijado a una mesa, de forma

que la altura total de la lente es de 1.10 metros respecto al piso.

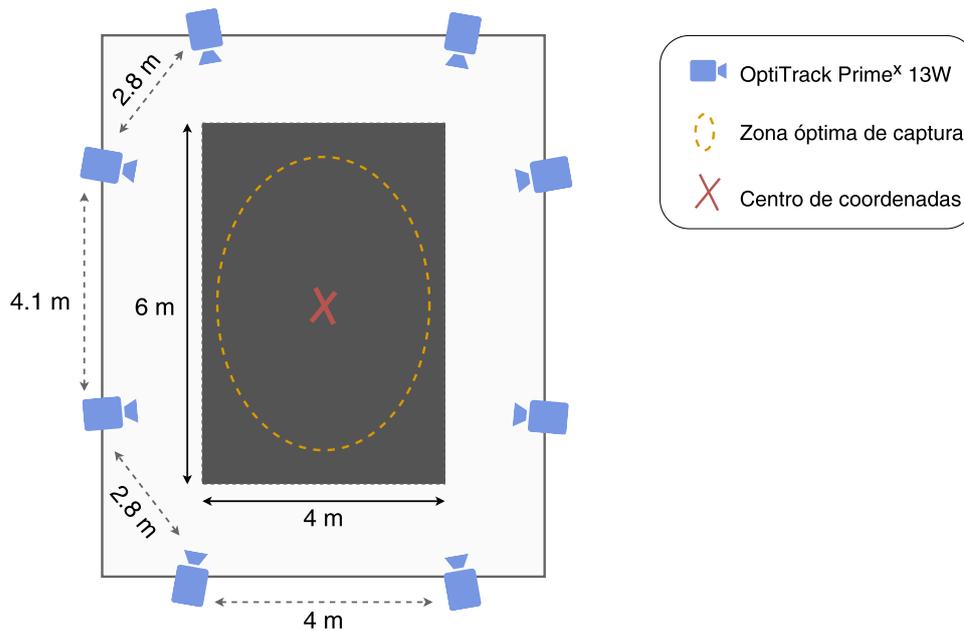


Figura 5.1: Configuración del laboratorio de captura de movimiento: vista esquemática



Figura 5.2: Configuración del laboratorio de captura de movimiento: vista real

Calibración del sistema

Como cualquier otro dispositivo de cuantificación, el sistema de cámaras *Clinical 3DMA* requiere de una calibración para aprender la posición relativa de las cámaras en relación con el entorno del laboratorio. En este sentido, para la calibración del sistema se han seguido las instrucciones del fabricante, utilizado una barra de calibración formada por marcadores uniformemente distribuidos a intervalos fijos de 300mm. Durante este proceso, el *software*

recoge varias instantáneas de la barra de calibración en diferentes posiciones del área de captura, y calcula automáticamente las posiciones de cada cámara. La Figura 5.3 muestra una captura de pantalla tomada de *Clinical 3DMA* durante el proceso de calibración, con la barra de marcadores ubicada en la zona central del laboratorio visible por las cámaras.

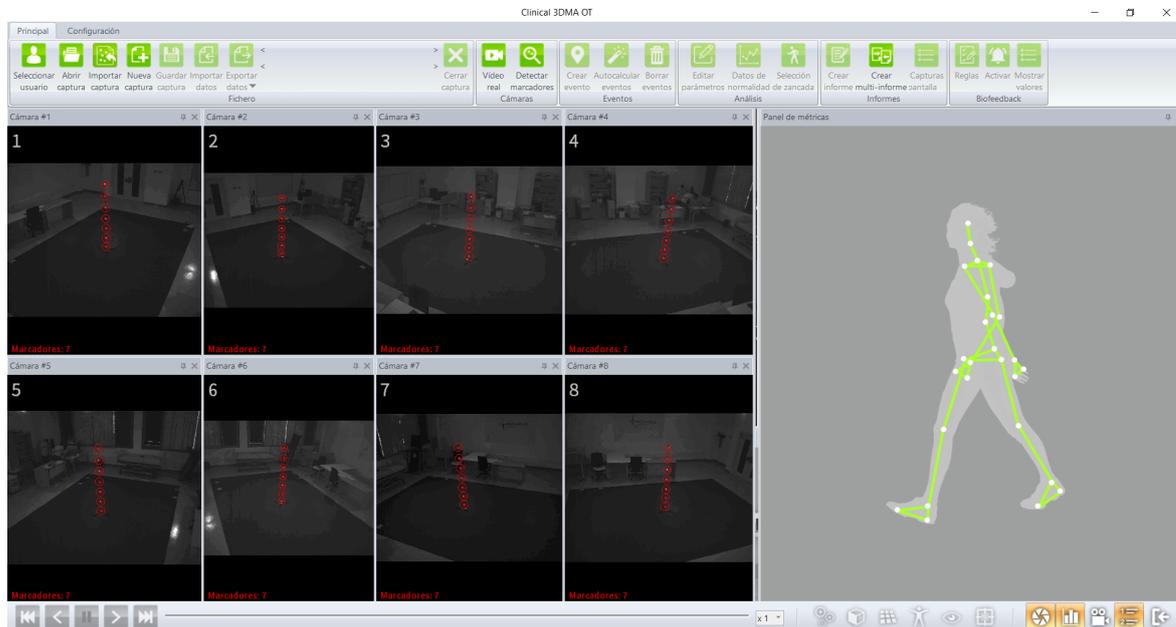


Figura 5.3: Calibración del sistema de captura

Teniendo en cuenta que las cámaras están fijadas al techo, se trata de una instalación quasi-permanente mucho más robusta, en términos de calibración, que la conseguida cuando las cámaras se colocan sobre trípodes portátiles. Por tanto, el proceso de calibración solo se ha llevado a cabo una vez al día antes de comenzar con las sesiones de grabación, independientemente de que en el mismo día se recopilarán datos de varios participantes.

Iluminación

En cuanto a la iluminación de la sala, se han seguido las recomendaciones del fabricante para limitar la entrada directa de luz solar a la zona visible por las cámaras durante las grabaciones de movimiento, ya que esto podría provocar problemas en la escena. En su lugar, se ha optado por la iluminación artificial habitual del laboratorio (fluorescentes, luz halógena).

5.1.2. Diseño del experimento y participantes

Para la construcción del *dataset* de marcha se ha capturado el movimiento de 10 participantes (7 hombres y 3 mujeres, con una media de edad de 29.8 ± 13.75 años). En todos los casos se trata de personas sanas sin alteraciones en la forma de caminar. En este sentido, es preciso mencionar que como consecuencia de la situación epidemiológica en la que se ha desarrollado esta fase de captura de datos, provocada por la COVID-19, el número de participantes ha sido más reducido de lo inicialmente deseado. Además, puesto que no se ha podido trabajar con personas ancianas que presentaran trastornos o inestabilidad en la marcha, para cada participante, además de capturas de marcha normal, se han tomado diferentes grabaciones simulando varias de las alteraciones consideradas en la escala de Tinetti.

Aunque no se ha impuesto ninguna restricción en cuanto al tipo de ropa utilizada por los sujetos capturados, se les ha pedido que eviten zapatillas o prendas con piezas reflectantes, así como anillos, pendientes o cinturones que puedan tener zonas brillantes. El protocolo seguido con cada participante se resume en los siguientes pasos, teniendo en cuenta que el tiempo requerido para la preparación y grabación de cada sujeto es de aproximadamente 60 minutos:

1. **Registrar información demográfica** (edad, género) y **medidas antropométricas** (altura, peso, ancho y alto de rodilla, ancho y alto de tobillo, longitud de pierna y longitud de pie). En los apartados posteriores se explica detalladamente el procedimiento seguido para la toma de estas medidas antropométricas.
2. **Registro en la base de datos del sistema.** *Clinical 3DMA* incorpora una base de datos que facilita el almacenamiento y organización de todas las capturas de movimiento. Para cada participante, se ha creado un nuevo usuario en la aplicación, cumplimentado un formulario con los datos recopilados en el paso anterior.
3. **Colocación de marcadores reflectantes** según el estándar Helen Hayes de cuerpo completo. En los siguientes apartados se describe la posición anatómica de 22 marcadores distribuidos según dicho protocolo.
4. **Grabación de movimientos.** Para cada participante se graban un total de 17 capturas de movimiento, con una duración de 30 segundos cada una, incluyendo marcha normal y varias simulaciones de alternaciones en la marcha (según la escala de Tinetti):
 - a) **Marcha normal.** Se comienza realizando cinco repeticiones de marcha normal. Para ello, se pide al sujeto que camine con su paso habitual a un ritmo ligero, procurando andar en línea recta (sin desviación de la trayectoria), con pasos continuos y con una longitud de paso aproximadamente similar en ambas extremidades.
 - b) **Longitud y altura de paso.**
 - Dos repeticiones simulando una marcha patológica en la que el pie derecho no sobrepasa al izquierdo con el paso en la fase de balanceo.
 - Dos repeticiones simulando una marcha patológica en la que el pie izquierdo no sobrepasa al derecho con el paso en la fase de balanceo.
 - Dos repeticiones simulando una marcha patológica en la que el pie derecho no se levanta completamente.
 - Dos repeticiones simulando una marcha patológica en la que el pie izquierdo no se levanta completamente.
 - c) **Discontinuidad entre pasos.** Dos repeticiones en las que se pide al participante que adopte una marcha discontinua e interrumpida, realizando breves paradas entre pasos y en la que se apoya completamente un pie antes de levantar el otro.
 - d) **Postura en la marcha.** Dos repeticiones simulando una marcha con los talones separados al caminar para conseguir un notable aumento de la base de sustentación.

Por tanto, por cada participante se han generado 17 grabaciones de 30 segundos: 5 caminando normal y el resto simulando distintos tipos de marchas patológicas, con la intención de que el conjunto de datos final sirva de línea base para la construcción de modelos que automaticen la escala de marcha en la prueba de Tinetti. Es decir, se cuenta con un total de 170 grabaciones de marcha, de las cuales 50 se etiquetan como normal y las 120 restantes abarcan 6 de los trastornos valorados en **POMA** (20 capturas en cada ítem patológico).

Además, en cumplimiento con lo dispuesto en el [Reglamento General de Protección de Datos \(RGPD\)](#), se ha elaborado una hoja de información y consentimiento informado (Anexo C) que ha sido entregada y firmada por todas las personas participantes en la investigación. En este documento se explica a los usuarios la finalidad y objetivos del estudio, se garantiza la confidencialidad con la que se tratarán los datos personales y se informa de la posibilidad de retiro voluntario en cualquier momento.

Medidas antropométricas

Con la finalidad de que *Clinical 3DMA* pueda obtener una estimación más precisa de la ubicación de los centros de rotación articulares, antes de comenzar la grabación de datos se ha preguntado a cada participante su edad y peso (en kilogramos), y se ha utilizado un *segmómetro* flexible disponible entre la instrumentación del laboratorio para medir la altura del sujeto y recopilar algunas medidas antropométricas del tren inferior. Además, cabe señalar que estas medidas antropométricas se han utilizado posteriormente en la fase de preprocesamiento para analizar y eliminar la influencia de las dimensiones corporales de cada participante sobre los parámetros estimados de la marcha (Sección 5.2.4).

Durante la toma de medidas antropométricas es importante seguir procedimientos estandarizados para incrementar la precisión, y, por tanto, la validez y exactitud de las mediciones. A continuación, se detallan las medidas corporales que han sido recopiladas, siguiendo las consideraciones de fuentes bibliográficas especializadas en antropometría para obtener datos más fiables [11, 132, 133]. Para complementar la explicación, la Figura 5.4 muestra la ubicación de los puntos anatómicos referenciados.

- **Longitud de pierna.** Distancia entre la primera eminencia del borde anterior del hueso coxal (espina iliaca anterosuperior) y el maléolo medial (interno) del tobillo.
- **Altura de rodilla.** Distancia vertical medida desde el suelo hasta la superficie superior de la rodilla (epicóndilo lateral del fémur).

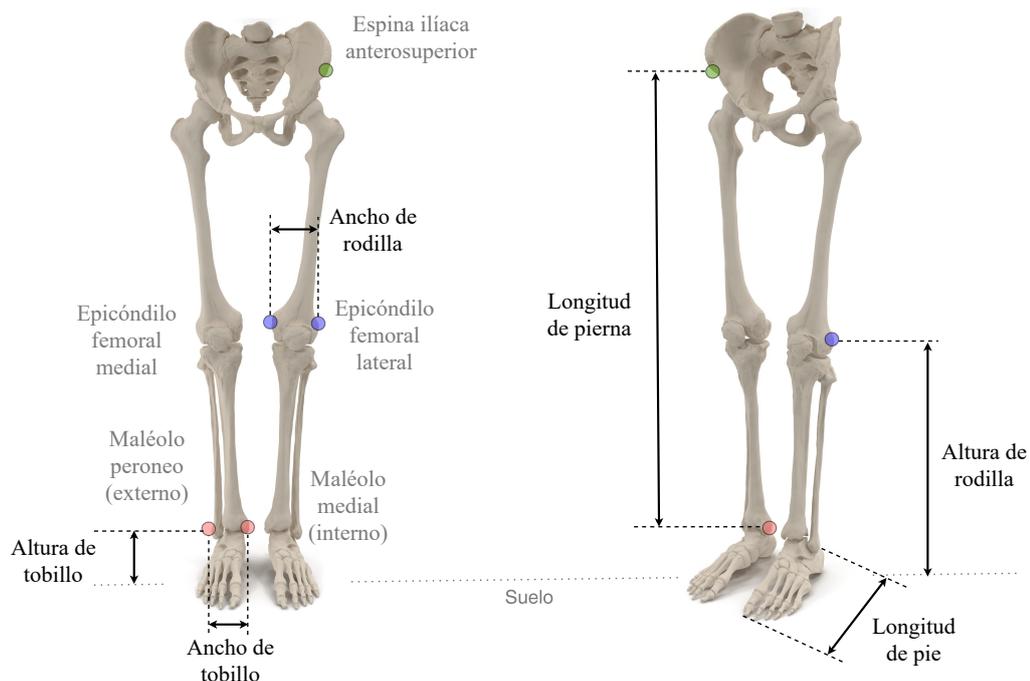


Figura 5.4: Puntos anatómicos y medidas antropométricas. Elaboración propia.

- **Ancho de rodilla.** Distancia horizontal entre el epicóndilo lateral del fémur y el epicóndilo femoral medial.
- **Altura de tobillo.** Distancia vertical medida desde el suelo hasta la zona exterior más prominente del tobillo (maléolo peroneo o externo).
- **Ancho de tobillo.** Distancia horizontal entre el maléolo peroneo (externo) y el maléolo medial (interno).
- **Longitud de pie.** Distancia horizontal desde el talón hasta la punta del dedo más prominente del pie.

Todas las medidas anteriores, excepto la longitud de pierna, se han tomado con la persona sentada con los pies apoyados al suelo, mirando hacia el frente y con el tronco erecto formando un ángulo de 90° con los muslos, al igual que la articulación de rodillas. Por su parte, para medir la longitud de la extremidad inferior y la estatura, el participante ha permanecido de pie en posición anatómica.

Ubicación de marcadores: protocolo Helen Hayes

Cuando se utilizan marcadores anatómicos es fundamental ubicarlos siempre siguiendo un estricto criterio, ya que pequeñas diferencias en su colocación pueden provocar datos imprecisos y errores significativos entre sujetos o entre sesiones [16]. En este trabajo se ha seguido el modelo de Helen Hayes de cuerpo completo modificado por STT [134], el cual utiliza 22 marcadores pasivos ubicados en determinados puntos anatómicos ilustrados en la Figura 5.5 y descritos en la Tabla 5.1.

La principal diferencia entre el conjunto de marcadores original [135] y esta versión modificada es que se sustituyen los marcadores ubicados en los trocánteres mayores (zona superior del hueso del cuerpo del fémur), por dos marcadores situados sobre los talones. Este cambio se debe a que los cálculos utilizados por el sistema de captura para determinar los centros de rotación de la cadera sólo requieren los marcadores colocados sobre la espina ilíaca anterosuperior (7 y 8 en la Figura 5.5) [134].

Tabla 5.1: Ubicación de los marcadores según el protocolo Helen Hayes modificado por STT

	Nombre del marcador	Posición anatómica
1	right_shoulder	Acromio de la escápula derecha
2	left_shoulder	Acromio de la escápula izquierda
3	right_elbow	Epicóndilo lateral de humero derecho
4	left_elbow	Epicóndilo lateral de humero izquierdo
5	right_wrist	Apófisis estiloides del cúbito derecho
6	left_wrist	Apófisis estiloides del cúbito izquierdo
7	right_asis	Espina ilíaca anterosuperior derecha
8	left_asis	Espina ilíaca anterosuperior izquierda
9	right_knee	Epicóndilo lateral de fémur derecho
10	left_knee	Epicóndilo lateral de fémur izquierdo
11	right_ankle	Maléolo lateral del peroné derecho
12	left_ankle	Maléolo lateral del peroné izquierdo
13	right_toes	Entre los metatarsos M2 y M3 del pie derecho
14	left_toes	Entre los metatarsos M2 y M3 del pie izquierdo
15	neck	Vértebra C7
16	back_hip	Union lumbosacral

Table 5.1 continued from previous page

	Nombre del marcador	Posición anatómica
17	left_lateral_thigh	Muslo izquierdo centrado; con extensor de marcador
18	right_lateral_thigh	Muslo derecho centrado; con extensor de marcador
19	left_calf	Pantorrilla izquierda centrada; con extensor de marcador
20	right_calf	Pantorrilla derecha centrada; con extensor de marcador
21	left_heel	Parte posterior de calcaneo pié derecho
22	right_heel	Parte posterior de calcaneo pié izquierdo

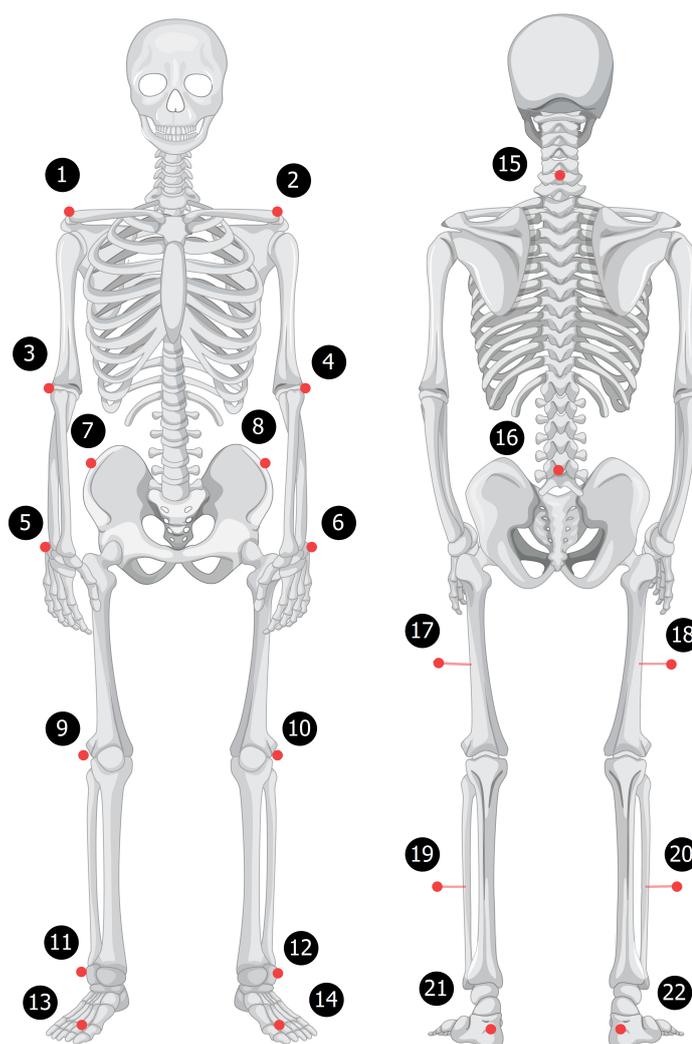


Figura 5.5: Modelo de ubicación de marcadores pasivos según el protocolo Helen Hayes. (Elaboración propia)

De los 22 marcadores utilizados, 18 se colocan directamente sobre la piel o la ropa del sujeto utilizando cintas adhesivas o velcros, mientras que los cuatro restantes (numerados como 17, 18, 19 y 20 en la Figura 5.5) se montan sobre una barras extensoras proyectadas lateralmente al segmento que permiten un mejor cálculo de las rotaciones articulares en el eje longitudinal [14]. Asimismo, en la documentación de *Clinical 3DMA* se recomienda una rigurosa colocación y ajuste de los marcadores en el eje anteroposterior para obtener medidas precisas, teniendo en cuenta las siguientes consideraciones:

- Los marcadores número 17 y 19 (muslo y gemelo izquierdo), así como los 18 y 20 (muslo y gemelo derecho) deben estar alineados.
- A su vez, el marcador 19 (gemelo izquierdo) debería estar alineado con los marcadores de la rodilla (10) y tobillo izquierdo (12). Lo mismo para los marcadores 20, 9 y 11 en la extremidad derecha.

En la Figura 5.6 se muestran los marcadores colocados en uno de los participantes en la adquisición de datos siguiendo el protocolo Helen Hayes explicado anteriormente.



Figura 5.6: Vista lateral (izquierda) y trasera (derecha) de los marcadores reflectantes colocados en uno de los participantes siguiendo el protocolo Helen Hayes

Una vez completadas las grabaciones de todos los participantes siguiendo el protocolo anteriormente explicado, el trabajo se ha centrado en exportar los datos desde la plataforma *Clinical 3DMA* en distintos formatos para generar un conjunto de datos flexible y reutilizable. En el siguiente apartado se documenta el proceso llevado a cabo para generar el *dataset* de marcha a partir de los datos grabados por el sistema óptico de captura de movimiento.

5.1.3. Exportación de datos y construcción del *dataset*

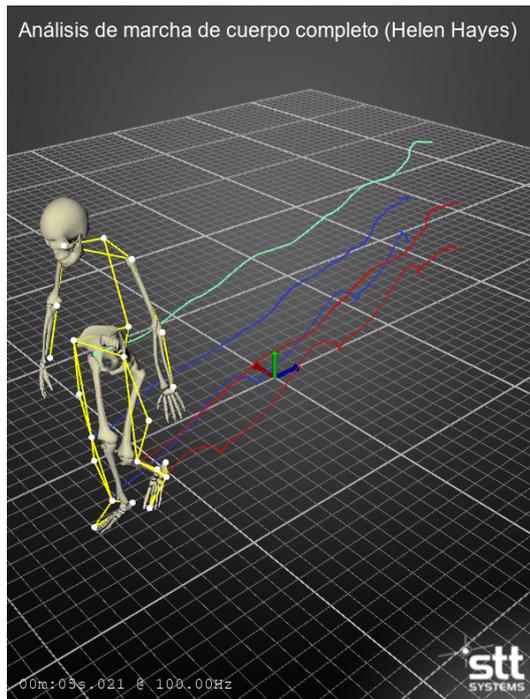
Para cada grabación de movimiento, *Clinical 3DMA* permite exportar los datos en una amplia variedad de formatos, desde archivos de muy bajo nivel (trayectorias de los marcadores) hasta información procesada de más alto nivel, como documentos con análisis de la marcha. En este sentido, aunque en el desarrollo de este trabajo solo se fueran a utilizar los datos en algunos de los formatos exportables, para la construcción del *dataset* se ha decidido almacenar las capturas de movimiento en todas las formas ofrecidas por *Clinical 3DMA*, con la intención de dotar al conjunto de datos resultante de una mayor flexibilidad y adaptabilidad a trabajos futuros.

De acuerdo con lo expresado, los formatos en los que se pueden exportar las capturas de movimiento desde *Clinical 3DMA* son los siguientes:

- **Archivo completo de captura 3MC.** Es un formato creado por *STT* y utilizado en los diferentes productos *3DMA* para almacenar todo tipo de información relacionada con una grabación, incluyendo datos en bruto, información de calibración y clips de vídeo sincronizados con el movimiento. Es el tipo de archivo que se utiliza cuando se necesita importar una captura en *Clinical 3DMA*.
- **Archivo de vídeo con la visualización del movimiento en 3D.** Los productos *3DMA* generan un archivo MP4 que contiene la vista en 3D de una captura de movimiento, tal y como se muestra en la captura de la Figura 5.7a.
- **Archivo de captura en formato C3D.** Es un formato estándar compatible con la gran mayoría de sistemas ópticos de captura de movimiento, que incluye toda la información necesaria para poder cargar, visualizar y analizar datos de movimiento en 3D con paquetes de *software* desarrollados por terceros como *Mokka*¹ o *MLSviewer*.
- **Trayectoria de marcadores en formato TRC.** Es un formato de datos creado por *Motion Analysis Corporation*², una de las compañías líderes en la fabricación de sistemas ópticos para la captura de movimiento, y que permite registrar los datos sin procesar (posiciones de los marcadores) durante una grabación de movimiento.
- **Trayectoria de marcadores en formato TXT.** *Clinical 3DMA* también permite exportar las trayectorias de los marcadores a un archivo TXT con una estructura más simple que el formato TRC. En estos archivos, la primera fila muestra los nombres de los marcadores, y las siguientes filas incluyen, para cada trama, las coordenadas espaciales de dichos marcadores, tal y como se muestra en el ejemplo de la Figura 5.19
- **Fichero TXT con tiempos de eventos.** Desde *3DMA* se puede exportar un archivo TXT con marcas de tiempo que indican en qué instantes de la grabación ha pasado algo relevante. Para el análisis de marcha, los eventos registrados automáticamente son el contacto con el suelo y despegue de cada pie.
- **Informes de análisis de marcha.** Son unos archivos de documento en formato Open XML de Microsoft Word que incluyen figuras, tablas y gráficas para, en este caso, componer un informe con parámetros espaciotemporales y cinemáticos de la marcha a partir de los datos grabados.
- **Archivos CSV con datos biomecánicos.** Los productos *3DMA* permiten generar archivos CSV con datos biomecánicos durante el movimiento. Más concretamente, desde *Clinical 3DMA* se puede exportar un archivo CSV por cada uno de los siguientes parámetros: ángulos del brazo, cadera, tronco, pelvis, pie, rodilla y tobillo, así como la posición del COG (ejemplo en la Figura 5.7b) y velocidad angular durante la marcha, de forma que cada CSV incluye una fila con los valores angulares para cada *frame* de la grabación. En este caso, 3000 filas (30 segundos de captura a 100Hz) por archivo.
- **Archivos CSV con valores cinemáticos.** A partir de los datos angulares anteriores, se pueden exportar hasta 26 archivos CSV que incluyen valores cinemáticos para las zancadas izquierda y derecha en los tres planos anatómicos: inclinación, oblicuidad y rotación de pelvis, flexión-extensión, abducción-aducción y rotación para las articulaciones de cadera, rodillas y tobillos, así como la dirección del pie durante la marcha. A diferencia de los archivos anteriores, en este caso los datos se proporcionan segmentados en zancadas, de forma que cada tabla contiene una columna por zancada y 100 filas que representan el instante dentro del ciclo de la marcha (0 a 100 %).

¹<https://biomechanical-toolkit.github.io/mokka/>

²<https://motionanalysis.com/>



(a) Vista 3D de una captura de movimiento

	A	B	C	D
1	Tiempo	CDG X	CDG Y	CDG Z
2	0	60,78877	88,95882	205,9523
3	0,01	60,8022	88,96023	205,93053
4	0,02	60,80463	88,96417	205,91052
5	0,03	60,81589	88,97406	205,87442
6	0,04	60,83459	88,97408	205,84155
7	0,05	60,85494	88,96501	205,81902
8	0,06	60,86965	88,96127	205,80925
9	0,07	60,87945	88,9584	205,7406
10	0,08	60,94053	88,95685	205,74974
11	0,09	60,95564	88,95951	205,7321
12	0,1	60,96525	88,95956	205,7014
13	0,11	60,98292	88,9554	205,67395
14	0,12	60,9987	88,9436	205,6339
15	0,13	61,01183	88,92315	205,59769
16	0,14	61,04355	88,91238	205,55594
17	0,15	61,07164	88,9087	205,52689
18	0,16	61,10975	88,9305	205,49217
19	0,17	61,15434	88,92349	205,45947
20	0,18	61,19247	88,92205	205,43272
21	0,19	61,25304	88,90744	205,37105
22	0,2	61,29194	88,89989	205,33304

(b) CSV con posición del centro de gravedad

Figura 5.7: Ejemplo de formatos exportados desde *Clinical 3DMA*

A partir de estos formatos de datos exportados desde *Clinical 3DMA*, se ha construido un conjunto de datos siguiendo la estructura de directorios y ficheros descrita a continuación.

Estructura del *dataset*

El conjunto de datos se ha organizado en una serie de archivos y directorios que ordenan los datos por participantes y grabaciones, tal y como se ilustra gráficamente en la Figura 5.8. En la raíz del *dataset* se incluyen tres archivos *CSV* (*summary*, *demographics* y *anthropometry*) con información común a todas las grabaciones. Más concretamente, *summary.csv* actúa como un índice general del conjunto de datos, especificando para cada grabación la etiqueta que le corresponde y su ubicación dentro de la estructura de directorios del *dataset*. Dichas etiquetas han sido codificadas según se indica en la Tabla 5.2. Además, en la Tabla D.1 del Anexo D se especifica la etiqueta asignada a cada grabación, junto con otra información de preprocesamiento explicada en la Sección 5.2.

Tabla 5.2: Etiquetado del conjunto de datos

Etiqueta	Grabaciones	Descripción
norm	rec_01 a rec_05	Marcha normal
lap1	rec_06 y rec_07	El pie derecho no sobrepasa al izquierdo en la fase de balanceo
lap2	rec_08 y rec_09	El pie izquierdo no sobrepasa al derecho en la fase de balanceo
lap3	rec_10 y rec_11	El pie derecho no se levanta completamente
lap4	rec_12 y rec_13	El pie izquierdo no se levanta completamente
dc	rec_14 y rec_15	Discontinuidad entre pasos. Es un marcha interrumpida en la que se realizan varias paras entre pasos
pm	rec_16 y rec_17	Postura en la marcha. Marcha en la que aumenta la base de sustentación (talones separados al caminar)

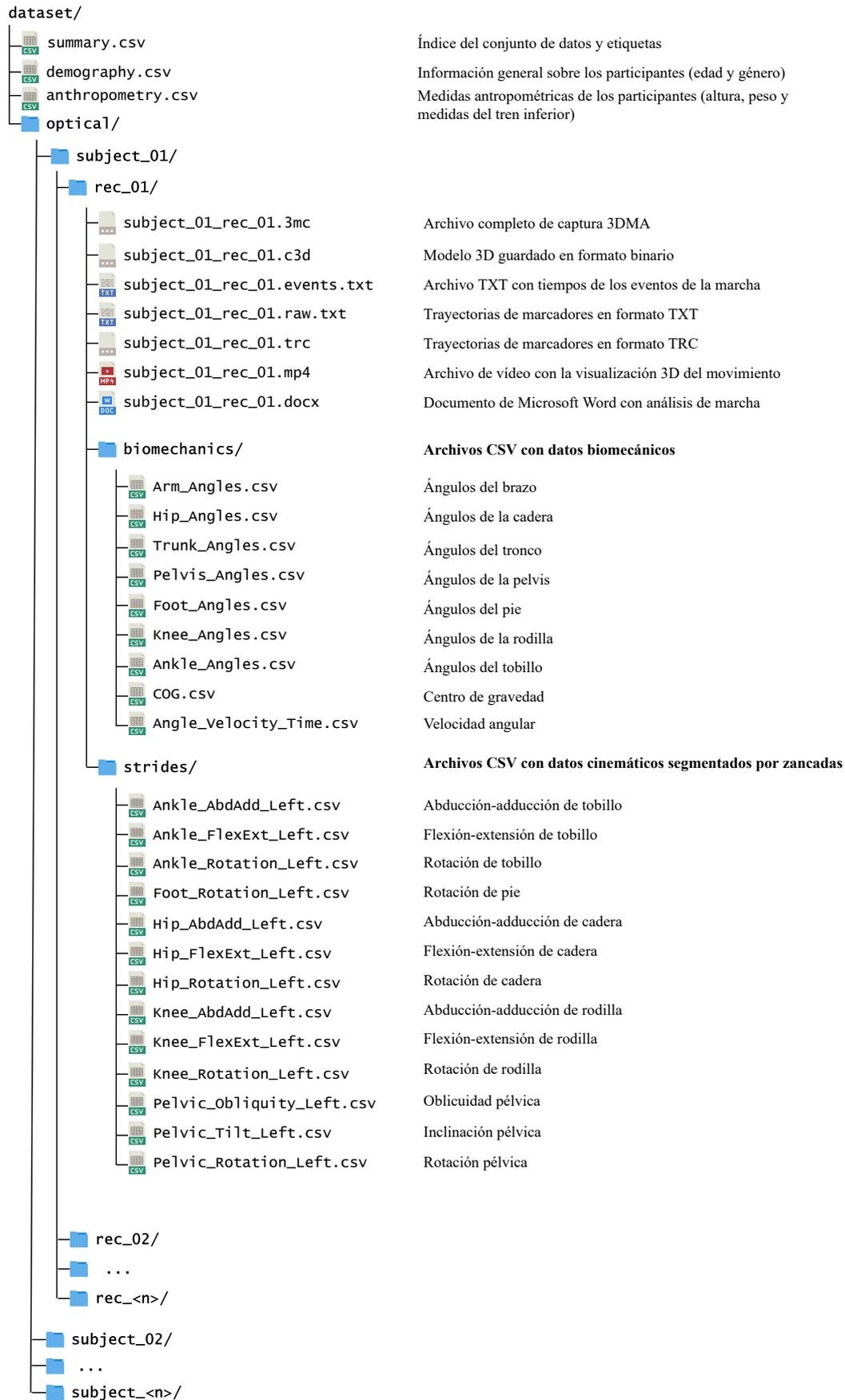


Figura 5.8: Organización en directorios y archivos del conjunto de datos

Por su parte, en el archivo *demographics.csv* se almacena información general sobre los participantes (edad y género), mientras que la tabla *anthropometry.csv* registra las medidas antropométricas tomadas para cada sujeto. En ambos archivos se tiene una fila por participante, incluyendo una primera columna con un identificador único de la forma *subject_<num>*, que sirve como enlace con el resto de los archivos del *dataset*.

Tabla 5.3: Columnas del archivo *anthropometry.csv*

Columna	Descripción
<i>subject</i>	Identificador único del participante. Es una cadena de texto de la forma " <i>subject_<num></i> "
<i>height</i>	Estatura del sujeto en centímetros.
<i>weight</i>	Peso en kilogramos.
<i>knee_width</i>	Anchura de rodilla medida en milímetros.
<i>knee_height</i>	Altura de rodilla medida en milímetros..
<i>ankle_width</i>	Anchura de tobillo medida en milímetros.
<i>ankle_height</i>	Altura de tobillo medida en milímetros.
<i>leg_length</i>	Longitud de pierna medida en milímetros.
<i>foot_length</i>	Longitud de pie medida en milímetros.

Sobre el directorio principal del *dataset* se encuentra un subdirectorio por cada participante, nombrado con el mismo identificador (*subject_<num>*) que se ha utilizado en los archivos *demographics* y *anthropometry*. A su vez, dentro de la carpeta de cada sujeto se ha definido un nuevo subdirectorio por cada grabación, dando así lugar a la estructura jerárquica que se ilustra gráficamente en la Figura 5.8. Estas carpetas asociadas a una grabación han sido nombradas como *rec_<num>*, de forma que el archivo *summary* global anteriormente mencionado especifica qué tipo de marcha contiene cada uno de estos subdirectorios.

Particularmente, puesto que en la adquisición de datos han participado 10 sujetos, el conjunto de datos contiene 10 carpetas nombradas como *subject_01* a *subject_10*, y dentro de cada uno de estos directorios, se han creado 17 carpetas (*rec_01* a *rec_17*) que contienen los datos exportados de las diferentes grabaciones realizadas por cada participante. Además, es importante remarcar que en el *dataset* los datos están anonimizados y únicamente identificados por estos códigos, de forma que no se proporciona información alguna que permita reconocer la identidad personal de los participantes.

En cuanto a los datos exportados para cada captura, tal y como se ha comentado anteriormente se ha decidido almacenar todos los formatos de archivo soportados por *Clinical 3DMA* para conseguir un *dataset* lo más flexible posible. Por tanto, dentro del directorio de cada grabación se incluye un archivo 3MC importable en *Clinical 3DMA* con toda la información de la captura; un archivo C3D con el modelo tridimensional guardado en formato binario; un archivo de vídeo en MP4 con la visualización 3D del movimiento; un documento de *Microsoft Word* con un informe de la marcha; un archivo TXT con los tiempos de los diferentes eventos durante la marcha, y dos archivos (TXT y TRC) con las trayectorias de los marcadores. Asimismo, en el directorio de cada captura se han creado dos nuevas carpetas, llamadas *biomechanics* y *stride*, que contienen una serie de archivos CSV almacenando, respectivamente, datos biomecánicos durante la captura de movimiento y datos cinemáticos segmentados por zancadas.

El conjunto de datos puede ser accedido desde el enlace proporcionado en el Anexo A. Su tamaño total, incluyendo todos los formatos descritos anteriormente, es de 16.1 GB.

No obstante, es conveniente señalar que cada grabación de movimiento tiene un peso aproximado de 92 MB, de los cuales cerca de 80MB se corresponden con el archivo 3MC de la captura *Clinical 3DMA* y el vídeo MP4 de la representación del movimiento en 3D. En este sentido, los archivos **CSV** individuales con valores cinemáticos de la marcha segmentados por zancada pesan, aproximadamente, 20KB, mientras que los **CSV** con datos biomecánicos ocupan unos 180KB.

Script para automatizar la exportación de datos

Exportar las capturas de movimiento desde *Clinical 3DMA*, en los diferentes formatos que soporta el sistema, es una tarea lenta y repetitiva, ya que el *software* no incluye ninguna herramienta de línea de comandos o algún mecanismo que permita agilizar o acelerar este trabajo. En su lugar, todo el proceso de exportación debe realizarse a través de la interfaz de usuario de la aplicación, y no ofrece ninguna funcionalidad que permita exportar varias grabaciones o formatos de manera simultánea, por lo que es necesario repetir una larga secuencia de clics y movimientos del ratón en cada grabación y cada formato a exportar.

Una forma de agilizar tareas con estas características consiste en utilizar herramientas que imitan automáticamente las acciones del ratón y teclado sobre las ventanas de una aplicación. Esta técnica se conoce como automatización de la interfaz gráfica de usuario (*GUI automation*), y aunque no se considera el método más elegante para la automatización de tareas, esta es la única alternativa para agilizar el proceso de exportación de datos en la versión actual de la herramienta *Clinical 3DMA*. En este sentido, se ha utilizado *PyAutoGUI*³, una librería de *Python* que permite controlar programáticamente las operaciones del ratón y teclado, de forma que se ha implementado un *script* que automatiza la interacción con *Clinical 3DMA* para exportar un conjunto de grabaciones en todos los formatos de datos soportados.

Al igual que la mayoría de las herramientas para la automatización **GUI**, esta librería de *Python* se basa en el sistema de coordenadas del dispositivo de visualización, de forma que los puntos de la pantalla se describen mediante pares de coordenadas x e y . En este sentido, la Figura 5.9 muestra el sistema de coordenadas para una pantalla con resolución 1920x1080 píxeles, similar a la utilizada durante la implementación de la herramienta.

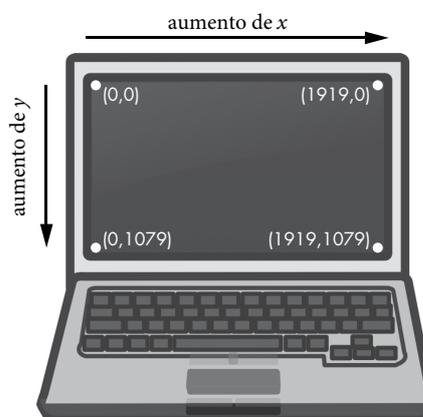


Figura 5.9: Sistema de coordenadas en una pantalla con resolución 1920x1080 píxeles. Imagen tomada de [136].

³<https://pyautogui.readthedocs.io/>

El punto de origen (0,0) se corresponde con la esquina superior izquierda, teniendo en cuenta que todos los puntos son valores positivos, ya que el valor de las coordenadas en el eje x aumenta hacia la derecha, y las coordenadas en y se incrementan hacia abajo. Como consecuencia del uso de este sistema de coordenadas, el primer paso para la implementación del *script* que automatice el proceso de exportación de datos ha consistido en el uso de *Mouse Point Viewer*⁴ para obtener la posición en la pantalla de los principales componentes de la GUI de *Clinical 3DMA*.

En cuanto al funcionamiento del *script* implementado, cuenta con un archivo de configuración en formato JSON que permite especificar el conjunto de grabaciones y formatos a exportar, así como el directorio en el que almacenar los ficheros generados. Partiendo de estos ajustes, el algoritmo seguido por el *script* de exportación consiste en los siguientes pasos ejecutados secuencialmente:

1. Comprobar que *Clinical 3DMA* (proceso llamado *MotionCaptor.exe*) se está ejecutando sobre un dispositivo con resolución 1920x1080 píxeles. En caso contrario, finaliza la ejecución del programa. Para determinar si el proceso indicado se está ejecutando, se ha utilizado la función `process_ite` de la librería *psutil*⁵, mientras que para obtener la resolución de la pantalla se ha empleado la función `size` de *PyAutoGUI*.
2. Cargar el archivo JSON de configuración y comprobar que el directorio de salida especificado es válido.
3. Maximizar la ventana de *Clinical 3DMA* y ponerla en primer plano. Para ello, se utiliza la función `getWindowsWithTitle` para buscar la ventana del *software* y los métodos `activate` y `maximize` de *PyAutoGUI*.
4. Por cada grabación definida en el archivo de configuración, repetir los siguientes pasos:
 - a) Si no existe, crear una carpeta para los datos de este usuario dentro del directorio de salida. El nombre de este subdirectorio se corresponderá con el código de usuario (*subject_<num>*).
 - b) Para cada una de las grabaciones del sujeto para la que se deseen exportar datos, según lo indicado en el JSON de ajustes, repetir los siguientes pasos:
 - 1) Si no existe, crear una carpeta para los datos de esta grabación dentro del directorio del usuario. El nombre de esta carpeta será de la forma *rec_<num>*.
 - 2) Haciendo uso de *PyAutoGUI*, interactuar con la pantalla de la base de datos en *Clinical 3DMA* para abrir la grabación correspondiente. Tal y como se muestra en la Figura 5.10, este proceso implica buscar y seleccionar el participante según el identificador de usuario (1), buscar la captura concreta utilizando el cuadro de búsqueda (2); pulsar sobre "Abrir captura" (3).
 - 3) Sobre la pestaña "Exportar datos" de la pantalla principal, (1) en la Figura 5.11, hacer clic secuencialmente sobre las siguientes opciones de formatos: captura completa en 3MC (2); trayectoria de marcadores en formato TRC (3) y TXT (5); modelo binario C3D (4); tiempos de eventos durante la marcha (6) y vídeo MP4 del movimiento en 3D (7). Además de hacer clic sobre cada una de estas opciones, es necesario gestionar el cuadro de dialogo de Windows que se abre para guardar los archivos.

⁴<https://github.com/ElektroStudios/Mouse-Point-Viewer/releases>

⁵<https://pypi.org/project/psutil/>

- 4) Para exportar los archivos **CSV** con datos biomecánicos, crear primero una carpeta llamada *biomechanichs* en el directorio de salida de la captura. Después, hacer clic sobre el icono (8) en la Figura 5.11, que permite guardar todas las curvas disponibles.
- 5) Para generar el informe Word con el análisis de la marcha hacer clic en la opción "Crear Informe" de la barra superior, (10) en la Figura 5.11. Se abrirá un cuadro de diálogo que permite seleccionar entre varios tipos de informe (marcha sobre cinta, marcha sobre suelo o resumido). Hacer clic sobre la opción "marcha sobre suelo" y esperar a que se abra un documento Microsoft Word con el informe generado.
- 6) Para exportar los archivos **CSV** con datos cinemáticos segmentados por zancadas, hacer clic en la opción "Selección de Zancada", (9) en la Figura 5.11. Se abre una nueva ventana desde la que se pueden guardar los datos cinemáticos con todas las zancadas registradas.

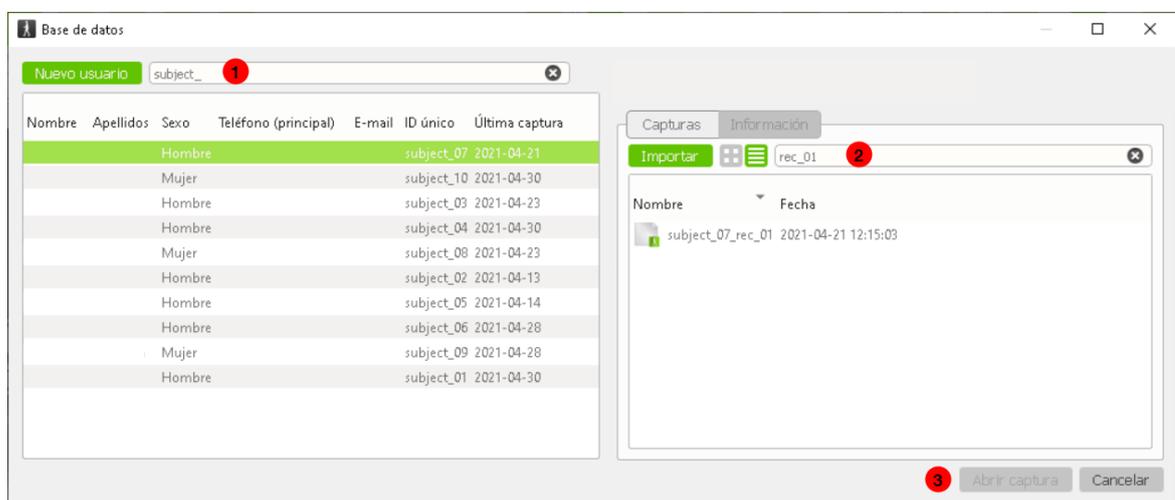


Figura 5.10: Pantalla de base de datos en *Clinical 3DMA*

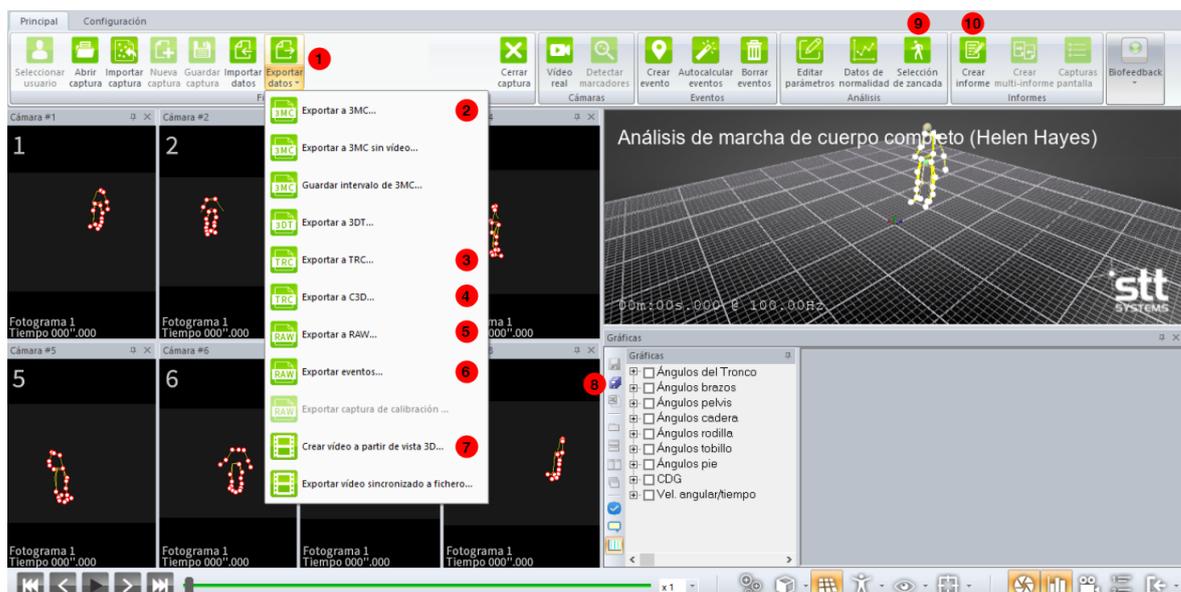


Figura 5.11: Pantalla principal en *Clinical 3DMA*

De esta forma, se ha conseguido implementar un *script* configurable que automatiza la exportación de datos desde el sistema óptico de captura de movimiento (*Clinical 3DMA*), reduciendo el tiempo necesario para llevar a cabo esta tarea, y generando un conjunto de datos que sigue la estructura y nomenclaturas de *dataset* explicada anteriormente. El código *Python* que implementa esta herramienta puede ser accedido desde el repositorio del proyecto, cuyo enlace se proporciona en el Anexo A, y en particular, dentro del directorio *3dma_exporter*.

5.1.4. Análisis exploratorio de datos sobre el *dataset* inicial

Una vez completada la fase de adquisición de datos siguiendo los criterios descritos en las secciones anteriores, como paso previo a la siguiente etapa de preprocesamiento y limpieza, es conveniente llevar a cabo un pequeño análisis exploratorio de los datos (EDA, por sus siglas en inglés), utilizando estadísticas descriptivas y técnicas de visualización que permitan comprender mejor los datos con los que se va a trabajar. Más concretamente, este estudio inicial se ha centrado en analizar las características demográficas y medidas antropométricas de los participantes, así como algunos de los formatos de datos exportados por el sistema de captura. En este sentido, conviene señalar que las variables predictoras específicamente utilizadas para analizar la marcha humana serán estimadas y calculadas mediante un proceso de ingeniería de características en la fase posterior (Sección 5.2).

Características demográficas y medidas antropométricas

Tal y como se ha comentado en la Sección 5.1.2, el *dataset* ha sido construido a partir de los movimientos capturados para 10 sujetos (7 hombres y 3 mujeres), con una media de edad de 29.8 ± 13.75 años. Para obtener un resumen más detallado de las características demográficas y antropométricas de los participantes en la captura, la Tabla 5.4 muestra una serie de estadísticas descriptivas para la edad y cada variable antropométrica, incluyendo media, mediana, máximo, mínimo, desviación estándar (*std*) y percentiles 25 y 75. En este sentido, conviene señalar que la estatura está representada en centímetros, el peso en kilogramos, y el resto de las medidas corporales en milímetros.

Tabla 5.4: Resumen de características demográficas y antropométricas

	edad	estatura	peso	anchura de rodilla	altura de rodilla	anchura de tobillo	altura de tobillo	longitud de pierna	longitud de pie
media	29.80	171.40	78.70	118.80	538.00	78.00	116.80	1009.50	265.70
mediana	24.50	171.50	74.00	115.00	535.00	75.00	115.00	977.50	260.00
std	13.75	7.23	17.19	15.70	35.83	9.18	17.67	69.65	24.72
mínimo	20	158.00	57.00	100.00	490.00	70.00	90.00	940.00	240.00
25 %	22.25	167.50	66.50	110.00	505.00	71.25	104.75	966.25	252.50
75 %	29.50	175.75	95.50	128.25	567.50	80.00	128.75	1025.00	268.00
máximo	66	184.00	104.00	145.00	590.00	100.00	150.00	1160.00	330.00

Por otro lado, una manera útil para explorar visualmente la relación entre cada par de variables numéricas consiste en trazar un histograma para cada atributo. La Figura 5.12 contiene 64 pequeños gráficos, resultado de emparejar dos a dos las variables antropométricas que han sido recogidas. En la diagonal se presentan una serie de histogramas que muestran la distribución de los valores de cada variable, y en cada gráfico se muestran los puntos diferenciados según el género.

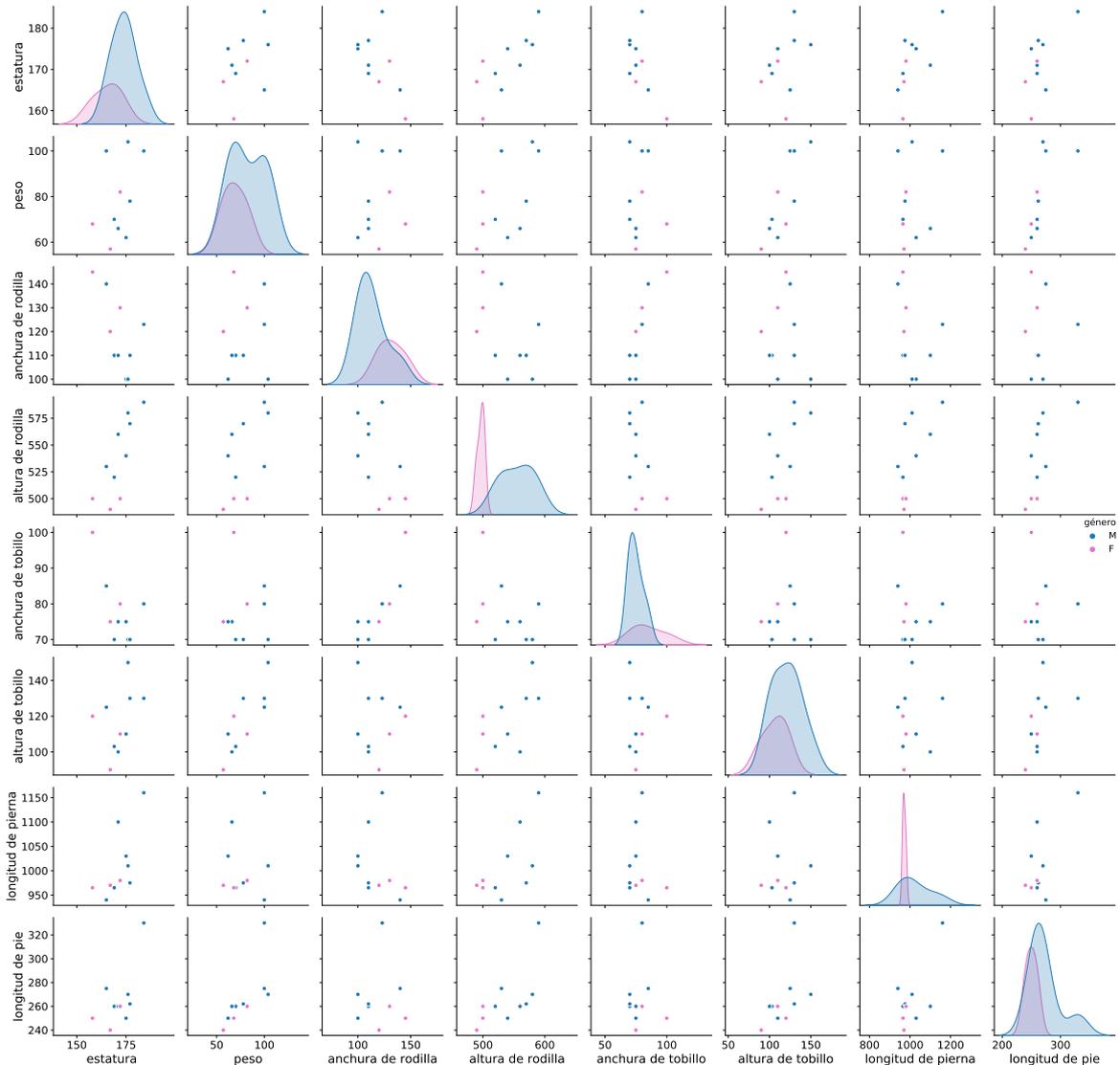


Figura 5.12: Visualización de las relaciones entre variables antropométricas del *dataset* diferenciadas por género

En estas gráficas se puede analizar cómo afecta cada variable a las demás. Por ejemplo, el gráfico de la cuarta fila y primera columna sugiere que existe una relación lineal entre la altura de rodilla y la estatura; mientras que en el gráfico de la penúltima fila y última columna parece que la longitud de pierna y pie están relacionadas. Sin embargo, en este contexto de análisis de marcha no es tan relevante la relación entre pares de características antropométricas, sino la influencia que estas variables tienen sobre los parámetros espaciotemporales de la marcha. Por tanto, en la Sección 5.2.4 se estudiará en detalle la relación entre estas medidas antropométricas y algunas variables propias del estudio de la marcha que serán estimadas durante la preparación de datos, aplicando las operaciones necesarias para tratar de reducir o eliminar tal influencia.

Trayectoria de marcadores

Para conocer y comprender mejor los datos adquiridos por el sistema óptico de captura de movimiento, durante este primer análisis exploratorio de los datos también resulta

interesante inspeccionar alguno de los ficheros con las trayectorias de los marcadores. En particular, el análisis se ha centrado en el formato TXT exportado desde *Clinical 3DMA*, y que de acuerdo con lo descrito en la Sección 5.1.3, incluye las coordenadas espaciales de dichos marcadores para cada trama de la grabación. Más concretamente, se trata de datos tabulares formados por dos columnas de índice (número de *frame* y marca de tiempo) más nueve columnas por cada marcador, correspondientes a la posición, velocidad y aceleración de cada punto en los tres ejes.

En este sentido, puesto que la captura de datos se ha realizado a una frecuencia de 100Hz y se han utilizado 22 marcadores, estos archivos de trayectorias contienen 3000 filas (30 segundos de grabación) y 200 columnas. En cualquier caso, la mejor manera de entender la estructura y contenido de estos datos es intentando visualizar la distribución de los marcadores en diferentes instantes de la grabación. Para ello, se ha importado uno de estos archivos junto con el fichero de eventos correspondiente a la misma grabación, y utilizando gráfico de dispersión (*scatter*) en tres dimensiones se ha representado visualmente la distribución de los marcadores al inicio de la captura (Figura 5.13a) y en dos instantes de tiempo relevantes: en un contacto inicial del talón con el suelo (HS) y en el despegue del pie (TO) (Figura 5.13b).

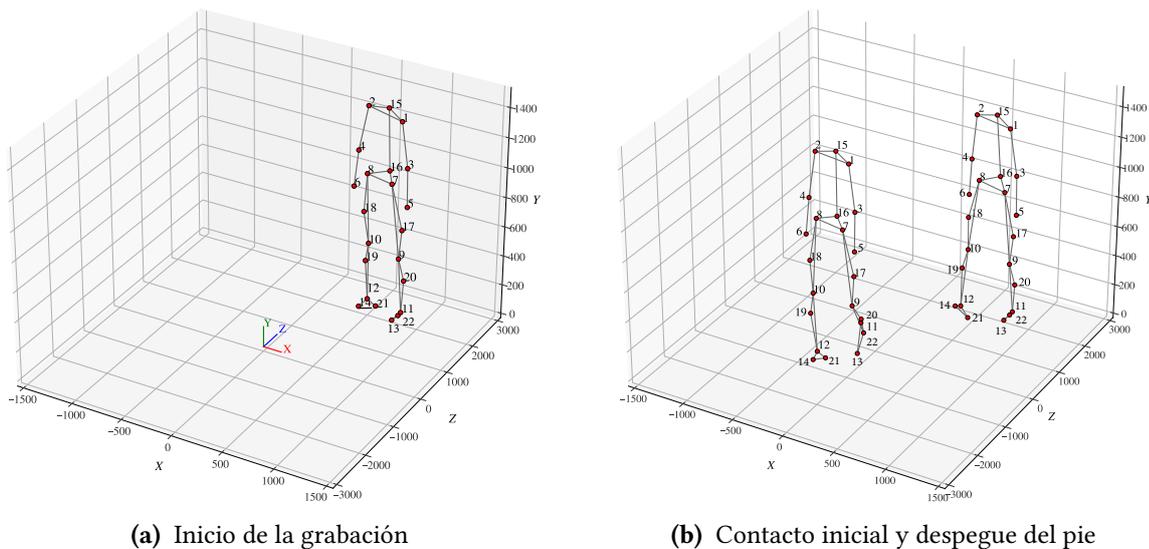


Figura 5.13: Distribución de los marcadores en diferentes instantes de la captura

A partir de estas gráficas de dispersión se deducen algunos aspectos relevantes relacionados con el sistema de referencia utilizado por la herramienta de captura. Tal y como se puede observar en las representaciones de la Figura 5.13, el centro de coordenadas (0,0,0) se ubica en el punto central de la zona de captura, y se corresponde con el lugar exacto donde se colocó la barra de marcadores durante el proceso de calibración. De esta forma, la posición espacial de cada marcador en un instante de la grabación viene dada por la distancia, en las tres dimensiones, a ese punto de calibración, medida en milímetros. Asimismo, se ha comprobado que los ejes X,Y,Z indican, respectivamente, el ancho (desplazamiento lateral), alto y dirección de progresión. Es fundamental conocer esta información para trabajar con este tipo de datos en las fases posteriores del proyecto.

Datos biomecánicos

Los archivos **CSV** con datos biomecánicos incluidos en el *dataset* son otro de los formatos de datos sobre los que resulta interesante llevar a cabo un pequeño análisis exploratorio inicial. Estos ficheros, con formato tabular, incluyen una fila para cada trama de la grabación (3000 filas para capturas de 30 segundos a 100Hz), y diferentes columnas dependiendo del segmento corporal y parámetro analizado. Como ejemplo, el archivo *COG.csv* de cada grabación incluye cuatro columnas, que se corresponden con el número de *frame* y la posición espacial del centro de gravedad (X,Y,Z).

En el gráfico de la Figura 5.14 se representa la trayectoria del centro de gravedad en sentido vertical (eje Y), durante una de las grabaciones de marcha normal incluidas en el conjunto de datos. Para relacionar esta visualización con lo comentado sobre el desplazamiento normal del **COG** en la Sección 3.1.3, se han trazado los eventos de la marcha, importados del fichero correspondiente, como líneas verticales. De esta forma, se puede comprobar que la posición vertical del centro de gravedad asciende desde el contacto inicial de una extremidad (líneas sólidas) hasta el despegue del pie opuesto (líneas punteadas), instante en el que comienza a descender hasta el contacto inicial siguiente. Además, se observa que, en este caso, el **COG** oscila a una altura comprendida entre 86 y 90 cm con respecto al suelo, coincidiendo así con los 4 cm de desplazamiento normal [6].

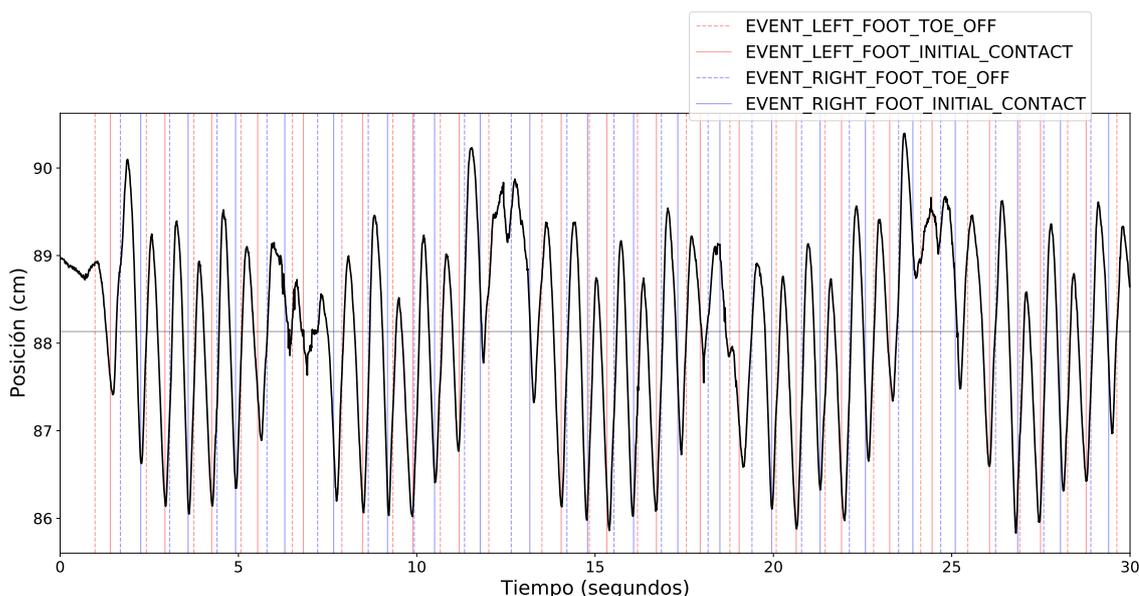


Figura 5.14: Desplazamiento vertical del centro de gravedad

Curvas cinemáticas

Dentro de la carpeta *strides* de cada grabación incluida en el *dataset*, se proporcionan unos ficheros **CSV**, exportados desde *Clinical 3DMA*, con valores cinemáticos para las zancadas izquierda y derecha en los tres planos anatómicos. A diferencia de los archivos anteriores, en este caso los datos están segmentados por zancadas, de forma que las filas representan la línea temporal normalizada de 0-100 % (100 filas), mientras que cada columna contiene los datos de un determinado ciclo de la marcha.

Con estos datos pueden trazarse los diagramas típicamente utilizados para analizar la cinemática articular de la marcha. Como ejemplo, en la Figura 5.15 se muestran las curvas cinemáticas en los planos sagital (primera columna), coronal (segunda columna) y transversal (tercera columna) para los casos de pelvis, cadera, rodilla, tobillo y pie.

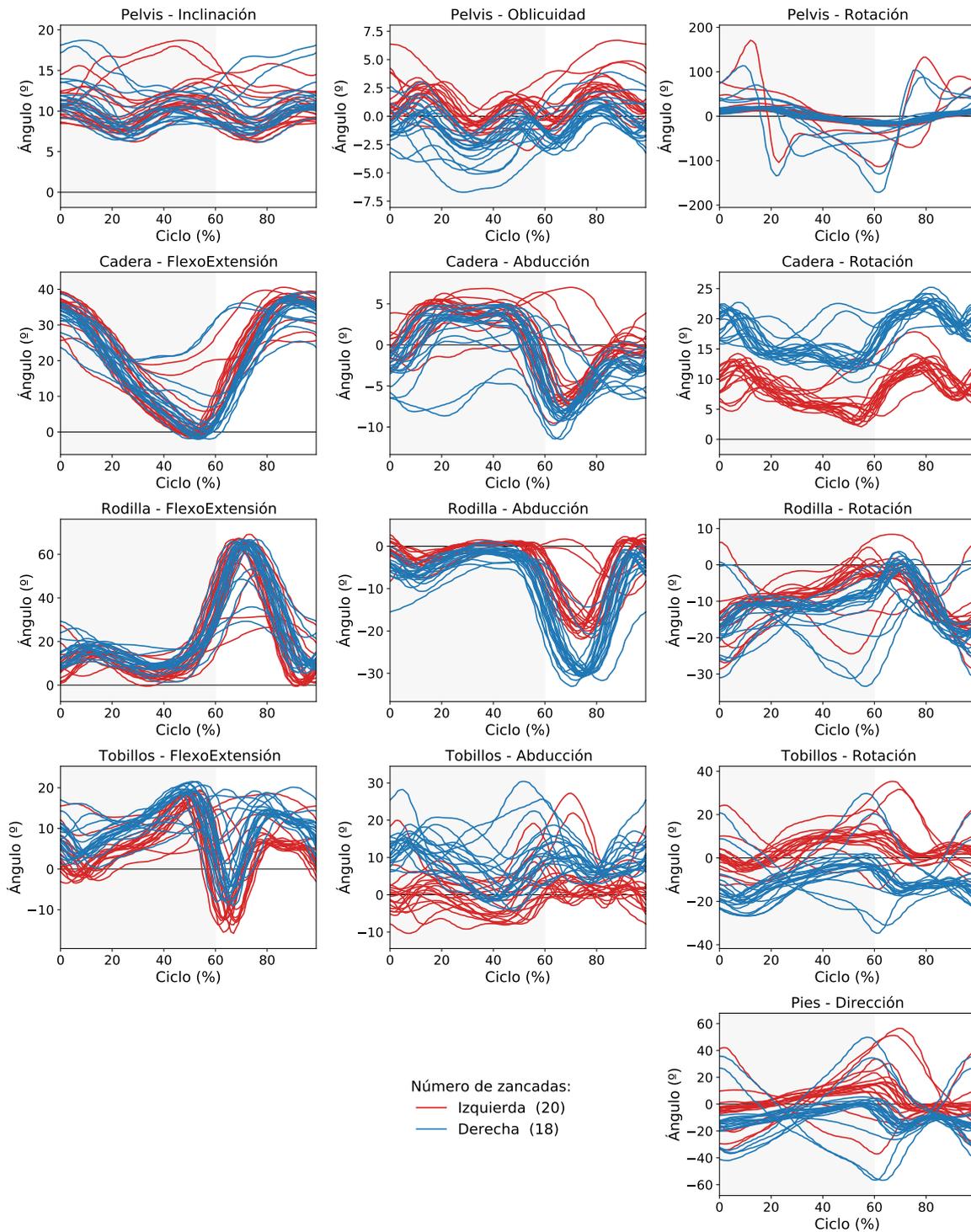


Figura 5.15: Curvas cinemáticas a partir de los datos exportados

En las gráficas anteriores se pueden observar algunas curvas significativamente desviadas del patrón definido por el resto de las trazas para esa grabación en particular. En este sentido, es importante llevar a cabo una fase de preprocesamiento para identificar datos, o zancadas en

este contexto, incompletas, incorrectas o inexactas producidas posiblemente por la oclusión de los marcadores durante el movimiento.

A modo de resumen de lo descrito hasta este punto, conviene señalar que durante esta fase de trabajo se ha construido un conjunto de datos formado por un total de 170 grabaciones (17 capturas por 10 participantes), incluyendo marcha normal y diferentes alteraciones durante el movimiento que han sido simuladas tomando como referencia la escala de Tinetti. Asimismo, se ha seguido el modelo Helen Hayes como protocolo validado para la ubicación de los marcadores, se han exportado los datos en una amplia variedad de formatos para dotar de una mayor flexibilidad al *dataset* resultante, y, finalmente, se ha implementado un *script* en *Python* que automatiza el proceso de exportación de los datos desde el sistema de captura. En la siguiente fase de trabajo se llevará a cabo el preprocesamiento, limpieza y preparación de estos datos, una etapa fundamental para obtener un conjunto de datos final sobre el que posteriormente aplicar técnicas de *Machine Learning*.

5.2. PREPARACIÓN DE DATOS

Una vez completada la recopilación de datos inicial, en esta sección se describen todas las tareas llevadas a cabo para construir el conjunto de datos preprocesado que se utilizará en la siguiente etapa de modelado. Entre las actividades de preprocesamiento abordadas en esta fase de trabajo, representadas de manera esquemática en la Figura 5.16, se encuentran la preparación de las trayectorias de marcadores (Sección 5.2.1) para, a partir de ellas, crear variables explicativas relacionadas con diferentes parámetros que caracterizan la marcha humana (Sección 5.2.2); la limpieza de datos, y en particular, el filtrado de zancadas en base a diferentes criterios, tales como giros durante la marcha, valores atípicos en los parámetros espaciotemporales, o problemas con las curvas cinemáticas (Sección 5.2.3); y, por último, la normalización de los datos de la marcha para eliminar la influencia de las características antropométricas de cada sujeto (Sección 5.2.4).

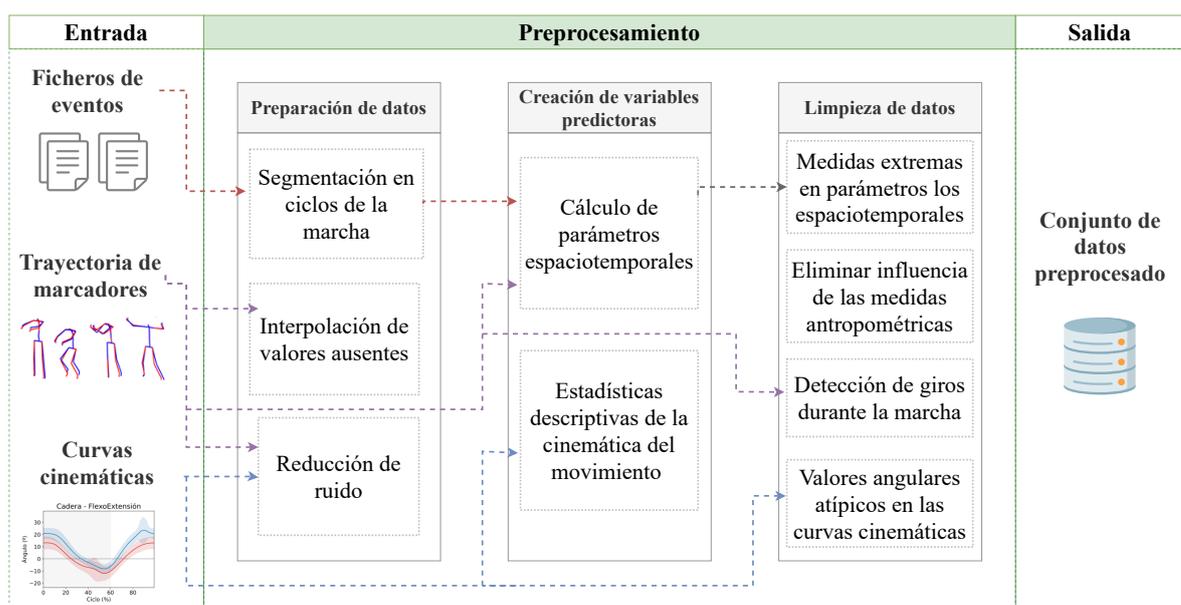


Figura 5.16: Esquema de las tareas abordadas en la fase de preprocesamiento

5.2.1. Eventos de marcha y trayectoria de marcadores

Se ha comenzado esta fase de preprocesamiento limpiando y consolidando los datos contenidos en dos de los ficheros más relevantes exportados desde *Clinical 3DMA*. En primer lugar, se han utilizado los eventos de la marcha registrados por el sistema de captura, para estructurar las trayectorias de los marcadores en zancadas, facilitando así el posterior análisis. Por otro lado, estas trayectorias de marcadores han sido preprocesadas para detectar y corregir valores faltantes.

Segmentación de datos en zancadas

De acuerdo con lo descrito en la Sección 3.1.1, generalmente se aprovecha la naturaleza periódica de la marcha humana para reducir la alta complejidad que conlleva su estudio, descomponiendo el movimiento en una secuencia cíclica de fases y eventos a los que se denomina ciclo de la marcha (zancadas). En este sentido, tomando como base los ficheros TXT con el tiempo de eventos, exportados desde *Clinical 3DMA*, es posible determinar en qué instantes de la grabación comienza y termina cada ciclo de la marcha para ambas extremidades. De esta forma, las trayectorias de los marcadores y los datos biomecánicos se pueden estructurar por ciclos de la marcha para facilitar su análisis.

El fichero TXT de eventos al que se hace referencia contiene marcas de tiempo que indican en qué instantes durante la captura se han producido acciones relevantes para el análisis. Tal y como se muestra en el archivo de ejemplo de la Figura 5.18, los eventos registrados son el contacto inicial del pie izquierdo y derecho con el suelo, o HS, (EVENT_LEFT_FOOT_INITIAL_CONTACT y EVENT_RIGHT_FOOT_INITIAL_CONTACT) y cuándo los dedos del pie dejan de contactar el piso, es decir, el comienzo de la fase de TO (EVENT_LEFT_FOOT_TOE_OFF y EVENT_RIGHT_FOOT_TOE_OFF). A partir de estos cuatro eventos se puede obtener el comienzo y fin de las fases de soporte y balanceo durante la marcha para ambas extremidades, así como el fin de respuesta a la carga y el inicio del prebalanceo, según se ilustra esquemáticamente en la Figura 5.17.

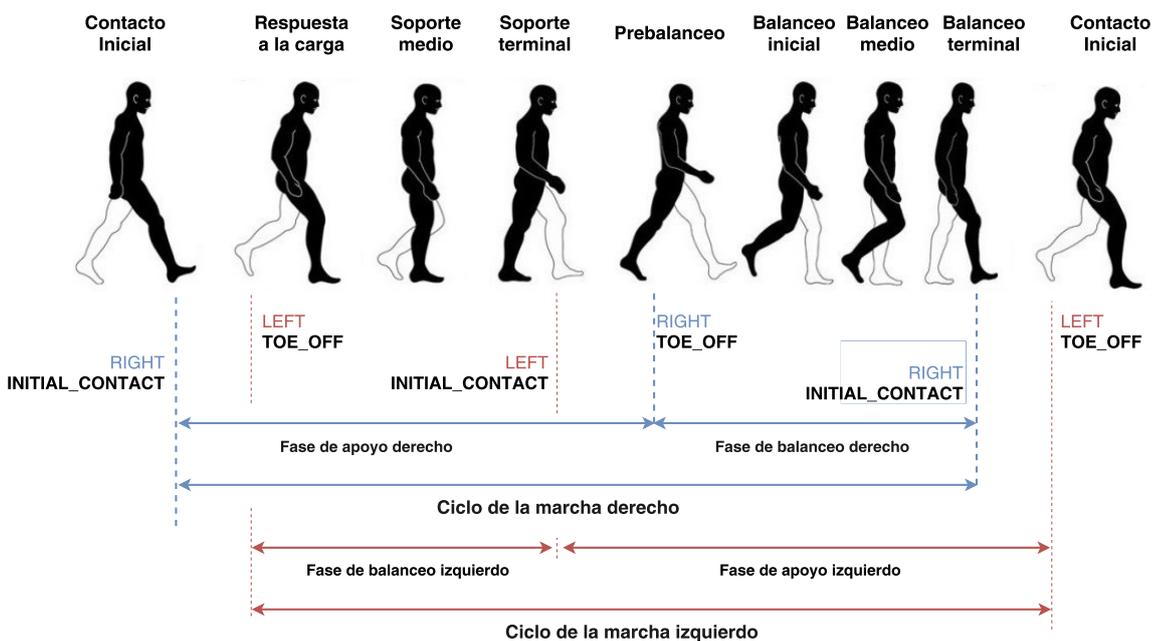


Figura 5.17: Ciclo de la marcha a partir de los eventos HS y TO. Elaboración basada en [12]

Aunque el ciclo de la marcha puede definirse y segmentarse a partir de cualquier evento relevante [9], se ha considerado el golpe de talón con el suelo HS como el inicio de cada zancada. De esta forma, los ciclos izquierdo y derecho en cada captura quedan delimitados, respectivamente, por dos eventos EVENT_LEFT_FOOT_INITIAL_CONTACT y EVENT_RIGHT_FOOT_INITIAL_CONTACT consecutivos, y están formados por tres eventos adicionales: el despegue y contacto de la extremidad opuesta, y el despegue de ese mismo pie. Así, el fichero de eventos durante la marcha se puede segmentar en ciclos, tal y como se muestra en la Figura 5.18, y, utilizando las referencias temporales en las que comienza y termina cada zancada, es posible organizar por ciclos los datos relativos a las trayectorias de marcadores y valores biomecánicos.

	0	EVENT_LEFT_FOOT_TOE_OFF	0.98	
Primera zancada izquierda	1	EVENT_LEFT_FOOT_INITIAL_CONTACT	1.41	Primera zancada derecha
	2	EVENT_RIGHT_FOOT_TOE_OFF	1.69	
	3	EVENT_RIGHT_FOOT_INITIAL_CONTACT	2.26	
Segunda zancada izquierda	4	EVENT_LEFT_FOOT_TOE_OFF	2.42	Segunda zancada derecha
	5	EVENT_LEFT_FOOT_INITIAL_CONTACT	2.94	
	6	EVENT_RIGHT_FOOT_TOE_OFF	3.07	
Tercera zancada izquierda	7	EVENT_RIGHT_FOOT_INITIAL_CONTACT	3.59	Tercera zancada derecha
	8	EVENT_LEFT_FOOT_TOE_OFF	3.75	
	9	EVENT_LEFT_FOOT_INITIAL_CONTACT	4.25	
	10	EVENT_RIGHT_FOOT_TOE_OFF	4.4	
	11	EVENT_RIGHT_FOOT_INITIAL_CONTACT	4.92	
	12	EVENT_LEFT_FOOT_TOE_OFF	5.07	
	13	EVENT_LEFT_FOOT_INITIAL_CONTACT	5.54	
	14	EVENT_RIGHT_FOOT_TOE_OFF	5.8	
	15	EVENT_RIGHT_FOOT_INITIAL_CONTACT	6.3	
	16	EVENT_LEFT_FOOT_TOE_OFF	6.51	

Figura 5.18: Extracción de zancadas en el fichero de eventos.

Durante esta extracción de zancadas se han detectado algunos eventos irrelevantes o incorrectos que han sido descartados. Por ejemplo, uno o varios eventos de despegue del pie antes del primer contacto inicial (como en el ejemplo de la Figura 5.18), o ciclos en los que faltaba o sobraba algún de los eventos, y que también han sido eliminados. De esta forma, se ha conseguido obtener la marca temporal del inicio y fin de cada zancada durante la marcha. En la siguiente sección, se explica cómo a partir de este fichero de eventos se han creado variables explicativas relacionadas con parámetros espaciotemporales.

Preprocesado de las trayectorias de los marcadores

En los sistemas ópticos de captura de movimiento es posible que las cámaras pierdan la visibilidad sobre uno o varios de los marcadores durante algunos instantes de la grabación. Cuando esto sucede, se habla de oclusión de marcadores porque no es posible reconstruir la posición espacial de algunos puntos para esas tramas de la captura.

En la Figura 5.19 se muestra un ejemplo del archivo que contiene los datos en crudo con las trayectorias de los marcadores, y que ha sido descrito en detalle en la Sección 5.1.4. Como se ha comentado, estos datos están estructurados en 3000 filas (30 segundos de grabación con una frecuencia 100 tramas por segundo) y 200 columnas (9 por cada uno de los 22 marcadores, correspondientes a la posición, aceleración y velocidad en las tres dimensiones, más dos índices que definen la marca temporal).

	A	B	C	D	E	F	G	H	I	J
1	right_asis	left_asis	s2	right_lateral	left_lateral	right_knee	left_knee	right_calf	left_calf	right_ankle
2		0 0.0000	716.40	912.24	2011.29	1.24	-0.78	-10.84	-407.71	556.34
3		1 0.0100	716.42	912.23	2011.18	-2.84	4.78	-18.73	608.19	1462.20
4		2 0.0200	716.35	912.34	2010.91	13.40	28.46	-70.73	938.01	709.51
5		3 0.0300	716.65	912.71	2009.94	15.33	16.67	-54.59	901.17	-1322.09
6		4 0.0400	716.69	912.67	2009.87	34.03	9.21	-24.27	2359.14	-209.70
7		5 0.0500	717.33	912.91	2009.38	54.01	9.42	-30.61	329.55	-793.51
8		6 0.0600	717.68	912.83	2009.32	39.75	-4.40	-2.95	-1846.53	-358.70
9		7 0.0700	718.10	912.87	2009.26	21.60	3.02	-3.45	-2438.39	521.32
10		8 0.0800	718.10	912.87	2009.26	-3.67	1.08	4.61	-1270.19	-758.82
11		9 0.0900	718.10	912.87	2009.26	-0.75	-9.29	-26.50	1166.20	-676.65
12		10 0.1000	718.12	912.70	2008.77	13.37	-11.89	-38.08	1071.68	-444.71
13		11 0.1100	718.35	912.63	2008.55	17.94	-19.54	-31.52	-66.02	-1146.72
14		12 0.1200	718.47	912.32	2008.13	17.23	-27.50	-32.60	1132.82	1549.56
15		13 0.1300	718.74	912.16	2007.87	38.04	3.40	-45.87	1119.83	699.57
16		14 0.1400	719.19	912.29	2007.26	38.25	-14.35	-34.67	505.07	-372.26
17		15 0.1500	719.53	911.94	2007.17	51.77	1.72	-25.98	1872.78	1306.33
18		16 0.1600	720.22	912.29	2006.67	66.46	3.77	-50.72	-766.72	-876.74
19		17 0.1700	720.76	911.99	2006.24	38.97	-8.43	-33.63	-2602.23	444.70
20		18 0.1800	721.04	912.19	2005.98	27.82	9.62	-28.93	1276.28	163.42
21		19 0.1900	721.41	912.11	2005.66	54.94	-4.77	-23.61	879.09	226.64
22		20 0.2000	722.05	912.15	2005.47	41.07	13.44	-29.94	-1136.69	135.52

Figura 5.19: Fichero con las trayectoria de marcadores en formato TXT

Más concretamente, este formato de datos se corresponde con una serie temporal multivariante (vectorial), ya que se dispone de una secuencia de 3000 observaciones ordenadas y equidistantes cronológicamente sobre varias características (posición, aceleración y velocidad de los 22 marcadores). Como ejemplo del problema mencionado, en la Figura 5.20 se representa gráficamente la trayectoria de los marcadores colocados en la espina iliaca (arriba) y talón (abajo) de la extremidad derecha, para una grabación en la que se ha producido una pequeña oclusión en el intervalo de tiempo comprendido entre los segundos 17 y 19 de la captura.

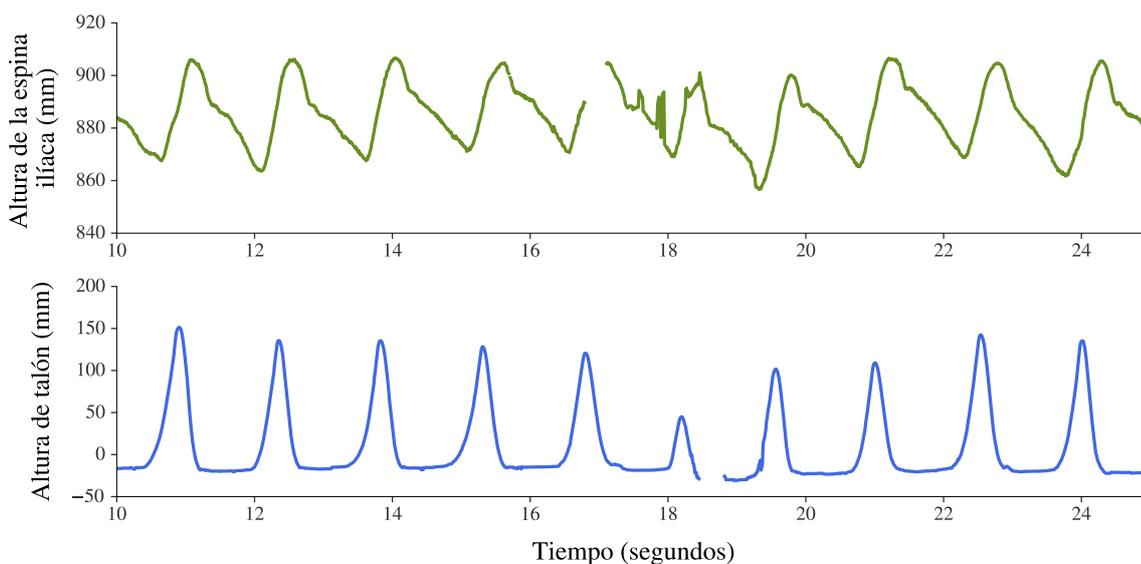


Figura 5.20: Datos faltantes en la trayectoria de los marcadores colocados en la espina iliaca (arriba) y en el talón (abajo) de la extremidad derecha

Para resolver el problema de la oclusión de marcadores se puede utilizar alguna técnica que permita la interpolación de los valores faltantes [137]. De esta forma, a pesar de la

ausencia de ciertos datos en la serie de un marcador cualesquiera, se puede construir su correlograma porque se dispone de muchas otras observaciones durante la marcha para ese mismo marcador. En particular, para sustituir los valores perdidos se ha utilizado una interpolación polinómica de segundo grado a través del método `interpolate` de *pandas*⁶

En la Figura 5.21 se muestra el resultado de interpolar los valores ausentes para los marcadores de la espina ilíaca y el talón derecho, en la misma captura de movimiento que la tomada en la Figura 5.20.

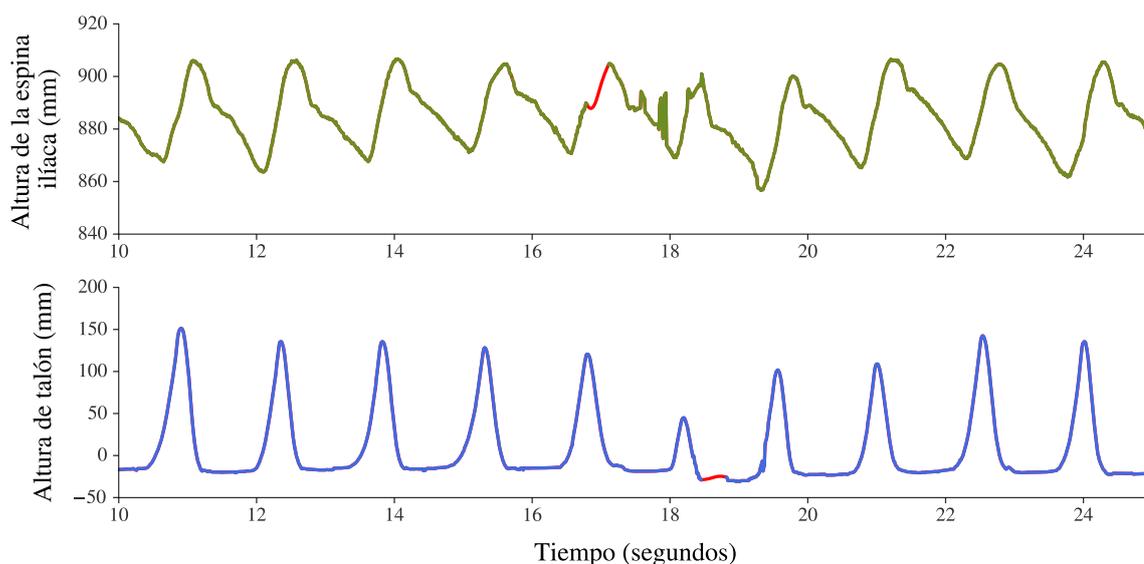


Figura 5.21: Interpolación de los datos faltantes en la trayectoria de los marcadores colocados en la espina ilíaca (arriba) y en el talón (abajo) de la extremidad derecha

Por otro lado, hay que tener en cuenta que a pesar de la alta precisión que ofrecen los sistemas ópticos para la captura de movimiento, las trayectorias de los marcadores pueden presentar cierto ruido, especialmente, producido cuando una extremidad obscurece a otra durante el movimiento. En este sentido, se ha utilizado una técnica bastante común para resolver este tipo de problemas: aplicar un filtro de paso bajo para reducir el ruido, dejando inalterada la señal original.

Más concretamente, se ha utilizado el **filtro de Savitzky–Golay**⁷ sobre las trayectorias de los marcadores para "suavizar" los datos y reducir el ruido [138-140]. Savitzky–Golay es un filtro de paso bajo que utiliza funciones locales polinomiales (de grado k) para determinar el nuevo valor de cada punto, de forma que se obtiene una función similar la original pero suavizada [141, 142]. Tomando como referencia [138], se han seleccionado como parámetros del filtro un orden de polinomio de 4 y un tamaño de ventana 17, ya que son valores ampliamente utilizados en el contexto de análisis biomecánico. Una vez reducido el ruido de estos datos, se describe a continuación cómo se han estimado algunos de los parámetros espaciotemporales de la marcha.

En la Figura 5.22 se representan las trayectorias de los marcadores de la espina ilíaca y el talón derecho después de aplicar el filtrado de Savitzky–Golay, sobre los datos de la misma captura de movimiento que la mostrada en las Figuras 5.20 y 5.21.

⁶<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.interpolate.html>

⁷La librería *SciPy* de *Python* implementa algunos filtrados de señales, incluyendo el filtro de *Savitzky–Golay*: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.savgol_filter.html

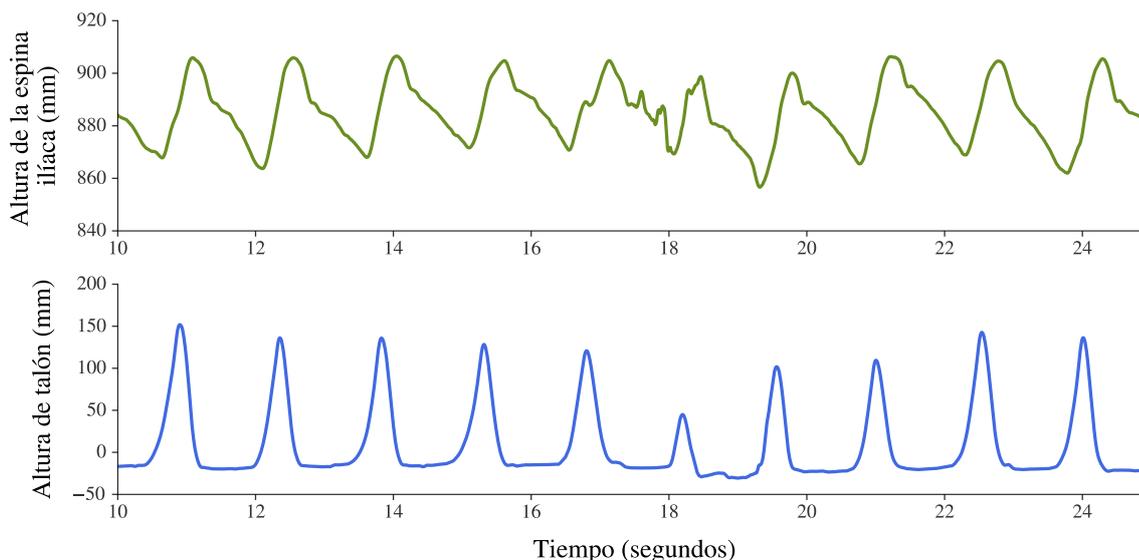


Figura 5.22: Suavizado mediante Savitzky–Golay de la la trayectoria de los marcadores colocados en la espina iliaca (arriba) y en el talón (abajo) de la extremidad derecha

Una vez interpolado los datos ausentes y reducido el ruido de estos datos, en la siguiente sección se describe cómo a partir de las trayectorias de marcadores y el fichero de eventos se han estimado algunos de los parámetros espaciotemporales de la marcha.

5.2.2. Creación de variables explicativas

A pesar del preprocesamiento llevado a cabo sobre las trayectorias de los marcadores, estos datos presentan una alta dimensionalidad incluyendo un elevado número de atributos irrelevantes. Por tanto, utilizar datos con estas características para el entrenamiento de modelos de *Machine Learning* puede degradar significativamente su rendimiento. Sin embargo, partiendo de estos datos es posible generar un conjunto de variables predictoras propias de los análisis de marcha humana. Más concretamente, estas nuevas características están basadas en los parámetros espaciotemporales y curvas cinemáticas de la marcha.

Estimar parámetros espaciotemporales

Tal y como se ha explicado en la Sección 3.1.2, a partir de los eventos temporales y espaciales del ciclo de la marcha se pueden deducir una serie de variables espaciotemporales que habitualmente se utilizan para llevar a cabo un análisis detallado de la marcha humana. En este sentido, partiendo del archivo TXT con los tiempos de eventos de la marcha, el fichero con las trayectorias de los marcadores y los archivos CSV con datos biomecánicos, incluidos para cada grabación del *dataset* (Figura 5.8), se han calculado diferentes parámetros espaciotemporales que posteriormente podrán ser utilizados como variables predictoras en la fase de modelado.

En primer lugar, tomando como base los ficheros TXT de eventos presentados en la Sección 5.2.1, se han calculado los siguientes parámetros relacionados con el tiempo, o temporales, de la marcha, para las zancadas de ambas extremidades:

- **Duración de zancada.** Tal y como se ha descrito en la Sección 3.1.2, se corresponde

con el intervalo de tiempo comprendido entre dos contactos iniciales consecutivos (HS, FOOT_INITIAL_CONTACT en el fichero de eventos) por parte de una misma extremidad (ver Figura 5.23).

- Duración de paso.** Determina el tiempo que transcurre entre el comienzo de una zancada y el comienzo de la zancada del pie contrario. Se estima, por tanto, a partir de dos eventos HS consecutivos para extremidades opuestas (ver Figura 5.23).
- Tiempo y porcentaje de apoyo.** El tiempo de apoyo hace referencia a la duración, en segundos, de la fase de soporte (apoyo) dentro del ciclo de la marcha. Teniendo en cuenta la definición de este parámetro temporal, presentada en la Sección 3.1.2, el tiempo de apoyo para la zancada de una determinada extremidad se ha calculado como el cronometraje comprendido entre el contacto inicial del pie de esa extremidad con el suelo (HS) y el próximo despegue (TO) de ese mismo pie (ver Figura 5.23). El porcentaje de tiempo que ocupa la fase de apoyo dentro del ciclo de captura (duración de zancada) define el porcentaje de apoyo.

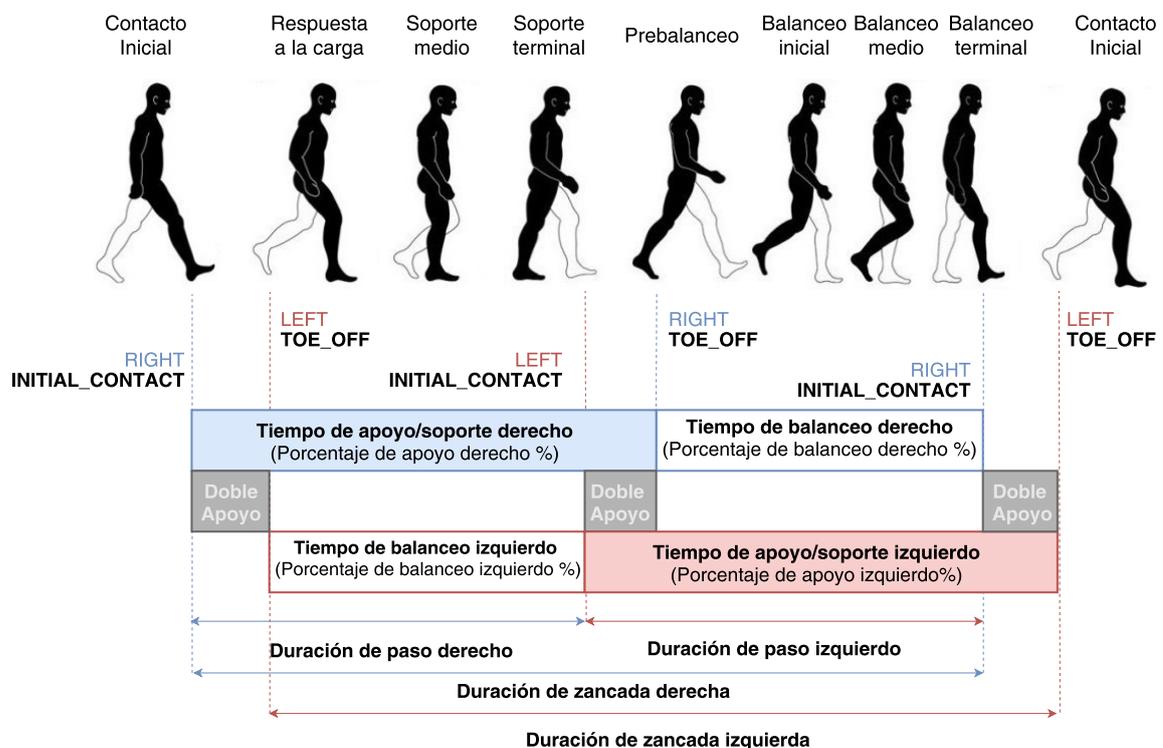


Figura 5.23: Parámetros temporales de la marcha estimados a partir de los eventos HS y TO. Elaboración basada en [12]

- Tiempo y porcentaje de balanceo.** Se corresponde con la duración, en segundos, de la fase de balanceo dentro del ciclo de la marcha. Equivale al intervalo de tiempo comprendido entre el despegue de talón de una extremidad (TO) y el siguiente contacto inicial de ese mismo pie (HS), tal y como se representa esquemáticamente en la Figura 5.23. Igualmente, el porcentaje de tiempo que ocupa la fase de balanceo en relación con la duración de zancada determina con el porcentaje de balanceo.
- Tiempos y porcentaje de doble apoyo.** Durante el ciclo de la marcha de cada extremidad se producen dos intervalos de apoyo bipodal en los que ambos pies están

simultáneamente en contacto con el suelo [143]. En el primer intervalo de doble apoyo, el pie analizado acaba de tocar el suelo mientras que la extremidad opuesta todavía no ha empezado su fase de balanceo [15]. Por tanto, este primer intervalo bipodal para una determinada zancada se corresponde con el cronometraje comprendido entre el contacto inicial del pie de esa extremidad con el suelo (HS) y el despegue (TO) del pie contrario. Por su parte, el segundo intervalo de doble apoyo comienza cuando el pie opuesto empieza su periodo de soporte, y se extiende hasta que la extremidad estudiada comienza la fase de balanceo (Figura 5.23).

El porcentaje de doble apoyo se corresponde con el porcentaje de tiempo que abarca la suma de la duración de ambos intervalos de doble apoyo dentro del ciclo de captura (duración de zancada).

- **Duración de la marcha.** Tomando como referencia [144], se ha calculado como el tiempo transcurrido, en segundos, desde el primer despegue del pie hasta el último golpe de talón.

Por otro lado, se ha combinado la información proporcionada en el fichero de eventos con la posición de ciertos marcadores en el instante en el que se producen dichos eventos, para estimar algunos parámetros espaciales de la marcha. Esta posición viene dada en milímetros, y se corresponde con la distancia del marcador hasta el punto central de la zona de captura (Figura 5.1).

- **Longitud de zancada.** Tal y como se ha explicado en la Sección 3.1.2, la longitud de zancada es la distancia lineal entre los contactos pie-suelo sucesivos de una misma extremidad. Teniendo en cuenta esta definición, se ha calculado como la diferencia entre la posición del marcador del talón (21 y 22 en la Figura 5.5) al principio del ciclo, y la posición de ese mismo marcador al final de la zancada, proyectada contra la dirección de avance [145].
- **Longitud de paso.** Es la distancia lineal, en la dirección de progresión, entre los contactos pie-suelo sucesivos de las dos extremidades. De manera similar a la longitud de zancada, se ha estimado como la distancia entre la posición del talón al principio del ciclo, y la posición del talón del pie contrario en el siguiente apoyo (HS) registrado.
- **Anchura de paso.** Generalmente se mide como la distancia entre las dos líneas paralelas trazadas sobre el punto medio de los dos talones [9]. Por tanto, se ha calculado como el valor medio de la distancia entre los marcadores de ambos talones proyectada contra el eje de las caderas.
- **Ángulo de paso.** De acuerdo con lo explicado la Sección 3.1.2, es el ángulo que se forma entre la línea que une los talones de pasos consecutivos de una misma extremidad y la línea que divide al pie en dos mitades [9]. Para su cálculo se han utilizado los marcadores colocados en el talón y metatarsos izquierdos (21 y 14 en la Figura 5.5) para el ángulo de paso izquierdo y, los marcadores 22 y 13 para el ángulo de paso derecho. De esta forma, es posible obtener el ángulo de paso mediante cálculos trigonométricos, y en particular, aplicando el teorema del coseno sobre el triángulo formado por los dos puntos de golpes de talón consecutivos para un mismo pie mismo pie, y el punto que divide el pie en dos mitades (definido por el marcador ubicado en los metatarsos). La Figura 5.24 ilustra gráficamente dicho triángulo para calcular el ángulo de paso en ambas extremidades.

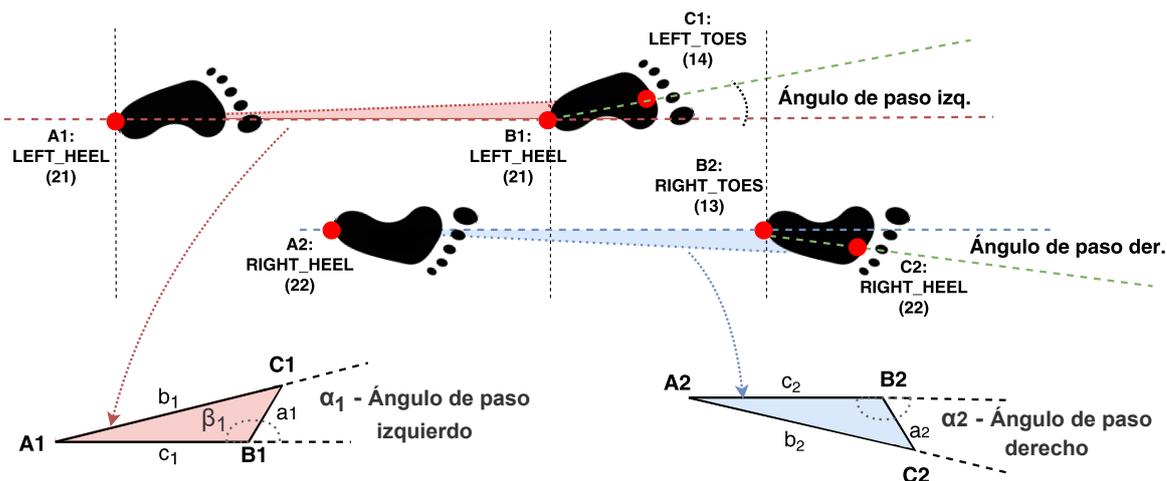


Figura 5.24: Cálculo del ángulo de paso. Elaboración propia.

Así, siendo A1 el punto dónde golpea el talón izquierdo en un primer paso; B1, la marca del suelo dónde golpea ese mismo talón en el paso siguiente; C1, la ubicación del marcador fijado en el pie izquierdo en el instante del golpeo de talón para ese segundo paso; β_1 el ángulo formado por B1 y C1; a_1 , la distancia entre los puntos C1 – B1; b_1 la distancia entre los puntos A1 – C1, y c_1 la distancia entre A1 – B1, el ángulo de paso izquierdo, α_1 viene dado por la Ecuación 5.1:

$$\begin{aligned}
 b_1^2 &= a_1^2 + c_1^2 - 2a_1c_1 \cdot \cos(\beta_1) \\
 \beta_1 &= \arccos \frac{a_1^2 + c_1^2 - b_1^2}{2a_1c_1} \\
 \alpha_1 &= 180 - \beta_1
 \end{aligned}
 \tag{5.1}$$

- Altura máxima del talón.** La altura del talón durante la marcha es un factor importante en el estudio de la postura y el equilibrio [146]. Por este motivo, se ha calculado dicho parámetro como la altura máxima de los marcadores colocados en el talón (21 y 22) durante el ciclo de la marcha de cada extremidad, tal y como se esquematiza en la Figura 5.25.
- Altura máxima del pie.** En relación con la altura máxima del talón, se ha estimado la altura máxima del pie como la altura más elevada de los marcadores colocados en los metatarsos (13 y 14) durante el ciclo de la marcha de cada zancada, según se muestra en la Figura 5.25.

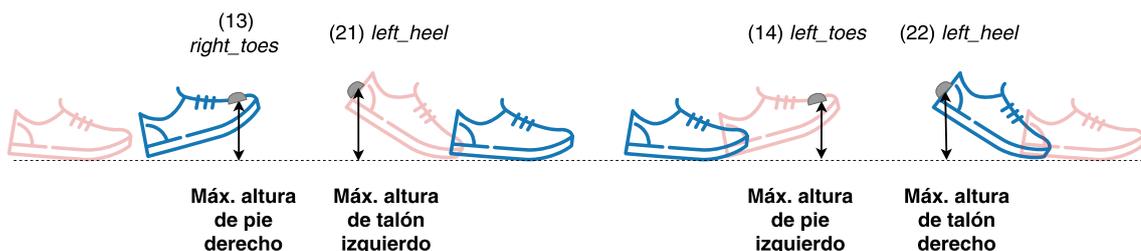


Figura 5.25: Cálculo de la altura de pie y talón durante la marcha. Elaboración propia.

- **Ángulo de contacto** (*strike-angle*). Es el ángulo que se forma entre el talón y el suelo cuando se produce el contacto inicial (HS) durante la marcha [147], siendo su valor normal, aproximadamente, 25 grados [10]. Para su cálculo se han utilizado los marcadores colocados en el talón y metatarsos (21 y 14 en la extremidad izquierdos; 22 y 13 para la derecha). El esquema izquierdo en la Figura 5.26 ilustra gráficamente cómo se ha determinado el ángulo de contacto a partir del triángulo formado por punto A1 en el marcador del talón (21-22), el punto B1 del marcador del pie (13-14) y un punto C1 auxiliar obtenido a partir de los dos anteriores.

De esta forma, calcular el ángulo de contacto (α_1) se reduce a aplicar el teorema del coseno (Ecuación 5.2), de forma similar a lo descrito para el ángulo de paso:

$$\alpha_1 = \arccos \frac{b_1^2 + c_1^2 - a_1^2}{2b_1c_1} \quad (5.2)$$

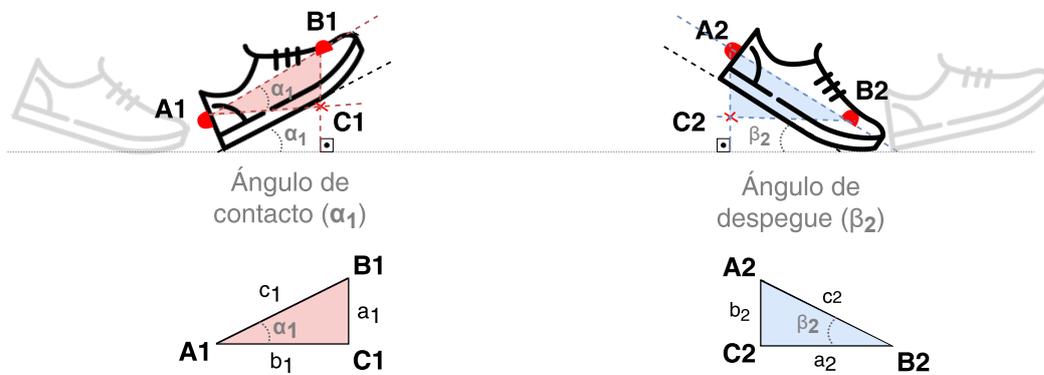


Figura 5.26: Cálculo del ángulo de contacto y despegue. Elaboración propia.

- **Ángulo de despegue**. Es el ángulo que se forma entre el pie de referencia y el suelo en el momento del despegue del antepié TO. Tal y como se ilustra en el gráfico derecho de la Figura 5.26, se obtiene utilizando los mismos marcadores que para el cálculo del ángulo de contacto. En este sentido, siendo A2 el marcador colocado en el talón; B2 el marcador ubicado en los metatarsos, y C2 un punto auxiliar derivado de los dos anteriores, el ángulo de despegue (β_1) viene dado por la Ecuación 5.3:

$$\beta_1 = \arccos \frac{a_1^2 + c_1^2 - b_1^2}{2a_1c_1} \quad (5.3)$$

Para completar el listado de variables predictoras basadas en parámetros espaciotemporales de la marcha, es interesante considerar los desplazamientos del centro de gravedad (COG) que se producen al caminar, ya que tal y como se ha explicado en la Sección 3.1.3, están directamente relacionados con el consumo energético y la eficiencia de la marcha humana. En este sentido, partiendo de los datos contenidos en el fichero *COG.csv*, descrito en la Sección 5.1.4, se ha obtenido para cada ciclo de la marcha el desplazamiento del centro de gravedad en dirección vertical (componente Y para el sistema de captura) y horizontal (eje X), según se ilustra en la Figura 5.27.

Como resultado de esta estimación de parámetros espaciotemporales, calculados con la intención de ser utilizados como variables predictoras durante la fase de modelado, se ha generado un archivo CSV dentro de la carpeta de cada captura de movimiento en el *dataset*.

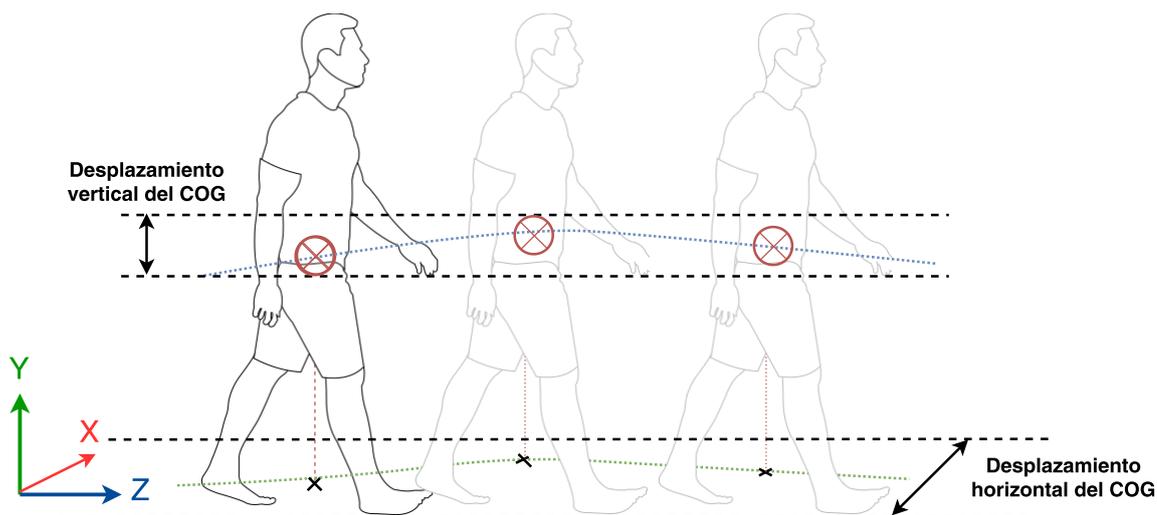


Figura 5.27: Desplazamiento vertical y horizontal del centro de gravedad. Elaboración propia.

Este fichero, así como el código *Python* que implementa esta funcionalidad, se describe en la Sección 5.2.5.

Resumen de datos cinemáticos

Puesto que las variables cinemáticas, presentadas en la Sección 3.1.3, son ampliamente utilizadas para caracterizar la marcha porque describen los movimientos realizados por las diferentes partes del cuerpo, resulta interesante disponer de ellas para ser utilizadas como variables explicativas en cualquier modelo orientado a valorar la marcha. En este sentido, el sistema de captura *Clinical 3DMA* proporciona esta información preprocesada como datos tabulares (un *CSV* por desplazamiento analizado y extremidad) donde las filas indican la marca de tiempo, normalizada para representar el ciclo de la marcha como un porcentaje (100 filas), y cada columna se corresponde con una zancada registrada durante la captura.

Esta estructura es apropiada para ser representada mediante los diagramas típicamente utilizados en el análisis de la cinemática articular, pero, sin embargo, presenta una alta dimensionalidad, es decir, consta de una gran cantidad de variables con respecto a las observaciones. Ante esta situación, y con el objetivo de resumir la información cinemática en un pequeño número de variables informativas, se ha optado por calcular una serie de medidas descriptivas para cada movimiento articular, así como tomar los valores angulares en ciertos puntos relevantes dentro del ciclo de la marcha. El listado de variables resultantes se muestra en la Tabla 5.5.

De esta forma, para cada captura de movimiento se ha generado un archivo *CSV* con los valores de las seis variables de resumen en cada una de las 26 curvas cinemáticas disponibles (13 movimientos para las extremidades izquierda y derecha). A partir de esta nueva estructura tabular se podrían deducir hasta 156 variables predictoras en cada grabación, por lo que en la fase de modelado será necesario estudiar qué subconjunto de ellas es más apropiado en cada problema individual. No obstante, se ha conseguido reducir la dimensionalidad de la estructura cinemática original, ya que para cada captura había 26 tablas con 100 filas y un número variable de columnas relativo al número de zancadas (comprendido entre 15 y 20). El *script* en *Python* utilizado para abordar esta tarea se describe en la Sección 5.2.5.

Tabla 5.5: Variables informativas calculadas para resumir los parámetros cinemáticos

Variable	Descripción
<i>min</i>	Mínimo valor angular. Es el promedio del menor ángulo del movimiento analizado en cada zancada
<i>max</i>	Máximo valor angular. Es el promedio del mayor ángulo del movimiento analizado en cada zancada
<i>mean</i>	Media aritmética del conjunto de valores angulares en el movimiento analizado
<i>range</i>	Rango articular. Es la amplitud del movimiento, calculada como la diferencia entre min y max.
<i>stride_start</i>	Valor angular al inicio de la zancada (0 % del ciclo). Promedio entre el ángulo inicial en cada ciclo.
<i>stride_end</i>	Valor angular al final del ciclo (100 %). Promedio entre el ángulo final en cada zancada.
<i>swing_angle</i>	Valor angular en el instante que comienza la fase de balanceo. Promedio entre todas las zancadas de la captura.

5.2.3. Limpieza de datos: filtrado de zancadas

La limpieza de datos (*data cleansing*) es una tarea crítica en cualquier proyecto de *Machine Learning*, ya que permite asegurar la calidad de los datos mediante el descubrimiento y corrección de valores incompletos, incorrectos, inexactos o no pertinentes. Cuando se trabaja con datos de movimiento relacionados con la marcha humana, la limpieza de datos puede abordarse a nivel de ciclos, es decir, identificando zancadas incorrectas dentro de la captura de movimiento. En este sentido, se ha decidido filtrar las zancadas de cada grabación en base a diferentes criterios.

Detección de giros durante la marcha

La zona de captura visible por las cámaras *OptiTrack* utilizadas para la grabación de los movimientos abarca un espacio limitado de, aproximadamente, 6 metros de largo y 4 metros de ancho (Sección 5.1.1). Esto implica que, durante los 30 segundos caminando de cada grabación, el participante se ve obligado a cambiar bruscamente el sentido de la marcha en repetidas ocasiones, realizando giros de 180 grados que pueden introducir una distorsión en los parámetros espaciotemporales y cinemáticos de la marcha. Por tanto, resulta especialmente interesante detectar y filtrar aquellas zancadas que se correspondan con giros al caminar.

En este sentido, en la literatura científica se han propuesto diferentes algoritmos para detectar automáticamente los cambios de sentido durante la marcha. Por un lado, [148], que utiliza sensores inerciales *IMUs* ubicados en la zona dorsal (empeine) de los pies, calcula la diferencia en la orientación de los sensores entre dos pasos consecutivos y considera que se ha producido un giro si dicha diferencia es superior a 20 grados alrededor del eje vertical. En cambio, en el enfoque presentado por [149], basado en sensorización inercial sobre la pelvis, cada giro puede prolongarse varios pasos, y se identifica cuando la velocidad angular medida alrededor del eje vertical supera los 15°/s. Como consecuencia de las limitaciones que presentan los dos métodos anteriores para la detección de giros durante la marcha, [150] propone un algoritmo mejorado que incorpora datos ópticos sobre la información inercial combinada de las dos técnicas anteriores.

Teniendo en cuenta que en este trabajo se ha utilizado un sistema óptico y que los

movimientos se han grabado dentro de una zona de captura cerrada con unas dimensiones constantes, se puede aprovechar el sistema de referencia fijado por las cámaras *OptiTrack* (ver Figura 5.1) para determinar en qué instantes se producen cambios en el sentido de la marcha. Más concretamente, la curva de la trayectoria que se genera por la variación de la posición del centro de gravedad (COG) en el eje de la dirección de avance de la marcha (Z), se puede interpretar como un indicador de la ubicación del participante sobre el área de grabación, medida como la distancia en centímetros desde el punto de referencia, fijado en la zona central del laboratorio por el sistema de captura.

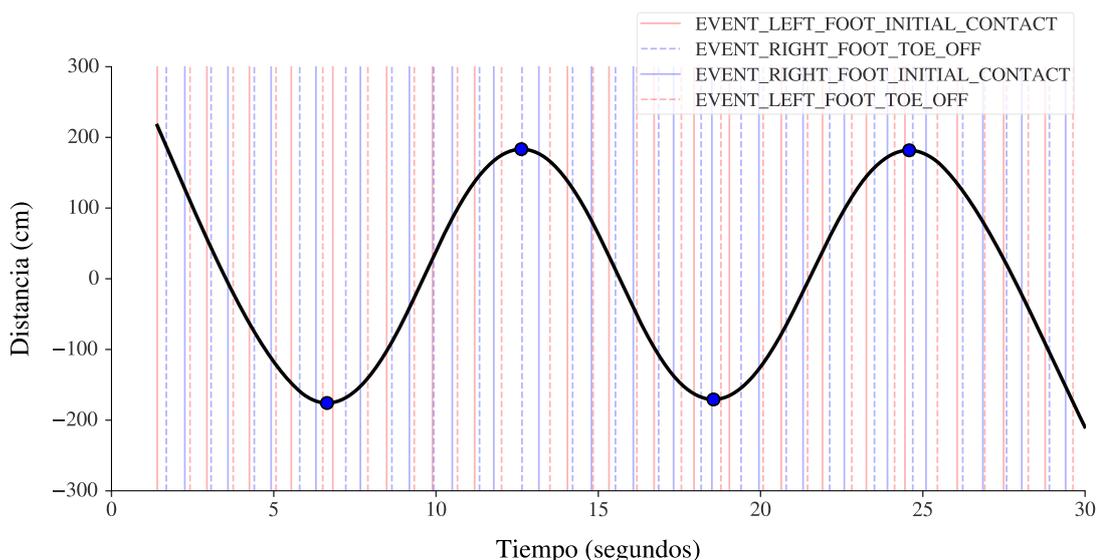


Figura 5.28: Posición del COG en el eje de avance durante la marcha

En la Figura 5.28 se muestra un ejemplo de la curva del COG en la dirección de progresión. El eje vertical de la gráfica representa la distancia, en el sentido de la marcha, medida en centímetros, de la posición del sujeto con respecto al punto de referencia del sistema de captura, mientras que el eje horizontal representa el instante, en segundos, de la grabación. En este sentido, los máximos y mínimos locales se corresponden con los instantes en los que el participante se encuentra en alguno de los extremos de la zona de captura, por lo que son los puntos en los que se produce un giro.

En estos datos relacionados con el desplazamiento del COG en el eje de avance de la marcha se han encontrado algunos valores atípicos que han sido eliminados antes de realizar cualquier tarea. Más concretamente, se han descartado aquellos valores superiores a los 500 cm e inferiores a -500 cm, ya que teniendo en cuenta las dimensiones de la zona de captura (6 metros de largo) deberían estar comprendidos, aproximadamente, entre -300 y 300 centímetros. Asimismo, durante el análisis de estos datos se han detectado inconsistencias en el formato de los diferentes archivos *COG.csv*, exportados desde *Clinical 3DMA*. En particular, se ha observado que en algunas de estas tablas aparecen columnas duplicadas para algunos de los ejes. Para garantizar la homogeneidad de estos datos, se ha optado por reescribir los *CSV* erróneos con las columnas apropiadas.

Al igual que lo comentado en la Sección 5.2.2 para las trayectorias de los marcadores, ha sido necesario aplicar un filtrado de Savitzky-Golay sobre estos datos para suavizar la curva del COG. Esta eliminación de ruido es especialmente necesaria en marchas patológicas, dado que en estos casos la trayectoria durante la deambulaci3n puede sufrir significativas

desviaciones. Como ejemplo, en la Figura 5.29 se muestra el desplazamiento del COG sin filtrar (Figura 5.29a) y tras aplicar el filtro de Savitzky-Golay (Figura 5.29b). En este ejemplo, se trata de una grabación etiquetada como *cd*, de acuerdo con la nomenclatura definida en la Tabla 5.2 (marcha discontinua con interrupciones entre pasos).

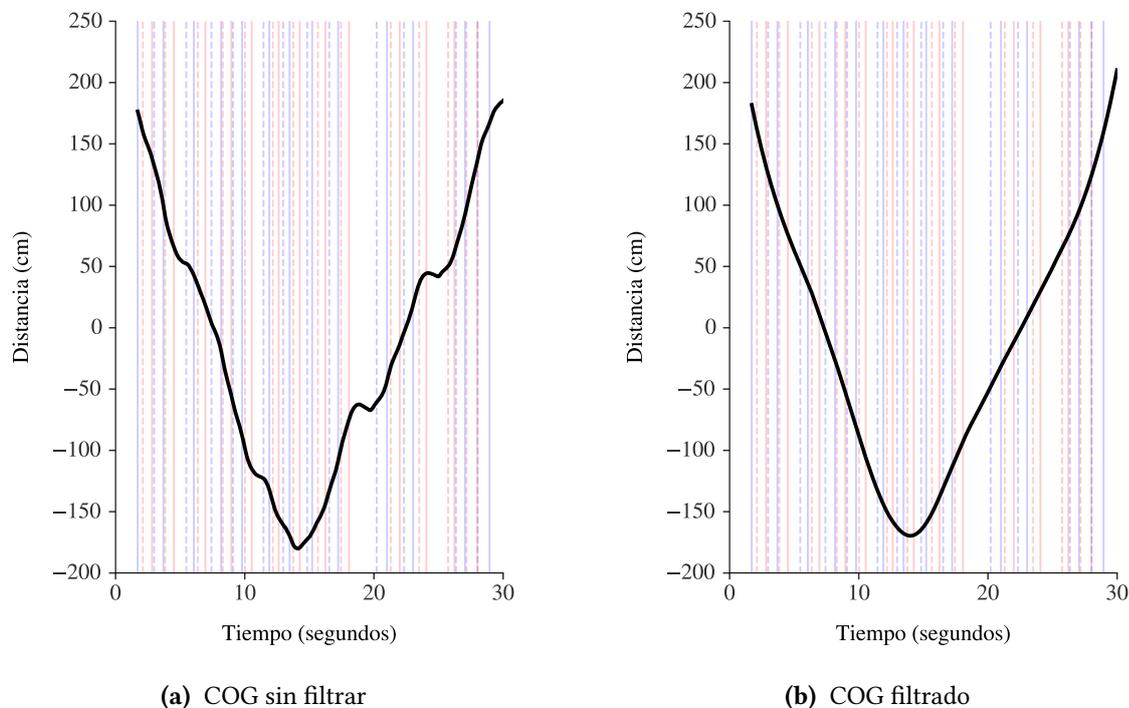


Figura 5.29: Trayectoria filtrada y sin filtrar del centro de gravedad, en la dirección de avance, para una marcha patológica

Una vez preparadas estos datos con el desplazamiento, en la dirección de progresión, del COG, se han seleccionado como zancadas potencialmente filtrables por estar afectadas por un giro aquellas cuyos eventos coinciden con alguno de los máximos y mínimos en la curva del COG. Más concretamente, para cada cambio de dirección detectado (máximo o mínimo) se han tomado los dos eventos de la marcha (anterior y posterior) más cercanos, seleccionando las zancadas a las que pertenecen dichos eventos. De esta forma, un giro puede afectar a una o varias zancadas, dependiendo de si los eventos más cercanos pertenecen o no a un mismo ciclo. Dichos eventos de la marcha se representan como líneas verticales en las gráficas de las Figuras 5.28 y 5.29.

Como resultado de esta detección de giros durante la marcha, se ha generado un archivo CSV, al que se ha llamado *stride_filtering* y se ha ubicado en la raíz del *dataset*, incluyendo una fila para cada captura y tres columnas que indican el número de giros detectados y las zancadas izquierda y derecha potencialmente eliminables por estar involucradas en un cambio de sentido durante la marcha. Este archivo se completa con la información sobre zancadas incorrectas explicada en los siguientes apartados.

Zancadas incorrectas según los parámetros espaciotemporales

El filtrado de zancadas basado en los giros durante la marcha, explicado en el apartado anterior, se puede complementar utilizando la información proporcionada por los parámetros espaciotemporales calculados en la Sección 5.2.2. En este sentido, mientras se produce un

cambio de dirección aumenta sustancialmente el ángulo de paso, para adaptar así el cuerpo a la nueva trayectoria, pero, además, la longitud de zancada y otros parámetros espaciales de la marcha también se ven alterados por estos ciclos peculiares.

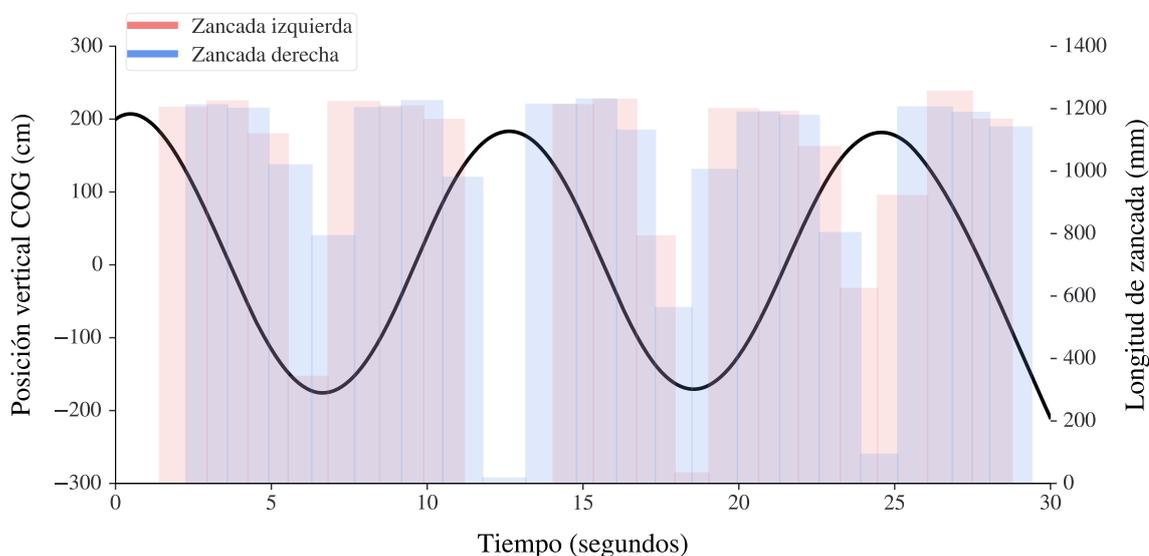


Figura 5.30: Relación entre la posición del COG en el eje de progresión y la longitud de zancada

Como ejemplo, en la Figura 5.30 se incluye un gráfico de eje doble que representa simultáneamente la curva trazada por el desplazamiento del COG, en el sentido de avance (eje izquierdo), y la longitud de cada zancada izquierda y derecha durante la marcha, en forma de barras verticales (eje derecho). De esta forma, es posible visualizar fácilmente que aquellas zancadas que presentan una longitud significativamente inferior al resto, se ubican sobre la marca temporal en la que se ha producido un cambio de dirección (máximos y mínimos de la curva), y, por tanto, conviene descartarlas.

En algunos trabajos se definen unos valores fijos que actúa como umbrales de los parámetros espaciotemporales para decidir qué ciclos deben ser eliminados [151]. Sin embargo, considerando que los valores de este parámetro están influenciados por la antropometría de cada persona, y que el conjunto de datos incluye marchas patológicas en las que la longitud de zancada puede verse alterada, se ha optado por seleccionar como ciclos a filtrar aquellas zancadas cuya longitud está excesivamente alejada de la mediana, calculada únicamente sobre esa misma captura de movimiento.

Por otro lado, los valores normales del ángulo de paso están comprendidos entre los 5° y 8° [11]. Sin embargo, tal y como se muestra visualmente en la gráfica de la Figura 5.31, los pasos que se producen durante un giro completo en el sentido de la marcha provocan que este parámetro espacial incremente notablemente, y por tanto, es un indicador claro de que estas zancadas deben filtrarse para eliminar ruido en los parámetros de la marcha. En este caso concreto, el ángulo de paso izquierdo durante los giros alcanza los 160°, mientras que el ángulo de paso derecho supera los 60°.

Teniendo en cuenta que algunas deficiencias en las extremidades pueden provocar un notable incremento del ángulo de paso [11], se ha seleccionado un umbral alto para marcar como zancadas filtrables aquellas cuyo ángulo de paso izquierdo o derecho supere los 40 grados. Como resultado, sobre el fichero CSV *stride_filtering* que contiene, para cada captura del *dataset*, las zancadas izquierda y derecha potencialmente pertenecientes a un giro, se han

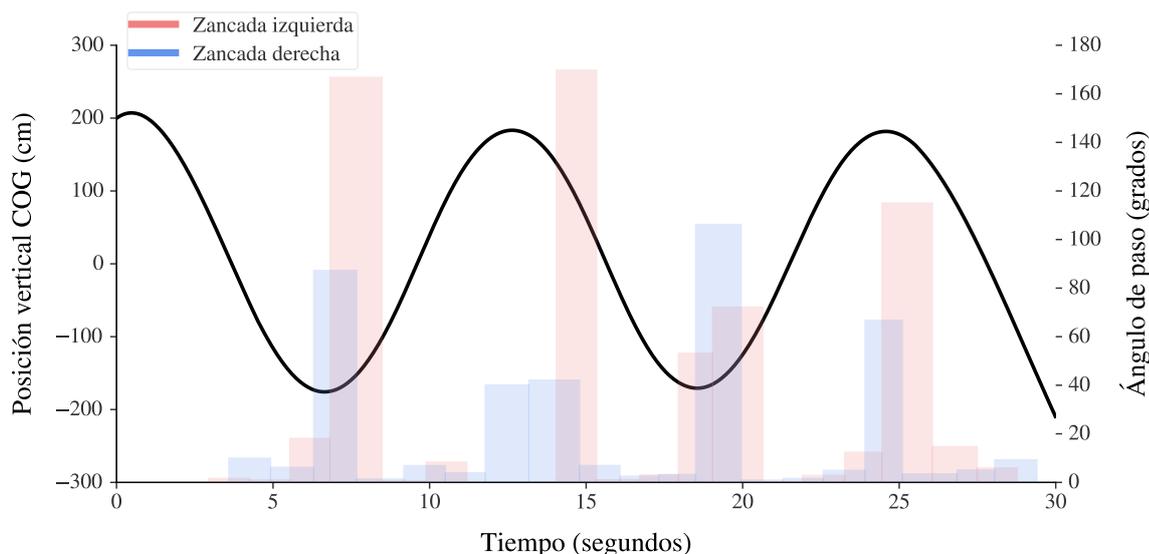


Figura 5.31: Relación entre la posición del COG en el eje de progresión y el ángulo de paso

añadido dos nuevas columnas que registran las zancadas que se deberían filtrar en base a los criterios de longitud de zancada y ángulo de paso. De esta forma, combinando la detección de giros basada en el desplazamiento del COG como los valores atípicos encontrados en estos parámetros espaciales de la marcha, se han eliminado todas aquellas zancadas que cumplan las condiciones de filtrado explicadas para ambos métodos.

Zancadas erróneas en los datos cinemáticos

La oclusión de ciertos marcadores durante determinados instantes de la grabación, además de afectar a los datos en crudo de las trayectorias de marcadores, tal y como se ha explicado en la Sección 5.2.1, tiene una influencia negativa sobre las curvas cinemáticas que se pueden exportar desde *Clinical 3DMA*. En estos casos, cuando el sistema de captura no puede reconstruir correctamente alguno de los segmentos corporales, se observan valores extremos (*outliers*) que no pueden ser eliminados aunque se reduzca el ruido mediante un filtro de paso bajo.

Para ejemplificar esta problemática, la Figura 5.32 representa gráficamente las curvas cinemáticas para la inclinación pélvica, oblicuidad de pelvis y abducción de cadera, en una captura en la que se observan dos zancadas, una izquierda y otra derecha, con valores excepcionalmente lejanos al patrón definido por el resto de los ciclos de la marcha. Con la finalidad de evitar que esta circunstancia influya en las variables explicativas que se han generado a partir de estos datos cinemáticos (Sección 5.2.2), conviene filtrar aquellas zancadas con estas características.

En este caso, para la detección de valores atípicos en las curvas cinemáticas se ha seguido un proceso que aprovecha el conocimiento sobre el dominio de los datos. En este sentido, para cada uno de los movimientos angulares, en el plano sagital, incluidos en los datos cinemáticos, se ha fijado un umbral inferior y superior de valores angulares permitidos (Tabla 5.6). Estos límites se han definido tomando como referencia datos de normalidad publicados en diferentes fuentes bibliográficas [9-11, 23, 152], y se ha optado por márgenes bastante amplios teniendo en cuenta que el envejecimiento [152], el género y las alteraciones

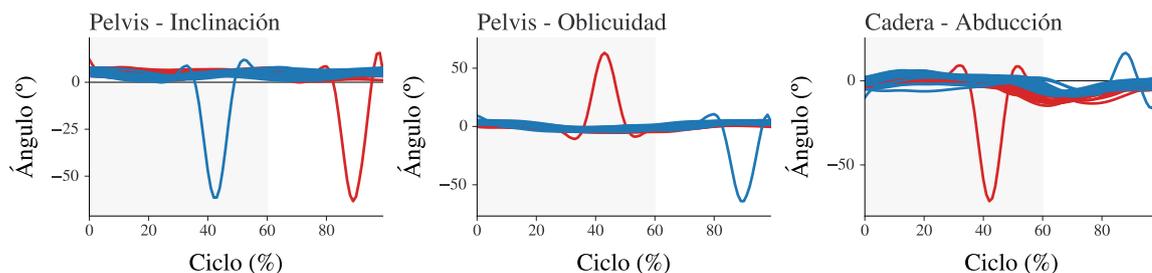


Figura 5.32: Valores atípicos en los datos cinemáticos

en la marcha influyen sobre los parámetros angulares. Dado que se han definido de umbrales extremos, se filtrarán todas aquellas zancadas que contengan valores angulares fuera de estos límites.

Tabla 5.6: Umbrales para el filtrado de zancadas en los datos cinemáticos

	Rango de movimiento normal en el plano sagital	Umbral de filtrado inferior	Umbral de filtrado superior
Pelvis	5° (inclinación inferior) 15° (inclinación superior)	-30°	60°
Cadera	-15° (Extensión máxima) 30° (Flexión)	-45°	60°
Rodilla	0° (Posición neutra) 60° (Flexión)	-30°	90°
Tobillo	-10° (Flexión dorsal) 20° (Flexión plantar)	-45°	60°

Al aplicar este filtrado sobre las capturas del *dataset*, se han detectado dos grabaciones (*subject_07_rec_03* y *subject_10_rec_05*) con una alta proporción de zancadas incorrectas según este criterio. Analizando más detalladamente estas capturas, se observa que un momento dado durante el movimiento, los valores cinemáticos se vuelven constantemente extremos e incoherentes, alcanzando, por ejemplo, puntos angulares inferiores a los -200 grados para la inclinación pélvica. En el primero de los casos, el problema afecta a 14 zancadas izquierda y 11 derecha (sobre un total de 20 ciclos), mientras que, en la segunda captura, el problema influye hasta en 16 de las 19 ciclos zancadas (con ambos pies). Por tanto, para evitar que estas grabaciones distorsionen el conjunto de datos, se ha optado por descartar completamente estos dos registros correspondientes a marchas normales.

De esta forma, el filtrado de zancadas basado en los errores de los datos cinemáticos, se ha combinado con las condiciones de giros durante la marcha y valores atípicos en los parámetros espaciotemporales descritas anteriormente. Como resultado, en el Anexo D se incluye una tabla que indica, para cada grabación, qué ciclos de la marcha deberían eliminarse del conjunto de datos. La siguiente y última actividad de preparación de datos ha consistido en la normalización de los parámetros espaciotemporales, calculados en la Sección 5.2.2, para eliminar la influencia que tienen las dimensiones corporales de cada participante sobre estos datos.

5.2.4. Escalar parámetros espaciotemporales

Los parámetros espaciotemporales que habitualmente se utilizan para caracterizar la marcha humana pueden estar influenciados por la antropometría de cada persona (ver Sección 3.1.5).

Una de las estrategias más utilizadas para eliminar esta dependencia, y poder comparar la forma de caminar de distintos sujetos, consiste en normalizar los valores espaciotemporales convirtiéndolos en variables adimensionales. Además, hay que tener en cuenta que los valores angulares no requieren de este proceso porque ya se trata de medidas adimensionales [26]. En este sentido, para decidir cómo abordar dicha normalización, en el presente trabajo se ha comenzado por estudiar la relación existente entre los parámetros espaciotemporales de la marcha, que han sido estimados según lo descrito en la Sección 5.2.2), y las medidas antropométricas de los participantes, recogidas durante la fase de adquisición de datos (Sección 5.1).

Considerado la altura del sujeto, longitud de pierna, longitud de pie y altura de rodilla, como dimensiones longitudinales potencialmente influyentes en los parámetros espaciales de la marcha, se ha utilizado un coeficiente de correlación para tratar de valorar la asociación lineal entre cada par de variables cuantitativas (parámetro espacial y medida antropométrica). En este contexto, los métodos de correlación de Pearson y Spearman son técnicas que se emplean para medir el grado de asociación entre dos variables, siendo el primero un test paramétrico que presupone normalidad en la distribución de las dos variables del análisis. Por su parte, Spearman es una técnica no paramétrica con supuestos menos estrictos y libre de asunciones sobre la distribución probabilística de los datos [153].

Por tanto, para decidir qué coeficiente utilizar en cada caso, se ha comenzado por contrastar la normalidad de cada grupo de datos utilizando el Test de Shapiro-Wilk⁸ con un nivel de confianza del 95 % ($\alpha = 0,05$). El resultado del estadístico en cada caso, junto con el p-valor asociado y la interpretación realizada, se muestra en la Tabla 5.7.

Tabla 5.7: Contraste de normalidad con el test de Shapiro-Wilk para los parámetros espaciales de la marcha (a) y medidas antropométricas (nivel de significancia del 0.05) (b)

(a) Normalidad de los parámetros espaciales

	Contraste de normalidad		
	W	p-valor	Hipótesis
Longitud de zancada	0.941	0.562	Se asume normalidad
Longitud de paso	0.924	0.387	Se asume normalidad
Ancho de paso	0.923	0.383	Se asume normalidad
Max. Altura de talón	0.942	0.581	Se asume normalidad
Max. Altura de pie	0.782	0.009	No hay normalidad
Despl. vertical COM	0.874	0.112	Se asume normalidad
Despl. horizontal COM	0.931	0.453	Se asume normalidad

(b) Normalidad de las medidas antropométricas

	Contraste de normalidad		
	W	p-valor	Hipótesis
Altura	0.989	0.996	Se asume normalidad
Longitud de pierna	0.824	0.028	No hay normalidad
Longitud de pie	0.753	0.004	No hay normalidad
Altura de rodilla	0.936	0.507	Se asume normalidad

Para tratar de valorar la asociación lineal entre cada uno de los pares parámetro espacial-medida antropométrica, se ha utilizado el coeficiente de correlación de Pearson en aquellos casos en los que ambas variables siguen una distribución normal, y un coeficiente de

⁸El Test de Shapiro-Wilk plantea la hipótesis nula de que la población está normalmente distribuida, por lo que si el p-valor asociado al estadístico es menor que el nivel de significancia se rechaza esa hipótesis nula y se concluye que los datos no siguen una distribución normal. En cambio, cuando el p-valor es mayor que el nivel de significancia no se puede rechazar dicha hipótesis, por lo que se asume normalidad.

correlación no paramétrico (coeficiente de Spearman) en caso contrario. El resultado se incluye en la Tabla 5.8, teniendo en cuenta que para los dos estadísticos se ha utilizado un nivel de significancia de 0.05, y se ha interpretado como una relación lineal moderada las puntuaciones comprendidas entre 0.36 y 0.67, y una fuerte correlación los valores entre 0.68 y 1 [154].

Tabla 5.8: Coeficiente de correlación entre los parámetros espaciotemporales de la marcha y variables antropométricas. Los datos mostrados representan el valor del coeficiente de correlación de Pearson o Spearman, según corresponda.

	Altura	Longitud de pierna	Longitud de pie	Altura de rodilla
Longitud de zancada	r=0.766 p=0.010	r=0.304 p=0.393	r=0.240 p=0.393	r=0.442 p=0.201
Longitud de paso	r=0.758 p=0.011	r=0.304 p=0.393	r=0.240 p=0.504	r=0.445 p=0.198
Ancho de paso	r=0.380 p=0.279	r=0.292 p=0.413	r=0.474 p=0.166	r=0.627 p=0.052
Max. Altura de talón	r=0.718 p=0.019	r=0.17 p=0.638	r=0.739 p=0.015	r=0.609 p=0.061
Max. Altura de pie	r=0.079 p=0.829	r=0.219 p=0.544	r=0.031 p=0.933	r=-0.097 p=0.789
Desplazamiento vertical del COM	r=0.539 p=0.108	r=0.000 p=1.000	r=0.665 p=0.036	r=0.626 p=0.053
Desplazamiento horizontal del COM	r=0.614 p=0.059	r=0.122 p=0.738	r=0.302 p=0.397	r=0.362 p=0.304
Velocidad	r=0.735 p=0.016	r=0.292 p=0.413	r=0.431 p=0.214	r=0.401 p=0.251
Cadencia	r=0.303 p=0.395	r=-0.267 p=0.455	r=0.215 p=0.550	r=0.126 p=0.730

A pesar de ser una muestra pequeña de datos, el estudio realizado sugiere que existe una relación lineal fuerte entre la longitud de zancada, longitud de paso y velocidad con la altura del participante, lo que concuerda con los resultados alcanzadas en otras investigaciones con muestras de mayor tamaño [155]. Por tanto, según se muestra en la Tabla 5.9, han normalizado estos tres parámetros de la marcha por la altura del sujeto para eliminar la influencia de esta variable antropométrica. En el caso concreto de la velocidad, se ha seguido el método de escalado propuesto por [28], basado en la condición de que las fuerzas inerciales y gravitacionales deben escalarse de forma proporcional.

Asimismo, aunque en este caso el coeficiente de correlación no indica ninguna relación lineal entre el ancho de paso y la estatura, se ha asumido que con una mayor altura se produce un incremento en la base de sustentación para mantener la estabilidad postural, ya que el centro de gravedad está en una posición más alta [156, 157]. Por otro lado, dada la fuerte relación lineal observada entre la longitud del pie con la altura máxima del talón y el desplazamiento vertical del COG, se ha eliminado la influencia de esta variable antropométrica sobre estos dos parámetros espaciales, así como sobre la máxima altura del pie, por ser un parámetro estrechamente relacionado con la máxima altura del talón.

En cuanto al desplazamiento horizontal del COG, ya que no se observa ninguna relación lineal con las variables antropométricas (significativamente distinta de cero con un 95 % de confianza), se ha decidido normalizar por la altura del participante, puesto que es un parámetro directamente relacionado con el ancho del paso. Igualmente, para la normalización

de la cadencia también se ha utilizado la altura, pues está relacionada con la velocidad de la marcha. gualmente, para la normalización de la cadencia también se ha utilizado la altura y el valor de la aceleración gravitacional, pues es un parámetro relacionado con la velocidad de la marcha.

A modo de resumen, en la Tabla 5.9 se indica el procedimiento llevado a cabo con cada variable espacial de la marcha para convertirla en una magnitud adimensional. Asimismo, los parámetros temporales de duración de la fase de soporte y balanceo, así como el tiempo de apoyo bipodal, se han normalizado por la duración de la zancada [17, 29, 34].

Tabla 5.9: Normalización adimensional de los parámetros espaciotemporales [26]. En la tabla, L indica que se trata de una variable longitudinal, mientras que T se refiere a temporal; l_0 es la estatura del sujeto; l_1 es la longitud del pie; g representa la aceleración de la gravedad ($9,81m^s^{-2}$) y t_0 expresa duración de zancada.

	Símbolo	Dimensión	Magnitud adimensional
Longitud de zancada	l	L	$\hat{l} = \frac{l}{l_0}$
Longitud de paso	l	L	$\hat{l} = \frac{l}{l_0}$
Ancho de paso	l	L	$\hat{l} = \frac{l}{l_0}$
Max. altura de talón	l	L	$\hat{l} = \frac{l}{l_1}$
Max. altura de pie	l	L	$\hat{l} = \frac{l}{l_1}$
Desplazamiento vertical del COM	l	L	$\hat{l} = \frac{l}{l_1}$
Desplazamiento horizontal del COM	l	L	$\hat{l} = \frac{l}{l_0}$
Velocidad	v	LT^{-1}	$\hat{v} = \frac{v}{\sqrt{gl_0}}$
Cadencia	c	T^{-1}	$\hat{c} = \frac{c}{\sqrt{\frac{g}{l_0}}}$
Duración fase de soporte	t	T	$\hat{t} = \frac{t}{t_0}$
Duración fase de balanceo	t	T	$\hat{t} = \frac{t}{t_0}$
Tiempo de doble apoyo	t	T	$\hat{t} = \frac{t}{t_0}$

De esta forma, se han normalizado los parámetros espaciotemporales previamente estimado para eliminar la influencia de las características antropométricas de cada sujeto, y en particular, la altura y longitud de pie. Además, es importante destacar que las dimensiones antropométricas pueden influir de forma diferente en los de la marcha para hombres y mujeres [158]. No obstante, en el presente TFM no se ha considerado este factor al no disponer de una amplia muestra que incluya datos significativos de ambos géneros.

5.2.5. Script de preprocesamiento y dataset resultante

Para aplicar las diferentes tareas de preprocesamiento, presentadas en las secciones anteriores, se ha implementado un *script* configurable en *Python* que ejecuta todo el proceso secuencialmente sobre el *dataset* original, dando lugar a un conjunto de datos estructurado, filtrado y adecuado para ser utilizado en la fase posterior de modelado. A continuación, se describen los detalles más relevantes de este *script*, cuyo código fuente es accesible desde el

repositorio del proyecto (Anexo A). Además, en el Anexo B se especifica la versión exacta de las librerías especializadas que se han utilizado para la manipulación de los datos (*pandas*, *NumPy*, *SciPy*, *Matplotlib*).

En cuanto al funcionamiento de este *script*, para llevar a cabo las tareas de preparación de datos parte de la configuración definida en un archivo **JSON** que permite especificar el conjunto de grabaciones a preprocesar, y los directorios de entrada y salida. Con la intención de fomentar la reusabilidad y mantenibilidad del programa, el código se ha organizado en diferentes clases y ficheros:

- ***turns_detector.py***. Incluye una clase `TurnsDetector` en la que se implementa el algoritmo para la detección de giros durante la marcha. Tal y como se ha explicado en la Sección 5.2.3, toma como entrada los ficheros con la trayectoria del **COG** en el sentido de progresión de la marcha, preprocesa estos datos y busca los máximos y mínimos sobre la curva resultante (utilizando la librería *peakdetect*⁹).
- ***spatiotemporal_estimator.py***. Define una clase llamada `SpatioTemporalEstimator` con métodos para calcular los parámetros espaciotemporales de la marcha a partir de las trayectorias de los marcadores y los eventos de la marcha (Sección 5.2.2), así como funciones para la preparación de estos datos en bruto (Sección 5.2.1), la normalización de las variables resultantes por las características antropométricas de cada participante (Sección 5.2.4), y la detección de valores atípicos sobre los parámetros espaciales de la marcha.
- ***kinematics.py***. Implementa una clase, `KinematicsPreprocessor`, con métodos para detectar *outliers* en las curvas cinemáticas (Sección 5.2.3), y construir variables informativas que resumen los movimientos articulares mediante medidas descriptivas calculadas sobre los datos cinemáticos (Sección 5.2.2).
- ***final_filtering.py***. Mediante una clase llamada `FinalFiltering`, se combina la información sobre el filtrado de zancadas generada por las tres clases anteriores para eliminar todos aquellos ciclos de la marcha imprecisos, inexactos o incompletos.
- ***utils.py***. Para evitar código duplicado, todas las funciones reutilizadas en varias de las clases anteriores se han agrupado dentro de un archivo *utils.py*. Por ejemplo, incluye las funciones para el segmentado en zancadas de los archivos en bruto, ya que esta operación se aplica tanto en la detección de giros como en el cálculo de parámetros espaciotemporales.

Para orquestar el funcionamiento ofrecido por estas clases y funciones, se incluye un archivo *main.py* que sirve como punto de entrada al *script* y ejecutar secuencialmente los siguientes pasos de preprocesamiento, ilustrados esquemáticamente en la Figura 5.33:

1. Carga el archivo **JSON** de configuración con el listado y ubicación de las capturas de movimiento a preprocesar.
2. Ejecuta el algoritmo para la detección de giros (`TurnsDetector`), segmentando previamente en zancadas los datos en crudo utilizados para la detección de los cambios de sentido durante la marcha (datos biomecánicos del **COG**). Como resultado, genera un archivo sobre la raíz del *dataset* llamado *stride_filtering.xlsx* (Figura 5.34), en el que se indica, para cada captura preprocesada, el número de giros detectados y los índices de zancada izquierda y derecha que deberían filtrarse en base al criterio de giros.
3. Mediante la clase `SpatioTemporalEstimator`, calcula los parámetros espaciotemporales de la marcha, tomando como entrada los archivos con las trayectorias de los marcadores, sobre los que previamente se han interpolado los valores ausentes,

⁹<https://pypi.org/project/peakdetect/>

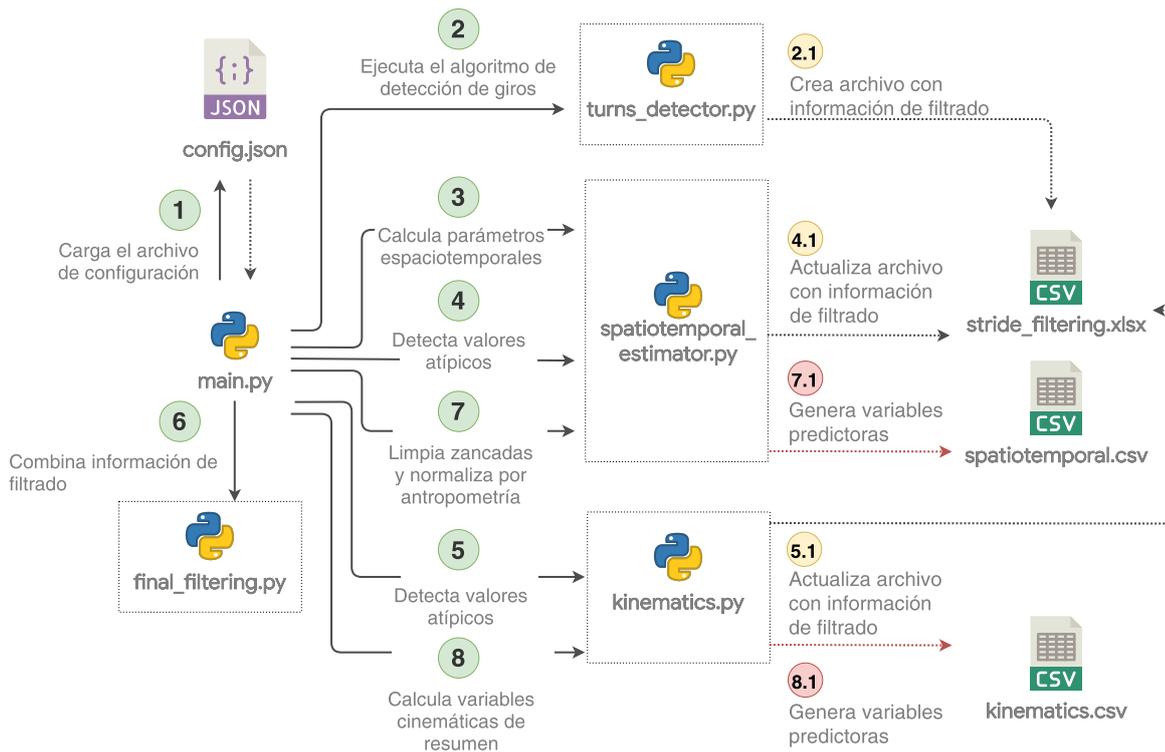


Figura 5.33: Esquema del flujo de operaciones llevadas a cabo por el *script* de preprocesamiento

reducido el ruido, y organizado en ciclos de la marcha.

4. Utiliza el método `find_outliers` de la clase `SpatioTemporalEstimator` para detectar valores atípicos en los parámetros espaciales estimados en el paso anterior. Como resultado, se actualiza el archivo `stride_filtering.xlsx` añadiendo dos nuevas columnas que especifican las zancadas izquierda y derecha que deberían filtrarse de acuerdo con la información de los parámetros espaciales de la marcha.
5. Con la clase `KinematicsPreprocessor` se cargan los archivos cinemáticos de la marcha y se analizan para detectar posibles *outliers* y zancadas inexactas o imprecisas. Como resultado, se actualiza nuevamente el archivo `stride_filtering.xlsx` añadiendo dos nuevas columnas con la información de filtrado generada a partir de las curvas cinemáticas.
6. El método `run` de la clase `FinalFiltering` carga el archivo `stride_filtering.xlsx` generado en los pasos anteriores para combinar la información de filtrado aportada por las diferentes estrategias, generando así dos nuevas columnas sobre esta tabla que indican los índices de las zancadas izquierda y derecha que finalmente deben ser eliminadas.
7. A través del método `generate_filtered_resume`, que se ha incluido dentro de la clase `SpatioTemporalEstimator`, se eliminan las zancadas filtradas sobre los parámetros espaciotemporales de la marcha estimados en el tercer paso. Los datos resultantes se normalizan con la intención de reducir la influencia de las características antropométricas de los sujetos, y se calculan estadísticas descriptivas para cada grabación a partir de los datos individuales de cada ciclo de la marcha. Como resultado, se genera un archivo `CSV` para cada captura de movimiento, incluyendo los parámetros temporales y espaciales, normalizados y sin normalizar, para cada una de las zancadas que no han sido filtradas. Además, se genera un archivo `CSV` global en la raíz del *dataset*, incluyendo los valores estructurados de las características explicativas basadas en estos parámetros de la marcha para cada una de las observaciones disponibles.

8. Finalmente, se llama al método `generate_filtered_resume` implementado sobre la clase `KinematicsPreprocessor` para generar el conjunto de variables informativas que resumen los datos cinemáticos (Tabla 5.5). Para ello, se calculan medidas descriptivas ignorando las zancadas que deben ser filtradas según lo descrito en el archivo `stride_filtering.xlsx`. Como resultado, se genera un archivo `CSV` dentro del subdirectorio de cada captura de movimiento, incluyendo el resumen cinemático de los diferentes movimientos articulares para esa grabación, y un archivo `CSV` global, en la raíz del `dataset`, con los valores de las variables predictoras construidas en este paso para cada una de las observaciones. Más adelante, se presenta en detalle la estructura de estos archivos (Figura 5.34).

Por tanto, durante la ejecución del script de preprocesamiento se generan una serie de archivos tabulares que contienen información sobre el filtrado de zancadas y, las variables explicativas que se derivan de este proceso de preparación de datos. En el siguiente apartado, se describe la estructura del `dataset` después de ejecutar estas tareas de preprocesamiento.

Conjunto de datos preprocesado

Sobre el conjunto de datos original, cuya estructura se mostró de manera esquematizada en la Figura 5.8, con las tareas de preparación de datos se han generado una serie de nuevos archivos preprocesados, tal y como se ilustra gráficamente en la Figura 5.34. Además, cabe mencionar que en el conjunto de datos públicamente accesible desde el enlace proporcionado en el Anexo A), también se han incluido estos archivos preprocesados.

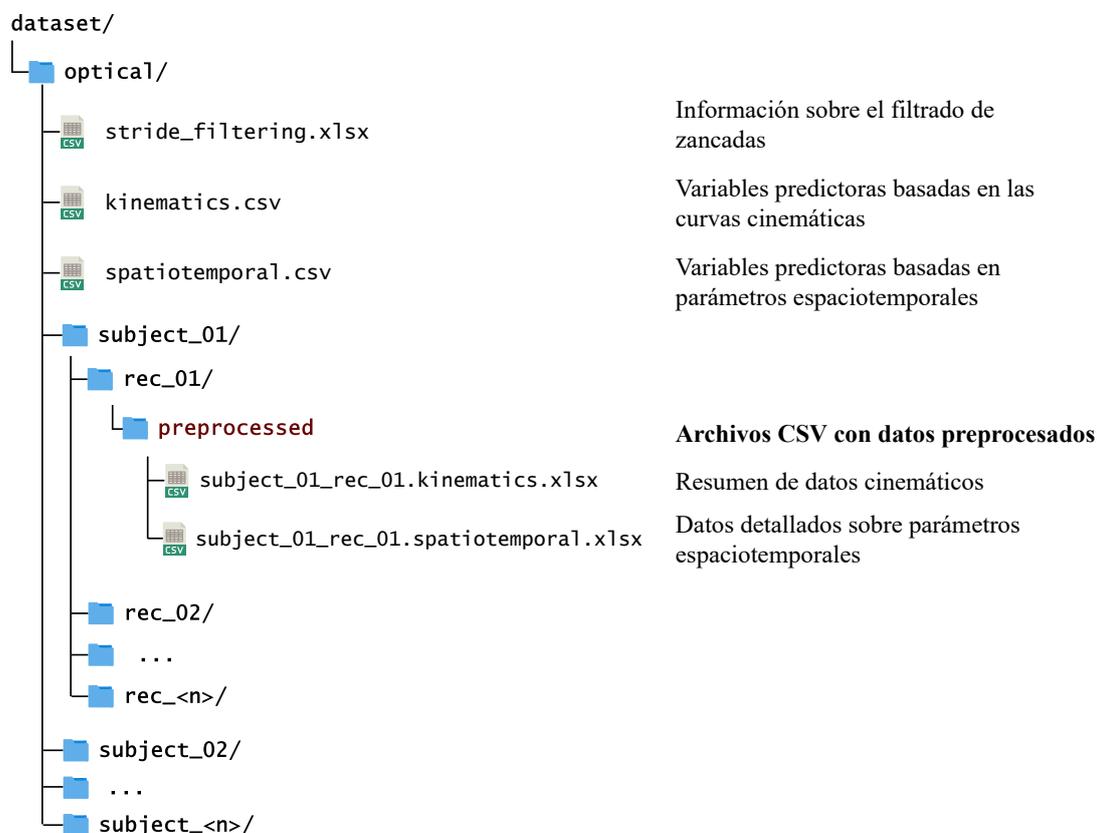


Figura 5.34: Conjunto de datos preprocesado

En primer lugar, se ha incluido un archivo `stride_filtering.xlsx` sobre la raíz del directorio

del *dataset* con información sobre el filtrado de zancadas. Este archivo tabular contiene una fila por cada captura de movimiento y 11 columnas con la siguiente información: *Subject* y *Record* identifican, respectivamente, el sujeto y grabación; *Turns_Count* es un valor entero con el número de giros detectados durante la marcha; *Turns_Left_Strides_Filtering* y *Turns_Right_Strides_Filtering* contienen un listado con los índices de zancadas izquierda y derecha que deben filtrarse en base al criterio de giros; con este mismo formato, las columnas *Spatiotemporal_Left_Strides_Filtering*, *Spatiotemporal_Right_Strides_Filtering*, *Kinematics_Left_Strides_Filtering* y *Kinematics_Right_Strides_Filtering*, determinan las zancadas a filtrar en base a la presencia de valores atípicos en los parámetros espaciotemporales y curvas cinemáticas. Las dos últimas columnas, *Left_Strides_Filtering* y *Right_Strides_Filtering*, combinan la información del resto de columnas para proporcionar dos listados con las zancadas a filtrar. La información de filtrado contenida en este archivo se proporciona en la Tabla D.1 del Anexo D.

En cuanto al cálculo y normalización de los parámetros espaciotemporales, dentro del subdirectorio de cada captura se ha incluido un archivo en formato **XLSX** formado por 6 hojas de cálculo con los datos relativos a los parámetros espaciales y temporales durante la marcha de dicha grabación: las hojas *Left_Gait_Cycle* y *Right_Gait_Cycle* contienen, de forma tabular, los parámetros temporales estimados para cada zancada durante la grabación, mientras que las hojas *Left_Spatial* e *Right_Spatial*, hacen lo propio para los parámetros espaciales. Asimismo, se incluye una hoja llamada *Overview* con los valores promedio de cada parámetro espaciotemporal de la marcha, calculado a partir de los valores individuales de cada zancada registrados en las hojas anteriores. Por su parte, en la hoja *Norm-Overview*, se proporcionan estos mismos parámetros normalizados por las características antropométricas del sujeto. Para facilitar la manipulación de estos datos, en la raíz del directorio se ha generado un archivo *spatiotemporal.csv*, que incluye las características predictoras obtenidas a partir de los archivos **XLSX** individuales. Este archivo contiene una fila por cada captura de movimiento en el *dataset*, y 36 columnas (más dos de índice) para las variables explicativas descritas en la Tabla 5.10.

De manera similar a lo explicado para los parámetros espaciotemporales, durante el preprocesamiento de los datos cinemáticos se ha generado un archivo en formato **XLSX** para cada captura del *dataset*. En este caso, se tiene una única hoja de cálculo con una tabla de datos que incluye una fila por cada captura en el *dataset*, y 182 columnas (más dos de índice) que se corresponden con las variables explicativa generadas como resumen de los desplazamientos angulares de las articulaciones de tobillo, rodilla, cadera, pelvis y pie, en los planos sagital, frontal y transversal. En la Tabla 5.11 se especifica cómo se forman estas características predictoras concatenando el prefijo indicado con las medidas descriptivas definidas en la Tabla 5.5 (*mean*, *min*, *max*, *range*, *stride_start*, *stride_end*, *swing_angle*).

Resumen de la fase de preparación de datos

Concluye así esta fase de preprocesamiento, en la se ha comenzado organizando los archivos con las trayectorias de los marcadores en ciclos de la marcha (zancadas), así como interpolando los valores ausentes y tratando de reducir el ruido mediante la aplicación de un filtro de paso bajo (Savitzky-Golay). También se ha llevado a cabo un proceso de ingeniería de características para transformar estos datos en bruto en un amplio conjunto de variables predictoras propias de los análisis de la marcha, que, además, han sido normalizadas para eliminar la influencia de las medidas antropométricas de las personas. Finalmente, se

han filtrado los datos eliminando aquellas zancadas inexactas, incompletas o imprecisas en base a diferentes criterios (giros durante la marcha, valores atípicos en los parámetros espaciotemporales y problemas en las curvas cinemáticas).

Como resultado, se dispone de un conjunto de datos preparado para construir modelos predictivos orientados a la valoración automática de la marcha humana. En este sentido, en la siguiente sección se explica cómo se han aplicado técnicas de *Machine Learning* para resolver diversos problemas derivados del criterio de evaluación de marcha en la prueba de Tinetti.

Tabla 5.10: Variables predictoras basadas en los parámetros espaciotemporales. En las unidades, *s* indica segundos, *m* metros y *mm* milímetros. Las variables para las que se indica un guión como unidad son adimensionales.

#	Variable	Descripción	Unidad
1	<code>gait_time</code>	Duración de la captura	s
2	<code>right_stride_duration</code>	Tiempo transcurrido entre el comienzo de una zancada y la siguiente	s
3	<code>left_stride_duration</code>		
4	<code>right_step_duration_norm</code>	Tiempo transcurrido entre el comienzo de una zancada y la contraria	- (s)
5	<code>left_step_duration_norm</code>		
6	<code>right_stance_phase_duration_norm</code>	Duración de la fase de apoyo dentro del ciclo de la marcha	- (s)
7	<code>left_stance_phase_duration_norm</code>		
8	<code>right_swing_phase_duration_norm</code>	Duración de la fase de balanceo dentro del ciclo de la marcha	- (s)
9	<code>left_swing_phase_duration_norm</code>		
10	<code>right_stance_percent</code>	Porcentaje de tiempo que ocupa la fase de apoyo dentro del ciclo de la marcha	% del ciclo de la marcha
11	<code>left_stance_percent</code>		
12	<code>right_swing_phase_percent</code>	Porcentaje de tiempo que ocupa la fase de balanceo en el ciclo de la marcha	% del ciclo de la marcha
13	<code>left_swing_phase_percent</code>		
14	<code>right_double_support_duration_norm</code>	Duración del apoyo bipodal dentro del ciclo de la marcha	- (s)
15	<code>left_double_support_duration_norm</code>		
16	<code>right_double_support_percent</code>	Porcentaje de tiempo que ambos pies están en contacto con el suelo	% del ciclo de la marcha
17	<code>left_double_support_percent</code>		
18	<code>right_stride_length_norm</code>	Distancia lineal entre los contactos pie-suelo sucesivos del mismo pie	- (mm)
19	<code>left_stride_length_norm</code>		
20	<code>right_step_length_norm</code>	Distancia lineal entre los contactos pie-suelo sucesivos del pie opuesto	- (mm)
21	<code>left_step_length_norm</code>		
22	<code>base_of_support_norm</code>	Distancia lineal entre el punto medio de los talones durante la marcha	- (mm)
23	<code>right_heel_height_norm</code>	Altura máxima del talón durante el ciclo de la marcha	- (mm)
24	<code>left_heel_height_norm</code>		
25	<code>right_toe_height_norm</code>	Altura máxima de los metatarsos del pie durante el ciclo de la marcha	- (mm)
26	<code>left_toe_height_norm</code>		
27	<code>com_vertical_displacement_norm</code>	Desplazamiento vertical del COG	- (mm)
28	<code>com_horizontal_displacement_norm</code>	Desplazamiento horizontal del COG	- (mm)
29	<code>right_step_angle</code>	Grado de giro del pie durante la marcha	grados
30	<code>left_step_angle</code>		
31	<code>right_strike_angle</code>	Ángulo formado entre el talón y el suelo cuando se produce el contacto inicial	grados
32	<code>left_strike_angle</code>		
33	<code>right_toe_off_angle</code>	Ángulo formado entre el pie y el suelo cuando se produce el despegue del pie	grados
34	<code>left_toe_off_angle</code>		
35	<code>speed_norm</code>	Relación entre la distancia recorrida por unidad de tiempo	- (m/s)
36	<code>cadence_norm</code>	Número de pasos por unidad de tiempo	-

Tabla 5.11: Variables predictoras basadas en los parámetros cinemáticos. El nombre completo de cada variable se obtiene concatenado al prefijo indicado las siguientes cadenas referentes a medidas descriptivas: *mean* (media), *min* (mínimo), *max* (máximo), *range* (rango articular), *stride_start* (ángulo al comienzo del ciclo), *stride_end* (ángulo al final del ciclo), *swing_angle* (ángulo al comienzo de la fase de balanceo).

#	Prefijo de variable	Plano	Descripción	Unidad
1 a 7	right_pelvis_tilt	Sagital	Inclinación pélvica inferior(-) y superior (+).	grados
8 a 14	left_pelvis_tilt			
15 a 21	right_pelvis_obliquity	Frontal	Oblicuidad pélvica posterior (-) y anterior (+).	grados
22 a 28	left_pelvis_obliquity			
29 a 35	right_pelvis_rotation	Transverso	Rotación pélvica interna (+) y externa (-)	grados
36 a 42	left_pelvis_rotation			
43 a 49	right_hip_flex_ext	Sagital	Flexión (+) y extensión (-) de cadera	grados
50 a 56	left_hip_flex_ext			
57 a 63	right_hip_add_abd	Frontal	Abducción (-) y adducción (+) de cadera	grados
64 a 70	left_hip_add_abd			
71 a 77	right_hip_rotation	Transverso	Rotación interna (+) y externa (-) de cadera	grados
78 a 84	left_hip_rotation			
85 a 91	right_knee_flex_ext	Sagital	Flexión (+) y extensión (-) de rodilla	grados
92 a 98	left_knee_flex_ext			
99 a 105	right_knee_val_var	Frontal	Valgo (-) y varo (+) de rodilla	grados
106 a 112	left_knee_val_var			
113 a 119	right_knee_rotation	Transverso	Rotación interna (+) y externa (-) de rodilla	grados
120 a 126	left_knee_rotation			
127 a 133	right_ankle_flex_ext	Sagital	Flexión (+) y extensión (-) de tobillo	grados
134 a 140	left_ankle_flex_ext			
141 a 147	right_ankle_add_abd	Frontal	Abducción (-) y adducción (+) de tobillo	grados
148 a 154	left_ankle_add_abd			
155 a 161	right_ankle_rotation	Transverso	Rotación interna (+) y externa (-) de tobillo	grados
162 a 168	left_ankle_rotation			
169 a 175	right_foot_rotation	Transverso	Dirección del pie durante la marcha	grados
176 a 182	left_foot_rotation			

5.3. MODELADO Y EVALUACIÓN

Las pruebas observacionales utilizadas tradicionalmente para valorar y diagnosticar la marcha humana están, en su mayoría, realizadas por medio de procesos rutinarios que requieren de una evaluación subjetiva por parte de uno o varios profesionales sanitarios. Aprovechando la precisión de los datos sensoriales previamente recogidos y preprocesados, durante esta fase de trabajo se pretende desarrollar y evaluar varios modelos predictivos, basados en *Machine learning*, que permitan generar automáticamente las respuestas a una de las pruebas observaciones para valoración de la marcha más extendidas y validadas actualmente: la prueba de Tinetti (POMA).

De acuerdo con lo presentado en la Sección 3.2.1, la escala de Tinetti evalúa diferentes aspectos de la marcha, tales como la longitud y altura de paso, simetría, fluidez, trayectoria o postura durante el movimiento. A cada uno de estos ítems, uno o varios observadores les asignan una puntuación de 0 o 1 (y 2 en algunos casos no considerados en este trabajo), dependiendo, respectivamente, de si se ha realizado de manera normal o se han percibido anomalías. Por tanto, el problema de generar automáticamente las respuestas a la prueba de Tinetti podría modelarse como una tarea supervisada de clasificación multietiqueta, asig-

nando a cada muestra un conjunto de puntuaciones binarias relacionadas con la valoración de los diferentes ítems de la escala. No obstante, para simplificar esta tarea, se ha optado por descomponer el problema general en varios subproblemas más sencillos, y resolver cada uno de ellos como un ejercicio supervisado de clasificación binaria independiente de las demás, tal y como se describe en la Sección 5.3.2. Como paso previo, en la siguiente sección se han utilizado diferentes gráficos y visualizaciones para explorar y analizar el conjunto de datos preprocesado.

5.3.1. Análisis exploratorio de datos sobre el *dataset* preprocesado

Con la finalidad de analizar la existencia de dependencias entre las variables del *dataset*, para tener así un mayor entendimiento de los datos antes de aplicar los algoritmos de Aprendizaje Automático, se ha realizado un análisis exploratorio (EDA) de las características predictoras basadas en los parámetros espaciotemporales y cinemáticos de la marcha contenidos en el conjunto de datos.

Comenzando por los parámetros espaciotemporales de la marcha, en la Tabla 5.12 se proporciona la media y desviación estándar de estos atributos, calculadas para cada tipo de grabación (según las etiquetas descritas en la Tabla 5.2).

Tabla 5.12: Media y desviación estándar de las variables espaciotemporales para cada tipo de marcha en el conjunto de datos

Variable	Etiqueta						
	norm	lap1	lap2	lap3	lap4	dc	pm
right_stride_length_norm	1.09 ± 0.09	0.56 ± 0.1	0.56 ± 0.1	0.62 ± 0.23	0.65 ± 0.23	0.44 ± 0.13	0.5 ± 0.24
left_stride_length_norm	1.09 ± 0.11	0.54 ± 0.09	0.59 ± 0.12	0.59 ± 0.21	0.66 ± 0.23	0.47 ± 0.19	0.5 ± 0.25
right_step_length_norm	0.52 ± 0.06	0.52 ± 0.09	0.06 ± 0.03	0.5 ± 0.12	0.13 ± 0.09	0.2 ± 0.11	0.15 ± 0.13
left_step_length_norm	0.59 ± 0.05	0.05 ± 0.03	0.61 ± 0.1	0.12 ± 0.1	0.6 ± 0.14	0.32 ± 0.11	0.37 ± 0.11
base_of_support_norm	0.13 ± 0.02	0.18 ± 0.03	0.07 ± 0.01	0.19 ± 0.03	0.09 ± 0.03	0.13 ± 0.03	0.39 ± 0.07
right_heel_height_norm	0.8 ± 0.07	0.53 ± 0.1	0.18 ± 0.13	0.35 ± 0.14	0.28 ± 0.2	0.31 ± 0.18	0.37 ± 0.19
left_heel_height_norm	0.79 ± 0.07	0.12 ± 0.08	0.57 ± 0.12	0.21 ± 0.17	0.39 ± 0.12	0.32 ± 0.19	0.36 ± 0.18
right_toe_height_norm	0.32 ± 0.03	0.06 ± 0.03	0.35 ± 0.08	0.04 ± 0.03	0.33 ± 0.1	0.11 ± 0.06	0.13 ± 0.09
left_toe_height_norm	0.33 ± 0.04	0.34 ± 0.05	0.07 ± 0.04	0.31 ± 0.1	0.04 ± 0.03	0.1 ± 0.07	0.12 ± 0.08
com_vertical_displ_norm	0.13 ± 0.01	0.14 ± 0.04	0.15 ± 0.05	0.15 ± 0.07	0.16 ± 0.09	0.07 ± 0.03	0.09 ± 0.03
com_horizontal_displ_norm	0.17 ± 0.02	0.09 ± 0.01	0.09 ± 0.02	0.11 ± 0.04	0.11 ± 0.04	0.08 ± 0.02	0.1 ± 0.03
left_step_duration_norm	0.49 ± 0.02	0.39 ± 0.03	0.58 ± 0.04	0.47 ± 0.06	0.52 ± 0.05	0.49 ± 0.11	0.5 ± 0.02
right_step_duration_norm	0.52 ± 0.01	0.6 ± 0.03	0.41 ± 0.04	0.52 ± 0.06	0.48 ± 0.05	0.49 ± 0.08	0.5 ± 0.02
right_stance_percent	62.57 ± 1.7	67.7 ± 2.52	63.56 ± 2.85	62.39 ± 3.9	67.06 ± 2.93	72.48 ± 3.37	65.76 ± 5.58
right_swing_phase_percent	37.43 ± 1.7	32.3 ± 2.52	36.44 ± 2.85	37.61 ± 3.9	32.94 ± 2.93	27.52 ± 3.37	34.24 ± 5.58
left_stance_percent	60.76 ± 3.41	62.81 ± 3.58	66.83 ± 2.38	67.52 ± 4.84	62.69 ± 4.14	72.86 ± 3.87	64.43 ± 5.69
left_swing_phase_percent	39.24 ± 3.41	37.19 ± 3.58	33.17 ± 2.38	32.48 ± 4.84	37.31 ± 4.14	27.14 ± 3.87	35.57 ± 5.69
right_double_support_percent	23.94 ± 4.27	30.65 ± 4.01	31.6 ± 2.88	29.87 ± 4.09	30.71 ± 5.05	44.29 ± 10.78	30.45 ± 11.2
left_double_support_percent	23.28 ± 3.98	31.98 ± 4.82	30.86 ± 3.34	31.08 ± 4.54	29.86 ± 5.19	48.24 ± 5.65	30.24 ± 10.79
right_toe_off_angle	19.33 ± 4.01	5.29 ± 2.67	11.6 ± 4.82	4.73 ± 3.67	12.44 ± 4.79	6.64 ± 3.78	7.49 ± 5.03
left_toe_off_angle	17.7 ± 3.43	10.34 ± 4.35	5.76 ± 1.93	10.94 ± 4.83	3.9 ± 2.19	7.12 ± 3.97	6.12 ± 4.68
right_strike_angle	70.67 ± 4.01	84.71 ± 2.67	78.4 ± 4.82	85.27 ± 3.67	77.56 ± 4.79	83.36 ± 3.78	82.51 ± 5.03
left_strike_angle	72.3 ± 3.43	79.66 ± 4.35	84.24 ± 1.93	79.06 ± 4.83	86.1 ± 2.19	82.88 ± 3.97	83.88 ± 4.68

Al analizar esta información se han observado varios aspectos relevantes:

- En los distintos tipos de marchas patológicas se percibe una reducción, cercana a la mitad, de la longitud de zancada. Consecuentemente, esto también afecta a la longitud de paso, y, especialmente, en las observaciones de tipo *lap1* y *lap2*, en las que se simula que una extremidad no supera a la otra durante la marcha. Además, en las marchas anómalas se produce un aumento del ángulo de contacto y una disminución el promedio del ángulo de despegue, así como una menor altura de talón y pie durante la marcha. Esto se representa visualmente en los gráficos de barras de la Figura 5.35.

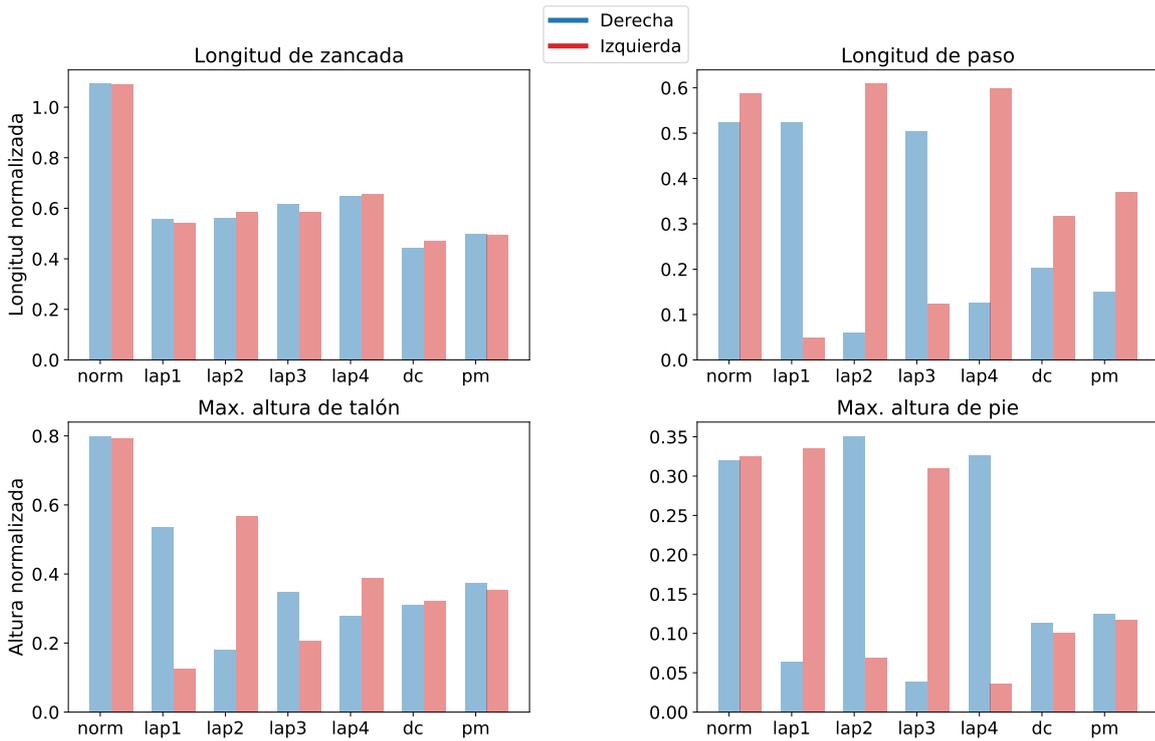


Figura 5.35: Diagramas de barras horizontales para representar la variación de la longitud promedio de zancada y paso, y las alturas máximas de talón y pie (normalizadas), en cada tipo de grabación

- Aunque en todas las grabaciones de marcha patológica aumenta ligeramente la duración porcentual de la fase de soporte dentro del ciclo de zancada, este crecimiento es aún más notable cuando se simula una marcha discontinua (*dc*), ya que la fase de apoyo asciende hasta más de un 70 % (siendo un 60 % en marchas normales). Por tanto, tal y como se ilustra en el diagrama de barras de la Figura 5.36, se reduce un 10 % la duración de la fase de balanceo. Además, en las capturas *dc* el aumento de los tiempos de doble apoyo también es más relevante: del 24 % del ciclo de la marcha en capturas normales, se pasa a valores cercanos al 30 % en el resto de las marchas patológicas, y más de un 44 % en marchas con interrupciones.

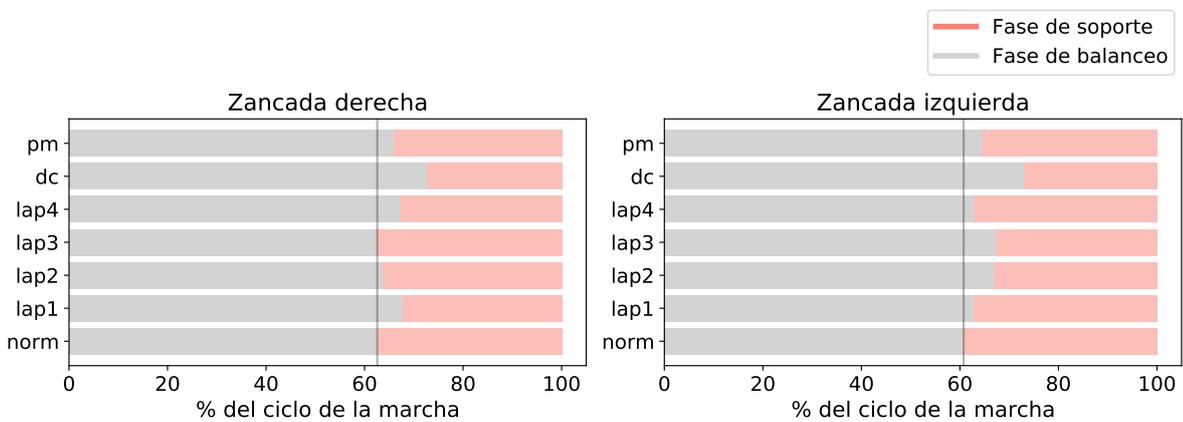


Figura 5.36: Diagramas de barras horizontales para representar la duración promedio de las fases de balanceo y soporte en cada tipo de grabación

- En las grabaciones en las que se aumenta la separación de talones al caminar (pm) se produce un significativo aumento de la base de sustentación ($base_of_support_norm$) con respecto al resto de grabaciones.

Por otro lado, teniendo en cuenta que algunos modelos de *Machine Learning* se ven perjudicados cuando incorporan predictores altamente correlacionados entre sí, conviene analizar el grado de correlación entre cada par de parámetros espaciotemporales. Para ello, en la Figura 5.37 se incluye un *heatmap* con la matriz de correlaciones calculada utilizando el método de Spearman.

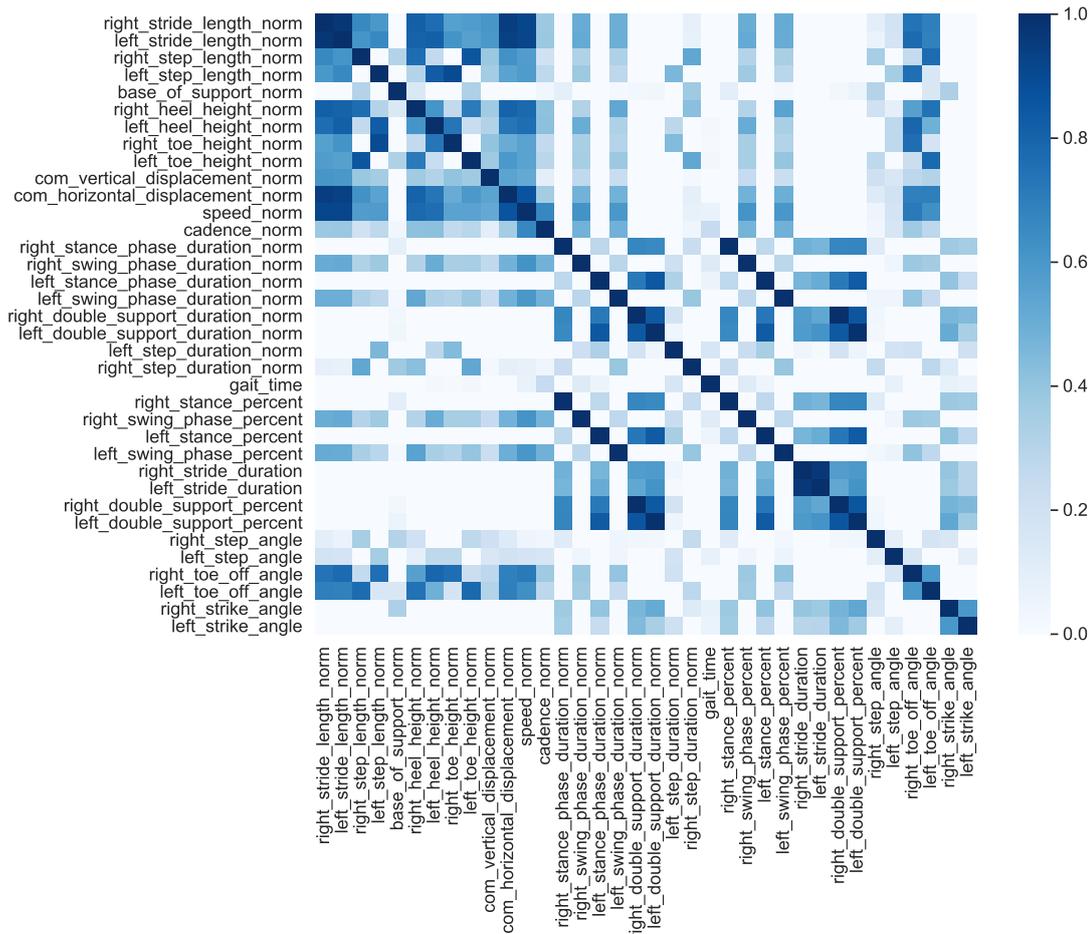


Figura 5.37: Mapa de calor con la correlación entre las variables espaciotemporales

En el *heatmap* anterior se observa una alta correlación entre los parámetros correspondientes a la longitud de zancada de ambas extremidades ($right_stride_length_norm$ y $left_stride_length_norm$) con el desplazamiento horizontal del centro de gravedad ($com_horizontal_displacement_norm$) y la velocidad de la marcha ($speed_norm$). Asimismo, las características que han sido incluidas tanto de forma temporal como porcentual están altamente correlacionadas y sería conveniente prescindir de una de ellas para no alimentar a los modelos de *Machine Learning* con información redundante. Es el caso, por ejemplo, de las duraciones de las distintas fases de la marcha, que también se proporcionan como el porcentaje que ocupan dentro del ciclo de zancada.

Por lo tanto, el conjunto de 36 predictores espaciotemporales presentados en la Tabla 5.10 se ha reducido a un listado de 28 variables, eliminando la velocidad y el desplazamiento

horizontal del COG, así como las duraciones de la fase de soporte, balanceo y el tiempo de doble apoyo para ambas extremidades, manteniendo únicamente las variables equivalentes representadas en forma de porcentaje.

En cuanto a las variables predictoras basadas en la cinemática de la marcha, puesto que se cuenta con un elevado número de atributos generados a partir de medidas descriptivas, se ha optado por analizar únicamente cómo varía el rango angular de las articulaciones en los diferentes tipos de marcha. En este sentido, la Tabla 5.13 proporciona la media y desviación estándar de la amplitud angular para cada movimiento cinemático incluido en el *dataset*.

Tabla 5.13: Media y desviación estándar del rango articular (variables cinemáticas) para cada tipo de marcha en el conjunto de datos

Variable	Etiqueta						
	norm	lap1	lap2	lap3	lap4	dc	pm
right_pelvis_tilt	4.65 ± 1.2	6.72 ± 1.87	6.7 ± 2.82	8.65 ± 3.73	8.76 ± 4.53	4.42 ± 0.96	3.99 ± 0.87
left_pelvis_tilt	4.4 ± 1.35	6.31 ± 1.69	6.78 ± 2.87	8.15 ± 3.16	8.81 ± 4.58	4.55 ± 1.18	3.88 ± 0.62
right_pelvis_obliquity	6.14 ± 2.97	7.31 ± 1.98	7.11 ± 1.44	7.24 ± 2.13	6.54 ± 1.77	6.82 ± 1.21	8.89 ± 1.59
left_pelvis_obliquity	5.99 ± 2.86	7.0 ± 1.86	7.14 ± 1.41	6.98 ± 2.27	6.74 ± 1.7	6.94 ± 1.25	8.88 ± 1.54
right_pelvis_rotation	35.76 ± 8.19	93.09 ± 28.39	84.99 ± 24.11	92.23 ± 31.9	84.45 ± 30.63	105.55 ± 24.93	96.64 ± 25.54
left_pelvis_rotation	35.17 ± 12.44	86.9 ± 22.98	84.6 ± 22.5	87.43 ± 31.46	84.63 ± 28.89	109.05 ± 27.13	94.35 ± 26.65
right_hip_flex_ext	38.23 ± 3.27	26.64 ± 4.43	21.3 ± 6.57	27.08 ± 7.75	28.67 ± 10.76	23.9 ± 6.7	20.82 ± 8.9
left_hip_flex_ext	41.9 ± 16.47	24.39 ± 19.14	27.57 ± 7.15	26.82 ± 12.17	27.77 ± 10.29	26.3 ± 8.4	21.56 ± 10.72
right_hip_add_abd	15.19 ± 4.13	11.15 ± 3.46	7.91 ± 2.81	11.66 ± 3.41	10.73 ± 4.93	9.72 ± 4.17	8.27 ± 2.9
left_hip_add_abd	16.82 ± 8.21	9.84 ± 7.66	11.7 ± 4.12	10.56 ± 5.49	11.85 ± 4.25	10.75 ± 4.85	9.24 ± 4.17
right_hip_rotation	13.18 ± 4.63	11.05 ± 4.54	13.18 ± 6.69	11.46 ± 5.65	15.31 ± 9.22	10.41 ± 4.57	10.2 ± 4.91
left_hip_rotation	20.5 ± 24.88	23.12 ± 34.15	15.25 ± 18.53	13.9 ± 6.13	11.12 ± 5.0	10.5 ± 4.07	9.68 ± 3.98
right_knee_flex_ext	62.06 ± 4.89	33.48 ± 4.55	38.59 ± 11.39	18.89 ± 5.98	40.28 ± 10.87	35.48 ± 11.06	25.91 ± 9.8
left_knee_flex_ext	56.53 ± 7.98	30.57 ± 6.01	31.41 ± 6.66	32.88 ± 9.53	18.27 ± 6.39	31.91 ± 9.39	22.26 ± 6.51
right_knee_val_var	27.94 ± 10.06	14.05 ± 6.17	19.83 ± 9.59	9.26 ± 4.69	20.48 ± 8.32	16.26 ± 7.72	9.67 ± 5.57
left_knee_val_var	33.8 ± 10.75	23.78 ± 12.43	17.01 ± 5.7	20.93 ± 9.72	9.88 ± 5.04	17.28 ± 8.17	10.93 ± 7.91
right_knee_rotation	20.72 ± 5.22	12.81 ± 3.6	12.21 ± 3.23	11.13 ± 3.51	13.42 ± 4.22	10.36 ± 2.88	10.69 ± 4.24
left_knee_rotation	22.33 ± 11.48	16.68 ± 16.67	14.57 ± 7.24	12.62 ± 5.43	12.24 ± 4.67	11.66 ± 4.1	9.98 ± 2.76
right_ankle_flex_ext	27.89 ± 4.21	20.35 ± 6.71	20.71 ± 3.12	17.33 ± 7.01	22.17 ± 3.74	17.94 ± 4.45	16.34 ± 7.58
left_ankle_flex_ext	30.54 ± 6.56	22.51 ± 4.84	22.67 ± 8.64	22.67 ± 5.97	22.62 ± 9.86	19.08 ± 6.36	16.59 ± 7.58
right_ankle_add_abd	15.23 ± 2.69	10.99 ± 1.98	14.58 ± 6.34	12.16 ± 2.94	14.01 ± 5.97	12.7 ± 3.25	12.13 ± 5.03
left_ankle_add_abd	18.97 ± 10.93	16.33 ± 9.34	14.63 ± 6.87	15.36 ± 8.92	16.27 ± 7.32	13.69 ± 5.58	12.58 ± 7.06
right_ankle_rotation	19.07 ± 2.98	11.39 ± 2.31	14.45 ± 4.0	12.74 ± 3.71	16.93 ± 4.61	13.11 ± 2.52	13.46 ± 2.81
left_ankle_rotation	21.68 ± 8.7	17.26 ± 11.59	12.95 ± 5.71	16.76 ± 5.94	15.96 ± 6.14	13.58 ± 4.67	13.72 ± 3.53
right_foot_rotation	28.86 ± 7.18	20.09 ± 6.92	17.84 ± 3.33	17.8 ± 6.5	20.7 ± 6.8	16.57 ± 3.74	17.02 ± 4.26
left_foot_rotation	29.64 ± 8.63	15.34 ± 4.88	20.6 ± 5.02	18.0 ± 8.57	20.36 ± 7.4	17.11 ± 5.96	17.4 ± 3.25

En estas estadísticas descriptivas se observa que la amplitud de la rotación pélvica sufre un importante aumento cuando la marcha no es normal, produciéndose el mayor incremento en las capturas con discontinuidad de pasos (*dc*). En este caso, el rango articular supera los 100° para ambas extremidades cuando en marchas no patológicas el promedio es cercano a 35°. También parece que los movimientos de rodilla en diferentes planos son sensibles a las alteraciones de marcha contenidas en el *dataset*. Por ejemplo, la amplitud de la flexión-extensión de rodilla se reduce cerca de 30°, experimentado los mayores decrementos en las grabaciones en las que uno de los pie no se levanta completamente del suelo. Para esta misma patología también se observa una importante alteración del rango articular de la rodilla en el plano frontal (valgo-varo).

Para analizar la correlación entre la amplitud angular de los diferentes movimientos articulares en los tres planos, se ha trazado el mapa de calor mostrado en la Figura 5.38, utilizando el coeficiente de Spearman como puntuación de correlación. En esta primera aproximación no se observan correlaciones importantes más allá de las generadas por los movimientos pélvicos de cada zancada en los tres planos. Puesto que, en este caso, la correlación se encuentran entre los atributos de un mismo movimiento pero referidos a distintas piernas (por ejemplo, rotación pélvica izquierda y derecha), se ha optado por reemplazar las

variables predictoras referidas a estos movimientos por una nuevas características generadas como el promedio del movimiento pélvico para las dos extremidades. Más concretamente, se ha detectado que esta correlación ocurre también con los atributos de máximo, mínimo, media, comienzo de zancada (*Stride_Start*) y fin de zancada (*Stride_end*) para la inclinación de la pelvis (*pelvis_tilt*).

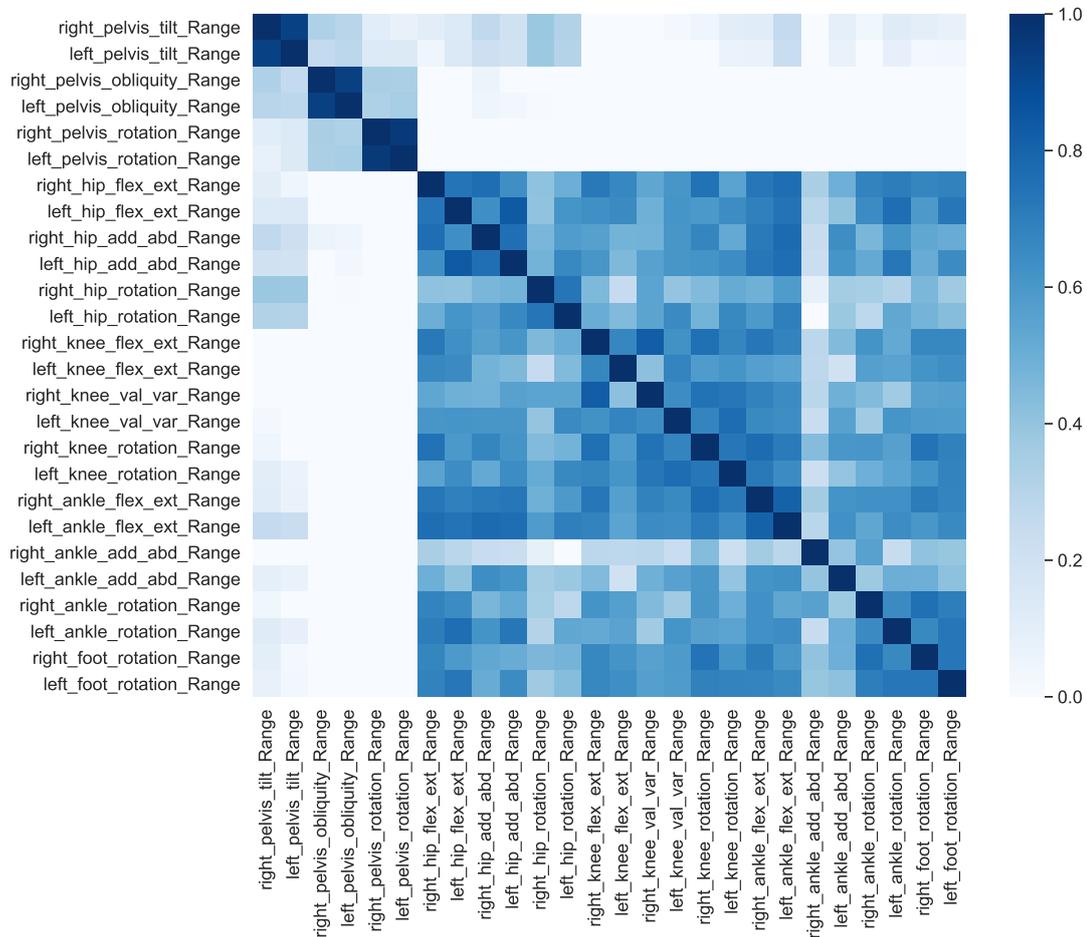


Figura 5.38: Mapa de calor con la correlación entre en los parámetros cinemáticos relacionados con la amplitud angular

Extendiendo el análisis de la correlación entre variables cinemáticas al resto de las medidas descriptivas, se han encontrado 121 pares de atributos con una puntuación de Spearman superior a 0.9, lo que sugiere una alta correlación. Se trata de casos en los que la naturaleza del movimiento angular provoca que dos medidas descriptivas para una mismo segmento corporal sean similares. Por ejemplo, durante la marcha se alcanza el mayor grado de flexión de cadera tanto al inicio como al final de la zancada (Figura 3.6), lo que conduce a que las características *right_hip_flex_ext_Stride_End* y *right_hip_flex_ext_Stride_Start* sean equivalentes y, consecuentemente, estén altamente correlacionadas. De nuevo, para no alimentar a los modelos de Aprendizaje Automático con información redundante, se ha optado eliminar una de las dos variables de cada par de atributos altamente correlacionados. De esta forma, el conjunto inicial de 182 características cinemáticas se ha reducido a un total de 105 atributos.

El código *Python* con el que se ha desarrollado este análisis exploratorio de datos está accesible desde el repositorio del proyecto, cuyo enlace se proporciona en el Anexo A. Más concretamente, se trata de un archivo de *Jupyter Notebook* llamado *eda-prepared.ipynb* y

ubicado dentro del directorio *notebooks*. En la siguiente sección se presenta cómo, utilizando clasificadores binarios, se ha modelado y resuelto el problema de generar automáticamente las respuestas a la prueba de Tinetti para valoración de la marcha.

5.3.2. Descripción del proceso de modelado

Partiendo del conjunto de datos etiquetado y preprocesado en las fases de trabajo anteriores, en esta etapa de modelado se han abordado, como tareas de clasificación binaria, los siguientes seis problemas individuales, que están relacionados con varios de los ítems más relevantes de la escala de Tinetti para valoración de la marcha. En todos estos problemas de clasificación, se asigna una etiqueta de 0 a las marchas normales para las que no se observan desviaciones, y una puntuación de 1 cuando se trata de una marcha anormal en la categoría analizada:

- **LAP1** (longitud y altura del paso 1). Otorga una puntuación de 1 a las marchas en las que el pie derecho no sobrepasa al izquierdo con el paso en la fase de balanceo (Sección 5.3.3).
- **LAP2** (longitud y altura del paso 2). Etiqueta con un 1 las capturas de movimiento en las que el pie izquierdo no sobrepasa al derecho con el paso en la fase de balanceo (Sección 5.3.4).
- **LAP3** (longitud y altura del paso 3). Considera como marcha patológica (1) aquellas en las que el pie derecho se arrastre o no se levante completamente del suelo en la fase de balanceo (Sección 5.3.5).
- **LAP4** (longitud y altura del paso 4). Puntúa como marcha anormal (1) si el pie izquierdo se arrastra o no se levanta completamente del suelo en la fase de balanceo (Sección 5.3.6).
- **DC** (discontinuidad entre pasos). Asigna una puntuación de 1 cuando el paciente adopta una marcha discontinua e interrumpida, en la que varía el largo del paso o en la que se apoya completamente un pie antes de levantar el otro (Sección 5.3.7).
- **PM** (postura en la marcha). Califica como marcha con alteraciones (1) aquellas en las que se camina con los talones separados, produciéndose un notable aumento en la base de sustentación (Sección 5.3.8).

El entrenamiento y evaluación de los modelos predictivos orientados a resolver los subproblemas anteriores se ha llevado a cabo sobre varios *notebooks* que utilizan como herramienta principal la librería *Scikit-learn* de *Python*. Estos archivos son accesibles desde el repositorio del proyecto, cuyo enlace se proporciona en el Anexo A. Para cada uno de los problemas abordados se ha llevado a cabo un proceso de Aprendizaje Automático pasando por las siguientes etapas:

1. **Cargar los datos.** espaciotemporales (*spatiotemporal.csv*) y cinemáticos (*kinematics.csv*), así como las etiquetas (*summary.xlsx*) del conjunto de datos preprocesado. De las 168 observaciones que forman el *dataset* (17 capturas por 10 participantes, menos dos que fueron filtradas según lo descrito en la Sección 5.2.3), en cada tarea se trabaja con un subconjunto de 68 muestras, correspondientes a las 48 grabaciones de marcha normal más las 20 específicas de cada subproblema, de acuerdo con las etiquetas definidas en la Tabla 5.2.

A partir de estos archivos se han generado tres estructuras de datos: una primera que contiene solo características predictoras relacionadas con los parámetros espaciotemporales de la marcha; otra basada solo en las características angulares de la cinemática,

y una última tabla generada como resultado de combinar las dos anteriores.

- Definir conjuntos de entrenamiento y prueba.** Para poder evaluar correctamente la capacidad de generalización de los modelos predictivos, el conjunto de datos inicial se ha dividido en dos partes: entrenamiento y prueba, dedicando dos tercios de los datos (66.6 %) para el entrenamiento, y el resto para la evaluación (33.3 %). En este sentido, puesto que se trabaja con un *dataset*, y se trata de problemas de clasificación en los que las clases de salida no están balanceadas (para cada participante hay cinco observaciones de marcha normal y dos con alteraciones), la división del conjunto de datos se ha llevado a cabo mediante un muestreo estratificado en base a las etiquetas de salida. De esta forma, se garantiza que el conjunto de prueba es representativo de los distintos tipos de marcha, y se reduce el riesgo de introducir un sesgo muestral significativo [159]. Para ello, se ha utilizado la clase `StratifiedShuffleSplit`¹⁰ de *Scikit-Learn*, que, además, antes de dividir mezcla aleatoriamente la muestra.
- Seleccionar las características más relevantes.** La selección de características es una parte crucial en cualquier proyecto de *Machine Learning*, ya que permite reducir el número de variables utilizadas para la construcción de modelos, seleccionando aquellas que se consideran más relevantes para predecir la variable objetivo. Consecuentemente, el coste computacional del entrenamiento disminuye, se consiguen modelos más sencillos e, incluso, en alguno casos, con mejor rendimiento.

De las distintas técnicas utilizadas habitualmente para seleccionar las características más relevantes de un conjunto de datos, en este trabajo se ha seguido un enfoque supervisado de filtrado basado en la correlación estadística entre cada variable explicativa y la etiqueta de salida. Más concretamente, puesto que todas las características de entrada en el conjunto de datos son numéricas, y la respuesta objetivo es una variable categórica, para determinar cómo de relacionado está cada predictor con la salida se ha utilizado el valor del estadístico F (ANOVA) [160]. Para ello, se ha utilizado la función `SelectKBest`¹¹ de *Scikit-Learn*, que permite seleccionar las n características más relevantes basándose en pruebas estadísticas univariantes, indicando como método de correlación la función `f_classif` que también facilita *Scikit-Learn*¹².

De esta forma, en cada uno de los problemas derivados de la valoración de Tinetti y modelados mediante modelos predictivos, se han obtenido las 15 variables más relevantes, para los casos en los que se considera como predictores solo el conjunto de características basadas en parámetros espaciotemporales, solo las variables relacionadas con la cinemática del movimiento, y una combinación de ambas.

- Escalar los datos de entrada.** Otra de las transformaciones más importantes que hay que aplicar a los datos de entrada es el escalado de características, ya que, generalmente, los algoritmos de Aprendizaje Automático son sensibles a las escalas de los atributos numéricos. Para decidir qué tipo de escalado aplicar, se ha tenido en cuenta tanto las transformaciones que realiza cada escalador como las propiedades de los algoritmos de *Machine Learning* que se van a utilizar en el siguiente paso.

¹⁰https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedShuffleSplit.html

¹¹https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html

¹²https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_classif.html

Los dos métodos más habituales para conseguir que todos los atributos tengan una misma escala, son la normalización y la estandarización. En este sentido, se ha descartado la estandarización (`StandardScaler` en *Scikit-Learn*) porque no comprime los valores en un rango específico, tal y como requiere el algoritmo `Complement Naive Bayes` utilizado posteriormente. Del resto de técnicas ofrecidas por *Scikit-Learn*, aunque la normalización `MinMax` sí limita los valores de entrada al intervalo $[0,1]$, se ve mucho más afectada por valores atípicos, lo que puede degradar el rendimiento de los modelos. Por su parte, el escalador `RobustScaler` es mucho más robusto a *outliers*, pero tampoco restringe los datos a un rango positivo válido. Así pues, el método que mejor se adapta a los datos concretos de estos problemas es `QuantileTransformer`¹³, que aplica una transformación no lineal de forma que la función de densidad de probabilidad de cada característica se mapea a una distribución uniforme. Por tanto, en todos los problemas abordados se ha utilizado este escalador, ya que es robusto a valores atípicos y reduce todos los valores al intervalo $[0,1]$.

5. **Entrenar y perfeccionar los modelos.** Una vez escalados los datos, se han entrenado varios modelos sobre el conjunto de entrenamiento, trabajando con diversos algoritmos supervisados de clasificación, y probando con diferentes tipos y número de las variables predictoras más informativas identificadas en el tercer paso. Además, se han ajustado los hiperparámetros de cada modelo para obtener el mejor rendimiento posible en cada caso.

Para el entrenamiento se ha utilizado la característica de validación cruzada de K iteraciones que ofrece *Scikit-Learn* a través de la función `cross_val_score`¹⁴. Con este método, el conjunto de entrenamiento se divide en k bloques, y se utiliza uno de ellos para validación y el resto para el entrenamiento. Más concretamente, se han realizado 10 iteraciones ($k=10$) utilizando como función de utilidad la exactitud de las predicciones. Por otro lado, se ha tratado de perfeccionar cada modelo buscando exhaustivamente la combinación de hiperparámetros que mejores resultados produzca. Para ello, se ha utilizado la clase `GridSearchCV`¹⁵ de *Scikit-Learn*, la cual utiliza validación cruzada para evaluar todas las combinaciones posibles de valores de hiperparámetros. Este procedimiento se ha llevado a cabo con los siguientes algoritmos:

- **K vecinos más cercanos**, o *K-Neighbors Classifier* (KNN). Es un algoritmo de tipo aprendizaje basado en instancias, que tal y como indica su nombre, clasifica a las observaciones según la clase predominante entre los K datos más próximos, basándose en el concepto de distancia entre puntos para cuantificar la vecindad de las instancias. Para crear modelos de este tipo se ha utilizado la clase `KNeighborsClassifier`¹⁶ de la librería *Scikit-Learn*.

Para afinar los hiperparámetros de este algoritmo, se ha buscado el valor óptimo de K teniendo en cuenta que un valor bajo provocará que el resultado de la predicción se ajuste mucho a los datos de entrenamiento, y un valor alto puede conducir a predicciones con muy poca variabilidad de una instancia a otra. En este sentido, se ha generado una lista de valores candidatos considerando que el

¹³<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.QuantileTransformer.html>

¹⁴https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_val_score.html

¹⁵https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

¹⁶<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

valor de K [87]: (1) no puede exceder el número de muestras en el conjunto de entrenamiento; y (2), no debería ser un número impar (o múltiplo del número de etiquetas) para evitar posibles empates entre clases. Con estas condiciones, se ha definido una rejilla de búsqueda (*grid search*) en la que también se ha explorado la función de distancia a considerar (Euclídea, Manhattan, Chebyshev o Minkowski).

- **Bayes ingenuo** (*Naive Bayes*). Son un conjunto de técnicas para clasificación supervisada basadas en la aplicación del teorema de Bayes¹⁷, asumiendo que las características son independientes entre sí. De los diferentes procedimientos de *Naive Bayes* disponibles, en este trabajo se ha utilizado **CNB**, ya que es especialmente adecuado para conjuntos de datos no balanceados, es decir, que el número de ejemplos de una determinada clase es significativamente mayor que del resto.

En este caso, no se ha realizado ningún ajuste de hiperparámetros, ya que durante el entrenamiento los algoritmos de la familia *Naive Bayes* aprenden las probabilidades de cada clase/etiqueta y las condiciones de cada etiqueta por cada clase, pero tienen muy pocos o ningún parámetro ajustable.

- **Árboles de decisión**, o *Decision Tree* (DT). Son unos potentes algoritmos que durante el entrenamiento construyen una estructura de árbol con preguntas de tipo *if-else*, de forma que cada nodo representa una elección entre varias alternativas. Se trata, por tanto, de modelos de tipo ‘caja blanca’ porque son intuitivos y fáciles de interpretar. La clasificación utilizando árboles de decisión se ha llevado a cabo mediante la clase `DecisionTreeClassifier`¹⁸ de *Scikit-Learn*.

Con este tipo de algoritmos no se ha aplicado el escalado de datos descrito en el cuarto paso, ya que los árboles de decisión no son sensibles a este aspecto y cambiar la escala de los datos puede dificultar la interpretación de los resultados obtenidos. Sin embargo, sí que se han ajustado los hiperparámetros de regularización para evitar que la estructura del árbol se adapte demasiado a los datos de entrenamiento (sobreajuste). En este sentido, se ha variado la profundidad máxima del árbol (`max_depth`) entre 1 y el número de predictores considerados; y el número mínimo de muestras que debe tener un nodo antes de poder dividirse (`min_samples_split`) y el número mínimo de muestras que debe tener un nodo terminal (`min_samples_leaf`) entre 1 y 20 [161].

- **Máquinas de vectores soporte**, o *Support Vector Machine* (SVM). Es un popular algoritmo de Aprendizaje Automático que ha demostrado ofrecer un muy buen rendimiento en problemas relacionados con análisis de marcha [102]. Intuitivamente, se basa en encontrar un hiperplano que separe de forma óptima los puntos de diferentes clases. De las distintas implementaciones de **SVM** para escenarios de clasificación que ofrece *Scikit-Learn*, en este trabajo se ha utilizado la clase `SVC`¹⁹ especificando como *kernel* la función de base radial gaussiana (**RBF**), ya que, aunque no existe ningún estudio analítico sobre la elección óptima de la función de *kernel* [102], **RBF** funciona bien con conjuntos de datos no demasiado grandes [159], y ha alcanzado buenos resultados en diferentes tareas de clasifica-

¹⁷https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.ComplementNB.html

¹⁸<https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>

¹⁹<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

ción de marcha [102-104, 162].

Para perfeccionar el modelo de SVC, se ha ajustado el hiperparámetro de regularización C , y el valor de gamma (γ), relacionado con el ancho de la curva RBF, y consecuentemente, con el rango de influencia de cada instancia. En este caso, se ha realizado una doble búsqueda con `GridSearchCV` de manera que la segunda refina a la anterior. En la primera búsqueda se ha utilizado una lista de valores equiespaciados en escala logarítmica ($C = 2^{-5}, 2^{-3}, \dots, 2^{15}$, $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$) y, después, se ha realizado una segunda búsqueda más exhaustiva sobre los valores cercanos a los mejores encontrados en la primera iteración [163].

- **Redes Neuronales Artificiales (RNA)**. Son una familia de versátiles y potentes algoritmos de *Machine Learning* inspirados en la forma en que las neuronas biológicas del cerebro procesan la información. En este trabajo, se ha construido una de las arquitecturas de RNA más sencillas, el **perceptrón multicapa**, o *Multi-layer Perceptron* (MLP), utilizando la API secuencial de *Keras*²⁰ sobre *TensorFlow 2*.

Más concretamente, se ha definido una arquitectura MLP formada por una primera capa de entrada con una neurona de tránsito por cada característica explicativa; una o varias capas profundas con un número variable de neuronas; y, puesto que se trata de problemas de clasificación binaria, la capa de salida formada por una única neurona con la función de activación logística, de forma que la respuesta será un número entero entre 0 y 1 que se puede interpretar como la probabilidad estimada de la clase positiva. Para compilar el modelo se ha utilizado el optimizador Adam con la entropía cruzada binaria como función de pérdida.

Sobre la base de esta arquitectura se han ajustado varios hiperparámetros utilizando la clase `Hyperband`²¹ de *Keras Tuner*²², una librería de optimización de hiperparámetros para modelos de *Keras*. En particular, se ha buscado la arquitectura que mejor rendimiento ofrece variando el número de capas ocultas (entre 1 y 3), el número de neuronas en cada capa oculta (entre 5 y 20) y la tasa de aprendizaje (probando con varios valores comprendidos entre 0.5 y 0.001).

Los mejores estimadores encontrados para cada algoritmo experimentado se han guardado utilizando el módulo `joblib` de *Python*, para poder así volver con facilidad a cualquiera de ellos. En este sentido, se han guardado tanto los modelos como sus hiperparámetros, los parámetros entrenados y la *pipeline* completa de transformación (selección de características y escalado de datos).

6. **Evaluar los modelos en el conjunto de prueba.** Después de entrenar y ajustar los clasificadores, se han calculado varias métricas de evaluación sobre el conjunto de prueba para determinar el rendimiento de estos modelos con datos desconocidos. Para ello, considerando que la exactitud no es la medida de evaluación más fiable cuando se trata con conjuntos de datos en los que algunas clases son mucho más frecuentes que otras, se han considerado también el valor F_1 y el área bajo la curva ROC:
 - **Exactitud** (*accuracy*). Es la proporción de predicciones correctas en relación con el número total de predicciones. Siendo VP el número de verdaderos positivos,

²⁰<https://keras.io/api/models/sequential/>

²¹https://keras.io/api/keras_tuner/tuners/hyperband/

²²https://keras.io/keras_tuner/

VN verdaderos negativos, FP falsos positivos y FN falsos negativos, la exactitud viene dada por la Ecuación 5.4, implementada en la función `accuracy_score`²³ de *Scikit-Learn*.

$$\text{exactitud} = \frac{VP + VN}{VP + FP + VN + FN} \quad (5.4)$$

En el contexto de este trabajo, los verdaderos positivos se pueden entender como el número de veces que el clasificador detecta correctamente una determinada alteración de la marcha, mientras que los falsos negativos son el número de desviaciones ignoradas por el modelo.

- **Valor F_1 .** Es la media armónica de las medidas de precisión y sensibilidad, y proporciona una forma sencilla de comparar varios modelos porque un clasificador solo obtendrá un valor F_1 elevado cuando tanto la sensibilidad como la precisión sean altas [159]. En este sentido, la precisión representa la exactitud de las predicciones positivas midiendo la relación entre el número de verdaderos positivos (VP) y el número total de positivos predichos (Ecuación 5.5), siendo FP el número de falsos positivos.

$$\text{precisión} = \frac{VP}{VP + FP} \quad (5.5)$$

Por su parte, la sensibilidad se define como el ratio de observaciones positivas clasificadas de manera correcta (Ecuación 5.6), donde FN el número de falsos negativos.

$$\text{sensibilidad} = \frac{VP}{VP + FN} \quad (5.6)$$

Así, la puntuación F_1 se obtiene como la media armónica de la sensibilidad y precisión (Ecuación 5.7).

$$F_1 = \frac{2}{\frac{1}{\text{precisión}} + \frac{1}{\text{sensibilidad}}} = \frac{VP}{VP + \frac{FN+FP}{2}} \quad (5.7)$$

Para calcular el valor F_1 se ha utilizado la función `f1_score` de *Scikit-Learn*²⁴.

- **Área bajo la curva (AUC por sus siglas en inglés).** La curva ROC (*Receiver Operating Characteristic* (ROC)) es una herramienta utilizada para evaluar los clasificadores binarios que, traza la sensibilidad frente a la tasa de falsos positivos (FPR, por su nombre en inglés). En este sentido, una forma de comparar clasificadores consiste en medir el área debajo de la curva, teniendo en cuenta que un clasificador perfecto tendrá un AUC igual a 1, mientras que un clasificador puramente aleatorio la tendrá igual a 0.5 [159].

Se ha utilizado la función `roc_auc_score`²⁵, ofrecida por *Scikit-Learn*, para calcular el área debajo de la curva ROC.

7. **Seleccionar los mejores modelos.** Una vez entrenados y evaluados un conjunto de clasificadores candidatos que resuelven un mismo problema, es necesario seleccionar un modelo final para su puesta en producción. Para ello, se ha seguido un doble criterio:

²³https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html

²⁴https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html

²⁵https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html

por un lado, la capacidad de generalización, medida con las métricas de evaluación calculadas en el paso anterior y, por otro lado, para elegir entre varios clasificadores con un rendimiento similar, se ha aplicado en el principio de parsimonia²⁶: seleccionar el modelo más simple o que requiere de un menor número de variables explicativas.

En este trabajo, para cada subproblema analizado se han seleccionado tres modelos basados en diferentes tipos de características de entrada: solo con parámetros espaciotemporales, solo con datos cinemáticos y con ambos tipos de variables. De esta forma, a los usuarios finales se les podrá ofrecer la posibilidad de seleccionar el tipo de características de la marcha utilizadas para generar automáticamente las respuestas a la escala de Tinetti.

Siguiendo estas consideraciones para cada ítem de la escala de Tinetti modelado como una tarea de Aprendizaje Automático, en las siguientes secciones se indican las variables espaciotemporales y cinemáticas más relevantes en cada problema individual; se proporciona una tabla con las métricas de evaluación, obtenidas con diferentes algoritmos y características de entrada, y, finalmente, se seleccionan los modelos que serán desplegados en la última fase de trabajo.

5.3.3. LAP1. El pie derecho no sobrepasa al izquierdo en la fase de balanceo

El primer subproblema modelado como una tarea de Aprendizaje Automático se corresponde con el ítem de la escala de Tinetti que valora si el pie derecho sobrepasa al izquierdo en la fase de balanceo, enmarcado dentro de la categoría que evalúa la longitud y altura de paso. Para abordar este primer caso, se ha tomado un subconjunto del *dataset* formado por las 48 instancias de marcha normal (etiquetándolas como 0), y las 20 capturas en las que se ha simulado esta patología (*lap1*), y que actúan como clase positiva. Siguiendo el proceso explicado en la sección anterior, esta subcolección de datos se ha dividido en un conjunto de entrenamiento (66.6%) y pruebas (33.3%), realizando un muestreo estratificado para tener la misma proporción de marchas anómalas en ambos grupos, tal y como se representa gráficamente en la Figura 5.39.

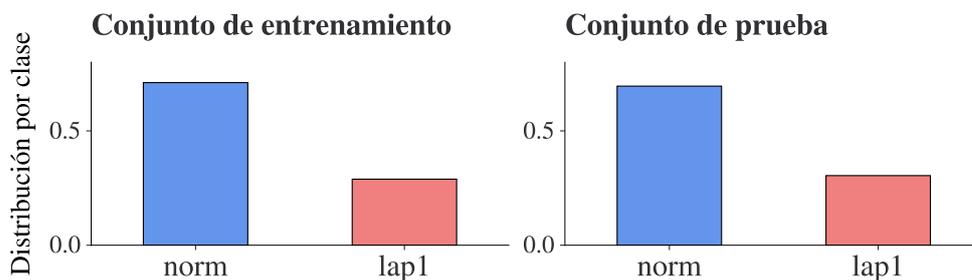


Figura 5.39: Comparativa de la proporción de instancias con clase *lap1* en el conjunto de entrenamiento y prueba

Para este problema concreto, la selección de características basada en ANOVA sugiere que los predictores espaciotemporales más informativos son la longitud de paso izquierdo, la altura del talón izquierdo y la altura del pie derecho. Por su parte, el rango articular de la flexión-extensión de rodilla derecha e izquierda son los atributos cinemáticos más

²⁶El principio de parsimonia, también conocido como la navaja de Ockham, es una formulación filosófica según la cual "en igualdad de condiciones, la explicación más sencilla suele ser la más probable".

importantes. En la Tabla 5.14 se muestra el listado de las 15 características espaciotemporales y cinemáticas más relevantes, ordenadas de mayor a menor relevancia según la puntuación F_1 . El significado de cada una de estas variables puede consultarse, respectivamente, en las Tablas 5.10 y 5.11.

Tabla 5.14: Selección de características espaciotemporales (a) y cinemáticas (b) para el problema de determinar si el pie derecho sobrepasa al izquierdo (LAP1)

(a) Variables espaciotemporales			(b) Variables cinemáticas		
#	Variable	F_1	#	Variable	F_1
1	left_step_length_norm	1317.96	1	right_knee_flex_ext_Range	324.52
2	left_heel_height_norm	763.17	2	left_knee_flex_ext_Range	237.14
3	right_toe_height_norm	671.92	3	right_knee_flex_ext_Max	178.23
4	right_step_duration_norm	347.64	4	left_knee_flex_ext_Max	135.73
5	left_step_duration_norm	286.79	5	pelvis_rotation_Range	124.37
6	right_stride_length_norm	273.90	6	right_ankle_rotation_Range	110.73
7	left_stride_length_norm	249.16	7	right_hip_flex_ext_Range	95.76
8	right_toe_off_angle	133.89	8	left_pelvis_rotation_Min	94.58
9	right_strike_angle	133.89	9	right_pelvis_rotation_Max	82.14
10	right_heel_height_norm	115.64	10	left_pelvis_rotation_Stride_Start	74.00
11	right_swing_phase_percent	79.66	11	left_pelvis_rotation_Stride_End	72.42
12	right_stance_percent	79.66	12	right_knee_flex_ext_Swing_Angle	63.62
13	left_toe_off_angle	61.76	13	left_hip_flex_ext_Swing_Angle	55.26
14	left_strike_angle	61.76	14	right_pelvis_rotation_Min	47.54
15	base_of_support_norm	55.47	15	left_pelvis_rotation_Max	42.06

Tanto para el listado de las variables espaciotemporales, como para las características cinemáticas, se han entrenado, ajustado y evaluado una serie de modelos alimentados con los 5,10 y 15 predictores más relevantes en cada caso. Además, se ha generado un tercer conjunto de variables explicativas combinando los dos anteriores (espaciotemporales + cinemáticas). De esta forma, se ha obtenido una subcolección de entradas formada por las 15 características representadas, por orden de importancia, en el diagrama de barras de la Figura 5.40.

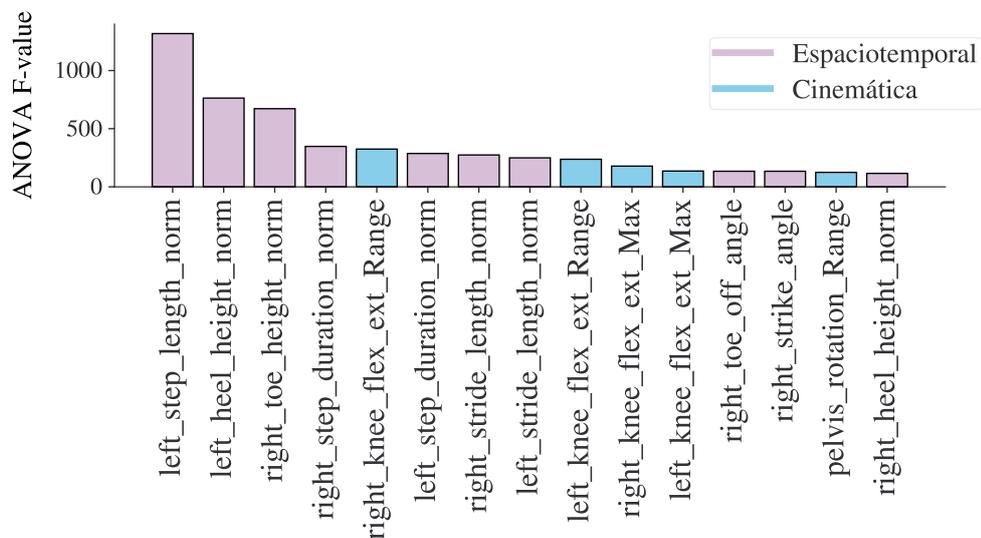


Figura 5.40: Representación de las 15 características predictoras más importantes (en base al valor F_1 de ANOVA) para el problema LAP1.

En la Tabla 5.15 se proporciona el rendimiento, estimado durante el entrenamiento con una validación cruzada, para cada algoritmo de explorado, tomando como entrada diferente número (5, 10 y 15) y tipo de variables predictoras (solo las espaciotemporales más relevantes, solo las cinemáticas más importantes, o una combinación de ambas). En cada caso se muestra la exactitud y tiempo de ajuste del mejor modelo encontrado durante el proceso de búsqueda exhaustiva. Por su parte, en la Tabla 5.16 se incluyen los resultados de evaluar cada uno de estos modelos sobre el conjunto de pruebas, utilizando las métricas de exactitud, F_1 y área bajo la curva ROC presentadas en la sección anterior.

Tabla 5.15: Rendimiento del entrenamiento, medido con la exactitud (%), y tiempo de entrenamiento (ms) promediado sobre una validación cruzada de 10 iteraciones, en el problema de determinar si el pie derecho sobrepasa al izquierdo (LAP1)

Algoritmo	Variables de entrada	Espaciotemporales		Cinemáticas		Espaciotemporales ± cinemáticas	
		Acc	Tiempo	Acc	Tiempo	Acc	Tiempo
Complement Naive Bayes	5	95.0 ± 10.0	4.8 ± 0.6	100	5.5 ± 3.7	100	3.6 ± 1.8
	10	100	5.3 ± 7.1	100	5.6 ± 0.5	100	5.8 ± 5.5
	15	100	6.4 ± 7.8	100	4.9 ± 4.7	100	5.4 ± 5.7
K-Nearest Neighbor	5	100	5.9 ± 3.1	100	6.1 ± 5.2	100	5.3 ± 0.8
	10	100	3.6 ± 5.6	100	6.0 ± 1.0	100	7.8 ± 7.8
	15	100	4.7 ± 7.2	100	3.1 ± 6.2	100	5.3 ± 7.0
Decision Tree	5	100	8.0 ± 4.0	97.5 ± 7.5	6.0 ± 0.5	100	5.0 ± 3.9
	10	97.5 ± 7.5	4.7 ± 7.2	100	8.2 ± 6.9	96.0 ± 8.0	6.2 ± 0.9
	15	97.5 ± 7.5	4.7 ± 7.2	100	6.1 ± 0.3	95.5 ± 9.1	2.6 ± 4.7
SVC	5	100	4.2 ± 4.1	100	8.1 ± 8.4	100	4.6 ± 4.7
	10	100	5.6 ± 4.3	100	6.4 ± 1.6	100	5.0 ± 3.9
	15	100	7.5 ± 5.3	100	7.7 ± 3.9	100	5.2 ± 4.1
Multilayer Perceptron	5	88.2 ± 12.1	37.8 ± 113.6	92.9 ± 13.7	41.5 ± 120.5	90.9 ± 17.6	38.8 ± 111.3
	10	97.3 ± 8.0	37.0 ± 105.9	97.1 ± 8.7	35.6 ± 106.8	95.3 ± 8.7	34.8 ± 99.2
	15	97.1 ± 8.7	35.4 ± 106.2	97.1 ± 8.7	34.7 ± 104.2	96.4 ± 8.7	33.5 ± 100.4

Tabla 5.16: Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC), expresadas como %, y medidas sobre el conjunto de pruebas para determinar si el pie derecho sobrepasa al izquierdo (LAP1). Se muestran remarcados los modelos seleccionados.

Algoritmo	Variables de entrada	Espaciotemporales			Cinemáticas			Espaciotemporales + cinemáticas		
		Acc	F_1	AUC	Acc	F_1	AUC	Acc	F_1	AUC
Complement Naive Bayes	5	91.30	87.50	93.75	100	100	100	100	100	100
	10	100	100	100	95.65	92.31	92.86	100	100	100
	15	95.65	93.33	96.88	100	100	100	100	100	100
K-Nearest Neighbor	5	95.65	93.33	96.88	100	100	100	100	100	100
	10	100	100	100	95.65	92.31	92.86	100.0	100	100
	15	95.65	93.33	96.88	100	100	100	100.0	100	100
Decision Tree	5	91.3	85.71	89.73	91.3	87.5	93.75	91.3	85.71	89.73
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100
SVC	5	100	100	100	95.65	92.31	92.86	100	100	100
	10	100	100	100	95.65	92.31	92.86	100	100	100
	15	100	100	100	95.65	92.31	92.86	100	100	100
Multilayer Perceptron	5	69.57	0.00	50.00	95.65	92.31	92.86	69.57	0.00	50.00
	10	91.30	83.33	85.71	95.65	92.31	92.86	69.57	0.00	50.00
	15	95.65	92.31	92.86	95.65	92.31	92.86	95.65	92.31	92.86

Sobre los resultados de evaluar los modelos en el conjunto de prueba (Tabla 5.16), se han seleccionado tres estimadores que resuelven, con una exactitud, sensibilidad y precisión del 100 %, y utilizando distintos tipos de variables explicativas, el problema de determinar si el pie derecho sobrepasa al izquierdo durante la fase de balanceo. En primer lugar, para el caso en el que se solo se consideran parámetros espaciotemporales de la marcha, se han obtenido varios modelos que ofrecen una exactitud, valor F_1 y AUC máximos. De entre todos estos candidatos, se ha seleccionado el estimador de SVC por ser el más parsimonioso, es decir, el que necesita un menor número de predictores (5) para ofrecer este rendimiento.

Por otro lado, cuando las características explicativas están basadas en la cinemática del movimiento, KNN y $Complement Naive Bayes$ son los estimadores que alcanzan una exactitud, F_1 y AUC igual al 100 % utilizando solo los cinco predictores más relevantes (Tabla 5.14). En este caso, se ha seleccionado el estimador basado en KNN porque es un método no paramétrico, a diferencia de los algoritmos de la familia $Naive Bayes$, que se basan en estrictas suposiciones sobre el conjunto de datos.

Por último, cuando los modelos se alimentan con una combinación de variables espaciotemporales y cinemáticas, tanto SVC como KNN y CNB alcanzan una mayor exactitud y F_1 (100 %) con un menor número de atributos de entrada. Aplicando el razonamiento anterior para seleccionar KNN antes que CNB , los candidatos se reducen a KNN y SVC . En este sentido, se ha seleccionado el estimador de SVC , ya que KNN es un modelo de aprendizaje basado en instancias (*lazy learning*), por lo que que necesita mayor memoria para almacenar los ejemplos de entrenamiento, y un coste computacional superior para generalizar a datos nuevos [164]. A modo de resumen, en la Tabla 5.17 se detallan las variables predictoras que alimentan los tres modelos que han sido seleccionados para abordar este primer ítem de la escala de Tinetti, así como los valores para sus hiperparámetros más relevantes: función de kernel, regularización C y γ para SVC , y número de vecinos más cercanos (k) y función de distancia (*metric*) para KNN .

Tabla 5.17: Modelos seleccionados para determinar si el pie derecho supera al izquierdo durante la fase de balanceo (LAP1)

Tipo de entradas	Predictores	Algoritmo	Hiperparámetros	Rendimiento
Espaciotemporales	left_step_length_norm left_heel_height_norm right_toe_height_norm right_step_duration_norm left_step_duration_norm	SVC	kernel = rbf C= 0.105 γ = 1.189	Acc = 100 % F_1 = 100 % AUC = 100 %
Cinemáticas	right_knee_flex_ext_Range left_knee_flex_ext_Range right_knee_flex_ext_Max left_knee_flex_ext_Max pelvis_rotation_Range	KNN	k=3 metric=euclidean	Acc = 100 % F_1 = 100 % AUC = 100 %
Espaciotemporales + cinemáticas	left_step_length_norm left_heel_height_norm right_toe_height_norm right_step_duration_norm right_knee_flex_ext_Range	SVC	kernel = rbf C= 0.105 γ = 1	Acc = 100 % F_1 = 100 % AUC = 100 %

La próxima cuestión de la prueba de Tinetti, modelada como una tarea de Aprendizaje Automático en la siguiente sección, está estrechamente relacionada con la que se acaba de resolver, pues el objetivo es el mismo pero para la extremidad opuesta.

5.3.4. LAP2. El pie izquierdo no sobrepasa al derecho en la fase de balanceo

El siguiente punto de la escala POMA evalúa si el pie izquierdo sobrepasa (1) o no (0) al pie derecho durante la marcha. Se trata, por tanto, de un problema estrechamente relacionado con el abordado en la sección anterior (LAP1), pero tomando como referencia la extremidad contraria. En este caso, el subconjunto de datos utilizado para la construcción de modelos predictivos que generen automáticamente las respuestas a dicha cuestión, está formado por las mismas 48 capturas de marcha normal, etiquetadas como 0, pero ahora importando como clase positiva las 20 grabaciones clasificadas como *lap2*. La división entre conjunto de entrenamiento (66 %) y pruebas (33 %) se ha llevado a cabo siguiendo el procedimiento descrito en el problema anterior.

Tal y como se muestra en la Tabla 5.18a, la selección de características basada en ANOVA indica que la longitud de paso derecho y, las alturas máximas del talón derecho y pie izquierdo durante la marcha, son las tres variables espaciotemporales más influyentes para determinar si el pie izquierdo sobrepasa al derecho al caminar. Estos resultados presentan una clara relación con los obtenidos para la valoración del ítem equivalente en la extremidad opuesta (LAP1), ya que coinciden los tres primeros parámetros, pero medidos sobre el pie contrario.

En cuanto a las características cinemáticas (Tabla 5.18b), las más importantes para este problema se corresponden con los rangos articulares de la flexión-extensión de ambas rodillas y cadera durante la zancada derecha. En la tarea anterior, el rango de flexión-extensión de las dos rodillas también eran los atributos más influyentes, pero en cambio, ahora se le da más importancia al referido a la pierna izquierda.

Tabla 5.18: Selección de características espaciotemporales (a) y cinemáticas (b) para el problema de determinar si el pie izquierdo sobrepasa al derecho (LAP2)

(a) Variables espaciotemporales

(b) Variables cinemáticas

#	Variable	F_1	#	Variable	F_1
1	right_step_length_norm	655.75	1	left_knee_flex_ext_Range	174.17
2	right_heel_height_norm	421.09	2	right_hip_flex_ext_Range	124.85
3	left_toe_height_norm	411.05	3	right_knee_flex_ext_Range	102.31
4	right_stride_length_norm	259.67	4	left_pelvis_rotation_Max	94.65
5	left_toe_off_angle	222.08	5	right_pelvis_rotation_Min	93.69
6	left_strike_angle	222.08	6	right_knee_flex_ext_Max	89.45
7	right_step_duration_norm	202.96	7	right_hip_flex_ext_Min	76.77
8	left_stride_length_norm	186.27	8	left_knee_flex_ext_Max	65.65
9	left_step_duration_norm	125.33	9	left_pelvis_rotation_Min	60.16
10	base_of_support_norm	123.35	10	right_pelvis_rotation_Stride_Start	53.04
11	left_heel_height_norm	68.86	11	right_hip_add_abd_Range	40.33
12	right_double_support_percent	66.74	12	left_ankle_flex_ext_Mean	38.59
13	left_swing_phase_percent	54.48	13	right_foot_rotation_Range	37.97
14	left_stance_percent	54.48	14	left_ankle_flex_ext_Min	36.56
15	left_double_support_percent	38.68	15	right_knee_rotation_Range	36.41

Combinando las variables espaciotemporales y cinemáticas se obtiene la lista ordenada de 15 predictores más relevantes representada en la Figura 5.41. Se observa que los 9 atributos más relacionados con la etiqueta de salida se corresponden con parámetros espaciotemporales de la marcha, de forma que el rango articular de la flexión extensión de rodilla izquierda, que es la variable cinemática más importante en esta tarea (según el ranking derivado del valor F_1 de ANOVA, mostrado en la Tabla 5.18b), ocupa la décima posición.

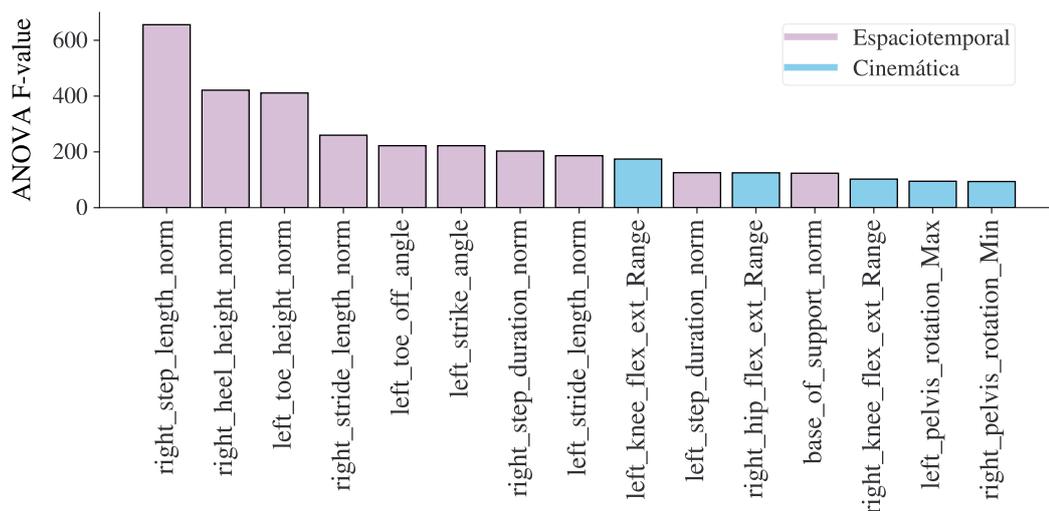


Figura 5.41: Representación de las 15 características predictoras más importantes (en base al valor F_1 de ANOVA) para el problema LAP2.

Por tanto, utilizando 5 variables predictoras solo se han entrenado los modelos basados en características espaciotemporales y cinemáticas de la marcha, ya que la combinación de ambas será equivalente a utilizar solo parámetros espaciotemporales. Para los casos en los que los modelos se alimentan con los 10 y 15 predictores más relevantes, sí que se han utilizado los tres subconjuntos de atributos: espaciotemporales, cinemáticas y ambas.

Los resultados del entrenamiento de los diferentes modelos para abordar esta tarea de clasificación binaria, afinados con una búsqueda exhaustiva para ajustar sus hiperparámetros, se proporcionan en la Tabla 5.15. En cada caso se incluye el tiempo de ajuste en milisegundos y el rendimiento (exactitud), promediados sobre una validación cruzada de 10 iteraciones. Asimismo, en la Tabla 5.20 se muestran las métricas de evaluación obtenidas sobre el conjunto de evaluación.

Tabla 5.19: Rendimiento del entrenamiento, medido con la exactitud (%), y tiempo de entrenamiento (ms) promediado sobre una validación cruzada de 10 iteraciones, en el problema de determinar si el pie izquierdo sobrepasa al derecho (LAP2)

Algoritmo	Variables de entrada	Espaciotemporales		Cinemáticas		Espaciotemporales ± cinemáticas	
		Acc	Tiempo	Acc	Tiempo	Acc	Tiempo
Complement Naive Bayes	5	55.5 ± 19.2	3.9 ± 4.5	100	4.6 ± 0.5	-	-
	10	100	6.3 ± 7.7	100	5.4 ± 0.5	100	4.6 ± 7.1
	15	100	5.0 ± 6.0	100	6.0 ± 0.4	100	6.0 ± 0.5
K-Nearest Neighbor	5	100	7.5 ± 4.0	100	5.2 ± 0.6	-	-
	10	100	5.6 ± 0.7	100	5.7 ± 0.6	100	4.7 ± 7.2
	15	100	6.1 ± 0.5	100	6.2 ± 0.6	100	6.1 ± 0.5
Decision Tree	5	100	4.7 ± 7.2	93.5 ± 10.0	2.6 ± 0.5	-	-
	10	95.0 ± 15.0	4.7 ± 7.2	93.5 ± 10.0	2.8 ± 0.4	95.5 ± 9.1	2.0 ± 3.6
	15	96.0 ± 8.0	1.6 ± 4.7	93.5 ± 10.0	2.0 ± 3.1	98.0 ± 6.0	2.9 ± 0.5
SVC	5	100	5.0 ± 4.3	100	4.2 ± 4.5	-	-
	10	100	7.1 ± 8.2	100	5.6 ± 0.5	100	5.0 ± 4.7
	15	100	8.0 ± 4.0	100	6.2 ± 0.4	100	8.5 ± 3.8
Multilayer Perceptron	5	94.4 ± 10.6	33.8 ± 101.3	80.4 ± 24.4	32.6 ± 94.1	-	-
	10	96.9 ± 7.9	37.3 ± 108.6	97.1 ± 8.7	38.2 ± 111.5	97.1 ± 8.7	36.2 ± 106.9
	15	97.6 ± 7.3	35.6 ± 104.0	96.4 ± 10.7	36.0 ± 102.8	97.3 ± 8.0	33.5 ± 97.2

Tabla 5.20: Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC), expresadas como %, y medidas sobre el conjunto de pruebas para determinar si el pie izquierdo sobrepasa al derecho (LAP2). Se muestran remarcados los modelos seleccionados.

Algoritmo	Variables de entrada	Espaciotemporales			Cinemáticas			Espaciotemporales + cinemáticas		
		Acc	F1	AUC	Acc	F1	AUC	Acc	F1	AUC
Complement Naive Bayes	5	56.52	44.44	56.7	100	100	100	-	-	-
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100
K-Nearest Neighbor	5	100	100	100	100	100	100	-	-	-
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100
Decision Tree	5	100	100	100	82.61	77.78	87.5	-	-	-
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	82.61	77.78	87.5	100	100	100
SVC	5	100	100	100	100	100	100	-	-	-
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100
Multilayer Perceptron	5	69.57	0.00	50.00	69.57	0.0	50.0	-	-	-
	10	86.96	72.73	78.57	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100

A partir de las métricas de evaluación obtenidas para los diferentes modelos, se ha seleccionado un estimador que toma únicamente entradas espaciotemporales y otro que se basa en los parámetros cinemáticos de la marcha. Tal y como se ha explicado anteriormente, el primero de ellos será también utilizado cuando se consideren como predictores la combinación de ambos tipos de variables, ya que para este problema, los 9 atributos más relevantes son parámetros espaciotemporales. En ambos casos, se obtiene una exactitud, precisión, sensibilidad y **AUC** igual al 100 %.

En primer lugar, para el caso en el que solo se consideran los parámetros espaciotemporales, se ha seleccionado el estimador basado en **SVC**. Los árboles decisión ofrecen un rendimiento similar con un mismo número de predictores (5), y son 'modelos de caja caja' fácilmente interpretables, pero, sin embargo, son más inestables que **SVC** ya que pequeñas variaciones en los datos de entrenamiento pueden conducir a la formación de un árbol completamente distinto. Por su parte, la elección entre **KNN** y el estimador de máquinas de vectores soporte se apoya sobre los criterios explicados para la tarea anterior: al ser un modelo basado en instancias, **KNN** necesita más memoria y tiempo para realizar las predicciones.

Por otro lado, utilizando como variables explicativas los atributos generados sobre los datos cinemáticos de la marcha, los modelos de **CNB**, **KNN** y **SVC** alcanzan una exactitud, F_1 y **AUC** del 100 % tomando como entrada las cinco características más influyentes (Tabla 5.18). Entre estos tres candidatos, se ha optado por el estimador de **SVC** siguiendo el mismo razonamiento que en el problema anterior (LAP1).

Para finalizar con este ítem de la prueba de Tinetti, en la Tabla 5.21 se resumen las variables predictoras que alimentan los dos modelos basados en máquinas de vectores soporte que han sido seleccionados, así como los valores para sus hiperparámetros más relevantes: función de kernel, regularización C y γ .

Tabla 5.21: Modelos seleccionados para determinar si el pie izquierdo supera al derecho durante la fase de balanceo (LAP2)

Tipo de entradas	Predictores	Algoritmo	Hiperparámetros	Rendimiento
Espaciotemporales	right_step_length_norm right_heel_height_norm	SVC	kernel = rbf C=0.125 γ =1.414	Acc = 100 % F_1 = 100 % AUC = 100 %
Espaciotemporales + cinemáticas	left_toe_height_norm right_stride_length_norm left_toe_off_angle			
Cinemáticas	left_knee_flex_ext_Range right_hip_flex_ext_Range right_knee_flex_ext_Range left_pelvis_rotation_Max right_pelvis_rotation_Min	SVC	kernel = rbf C= 9.514 γ =0.42	Acc = 100 % F_1 = 100 % AUC = 100 %

5.3.5. LAP3. El pie derecho no se levanta completamente

Continuando con la categoría que valora la longitud y altura del paso en la escala de Tinetti, la cuestión tratada en esta sección como un problema de clasificación binaria consiste en determinar si el pie derecho se levanta completamente con el paso durante la marcha (normal) o no (patológica). Para ello, se ha utilizado un subconjunto de datos que contiene las 48 grabaciones normales (*norm*), etiquetadas como 0, y las 20 capturas de movimiento en las que se ha simulado esta alteración de la marcha (*lap3*), etiquetadas como 1.

Tabla 5.22: Selección de características espaciotemporales (a) y cinemáticas (b) para el problema de determinar si el pie derecho se levanta completamente del suelo (LAP3)

(a) Variables espaciotemporales

(b) Variables cinemáticas

#	Variable	F_1	#	Variable	F_1
1	right_toe_height_norm	739.95	1	right_knee_flex_ext_Range	652.94
2	left_step_length_norm	396.87	2	right_knee_flex_ext_Max	495.11
3	left_heel_height_norm	275.16	3	right_knee_flex_ext_Swing_Angle	155.87
4	right_heel_height_norm	256.38	4	left_knee_flex_ext_Range	139.62
5	cadence_norm	120.16	5	left_knee_flex_ext_Max	94.98
6	left_stride_length_norm	116.04	6	left_pelvis_rotation_Stride_End	84.67
7	right_strike_angle	113.30	7	pelvis_rotation_Range	81.62
8	right_toe_off_angle	113.30	8	right_pelvis_rotation_Min	76.13
9	left_stride_duration	104.23	9	left_pelvis_rotation_Stride_Start	65.16
10	right_stride_length_norm	104.11	10	right_knee_val_var_Range	64.75
11	right_stride_duration	88.56	11	left_pelvis_rotation_Max	60.04
12	base_of_support_norm	76.66	12	right_pelvis_rotation_Max	55.30
13	left_toe_off_angle	47.00	13	right_pelvis_rotation_Stride_Start	53.84
14	left_strike_angle	47.00	14	right_hip_flex_ext_Range	50.25
15	right_double_support_percent	33.77	15	left_pelvis_rotation_Min	48.03

De acuerdo con la selección de características basada en ANOVA para este ítem de la prueba de Tinetti (Tabla 5.22a), la variable espaciotemporal más relevante es la altura máxima del pie derecho durante el ciclo de la marcha, y que tal y como se describió en la Sección 5.2.2, se corresponde con la posición más elevada del marcador colocado en los metatarsos del pie en una zancada. Este resultado está relacionado con algunas de las conclusiones que fueron deducidas a partir del análisis exploratorio previo (Sección 5.3.1), ya que como se

mostró en la Figura 5.35, en este tipo de capturas de marcha se produce una importante disminución del valor medio de este atributo. Además, entre los parámetros espaciales más significativos aparece también la altura máxima de ambos talones, así como la longitud del paso izquierdo, que en este caso, es la extremidad opuesta a la analizada. Este listado de variables predictoras es parecido al obtenido en el problema de determinar si el pie derecho sobrepasa al izquierdo (Sección 5.3.3), pero, en este caso, se le da más importancia a la altura del pie que a la longitud del paso.

Por otro lado, en lo que se refiere a parámetros cinemáticos, los resultados mostrados en la Tabla 5.22b sugieren que la amplitud de la flexión-extensión de la rodilla derecha es la variable más relacionada con la etiqueta *lap3* asignada a la marcha. En particular, los cinco atributos cinemáticos más relevantes se corresponden con diferentes medidas de la flexión-extensión de ambas rodillas: el rango angular, el valor máximo y el ángulo de flexión/extensión de la rodilla analizada cuando comienza la fase de balanceo (*right_knee_flex_ext_Range*). De nuevo, estos resultados son similares a los obtenidos en el problema LAP1, pero ahora, otorgando más importancia al ángulo máximo de flexión en la rodilla analizada que a la amplitud angular de la extremidad opuesta.

Tal y como se muestra en el gráfico de la Figura 5.42, cuando se combinan las propiedades cinemáticas y espaciotemporales en un único conjunto de predictores, las cinco variables más relevantes para determinar si el pie derecho se levanta completamente del suelo son, en este orden, la altura máxima del pie derecho, la amplitud y ángulo máximo de la flexión-extensión de rodilla derecha y, la longitud del paso y altura máxima del talón izquierdo durante la marcha. Siguiendo el mismo procedimiento que en los problemas anteriores, se han entrenado, ajustado y evaluado una serie de modelos de Aprendizaje Automático que toman como entrada los 5,10 y 15 predictores más relevantes en cada caso.

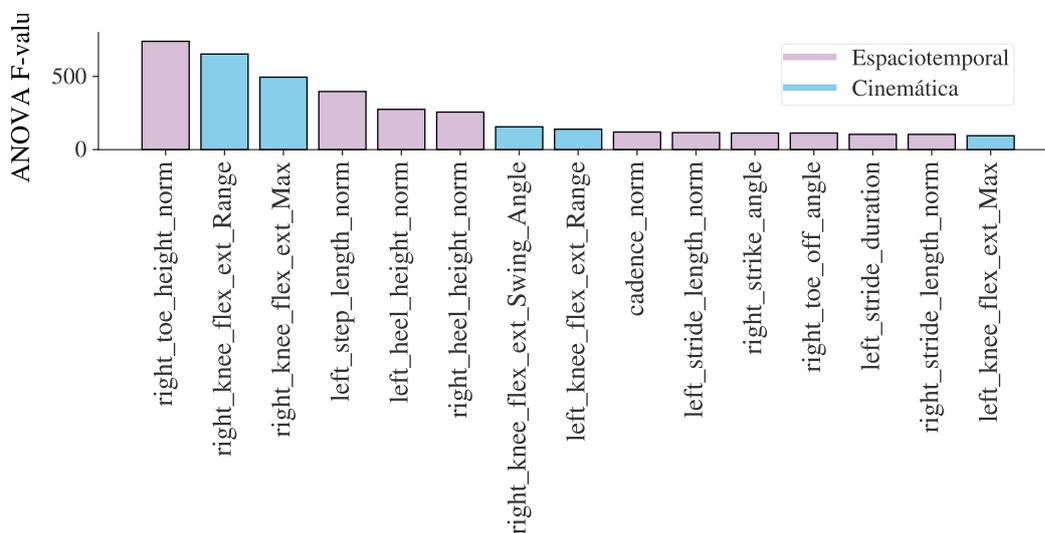


Figura 5.42: Representación de las 15 características predictoras más importantes (en base al valor F_1 de ANOVA) para el problema LAP3.

Los resultados del entrenamiento se proporcionan en la Tabla 5.23, incluyendo la exactitud y tiempo de ajuste promediados sobre una validación cruzada de 10 iteraciones. En cada caso, el resultado mostrado es el correspondiente al modelo con el mejor ajuste de parámetros encontrado mediante un proceso de búsqueda exhaustiva. Por otro lado, cada uno de estos modelos han sido evaluados sobre el conjunto de pruebas, obteniendo las métricas de exactitud, F_1 y área bajo la curva ROC que se muestran en la Tabla 5.24.

Tabla 5.23: Rendimiento del entrenamiento, medido con la exactitud (%), y tiempo de entrenamiento (ms) promediado sobre una validación cruzada de 10 iteraciones, en el problema de determinar si el pie derecho se levanta completamente (LAP3).

Algoritmo	Variables de entrada	Espaciotemporales		Cinemáticas		Espaciotemporales ± cinemáticas	
		Acc	Tiempo	Acc	Tiempo	Acc	Tiempo
Complement Naive Bayes	5	39.5 ± 23.0	5.2 ± 5.5	57.5 ± 16.2	4.1 ± 1.5	50.5 ± 19.9	4.0 ± 1.4
	10	100	6.3 ± 7.7	100	4.2 ± 4.4	49.5 ± 21.8	5.1 ± 5.7
	15	98.0 ± 6.0	6.9 ± 7.4	100	6.3 ± 4.1	100	3.9 ± 4.5
K-Nearest Neighbor	5	100	6.6 ± 8.1	100	3.1 ± 6.3	100	0.0 ± 0.0
	10	100	3.8 ± 6.2	100	5.6 ± 6.8	100	8.3 ± 7.5
	15	100	6.2 ± 3.8	98.0 ± 6.0	9.4 ± 7.7	100	6.0 ± 0.4
Decision Tree	5	100	4.7 ± 7.2	100	1.6 ± 4.7	100	3.0 ± 3.8
	10	97.5 ± 7.5	1.6 ± 4.7	100	4.7 ± 7.2	98.0 ± 6.0	4.7 ± 7.2
	15	100	5.3 ± 6.0	98.0 ± 6.0	4.1 ± 6.5	98.0 ± 6.0	1.8 ± 3.2
SVC	5	100	4.8 ± 4.1	100	6.4 ± 5.9	100	6.0 ± 6.2
	10	100	6.1 ± 1.0	100	6.9 ± 3.7	100	5.0 ± 3.4
	15	98.0 ± 6.0	8.1 ± 3.7	98.0 ± 6.0	6.3 ± 4.5	100	7.5 ± 5.5
Multilayer Perceptron	5	93.3 ± 10.1	36.3 ± 103.9	92.7 ± 10.1	41.0 ± 119.9	92.2 ± 9.5	41.3 ± 118.9
	10	90.0 ± 19.0	36.8 ± 110.5	97.1 ± 8.7	36.1 ± 104.8	88.0 ± 13.1	33.8 ± 96.3
	15	97.1 ± 8.7	33.2 ± 99.6	94.9 ± 8.8	34.7 ± 104.2	93.6 ± 9.1	34.5 ± 101.0

Tabla 5.24: Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC), expresadas como %, y medidas sobre el conjunto de pruebas para determinar si el pie derecho se levanta completamente (LAP3). Se muestran remarcados los modelos seleccionados.

Algoritmo	Variables de entrada	Espaciotemporales			Cinemáticas			Espaciotemporales + cinemáticas		
		Acc	F_1	AUC	Acc	F_1	AUC	Acc	F_1	AUC
Complement Naive Bayes	5	60.87	52.63	63.84	65.22	50.0	62.95	47.83	40.00	50.45
	10	100	100	100	100	100	100	73.91	57.14	69.20
	15	100	100	100	100	100	100	100	100	100
K-Nearest Neighbor	5	100	100	100	100	100	100	100	100	100
	10	100	100	100	95.65	93.33	96.88	100	100	100
	15	95.65	93.33	96.88	100	100	100	100	100	100
Decision Tree	5	100	100	100	100	100	100	100	100	100
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100
SVC	5	100	100	100	100	100	100	100	100	100
	10	100	100	100	100	100	100	100	100	100
	15	95.65	92.31	92.86	95.65	92.31	92.86	100	100	100
Multilayer Perceptron	5	69.57	0.0	50.0	100	100	100	69.57	0.0	50.0
	10	100	100	100	100	100	100	69.57	0.0	50.0
	15	100	100	100	86.96	72.73	78.57	69.57	0.0	50.0

Para esta tarea, los estimadores basados en **KNN**, **SVC** y **DT** ofrecen una exactitud, precisión, sensibilidad y **AUC** del 100 % para los tres casos en los que se utilizan cinco variables predictoras: solo parámetros espaciotemporales, solo cinemáticos y una combinación de ambas. Por lo tanto, siguiendo la misma justificación que en los problemas anteriores, se ha optado por seleccionar los tres modelos de **SVC**. No obstante, puesto que los árboles de decisión también alcanzan una exactitud máxima, es interesante analizar la estructura de árbol basada en reglas condicionales que ha sido construida durante el entrenamiento. Como ejemplo, en la Figura 5.43 se muestra el árbol de decisión que determina si el pie derecho se levanta completamente del suelo utilizando solo características espaciotemporales.

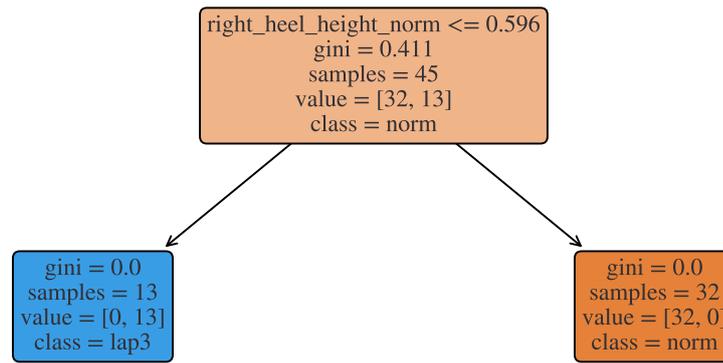


Figura 5.43: Árbol de decisión para el problema LAP3 con parámetros espaciotemporales.

Se trata de un árbol con profundidad uno, por lo que la decisión de asignar una etiqueta u otra se fundamenta en una única propiedad: la altura máxima del talón derecho durante la marcha. Más concretamente, el árbol de decisión ha determinado que 0.596 es el valor de la altura de talón que actúa como umbral para decidir si una captura de movimiento se etiqueta como marcha normal o *lap3* (no se levanta el pie derecho). En este sentido, hay que tener en cuenta que se trata de un atributo normalizado por la longitud del pie, tal y como se explicó en la Sección 5.2.4. Por tanto, si se deshace esta normalización para una persona con una longitud de pie de, por ejemplo, 27 centímetros, según este árbol de decisión se determinaría que no se levanta el pie completamente del suelo si la altura máxima del talón durante la marcha es igual o inferior a 16.092 cm.

Como resumen, en la Tabla 5.24 se describen las características de los tres modelos SVC que han sido seleccionados para determinar si el pie derecho se levanta completamente del suelo durante la marcha. Se incluye el listado de atributos que toman como entrada y el ajuste de sus hiperparámetros más relevantes. En la siguiente sección se trabaja sobre un ítem muy similar al que se acaba de explicar, ya que el objetivo será determinar si el pie izquierdo se levanta, o no, completamente del suelo.

Tabla 5.25: Modelos seleccionados para determinar si el pie derecho se levanta completamente del suelo (LAP3)

Tipo de entradas	Predictores	Algoritmo	Hiperparámetros	Rendimiento
Espaciotemporales	right_toe_height_norm left_step_length_norm left_heel_height_norm right_heel_height_norm cadence_norm	SVC	kernel = rbf C= 0.125 $\gamma = 1$	Acc = 100 % $F_1 = 100 %$ AUC = 100 %
Cinemáticas	right_knee_flex_ext_Range right_knee_flex_ext_Max right_knee_flex_ext_Swing_Angle left_knee_flex_ext_Range left_knee_flex_ext_Max	SVC	kernel = rbf C= 0.177 $\gamma = 1$	Acc = 100 % $F_1 = 100 %$ AUC = 100 %
Espaciotemporales + cinemáticas	right_toe_height_norm right_knee_flex_ext_Range right_knee_flex_ext_Max left_step_length_norm left_heel_height_norm	SVC	kernel = rbf C= 0.125 $\gamma = 1$	Acc = 100 % $F_1 = 100 %$ AUC = 100 %

5.3.6. LAP4. El pie izquierdo no se levanta completamente

En este último ítem perteneciente a la categoría de longitud y altura del paso en la escala de Tinetti, el objetivo es determinar automáticamente si el pie izquierdo se levanta completamente del suelo con el paso. Se trata, por tanto, de un problema de clasificación binaria similar al resuelto en la sección anterior, pero ahora, analizando el movimiento de la pierna opuesta. Para ello, se ha utilizado como subconjunto de datos las 48 muestras de marcha normal, en las que el pie izquierdo sí se levanta completamente del suelo, más las 20 capturas en las que se ha simulado esta patología (*lap4*), etiquetadas como clase positiva (1). La división de este subconjunto de datos en entrenamiento y pruebas se ha realizado, al igual que en los casos anteriores, según lo descrito en la Sección 5.3.2: un 66 % de los datos para la construcción de modelos y el 33 % restante para la evaluación.

Para este ítem, la selección de características basada en ANOVA propone que el parámetro espaciotemporal más relacionado con la variable respuesta es la altura máxima del pie izquierdo durante el ciclo de la marcha (Tabla 5.26). Este resultado coincide con el obtenido en el problema anterior (LAP3), ya que para determinar si el pie derecho se levanta completamente del suelo, el atributo espaciotemporal más importante era la altura máxima de ese mismo pies. Además, al igual que en el caso anterior, también está bastante relacionada con la etiqueta objetivo la longitud de paso para la extremidad opuesta. Sin embargo, en esta tarea aparecen entre los cinco atributos más relevantes los ángulos de despegue y contacto del pie izquierdo, relacionados con el ángulo que se forma entre ese pie y el suelo en los instantes de contacto inicial y comienzo de la fase de balanceo.

En cuanto a los parámetros relacionados con la cinemática del movimiento, la amplitud y el máximo valor angular de la flexión-extensión de rodilla izquierda son las características más relevantes para determinar si el pie de esa pierna se levanta completamente del suelo (Tabla 5.26b). Además, al igual que en la tarea anterior (LAP3), también son importantes estos mismos atributos pero referidos a la extremidad opuesta a la analizada.

Tabla 5.26: Selección de características espaciotemporales (a) y cinemáticas (b) para el problema de determinar si el pie izquierdo se levanta completamente del suelo (LAP4)

(a) Variables espaciotemporales

(b) Variables cinemáticas

#	Variable	F_1	#	Variable	F_1
1	left_toe_height_norm	599.82	1	left_knee_flex_ext_Range	452.45
2	right_step_length_norm	279.42	2	left_knee_flex_ext_Max	265.09
3	left_toe_off_angle	274.76	3	right_knee_flex_ext_Range	89.09
4	left_strike_angle	274.76	4	right_knee_flex_ext_Max	63.73
5	left_heel_height_norm	202.94	5	left_knee_val_var_Range	56.67
6	right_heel_height_norm	167.73	6	left_pelvis_rotation_Max	56.40
7	cadence_norm	93.71	7	left_pelvis_rotation_Min	54.75
8	right_stride_length_norm	87.21	8	left_knee_val_var_Min	53.73
9	right_stride_duration	84.13	9	right_pelvis_rotation_Min	49.34
10	left_stride_duration	83.99	10	left_pelvis_rotation_Swing_Angle	47.17
11	left_stride_length_norm	80.39	11	right_pelvis_rotation_Stride_Start	43.47
12	right_swing_phase_percent	47.00	12	right_hip_flex_ext_Min	40.36
13	right_stance_percent	47.00	13	right_ankle_flex_ext_Swing_Angle	32.65
14	right_double_support_percent	38.58	14	right_pelvis_rotation_Max	31.60
15	right_toe_off_angle	36.69	15	left_pelvis_rotation_Stride_End	31.48

El listado que combina las variables espaciotemporales y cinemáticas más relevantes

queda definido según lo mostrado en el gráfico de la Figura 5.44. Por tanto, para este ítem, las cinco variables más relevantes, teniendo en cuenta ambos tipos de atributos, son la altura máxima del pie izquierdo, el rango de amplitud angular de la flexión-extensión de rodilla izquierda, la longitud normalizada del paso derecho, y los ángulos de despegue y contacto del pie izquierdo.

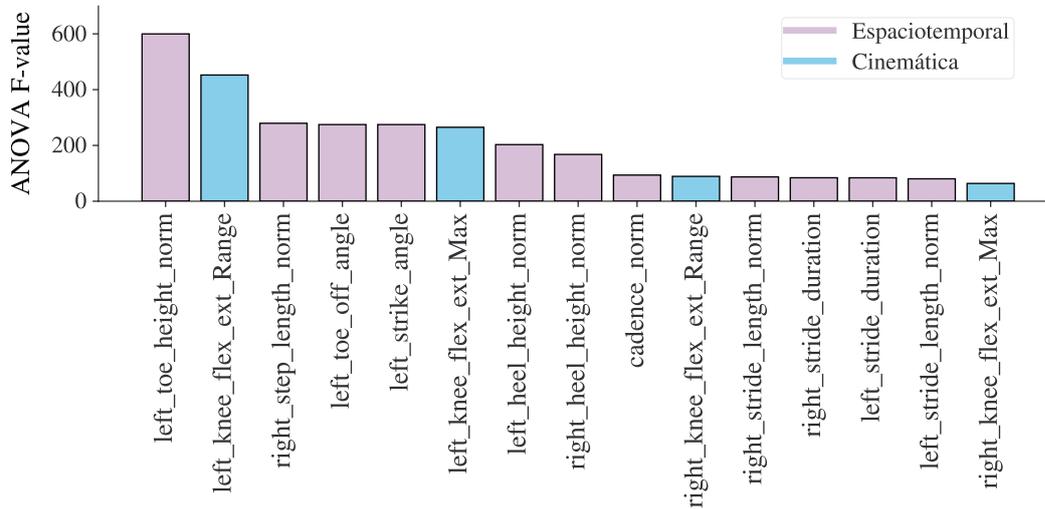


Figura 5.44: Representación de las 15 características predictoras más importantes (en base al valor F_1 de ANOVA) para el problema LAP4.

En la Tabla 5.27 se muestran los resultados del rendimiento y tiempo de ajuste, estimados durante el entrenamiento con una validación cruzada, para cada algoritmo entrenado con diferente número y tipo de variables predictoras. Por su parte, en la Tabla 5.28 se proporcionan las métricas de evaluación obtenidas al utilizar cada uno de estos modelos con datos desconocidos.

Tabla 5.27: Rendimiento del entrenamiento, medido con la exactitud (%), y tiempo de entrenamiento (ms) promediado sobre una validación cruzada de 10 iteraciones, en el problema de determinar si el pie derecho se levanta completamente (LAP4).

Algoritmo	Variables de entrada	Espaciotemporales		Cinemáticas		Espaciotemporales ± cinemáticas	
		Acc	Tiempo	Acc	Tiempo	Acc	Tiempo
Complement Naive Bayes	5	100	5.2 ± 5.1	65.5 ± 15.7	4.7 ± 4.1	98.0 ± 6.0	4.8 ± 0.4
	10	100	9.5 ± 0.9	100	3.1 ± 6.3	100	4.1 ± 4.7
	15	100	5.3 ± 7.0	98.0 ± 6.0	5.6 ± 5.7	100	6.0 ± 0.6
K-Nearest Neighbor	5	100	9.6 ± 11.9	100	4.7 ± 0.5	100	5.4 ± 0.7
	10	100	3.2 ± 4.1	100	4.7 ± 7.2	100	4.0 ± 6.1
	15	100	6.1 ± 0.7	98.0 ± 6.0	6.0 ± 6.8	100	6.1 ± 0.3
Decision Tree	5	100	2.4 ± 3.9	100	2.4 ± 3.5	100	3.1 ± 6.3
	10	95.5 ± 9.1	3.2 ± 3.2	96.0 ± 8.0	4.7 ± 7.2	100	2.7 ± 0.5
	15	100	0.0 ± 0.0	100	3.1 ± 6.3	95.5 ± 9.1	1.8 ± 2.8
SVC	5	100	5.9 ± 4.1	100	11.5 ± 11.7	100	4.7 ± 4.0
	10	100	5.1 ± 2.3	100	5.8 ± 3.5	100	8.3 ± 8.3
	15	100	13.8 ± 16.3	98.0 ± 6.0	4.6 ± 4.3	100	5.9 ± 3.9
Multilayer Perceptron	5	92.7 ± 12.8	40.8 ± 119.4	93.1 ± 10.0	40.1 ± 118.1	92.4 ± 13.9	36.4 ± 109.2
	10	97.3 ± 8.0	35.4 ± 104.0	96.4 ± 8.5	36.9 ± 105.5	95.3 ± 8.7	35.8 ± 104.1
	15	96.9 ± 9.3	36.8 ± 101.1	96.7 ± 8.6	36.3 ± 102.9	96.2 ± 8.4	37.0 ± 105.8

Tabla 5.28: Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC), expresadas como %, y medidas sobre el conjunto de pruebas para determinar si el pie izquierdo se levanta completamente (LAP4). Se muestran remarcados los modelos seleccionados.

Algoritmo	Variables de entrada	Espaciotemporales			Cinemáticas			Espaciotemporales + cinemáticas		
		Acc	F_1	AUC	Acc	F_1	AUC	Acc	F_1	AUC
Complement Naive Bayes	5	100	100	100	43.48	31.58	43.3	91.3	87.5	93.75
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100
K-Nearest Neighbor	5	100	100	100	100	100	100	100	100	100
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100
Decision Tree	5	91.3	87.5	93.75	100	100	100	91.3	87.5	93.75
	10	91.3	87.5	93.75	91.3	87.5	93.75	100	100	100
	15	100	100	100	91.3	87.5	93.75	100	100	100
SVC	5	100	100	100	100	100	100	100	100	100
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	95.65	92.31	92.86	100	100	100
Multilayer Perceptron	5	100	100	100	69.57	0.00	50.00	95.65	92.31	92.86
	10	100	100	100	95.65	92.31	92.86	69.57	0.00	50.00
	15	100	100	100	95.65	92.31	92.86	69.57	0.00	50.00

Tal y como se muestra en los resultados obtenidos al evaluar los modelos en el conjunto de prueba (Tabla 5.28), de nuevo ha sido posible seleccionar tres estimadores que, tomando como entrada diferentes tipos de variables predictoras, resuelven con una exactitud, F_1 y AUC del 100 %, el problema de determinar si el pie izquierdo se levanta completamente del suelo durante la marcha. Para el caso en el que solo se consideran cinco atributos espaciotemporales, hasta cuatro de los modelos ajustados alcanzan las máximas puntuaciones en las métricas de evaluación: CNB, KNN, SVC y, por primera vez con este número de entradas, el modelo basado en redes neuronales (MLP). En particular, la afinación de hiperparámetros realizada mediante *keras-tuner* ha conducido a una arquitectura de MLP con dos capas profundas formadas, respectivamente, por 20 y 35 unidades neuronales. Por tanto, teniendo en cuenta las variables explicativas consideradas y lo descrito en la Sección 5.3.2 para los modelos de MLP, la estructura de red neuronal construida tiene el aspecto ilustrado gráficamente en la Figura 5.45.

No obstante, se ha optado por seleccionar el modelo de SVC utilizando cinco variables espaciotemporales, ya que, aunque en esta tarea MLP ofrece un rendimiento similar, la arquitectura neuronal es mucho más compleja. Por tanto, en general los modelos de SVC son más eficientes en las predicciones [165].

Por otro lado, para el caso en el que toman como entradas solo atributos relacionados con la cinemática del movimiento, los modelos de KNN, árboles de decisión (DT) y SVC, ofrecen un rendimiento similar utilizando solo cinco predictores. Por su parte, cuando se combinan atributos espaciotemporales y cinemáticos, son los estimadores basados en KNN y máquinas de vectores soporte los que alcanzan la exactitud, F_1 y AUC máximos. En ambos casos, se ha seleccionado el modelo de SVC siguiendo el razonamiento expuesto en los problemas anteriores: KNN es un aprendizaje basado en instancias, por lo que las predicciones requieren más memoria y coste computacional, mientras que los árboles de decisión son más inestables que SVC a pequeños cambios en el conjunto de datos.

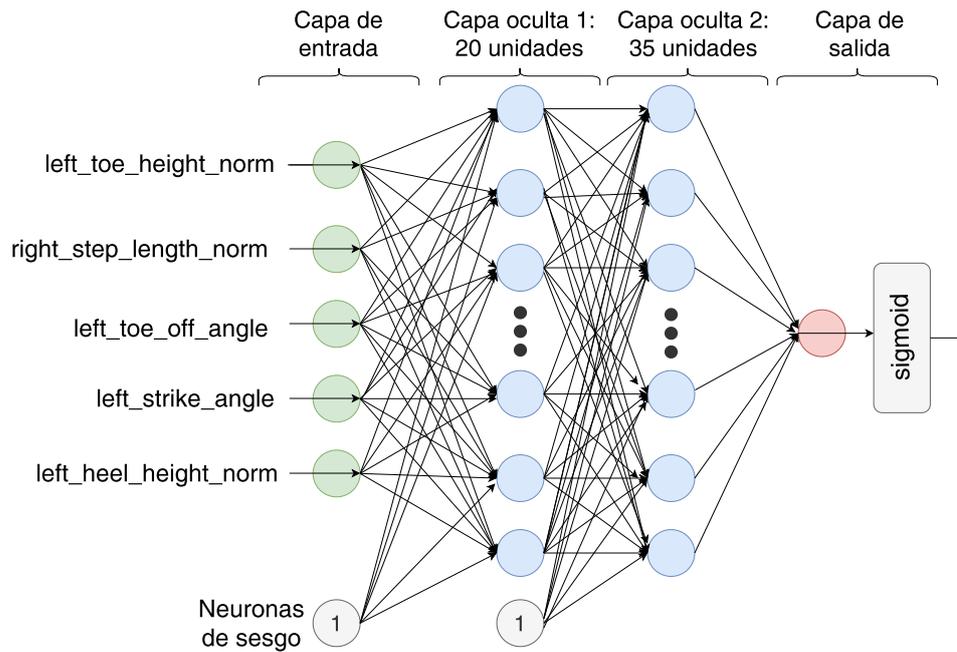


Figura 5.45: Arquitectura de PMC para el problema de clasificación LAP4

Finalmente, en la Tabla 5.29 se muestra un resumen de los tres modelos *SVC* seleccionados para determinar si el pie izquierdo se levanta completamente del suelo durante la marcha, utilizado como predictores cinco atributos espaciotemporales, cinemáticos y una combinación de ambos.

Tabla 5.29: Modelos seleccionados para determinar si el pie izquierdo se levanta completamente (LAP4)

Tipo de entradas	Predictores	Algoritmo	Hiperparámetros	Rendimiento
Espaciotemporales	left_toe_height_norm right_step_length_norm left_toe_off_angle left_strike_angle left_heel_height_norm	SVC	kernel = rbf C = 0.125 $\gamma = 1.414$	Acc = 100 % $F_1 = 100 %$ AUC = 100 %
Cinemáticas	left_knee_flex_ext_Range left_knee_flex_ext_Max right_knee_flex_ext_Range right_knee_flex_ext_Max left_knee_val_var_Range	SVC	kernel = rbf C = 0.707 $\gamma = 1.414$	Acc = 100 % $F_1 = 100 %$ AUC = 100 %
Espaciotemporales + cinemáticas	left_toe_height_norm left_knee_flex_ext_Range right_step_length_norm left_toe_off_angle left_strike_angle	SVC	kernel = rbf C = 0.125 $\gamma = 1.682$	Acc = 100 % $F_1 = 100 %$ AUC = 100 %

5.3.7. DC. Discontinuidad entre pasos

La siguiente cuestión abordada en la escala de valoración de la marcha de Tinetti está relacionada con la fluidez de los pasos: asigna una puntuación de 1 (anormal) cuando el paciente adopta una marcha discontinua e interrumpida, en la que varía el largo del

paso o en la que se apoya completamente un pie antes de levantar el otro. Por tanto, se abordará como una nueva tarea supervisada de clasificación binaria, en la que el objetivo es determinar automáticamente si una persona camina de forma discontinua (1) o no (0) a partir de parámetros espaciotemporales y/o cinemáticos de la marcha. Para ello, se ha utilizado un subconjunto del *dataset* formado por las 48 instancias de marcha normal más las 20 en las que se ha simulado una marcha discontinua e interrumpida (*dc*), etiquetadas como clase positiva (1). Al igual que en los problemas anteriores, este subconjunto de datos se ha dividido en dos partes estratificadas, una para el entrenamiento (66 %) y otra para la evaluación (33 %).

Tal y como se muestra en la Tabla 5.30a entre las variables espaciotemporales más relevantes se encuentran características relacionadas tanto con la altura y longitud del paso como con la distribución porcentual de la fase de doble apoyo dentro del ciclo de la marcha: longitud de zancada y altura del pie derecho, porcentaje de apoyo bipodal en la zancada izquierda y altura máxima de ese mismo pie. Además, la cadencia, que representa el número de pasos ejecutados por unidad de tiempo, también aparece entre las cinco características que mejor explican si se trata de una marcha discontinua o interrumpida. Por su parte, las características cinemáticas más influyentes en este problema (Tabla 5.30b), según la selección de características basada en ANOVA, están relacionadas con el movimiento de rotación de la pelvis. En particular, entre las cinco variables explicativas más relevantes para esta tarea de clasificación, se encuentran el rango angular y los valores mínimos de rotación pélvica en ambas zancadas, así como el máximo ángulo de rotación de la pelvis durante la zancada izquierda.

Tabla 5.30: Selección de características espaciotemporales (a) y cinemáticas (b) para el problema de valorar si se trata de una marcha discontinua o interrumpida (DC)

(a) Variables espaciotemporales			(b) Variables cinemáticas		
#	Variable	F_1	#	Variable	F_1
1	right_stride_length_norm	356.78	1	pelvis_rotation_Range	245.36
2	right_toe_height_norm	283.46	2	left_pelvis_rotation_Min	171.70
3	cadence_norm	279.58	3	left_pelvis_rotation_Max	170.29
4	left_double_support_percent	277.13	4	right_pelvis_rotation_Min	168.68
5	left_toe_height_norm	235.38	5	left_knee_flex_ext_Range	142.19
6	right_heel_height_norm	227.26	6	right_knee_flex_ext_Range	139.56
7	left_stride_length_norm	216.92	7	right_pelvis_rotation_Max	127.13
8	left_swing_phase_percent	200.04	8	right_hip_flex_ext_Range	116.55
9	left_stance_percent	200.04	9	left_knee_flex_ext_Max	104.27
10	right_step_length_norm	197.29	10	right_knee_flex_ext_Max	101.78
11	left_heel_height_norm	189.55	11	right_ankle_rotation_Range	67.28
12	left_step_length_norm	179.50	12	right_knee_rotation_Range	60.77
13	right_swing_phase_percent	173.25	13	right_foot_rotation_Range	59.03
14	right_stance_percent	173.25	14	left_pelvis_rotation_Stride_End	55.39
15	left_toe_off_angle	140.36	15	right_pelvis_rotation_Stride_Start	51.83

Combinando y ordenando los atributos relacionados con los parámetros espaciotemporales y cinemáticos de la marcha, se obtiene el listado de características más relevantes que se muestra en el diagrama de barras de la Figura 5.46. En este caso, se observa que entre las 15 variables explicativas que mejor separan entre los dos tipos de marcha solo aparece un atributo cinemático: el rango angular del movimiento de rotación de la pelvis. No obstante, esta variable cinemática se encuentra entre las primeras 5 características, por

lo que, nuevamente, se han entrenado varios modelos alimentados con las 5,10 y 15 variables más importantes en cada uno de estos tres subconjuntos de predictores.

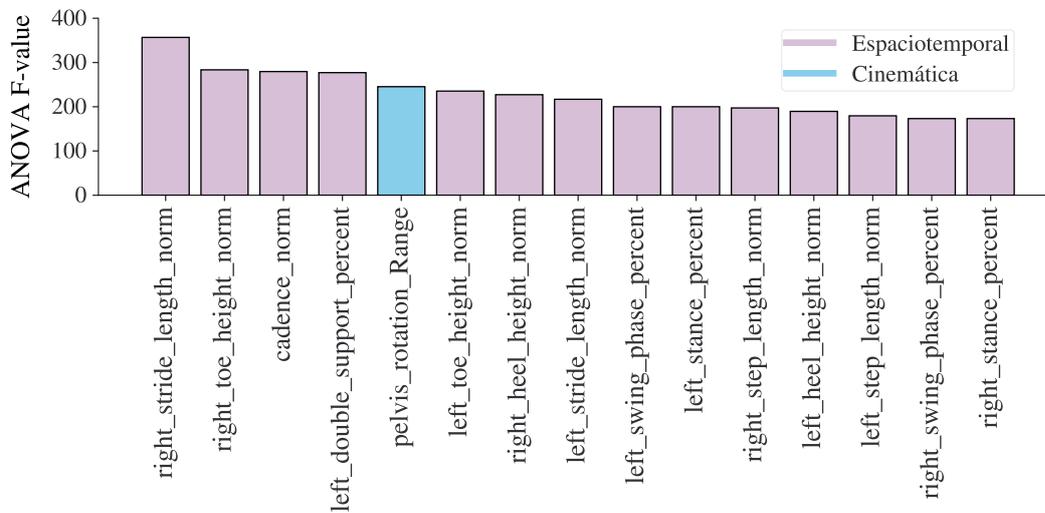


Figura 5.46: Representación de las 15 características predictoras más importantes (en base al valor F_1 de ANOVA) para el problema DC.

Los resultados obtenidos durante el entrenamiento de los diferentes modelos se proporcionan en la Tabla 5.31, incluyendo el rendimiento medido con la exactitud y el tiempo de ajuste, promediados sobre una validación cruzada de 10 iteraciones. Por su parte, en la Tabla 5.32 se incluyen las puntuaciones de exactitud, F_1 y AUC obtenidas al evaluar cada uno de estos modelos en el conjunto de pruebas.

Tabla 5.31: Rendimiento del entrenamiento, medido con la exactitud (%), y tiempo de entrenamiento (ms) promediado sobre una validación cruzada de 10 iteraciones, en el problema de determinar si se trata de una marcha discontinua (DC)

Algoritmo	Variables de entrada	Espaciotemporales		Cinemáticas		Espaciotemporales ± cinemáticas	
		Acc	Tiempo	Acc	Tiempo	Acc	Tiempo
Complement Naive Bayes	5	100	4.8 ± 0.4	96.0 ± 8.0	4.7 ± 0.5	100	4.7 ± 0.6
	10	98.0 ± 6.0	5.3 ± 0.5	100	5.0 ± 0.0	100	5.1 ± 0.3
	15	100	6.0 ± 0.6	100	6.0 ± 0.6	96.0 ± 8.0	5.8 ± 0.4
K-Nearest Neighbor	5	100	5.5 ± 0.5	100	5.0 ± 0.6	100	4.8 ± 0.9
	10	100	5.6 ± 0.5	100	5.6 ± 0.5	100	5.1 ± 0.3
	15	100	5.7 ± 0.5	100	6.1 ± 0.5	100	5.9 ± 0.5
Decision Tree	5	100	3.2 ± 0.6	95.5 ± 9.1	2.7 ± 0.5	100	2.5 ± 0.5
	10	97.5 ± 7.5	3.1 ± 6.3	93.5 ± 10.0	2.8 ± 0.4	97.5 ± 7.5	2.7 ± 0.5
	15	95.5 ± 9.1	2.5 ± 0.5	98.0 ± 6.0	2.8 ± 0.4	95.5 ± 9.1	2.7 ± 0.5
SVC	5	100	5.1 ± 0.5	100	8.4 ± 10.2	100	5.7 ± 2.1
	10	100	5.6 ± 0.5	100	5.5 ± 0.7	100	5.3 ± 0.5
	15	100	6.5 ± 0.5	100	6.3 ± 0.6	100	7.2 ± 3.3
Multilayer Perceptron	5	92.9 ± 13.0	43.7 ± 128.0	99.1 ± 2.7	36.2 ± 105.6	93.3 ± 12.6	37.3 ± 108.6
	10	97.6 ± 7.3	34.8 ± 100.4	99.8 ± 0.7	37.5 ± 108.6	90.2 ± 11.4	37.1 ± 107.6
	15	89.3 ± 9.0	37.4 ± 108.5	97.3 ± 8.0	35.4 ± 103.5	88.2 ± 15.2	37.0 ± 107.7

Tal y como muestran los resultados de la evaluación (Tabla 5.32), los estimadores basados en KNN, CNB, DT y SVC ofrecen un rendimiento similar utilizando solo cinco predictores espaciotemporales: 100% de exactitud, F_1 y AUC. Por tanto, siguiendo los criterios de selección descritos para los problemas anteriores, se ha escogido el modelo de SVC. Estos

Tabla 5.32: Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC), expresadas como %, y medidas sobre el conjunto de pruebas para determinar si se trata de una marcha discontinua (DC). Se muestran remarcados los modelos seleccionados.

Algoritmo	Variables de entrada	Espaciotemporales			Cinemáticas			Espaciotemporales + cinemáticas		
		Acc	F_1	AUC	Acc	F_1	AUC	Acc	F_1	AUC
Complement Naive Bayes	5	100	100	100	91.3	87.5	93.75	100	100	100
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100
K-Nearest Neighbor	5	100	100	100	95.65	92.31	92.86	100	100	100
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100
Decision Tree	5	100	100	100	91.30	85.71	89.73	100	100	100
	10	95.65	92.31	92.86	95.65	92.31	92.86	95.65	92.31	92.86
	15	100	100	100	95.65	93.33	96.88	100	100	100
SVC	5	100	100	100	95.65	92.31	92.86	100	100	100
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100
Multilayer Perceptron	5	69.57	0.0	50.0	91.3	83.33	85.71	69.57	0.0	50.0
	10	100.00	100.0	100.0	91.3	83.33	85.71	69.57	0.0	50.0
	15	69.57	0.0	50.0	91.3	83.33	85.71	69.57	0.0	50.0

misimos resultados se obtienen al utilizar la combinación de ambos tipos de predictores. Sin embargo, cuando los modelos se alimentan solo con cinco atributos cinemáticos, ninguno alcanza la máxima puntuación en las métricas de evaluación. En este caso, hay que recurrir a los modelos que han sido entrenados con 10 variables explicativas, ya que entonces, **CNB** y **SVC** sí que ofrecen una exactitud, F_1 y **AUC** del 100 %. Entre estos dos candidatos se ha seleccionado el estimador **SVC**. En la Tabla 5.33 se resumen los tres modelos seleccionados.

Tabla 5.33: Modelos seleccionados para determinar si se trata de una marcha discontinua (DC)

Tipo de entradas	Predictores	Algoritmo	Hiperparámetros	Rendimiento
Espaciotemporales	right_stride_length_norm right_toe_height_norm cadence_norm left_double_support_percent left_toe_height_norm	SVC	kernel = rbf C= 0.105 $\gamma = 1$	Acc = 100 % $F_1 = 100 %$ AUC = 100 %
Cinemáticas	pelvis_rotation_Range left_pelvis_rotation_Min left_pelvis_rotation_Max right_pelvis_rotation_Min left_knee_flex_ext_Range right_knee_flex_ext_Range right_pelvis_rotation_Max right_hip_flex_ext_Range left_knee_flex_ext_Max right_knee_flex_ext_Max	SVC	kernel = rbf C= 0.149 $\gamma = 0.42$	Acc = 100 % $F_1 = 100 %$ AUC = 100 %
Espaciotemporales + cinemáticas	right_stride_length_norm right_toe_height_norm cadence_norm left_double_support_percent pelvis_rotation_Range	SVC	kernel = rbf C= 0.105 $\gamma = 1.189$	Acc = 100 % $F_1 = 100 %$ AUC = 100 %

5.3.8. PM. Postura en la marcha

El último ítem de la escala de Tinetti, abordado como un problema supervisado de clasificación binaria, trata de valorar la postura adoptada por el paciente al caminar. Para ello, evalúa si durante la marcha los talones se mantienen juntos, de forma que casi se tocan con los pasos (anormal), o, si por el contrario, los talones permanecen separados (normal). Para resolver este nuevo problema de clasificación binaria, se ha utilizado un subconjunto de datos formado por las 48 muestras de marcha normal, etiquetadas como 0, más las 20 observaciones de clase positiva (1) en las que se ha simulado un aumento del ángulo del paso (hacia el exterior), para forzar así que los talones marchen juntos al caminar (*pm*).

Para el problema de determinar si los talones están juntos o separados al caminar, con el análisis de ANOVA (Tabla 5.34a) se obtiene que la variable espaciotemporal más influyente en la variable respuesta es la base de sustentación, y que tal y como se describió en la Sección 5.2.2, se corresponde con el promedio de la distancia lineal entre los marcadores colocados en el talón de cada pie durante una zancada. Otros atributos espaciotemporales relevantes en esta tarea son la longitud de paso y zancada, así como la máxima altura de ambos talones durante la marcha.

En cuanto a las características cinemáticas (Tabla 5.34b), el rango de amplitud angular para la flexión-extensión de ambas rodillas son las variables predictoras más relacionadas con la etiqueta de salida. Estos resultados coinciden con los obtenidos en otros problemas anteriores orientados a valorar la altura y longitud de paso. Sin embargo, en este caso, aparecen entre los cinco atributos más influyentes el ángulo promedio de rotación del pie derecho y el mínimo valor en el movimiento de rotación pélvica durante la zancada izquierda.

Tabla 5.34: Selección de características espaciotemporales (a) y cinemáticas (b) para el problema de valorar la postura durante la marcha (PM)

(a) Variables espaciotemporales

(b) Variables cinemáticas

#	Variable	F_1	#	Variable	F_1
1	base_of_support_norm	395.30	1	left_knee_flex_ext_Range	351.92
2	right_step_length_norm	218.68	2	right_knee_flex_ext_Range	306.81
3	right_stride_length_norm	171.89	3	left_pelvis_rotation_Min	160.75
4	left_toe_height_norm	168.60	4	right_foot_rotation_Stride_Start	160.05
5	right_heel_height_norm	168.32	5	right_foot_rotation_Mean	138.45
6	left_heel_height_norm	166.96	6	right_pelvis_rotation_Max	138.27
7	left_stride_length_norm	146.38	7	right_knee_flex_ext_Max	129.47
8	right_toe_height_norm	145.28	8	right_hip_flex_ext_Range	120.50
9	left_toe_off_angle	133.09	9	right_pelvis_rotation_Min	115.45
10	left_strike_angle	133.09	10	left_pelvis_rotation_Stride_Start	112.97
11	left_step_length_norm	99.52	11	left_pelvis_rotation_Max	112.10
12	right_toe_off_angle	73.23	12	left_pelvis_rotation_Stride_End	108.22
13	right_strike_angle	73.23	13	left_knee_flex_ext_Max	102.60
14	com_vertical_displacement_norm	42.33	14	right_knee_val_var_Range	65.85
15	right_step_duration_norm	11.23	15	right_ankle_rotation_Range	60.64

El listado de variables predictoras más relevantes para determinar si los talones permanecen juntos durante la marcha, considerando tanto parámetros espaciotemporales como medidas cinemáticas, se ilustra en el diagrama de barras de la Figura 5.47. En este caso, las cinco primeras características explicativas son el ancho de la base de sustentación, la amplitud de la flexión-extensión de ambas rodillas y, las longitudes de paso y zancada para

la extremidad derecha.

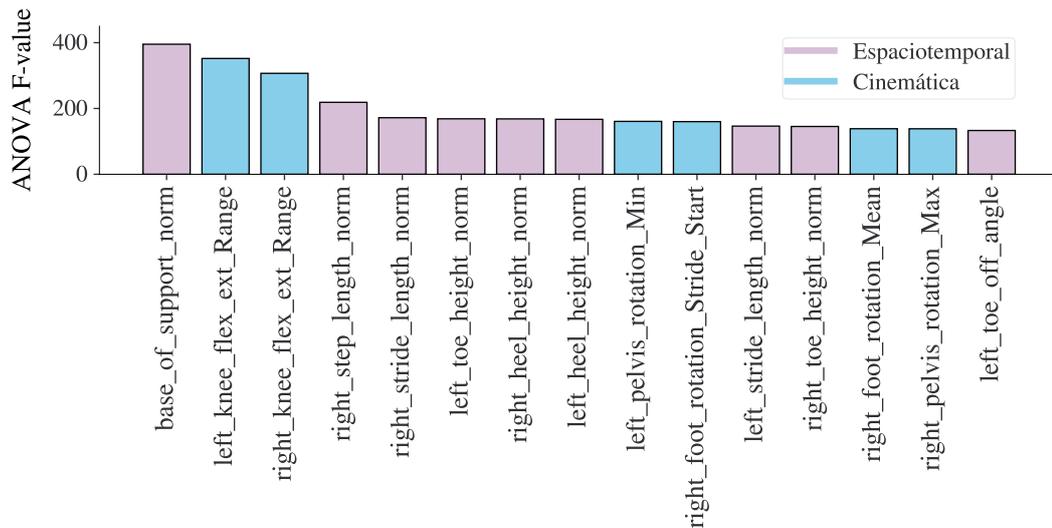


Figura 5.47: Representación de las 15 características predictoras más importantes (en base al valor F_1 de ANOVA) para el problema PM.

Siguiendo el mismo procedimiento que en el resto de problemas, explicado en la Sección 5.3.2, se han entrenado varios modelos basados en diferentes algoritmos y utilizando distinto número y tipo de variables predictoras: las 5, 10 y 15 más relevantes considerando solo parámetros espaciotemporales, solo atributos cinemáticos o una combinación de ambos. Los resultados del entrenamiento se muestran en la Tabla 5.35, incluyendo el tiempo de ajuste y rendimiento (exactitud), promediados sobre una validación cruzada de 10 iteraciones. Los valores proporcionados para cada caso se corresponden con el mejor ajuste de hiperparámetros encontrado mediante una búsqueda exhaustiva. Por su parte, en la Tabla 5.36 se incluyen las métricas de exactitud, F_1 , y área bajo la curva ROC obtenidas al evaluar los diferentes estimadores sobre el conjunto de pruebas.

Tabla 5.35: Rendimiento del entrenamiento, medido con la exactitud (%), y tiempo de entrenamiento (ms) promediado sobre una validación cruzada de 10 iteraciones, en el problema de valorar la postura durante la marcha (PM)

Algoritmo	Variables de entrada	Espaciotemporales		Cinemáticas		Espaciotemporales ± cinemáticas	
		Acc	Tiempo	Acc	Tiempo	Acc	Tiempo
Complement Naive Bayes	5	97.5 ± 7.5	5.3 ± 5.5	98.0 ± 6.0	4.1 ± 1.4	100	5.8 ± 3.8
	10	97.5 ± 7.5	4.7 ± 7.2	100	3.8 ± 6.2	97.5 ± 7.5	5.4 ± 6.4
	15	97.5 ± 7.5	6.3 ± 7.7	97.5 ± 7.5	5.4 ± 5.7	97.5 ± 7.5	3.8 ± 4.8
K-Nearest Neighbor	5	97.5 ± 7.5	5.0 ± 3.9	100	4.9 ± 0.7	100	10.5 ± 9.7
	10	97.5 ± 7.5	3.8 ± 6.2	100	4.4 ± 6.1	97.5 ± 7.5	3.1 ± 6.2
	15	97.5 ± 7.5	7.8 ± 7.8	97.5 ± 7.5	6.4 ± 7.6	97.5 ± 7.5	3.1 ± 6.3
Decision Tree	5	97.5 ± 7.5	2.9 ± 0.5	93.0 ± 10.8	6.3 ± 7.7	95.5 ± 9.1	2.0 ± 3.6
	10	97.5 ± 7.5	1.0 ± 3.0	88.5 ± 11.6	1.6 ± 4.7	95.5 ± 9.1	3.1 ± 6.3
	15	100	1.7 ± 4.7	95.5 ± 9.1	3.0 ± 3.4	95.5 ± 9.1	4.7 ± 7.2
SVC	5	97.5 ± 7.5	4.4 ± 4.3	100	5.6 ± 4.8	100	5.1 ± 3.1
	10	97.5 ± 7.5	5.9 ± 4.5	97.5 ± 7.5	5.8 ± 4.1	97.5 ± 7.5	3.4 ± 4.0
	15	97.5 ± 7.5	5.6 ± 4.6	97.5 ± 7.5	12.2 ± 11.4	97.5 ± 7.5	8.3 ± 3.2
Multilayer Perceptron	5	96.0 ± 2.4	31.8 ± 90.4	92.7 ± 12.5	38.1 ± 108.4	96.4 ± 2.8	31.6 ± 89.8
	10	93.8 ± 6.8	34.8 ± 104.3	94.7 ± 10.9	40.0 ± 114.8	90.7 ± 17.4	37.0 ± 111.0
	15	94.4 ± 6.5	34.3 ± 100.4	92.2 ± 15.2	40.1 ± 115.3	96.4 ± 3.7	38.5 ± 110.4

Tabla 5.36: Exactitud (Acc), valor F_1 y área bajo la curva ROC (AUC), expresadas como %, y medidas sobre el conjunto de pruebas para valorar la postura durante la marcha (PM). Se muestran remarcados los modelos seleccionados.

Algoritmo	Variables de entrada	Espaciotemporales			Cinemáticas			Espaciotemporales + cinemáticas		
		Acc	F_1	AUC	Acc	F_1	AUC	Acc	F_1	AUC
Complement Naive Bayes	5	95.65	92.31	92.86	100	100	100	100	100	100
	10	91.30	85.71	89.73	100	100	100	100	100	100
	15	95.65	92.31	92.86	100	100	100	100	100	100
K-Nearest Neighbor	5	95.65	92.31	92.86	100	100	100	100	100	100
	10	95.65	92.31	92.86	100	100	100	100	100	100
	15	95.65	92.31	92.86	100	100	100	95.65	92.31	92.86
Decision Tree	5	100	100	100	100	100	100	91.3	87.5	93.75
	10	100	100	100	100	100	100	100	100	100
	15	100	100	100	91.3	87.5	93.75	100	100	100
SVC	5	95.65	92.31	92.86	95.65	92.31	92.86	100	100	100
	10	95.65	92.31	92.86	95.65	92.31	92.86	95.65	92.31	92.86
	15	91.30	83.33	85.71	95.65	92.31	92.86	95.65	92.31	92.86
Multilayer Perceptron	5	69.57	0.0	50.0	69.57	0.0	50.0	69.57	0.00	50.00
	10	69.57	0.0	50.0	69.57	0.0	50.0	78.26	44.44	64.29
	15	69.57	0.0	50.0	69.57	0.0	50.0	69.57	0.00	50.00

Para esta tarea de determinar si los talones permanecen juntos durante la marcha, el modelo basado en árboles de decisión **DT** es el único que alcanza una exactitud, precisión, sensibilidad y **AUC** del 100 % cuando se alimenta solo con variables espaciotemporales. Puesto que este tipo de algoritmos ofrecen una fácil interpretación desde el punto de vista de las predicciones, es posible representar visualmente la estructura de árbol de decisión construida durante el entrenamiento para analizar cómo este modelo puede automatizar la valoración de la postura de la marcha en la escala de Tinetti.

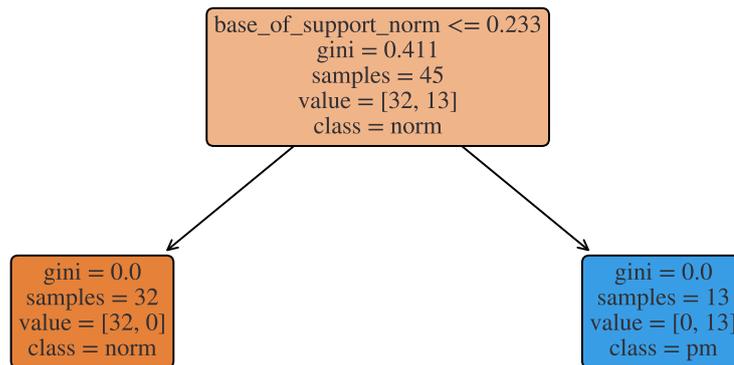


Figura 5.48: Árbol de decisión para el problema PM con parámetros espaciotemporales.

Tal y como se observa en el árbol de la Figura 5.48, se trata de un árbol con un único nivel de reglas que fundamenta las decisiones de la predicción en el atributo `base_of_support_norm` (ancho de la base de sustentación). En particular, asigna una puntuación de 1 (marcha anormal) todas aquellas instancias en las que el valor de esta variable predictora es superior a 0,233. En este sentido, conviene señalar que el ancho de la base de sustentación fue normali-

zado por la estatura del sujeto para eliminar la influencia de esta medida antropométrica (Sección 5.2.4). Por tanto, este umbral de decisión depende de la altura del paciente para el que se evalúe la marcha. Por ejemplo, para una persona con una estatura de 1.70 metros, el árbol de decisión determinaría que se camina con los talones juntos si el ancho de paso es superior a los 40 centímetros (el promedio para una marcha normal es de 10 cm [9]).

Para el caso en el que solo se consideran variables cinemáticas, además de **DT**, los modelos de **KNN** y **CNB** también ofrecen una exactitud, F_1 y **AUC** del 100 % tomando como entrada los cinco predictores más relevantes de este tipo (Tabla 5.34b). Entre estos tres candidatos se ha seleccionado el clasificador de árboles de decisión porque son más rápidos en las predicciones y, a través de su proceso automático de selección de características, necesita menos variables explicativas que **KNN** y **CNB**. Por otro lado, cuando se consideran tanto parámetros espaciotemporales como variables cinemáticas, el modelo de **SVC** sí que consigue un rendimiento del 100 % (en exactitud, F_1 y **AUC**) con solo cinco predictores. Por tanto, se ha seleccionado este modelo antes que el de **KNN** y **CNB** que ofrecen unos resultados similares, siguiendo los criterios descritos para los problemas anteriores.

Como resumen, en la Tabla 5.37 se muestran las características de los tres clasificadores seleccionados para valorar la postura durante la marcha, incluyen las variables predictoras que los alimentan en cada caso y los hiperparámetros más relevantes: profundidad máxima del árbol (**max_depth**), número máximo de nodos terminales (**max_leaf_nodes**) y medida de impureza (**criterion**) para árboles de decisión, y función de kernel, regularización C y γ para **SVC**.

Tabla 5.37: Modelos seleccionados para valora la postura durante la marcha (PM)

Tipo de entradas	Predictores	Algoritmo	Hiperparámetros	Rendimiento
Espaciotemporales	base_of_support_norm right_step_length_norm right_stride_length_norm left_toe_height_norm right_heel_height_norm	DT	max_depth=1 max_leaf_nodes=2	Acc = 100 % F1 = 100 % AUC = 100 %
Cinemáticas	left_knee_flex_ext_Range right_knee_flex_ext_Range left_pelvis_rotation_Min right_foot_rotation_Stride_Start right_foot_rotation_Mean	DT	max_depth=1 max_leaf_nodes=2	Acc = 100 % F1 = 100 % AUC = 100 %
Espaciotemporales + cinemáticas	base_of_support_norm left_knee_flex_ext_Range right_knee_flex_ext_Range right_step_length_norm right_stride_length_norm	SVC	kernel = rbf C = 0.595 γ = 0.595	Acc = 100 % F1 = 100 % AUC = 100 %

5.3.9. Resumen y discusión de resultados

Los resultado obtenidos en el presente trabajo demuestran que los modelos basados en máquinas de vectores soporte, y en particular, la implementación **SVC** de *Scikit-Learn* con función de kernel **RBF**, ofrecen un buen desempeño en este tipo de tareas supervisadas de clasificación binaria, por lo que parecen tener un gran potencial en futuras aplicaciones relacionadas con la identificación de patrones de marcha normal y patológica. Tanto es así, que en los 6 problemas abordados, **SVC** alcanza las máximas puntuaciones en las métricas de

evaluación cuando es alimentado con una combinación de 5 variables espaciotemporales y cinemáticas, mientras que en 5 de las 6 tareas de clasificación, ofrece este mismo rendimiento tomando como entradas únicamente 5 atributos espaciotemporales o cinemáticos. Asimismo, aunque se han seleccionado los modelos de *SVM*, conviene remarcar que los clasificadores basados en los *k*-vecinos más cercanos ofrecen un rendimiento similar, alcanzando también una exactitud, precisión, sensibilidad y *AUC* del 100 % en todas las tareas en las que han sido entrenados con los 5 predictores espaciotemporales y cinemáticos más relevantes.

Para complementar este análisis del rendimiento ofrecido por los distintos algoritmos de *Machine Learning*, en la Figura 5.49 se incluye un diagrama de barras que representa visualmente la exactitud, medida sobre el conjunto de pruebas, de los distintos modelos entrenados con una combinación de los cinco predictores espaciotemporales y cinemáticos más relevantes para cada tarea.

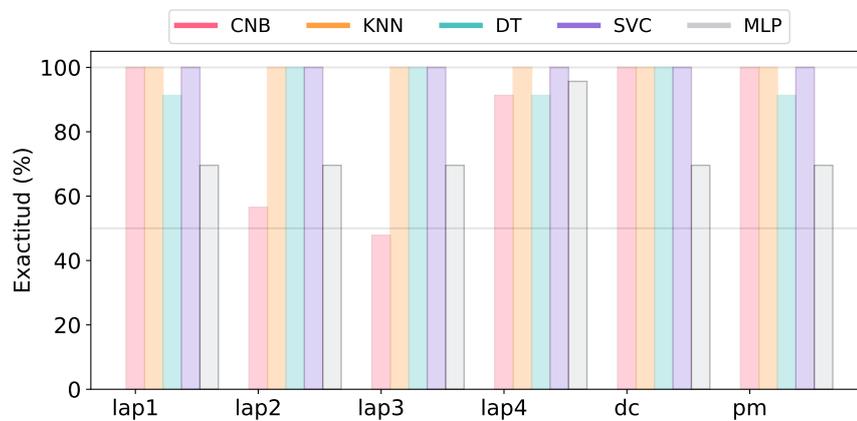


Figura 5.49: Evaluación en el conjunto de prueba (exactitud) de cada familia de algoritmos utilizando los cinco predictores espaciotemporales y cinemáticos más relevantes

En el gráfico se observa que los estimadores de *Complement Naive Bayes* ofrecen unos resultados globales algo inestables: aunque en las tareas *lap1*, *dc* y *pm*, sí que tienen un muy buen desempeño (exactitud y F_1 del 100 %), en los problemas de determinar si el pie izquierdo sobrepasa al derecho (*lap2*) y si el pie derecho se levanta completamente del suelo (*lap3*), las puntuaciones de evaluación no superan el 60 %. Por su parte, los modelos basados en redes neuronales (*MLP*) son los clasificadores que peor rendimiento han ofrecido, ya que en 5 de las 6 tareas no alcanzan el 70 % de exactitud y F_1 . El pobre rendimiento ofrecido por las arquitecturas de redes neuronales en este trabajo podría deberse a una cantidad escasa de ejemplos de aprendizaje, así como a un tiempo insuficiente de entrenamiento.

Concluye así esta fase de modelado, en la que se han entrenado, ajustado y evaluado una serie de clasificadores binarios basados en diferentes algoritmos de Aprendizaje Automático, número y tipo de variables predictoras. Como resultado, para seis ítems de la escala de Tinti que valoran distintos aspectos de la marcha humana, se han seleccionado tres modelos que generan automáticamente la puntuación asignada. En cada caso, se dispone de un clasificador que se basa únicamente en los parámetros espaciotemporales de la marcha, otro estimador que toma como entradas las propiedades cinemáticas del movimiento, y un tercero, que combina ambos tipos de información. Para todos los problemas abordados, los tres modelos seleccionados han alcanzado una exactitud, F_1 y área bajo la curva *ROC* del 100 %, evaluados sobre datos no vistos durante el entrenamiento. En la siguiente y última etapa del trabajo, estos clasificadores han sido desplegados sobre una herramienta final que genera informes con detalladas valoraciones de la marcha humana.

5.4. IMPLEMENTACIÓN

Una vez que los modelos de *Machine Learning* han sido satisfactoriamente evaluados, es necesario implementarlos en un entorno de producción, o en un ambiente de pruebas de forma limitada hasta que su rendimiento se haya evaluado completamente. En esta fase de trabajo, los clasificadores binarios que fueron seleccionados durante la etapa de modelado (sección anterior), han sido desplegados e integrados en una herramienta *software*, basada en una interfaz de línea de comandos (CLI), que genera automáticamente las respuestas a los ítems de la escala de Tinetti para valoración de la marcha. Para ello, se ha desarrollado un *script* en *Python* que produce informes en formato PDF, en los que se incluye, además de las respuestas predichas para POMA, unos detallados gráficos y tablas con información cinemática y espaciotemporal relevante en este tipo de estudios.

Tal y como se ilustra esquemáticamente en la Figura 5.50, la herramienta desarrollada toma como entrada datos en crudo exportados desde el sistema de captura de movimiento (*Clinical 3DMA*), y algunas medidas antropométricas del sujeto analizado. Sobre estos datos, aplica todas las operaciones de preprocesamiento y limpieza explicadas en la Sección 5.2, para poder utilizar los modelos predictivos de *Machine Learning* que devuelven la puntuación a las diferentes categorías consideradas en la escala de Tinetti. Finalmente, se combina la información de los parámetros espaciotemporales y cinemáticos, obtenida durante la fase de preprocesamiento, con las predicciones de los clasificadores, generando así unos documentos que contienen una valoración automática de la marcha humana. En el Anexo E, se proporciona un ejemplo de uno los informes producidos por esta herramienta.

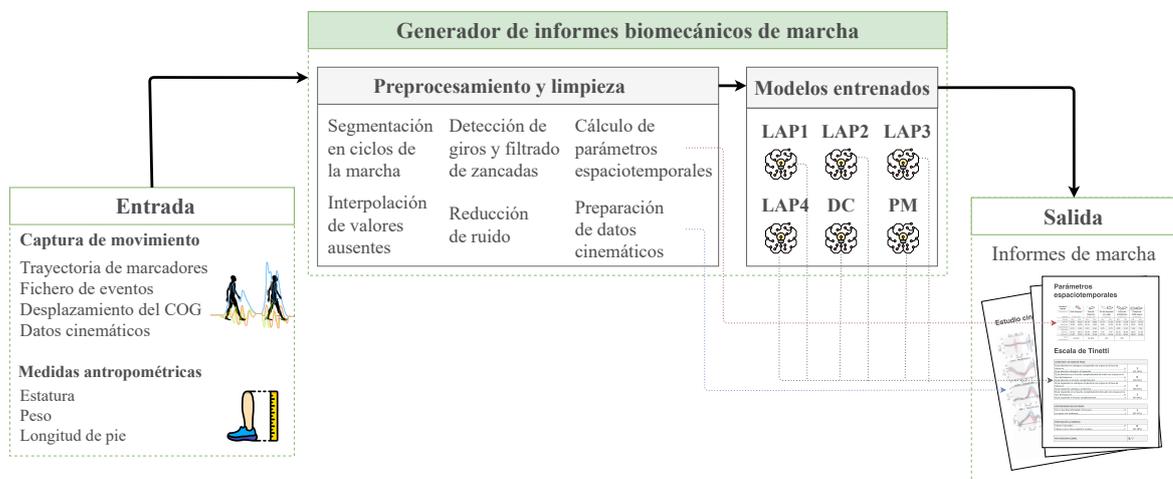


Figura 5.50: Esquema de la generación de informes con valoración de la marcha

En las siguientes secciones se describe, en detalle, el funcionamiento del *script* en *Python* que implementa esta funcionalidad, así como, el contenido de los informes de marcha producidos por la herramienta desarrollada durante esta fase de trabajo.

5.4.1. Script para la generación de informes con valoración de la marcha

El código fuente del *script* en *Python* que ensambla los diferentes modelos predictivos y genera los informes con una valoración de la marcha, está disponible en el repositorio del proyecto, cuyo enlace se proporciona en el Anexo A, y en particular, dentro del directorio *report-generator*. A continuación, se describen los pasos que ejecuta secuencialmente esta

herramienta, desde la entrada de datos hasta la producción del documento PDF con los resultados. Asimismo, para más detalles sobre las librerías de terceros que han sido utilizadas en el desarrollo del *script*, en el Anexo B se indica la versión exacta junto con un enlace al repositorio del que han sido descargadas.

- **Entrada de datos.** Como ya se ha comentado, la aplicación desarrollada toma como entrada una serie de medidas corporales y ficheros con datos relativos a la captura de marcha que se pretende valorar. En concreto, teniendo en cuenta las operaciones de transformación llevadas a cabo sobre los datos en crudo para generar las variables explicativas que requieren los modelos predictivos, el *script* necesita los siguientes ficheros exportados desde *Clinical 3DMA*:
 - Archivo con trayectoria de marcadores, utilizado para estimar los parámetros espaciotemporales de la marcha según lo descrito en la Sección 5.2.2.
 - Fichero de eventos, necesario para segmentar el resto de datos en ciclos de la marcha (zancadas).
 - CSV con datos relativos al desplazamiento del COG, importado por el *script* que detecta los giros durante la grabación (Sección 5.2.3), y necesario para obtener las variables espaciotemporales relacionadas con el movimiento del centro de gravedad al caminar.
 - Datos cinemáticos, utilizados para completar el filtrado de zancadas, y generar las características predictoras basadas en este tipo de información de la marcha (Tabla 5.11).

Por su parte, las dos características antropométricas fundamentales para obtener unas correctas predicciones son la estatura del sujeto y la longitud del pie, ya que los parámetros espaciales de la marcha con los que se entrenaron los modelos estaban normalizados por estas medidas longitudinales (Sección 5.2.4). No, obstante, para complementar los informes generados, el *script* pide también que se especifique la edad, género y peso del paciente. Adicionalmente, puesto que para cada categoría de Tinetti se han seleccionado tres modelos basados en diferentes tipos de predictores, la aplicación da opción a señalar el tipo de variables utilizadas para generar las respuestas a la escala: solo espaciotemporales, solo cinemáticas o ambas.

La introducción de estos datos se lleva a cabo mediante una interfaz de usuario basada en línea de comandos. No obstante, para que la herramienta sea ágil y fácil de usar, se han empleado los paquetes *PyInquirer*²⁷ y *easygui*²⁸. El primero de ellos, *PyInquirer*, permite obtener la información de entrada mediante un procedimiento interactivo basado en preguntas, en lugar de argumentos u opciones como típicamente se caracterizan las aplicaciones con este tipo de interfaz de usuario. Como ejemplo del funcionamiento, en la Figura 5.51 se muestra la introducción en la herramienta de las características generales y antropométricas del usuario, así como la definición del tipo de parámetros utilizados para realizar las predicciones.

Por otro lado, para facilitar la especificación de la ubicación de los archivos con los datos en crudo, se han utilizado las funciones `fileopenbox` y `diropenbox` de *easygui*, las cuales se utilizan, respectivamente, para abrir un cuadro de dialogo y obtener la ruta al archivo o directorio seleccionado por el usuario.

²⁷<https://pypi.org/project/PyInquirer/>

²⁸<https://pypi.org/project/easygui/>

```
(base) PS D:\OneDrive - UNED\Master en Ciencia de Datos\TFM\CODIGO\Gait_Optical_Sensors_Tinetti\report-generator> python .\main.py
Características generales del paciente
? Nombre Juan
? Edad 72
? Género Hombre
? Altura (en centímetros) 172
? Peso (en kilogramos) 80
? Longitud de pie (en milímetros) 265

Configuración para la generación de informes
? Tipo de parámetros de la marcha utilizados para generar automáticamente las respuestas a la prueba de Tinetti (Use arrow keys)
  Espaciotemporales
  Cinematicas
  > Espaciotemporales y cinematicas
```

Figura 5.51: Entrada de datos en la herramienta de generación de informes

- **Preparación y limpieza de datos.** Una vez conocida la ubicación de los datos de marcha que se pretenden clasificar, el objetivo es generar el conjunto de características predictivas, basadas en los parámetros espaciotemporales y/o cinemáticos de la marcha, que necesitan los modelos para realizar las predicciones. En este sentido, se ha reutilizado el *script* implementado durante la fase de preparación de datos (Sección 5.2) con la intención de que los nuevos datos en crudo pasen por toda la *pipeline* de preprocesamiento, limpieza y transformación por la que pasaron los ejemplos utilizados en el entrenamiento.

Dado que el *script* de preprocesamiento requiere que los datos exportados desde *Clinical 3DMA* estén organizados en una serie de directorios específicos (Figura 5.8), se ha utilizado el módulo *tempfile*²⁹ para copiar los archivos de datos, siguiendo esa estructura específica, desde las ubicaciones indicadas por el usuario hasta un directorio temporal que es limpiado automáticamente cuando finaliza la ejecución del programa. De esta forma, se aplican secuencialmente todas las operaciones descritas en la Sección 5.2.5 (Figura 5.33) hasta generar los archivos *CSV* *spatiotemporal.csv* y *kinematics* con el conjunto de variables explicativas sobre las que realizar las predicciones.

- **Predicción de las respuestas a la escala de Tinetti.** Durante la fase de modelado, los clasificadores seleccionados fueron guardados usando *joblib* en formato *.pkl*, de forma que contienen las tareas de preparación (selección de características y escalado de datos), y la *pipeline* completa de predicción. Estos modelos, que se pueden descargar desde la carpeta *models* incluida en el repositorio del proyecto (Anexo A), son importados por el *script* para predecir las diferentes categorías de Tinetti en función de las variables explicativas generadas en el punto anterior. En este sentido, conviene señalar que dependiendo de la configuración especificada por el usuario, se importan unos clasificadores u otros (solo con parámetros espaciotemporales, solo con atributos cinemáticos, o con ambos).

Para realizar las predicciones se han utilizado los métodos `predict_proba` de los estimadores, que devuelven la probabilidad de que la instancia dada pertenezca a cada una de las clases. De esta forma, se ha elegido la etiqueta con probabilidad más alta como resultado del proceso de clasificación y, además, en los informes se ha incluido este valor de probabilidad como un indicador de la certeza que tiene el modelo para asignar una determinada puntuación. En la Figura 5.52 se muestra un ejemplo del funcionamiento de la herramienta desarrollada, desde la introducción de datos hasta

²⁹<https://docs.python.org/es/3.10/library/tempfile.html>

predecir y mostrar las respuestas automáticas a los diferentes ítems de la prueba de Tinetti.

```
(base) PS D:\OneDrive - UNED\Master en Ciencia de Datos\TFM\CODIGO\Gait_Optical_Sensors_Tinetti\report-generator> python .\main.py
Características generales del paciente
? Nombre Juan
? Edad 72
? Género Hombre
? Altura (en centímetros) 172
? Peso (en kilogramos) 88
? Longitud de pie (en milímetros) 265

Configuración para la generación de informes
? Tipo de parámetros de la marcha utilizados para generar automáticamente las respuestas a la prueba de Tinetti Espaciotemporales y cinematicas
Seleccione el archivo con la trayectoria de los marcadores (exportado desde Clinical 3DMA)
D:\datos\trayectorias.raw
Seleccione el archivo de eventos de la marcha (exportado desde Clinical 3DMA)
D:\datos\eventos.TXT
Seleccione el archivo con el desplazamiento del centro de gravedad (exportado desde Clinical 3DMA)
D:\datos\biomechanics\COG.csv
Seleccione el directorio con los archivos CSV de datos cinemáticos (exportados desde el diálogo de zancadas de Clinical 3DMA)
D:\datos\strides

LAP1: ¿El pie derecho sobrepasa al izquierdo con el paso en la fase de balanceo?
0 (1.75%) 1 (98.25%)
Puntuación: 1

LAP2: ¿El pie izquierdo sobrepasa al derecho con el paso en la fase de balanceo?
0 (92.92%) 1 (7.08%)
Puntuación: 0

LAP3: ¿El pie derecho se levanta completamente del suelo con el paso en la fase del balanceo?
0 (3.09%) 1 (96.91%)
Puntuación: 1

LAP4: ¿El pie izquierdo se levanta completamente del suelo con el paso en la fase del balanceo?
0 (93.88%) 1 (6.12%)
Puntuación: 0

DC: ¿Para o hay discontinuidad entre pasos?
0 (33.11%) 1 (66.89%)
Puntuación: 1

PM: ¿Talones casi se tocan mientras camina?
0 (67.64%) 1 (32.36%)
Puntuación: 0

Puntuación total: 3/6
```

Figura 5.52: Funcionamiento de la herramienta para la generación de informes con valoración de la marcha

- Generación de informe con resultados.** Combinando todos los datos generados en los pasos anteriores, el *script* genera un informe en formato PDF con tablas y gráficos que facilitan el análisis de esta información, y consecuentemente, el diagnóstico y valoración de la marcha humana. Para ello, se ha utilizado *docxtpl*³⁰, una librería que permite crear documentos de *Microsoft Word* a partir de plantillas que contienen variables delimitadas por una doble llave ('{}'). De esta forma, la herramienta sustituye todas las etiquetas de la plantilla por la información generada, guardando el documento de *Microsoft Word* resultante sobre el directorio temporal que fue creado en el segundo paso. Finalmente, este documento es convertido a formato PDF empleando la librería *pywin32*³¹, y se elimina todo el contenido del directorio temporal.

En la siguiente sección se detalla la estructura y contenido de los informes producidos.

³⁰<https://docxtpl.readthedocs.io/>

³¹<https://pypi.org/project/pywin32/>

5.4.2. Informe biomecánico de la marcha humana

Tal y como se explicó en la Sección 3.2, existen dos tipos de exámenes para valorar la marcha humana: el análisis cuantitativo, basado en el estudio de parámetros espaciotemporales y cinemáticos del movimiento y, el examen observacional, que consiste en pruebas clínicas y escalas que permiten conocer alteraciones al caminar. Con la intención de facilitar el diagnóstico de la marcha a los profesionales sanitarios, los informes biomecánicos generados por la herramienta desarrolla en este trabajo, ofrecen un enfoque híbrido bajo el cual se incluyen tanto datos cuantitativos de la marcha (parámetros espaciotemporales y cinemáticos) como los resultados, generados automáticamente, de una de las pruebas observacionales más ampliamente utilizadas (Tinetti).

Por tanto, estos informes están estructurados en una serie de secciones y apartados que se presentan a continuación: características generales, análisis del ciclo de la marcha, parámetros espaciotemporales, valoración según la escala de Tinetti y estudio cinemático. Se puede consultar un ejemplo completo en el Anexo E.

- Características generales.** La primera sección incluida en los informes generados, proporciona un resumen con información sobre el paciente y la captura de movimiento analizada. Tal y como se muestra en el ejemplo de la Figura 5.53, se indica la edad, género, peso, estatura y longitud del pie del sujeto. Estas dos últimas medidas corporales son especialmente relevantes para la herramienta, ya que los parámetros espaciales utilizados para generar las respuestas a la prueba de Tinetti, son normalizados por dichas medidas antropométricas.

Además, se informa sobre la duración de la grabación, velocidad de marcha, cadencia (número de pasos por minuto), número de zancadas detectadas para cada extremidad, número de zancadas válidas (es decir, que no han sido filtradas por el *script* de preprocesamiento), y finalmente, la cantidad de giros detectados por el algoritmo presentado en la Sección 5.2.3.

Características generales

72	Hombre	1.72 m	80 kg	265 cm	27.51s
Edad	Género	Altura	Peso	Longitud de pie	Duración
0.7 m/s	82.3 pasos/min	14 Izq.	13 Der.	13 Izq. 12 Der.	2
Velocidad	Cadencia	Zancadas totales	Zancadas procesadas	Giros	

Figura 5.53: Información general incluida en los informes de marcha

- Análisis del ciclo de la marcha.** Para facilitar el estudio de la división en fases del ciclo de la marcha, se incluye un gráfico similar al del la Figura 5.23, que muestra visualmente, el tiempo en segundos y el porcentaje que ocupan las fases de apoyo y balanceo dentro del ciclo de las zancadas de ambas extremidades, así como la duración, también en segundos, de cada uno de los dos períodos de doble apoyo que tienen lugar en cada zancada.

Asimismo, estas medidas se proporcionan detalladamente en formato tabular, de forma que, para cada una de ellas, se especifica la media, mediana, desviación estándar, valor mínimo y valor máximo, calculados sobre el total de zancadas de cada pie. Además, tal y como se muestra en el ejemplo de la Figura 5.54, se proporcionan unos

valores de normalidad, obtenidos de bibliografía especializada [11], para agilizar la interpretación de esa información.

Ciclo de la marcha										
Parámetro	Fase de apoyo		Fase de balanceo		Fin de respuesta a la carga	Inicio del prebalanceo		Tiempo de doble apoyo		
Unidad	% del ciclo		% del ciclo		% del ciclo		% del ciclo		% del ciclo	
Pierna	lzk.	Der.	lzk.	Der.	lzk.	Der.	lzk.	Der.	lzk.	Der.
Media	68.76	64.47	31.24	35.53	6.33	27.06	40.69	57.93	34.40	33.60
Mediana	67.51	66.16	32.49	33.84	7.22	25.52	40.62	58.40	32.63	33.16
Desviación Estándar	4.24	3.90	4.24	3.90	2.04	4.82	2.58	3.13	5.72	6.18
Mínimo	64.42	53.71	19.07	32.63	1.11	19.43	35.29	52.13	23.89	20.57
Máximo	80.93	67.37	35.58	46.29	9.52	38.27	46.11	61.34	45.88	47.45
Valor orientativo ¹	62%		38%		10%		50%		25%	

Figura 5.54: Análisis detallado del ciclo de la marcha en los informes generados

- Parámetros espaciotemporales.** La presentación de los parámetros espaciotemporales calculados a partir de la captura de movimiento, se ha organizado en una serie de apartados con información relativa a diferentes aspectos de la marcha. Así, se incluye en primer lugar un grupo de gráficos y tablas con datos sobre la duración y longitud de las zancadas y pasos. Para ello, se ha utilizado un gráfico (Figura 5.55) que representa de manera esquematizada la longitud y duración de los pasos y zancadas de ambas extremidades, así como los ángulos y anchura de paso (base de sustentación).

Además, siguiendo el mismo formato que el presentado para el ciclo de la marcha en el punto anterior, se incluyen dos tablas con las medidas descriptivas de los principales parámetros espaciotemporales: duración de zancada y paso, duración de la fase de apoyo y balanceo, longitud de zancada y paso, y la anchura y ángulo de paso. De igual forma, junto con estos datos se proporcionan unos valores orientativos que han sido extraídos de varias fuentes bibliográficas [9, 166].

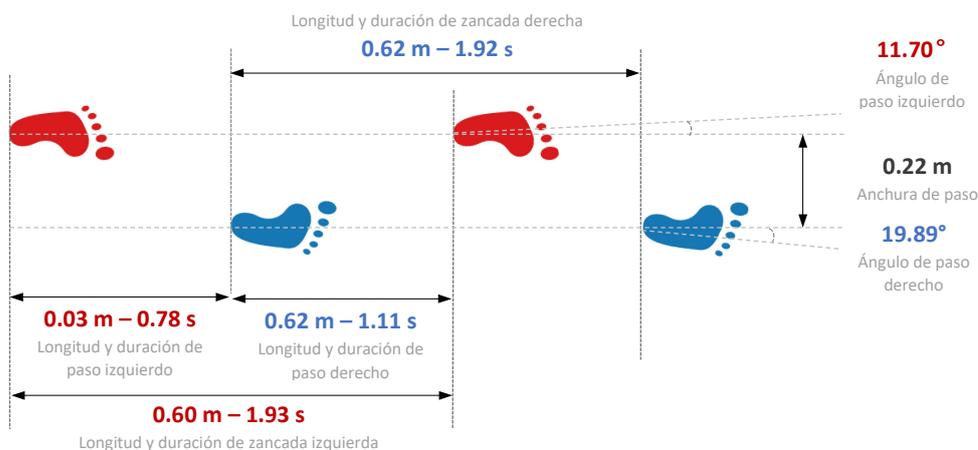


Figura 5.55: Esquema para representar visualmente la longitud de paso y zancada en los informes generados

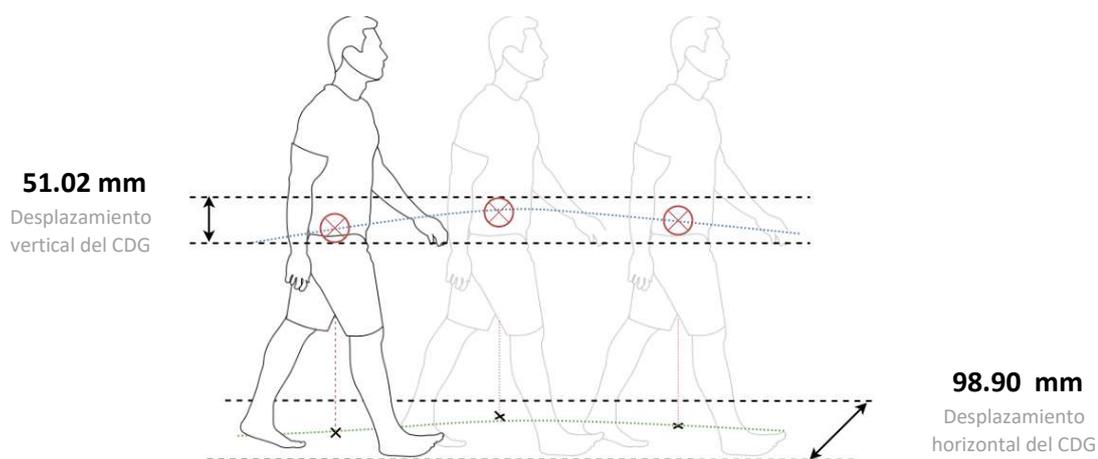
Dentro de la sección de parámetros espaciotemporales, se incluye también un apartado con datos sobre la altura y ángulos del tobillo y pie durante la marcha. En concreto, se proporciona un gráfico similar al mostrado en la Figura 5.56 para presentar visualmente

la máxima altura de talón y pie, y los ángulos de contacto y despegue para ambas extremidades. Asimismo, se incluye una tabla con las medidas descriptivas comentadas anteriormente para estas variables espaciales de la marcha.



Figura 5.56: Esquema para representar visualmente la altura y ángulos de tobillo y pie en los informes generados

Finalmente, teniendo en cuenta que el desplazamiento del centro de gravedad (COG) puede ser un importante indicador de alteraciones en la marcha (Sección 3.1.3), se ha incluido un esquema y una tabla con información sobre el movimiento del COG, en sentido vertical y horizontal, tal y como se muestra en el ejemplo de la Figura 5.57.



Parámetro	Desplazamiento vertical		Desplazamiento horizontal	
	milímetros		milímetros	
Unidad				
Pierna	Izq.	Der.	Izq.	Der.
Media	53.31	48.73	85.02	112.77
Mediana	55.42	52.00	82.38	115.72
Desviación Estándar	6.50	11.34	8.57	17.76
Mínimo	37.63	18.71	72.99	68.46
Máximo	59.76	59.72	95.89	134.86

Figura 5.57: Esquema y tabla para representar el desplazamiento del centro de gravedad durante la marcha en los informes generados

- **Escala de Tinetti.** Las respuestas a la prueba de Tinetti, generadas automáticamente por los modelos predictivos, se presentan en forma de tabla, incluyendo para cada

ítem analizado, la puntuación asignada por el clasificador. Además, tal y como se muestra en el ejemplo de la Figura 5.58, se indica también la probabilidad de clase, que puede ser interpretada como un indicador de la certeza del modelo para asignar esa puntuación. Con la intención de aclarar este aspecto para los usuarios finales, bajo la tabla se incluye un breve texto explicativo en el que también se informa del tipo de variables de la marcha utilizadas para realizar la predicción.

LONGITUD Y ALTURA DE PASO	
El pie derecho no sobrepasa al izquierdo con el paso en la fase de balanceo.....0	1
El pie derecho sobrepasa al izquierdo.....1	(98.25%)
El pie derecho no se levanta completamente del suelo con el paso en la fase del balanceo.....0	1
El pie derecho se levanta completamente.....1	(96.91%)
El pie izquierdo no sobrepasa al derecho con el paso en la fase de balanceo.....0	0
El pie izquierdo sobrepasa al derecho.....1	(92.92%)
El pie izquierdo no se levanta completamente del suelo con el paso en la fase del balanceo.....0	0
El pie izquierdo se levanta completamente.....1	(93.88%)
CONTINUIDAD DE LOS PASOS	
Para o hay discontinuidad entre pasos.....0	1
Los pasos son continuos.....1	(66.89%)
POSTURA EN LA MARCHA	
Talones separados.....0	0
Talones casi se tocan mientras camina.....1	(67.64%)
PUNTUACION GLOBAL	
	3 / 6

Respuestas generadas automáticamente en base a las características espaciotemporales y cinemáticas de la marcha. El valor porcentual incluido para cada categoría es una probabilidad que indica la certeza de la predicción.

Figura 5.58: Generación automática de las respuestas a la escala de Tinetti

- Análisis cinemático.** Finalmente, la información sobre la cinemática articular de la marcha se representa mediante una serie de diagramas con el formato generalmente usado para analizar estos datos: el eje horizontal indica la línea de tiempo, como un porcentaje del ciclo, mientras que el eje vertical representa los valores que toman las variables cinemáticas (similar al gráfico de la Figura 5.15). De manera adicional, se incluyen una serie de tablas que resumen los diferentes movimientos cinemáticos mediante medidas descriptivas: media, mínimo, máximo, rango articular, ángulo al inicio y fin del ciclo, y valor angular al comienzo de la fase de despegue.

Conclusiones y Trabajo Futuro

En este capítulo se realiza una valoración crítica sobre el grado de cumplimiento de los objetivos inicialmente planteados para este TFM (Capítulo 2). Asimismo, se propone una serie de posibles ampliaciones futuras que permitirían mejorar el sistema desarrollado y se incluye una valoración en términos de costes (tiempo y dinero) del trabajo realizado.

6.1. OBJETIVOS ALCANZADOS

El trabajo realizado en el presente TFM ha permitido cumplir con el objetivo principal: partiendo de un conjunto de datos basado en un sistema óptico de captura de movimiento, se han entrenado, evaluado y desplegado varios modelos predictivos de Aprendizaje Automático orientados a la valoración cuantitativa de la marcha humana, siguiendo las cuestiones consideradas en la escala de Tinetti.

La consecución del objetivo principal implica que se ha cumplido todos los objetivos específicos inicialmente planteados, descritos en el Capítulo 2:

- **O1. Construir un *dataset* que sirva de línea base para la valoración automática de la marcha.** Con el trabajo realizado en la fase de recopilación de datos (Sección 5.1), se ha conseguido un *dataset* formado por un total de 170 grabaciones (17 capturas por 10 participantes), incluyendo marcha normal y diferentes alteraciones durante el movimiento que han sido simuladas tomando como referencia la escala de Tinetti.

Con la finalidad de garantizar la homogeneidad de los datos adquiridos, se ha seguido el modelo Helen Hayes como protocolo estandarizado para la ubicación de los marcadores, mientras que para dotar de una mayor flexibilidad al *dataset* resultante, se han exportado los datos en una amplia variedad de formatos. Finalmente, el conjunto de datos construido ha sido publicado en la plataforma *Kaggle*, cuyo enlace se proporciona en el Anexo A.

- **O2. Preprocesar y limpiar los datos.** En la segunda fase de trabajo (Sección 5.2) se ha implementado un *script* en *Python* que ejecuta secuencialmente una serie de tareas de preprocesamiento sobre los datos en crudo del *dataset* original, dando lugar a un conjunto de datos final listo para ser utilizado como fuente para las técnicas de minería de datos.

Entre las actividades de preprocesamiento llevadas a cabo, se ha comenzado segmentando los ficheros con las trayectorias de los marcadores para facilitar el análisis de la

marcha, además de interpolar los valores ausentes y suavizar estas series para reducir el ruido (Sección 5.2.1). A partir de estos datos y de las curvas cinemáticas generadas por el sistema de captura, se ha llevado a cabo un proceso de ingeniería de características para crear un conjunto de variables predictoras propias de los análisis de marcha (Sección 5.2.2), que, además, han sido normalizadas para eliminar la influencia de las medidas antropométricas de cada sujeto (Sección 5.2.4). Finalmente, se han limpiado los datos eliminando zancadas distorsionadas, incorrectas o inexactas en base a diferentes criterios: giros durante la marcha, valores atípicos en los parámetros espaciotemporales y, problemas en las curvas cinemáticas (Sección 5.2.3).

- **O3. Aplicar diversas técnicas de *Machine Learning*.** En la Sección 5.3 se ha descrito cómo se han abordado seis tareas supervisadas de clasificación binaria, relacionadas con varios de los ítems más relevantes de la escala de Tinetti para valoración de la marcha.

En cada uno de los subproblemas se ha seguido un procedimiento similar, comenzando por dividir el conjunto de datos en entrenamiento (66 %) y prueba (33 %) mediante un muestreo estratificado en base a la etiqueta de salida; seleccionar las características más relevantes para cada clasificador individual, aplicando un enfoque supervisado de filtrado basado en la correlación estadística entre cada predictor y la variable de respuesta (*ANOVA*); escalar los datos de entrada; entrenar varios modelos basados en diversos algoritmos (*Complement Naive Bayes*, *KNN*, árboles de decisión, *SVM* y redes neuronales), tipo y número de predictores (solo parámetros espaciotemporales, solo atributos cinemáticos o una combinación de ambos). Además, para cada modelo entrenado se han ajustado sus hiperparámetros mediante un proceso de búsqueda exhaustiva.

- **O4. Evaluar la capacidad predictiva de los modelos de Aprendizaje Automático.** Para evaluar el rendimiento que ofrecen los modelos entrenados y ajustados en O3 con datos desconocidos, se han calculado varias métricas de evaluación sobre el conjunto de prueba. En particular, se ha utilizado la exactitud (*accuracy*), valor F_1 (media armónica de sensibilidad y precisión) y área bajo la curva *ROC* (*AUC*). En las Secciones 5.3.3 a 5.3.8 se han presentado los resultados obtenidos para cada una de las tareas de Aprendizaje Automático abordadas, seleccionado en cada caso, aquellos clasificadores que ofrecen un mejor rendimiento en base a estas métricas de evaluación.
- **O5. Desplegar e integrar los modelos de *Machine Learning* elaborados para la valoración automática de la marcha.** En la última fase del trabajo (Sección 5.4), se ha diseñado y desarrollado una herramienta que integra e implementa los modelos predictivos seleccionados en las etapas anteriores, de forma que, partiendo de los datos en crudo de una captura de movimiento, genera automáticamente un informe en formato *PDF* con las respuestas a la escala de Tinetti. Además, estos informes de resultados incluyen gráficos y tablas para comunicar y presentar todos los datos espaciotemporales y cinemáticos que han sido generados durante el proceso de preparación de datos. Como ejemplo, en el Anexo E se adjunta un informe generado por esta aplicación.

6.2. APLICACIÓN DE COMPETENCIAS ADQUIRIDAS EN EL MÁSTER

Para lograr la consecución de objetivos descrita en la sección anterior, en el presente TFM se han trabajado las siguientes **Competencias Generales (CG)** y **Competencias Específicas (CE)** correspondientes al Máster Universitario en Ingeniería y Ciencia de Datos, impartido por la **Universidad Nacional de Educación a Distancia (UNED)** (Tabla 6.1):

Tabla 6.1: Aplicación de competencias adquiridas en el Máster

Competencia	Justificación
CE2. Desarrollar aplicaciones/servicios/scripts orientados a la analítica de datos y analizar el uso de diferentes librerías para el desarrollo e implementación de métodos numéricos, algoritmos y modelos asociados a los datos.	Desarrollo de un <i>script</i> en <i>Python</i> para preprocesar y limpiar los datos en crudo proporcionados por el sistema de captura de movimiento (Sección 5.2.5), haciendo uso de las funciones y estructuras ofrecidas por <i>pandas</i> , <i>numpy</i> y <i>scipy</i> . Implementación de modelos predictivos para la automatización de la prueba de Tinetti (Sección), utilizando las librerías de <i>sklearn</i> , <i>tensorflow</i> y <i>keras</i> .
GC2. Ser capaz de aplicar diferentes técnicas de aprendizaje máquina, seleccionando el algoritmo óptimo que genere modelos precisos y permita el desarrollo de soluciones predictivas en diferentes ámbitos de uso.	Entrenamiento supervisado de modelos predictivos basados en los algoritmos de CNB, KNN, SVM, árboles de decisión y redes neuronales (MLP), descrito en la Sección 5.3. Asimismo, se han perfeccionado estos modelos ajustando sus hiperparámetros mediante un proceso de búsqueda exhaustiva, y se han calculado diferentes métricas de evaluación para seleccionar los clasificadores con mejor rendimiento
CE6. Diseñar mecanismos de evaluación de modelos de aprendizaje y comprender las métricas usadas para dicha evaluación.	Definición y justificación de un procedimiento para evaluar la capacidad predictiva de los modelos de ML (Sección 5.3.2), basado en la división estratificada del conjunto de datos en entrenamiento y prueba, y el cálculo de la exactitud, F_1 (sensibilidad y precisión) y el área bajo la curva ROC (AUC) sobre datos no vistos durante el entrenamiento.
GC4. Aplicar las técnicas de visualización sobre grandes volúmenes de datos para comunicar y presentar los resultados más relevantes del análisis de dicha información a diferentes roles de la organización.	Uso de diferentes técnicas de visualización para comprender mejor los datos exportados desde el sistema de captura de movimiento (Sección 5.1.4), así como para llevar a cabo un análisis exploratorio inicial de las características preprocesadas (Sección 5.3.1) Además, se han utilizado las herramientas de visualización más apropiadas para representar los resultados de la valoración de la marcha en unos informes PDF (Sección 5.4)

6.3. TRABAJO FUTURO

Tal y como se ha justificado en la Sección 6.1, el desarrollo del proyecto ha cumplido con todos los objetivos que se plantearon al inicio del presente TFM. No obstante, el trabajo realizado se podría extender considerando otras escalas y pruebas para la valoración de la

marcha, ampliando el tamaño del *dataset* o utilizando otras técnicas de *Machine Learning* que no han sido exploradas en el presente trabajo. A continuación, se describen brevemente algunas de estas potenciales líneas de trabajo futuro o ampliaciones.

- **Ampliar del conjunto de datos.** La mayoría de algoritmos de *Machine Learning* requieren una gran cantidad de ejemplos de entrenamiento para funcionar bien, y que además, estos datos sean representativos de los casos nuevos a los que se pretende generalizar. En este sentido, teniendo en cuenta que el deterioro físico inherente al envejecimiento conlleva trastornos y alteraciones en la marcha, se propone ampliar el *dataset* con datos de adultos mayores, incluyendo personas en estado de pre-fragilidad y/o fragilidad.
- **Completar la prueba de Tinetti,** automatizando los ítems para valoración de la marcha que no han sido abordados en el presente TFM (por ejemplo, la desviación de la trayectoria o la estabilidad del tronco). Asimismo, se propone ampliar este trabajo para cubrir también la subescala de POMA que valora el equilibrio estático. De esta forma, se ofrecería una puntuación global que podría ser interpretada, por ejemplo, para determinar el riesgo de caída en una persona anciana.
- **Evaluar y validar el trabajo desarrollado con otros sistemas de captura y análisis de movimiento** disponibles en el laboratorio *WeCareLab* (sensores inerciales, plantillas instrumentadas o electromiografía), para determinar así la precisión y calidad de los parámetros espaciotemporales y cinemáticos de la marcha que han sido calculados en este trabajo.
- **Dotar a la herramienta de una interfaz gráfica de usuario,** ya que ha sido implementada en forma de interfaz por línea de comandos, por lo que es poco intuitiva y su uso puede resultar complejo para ciertos usuarios. Por tanto, se propone añadir al *script* desarrollado una interfaz gráfica de usuario (GUI) con un diseño intuitivo, visualmente atractivo y que resulte familiar y fácil de usar para cualquier persona.
- **Implementar el sistema en un entorno de producción real.** Como trabajo futuro relacionado con los resultados obtenidos en las ampliaciones anteriores, se propone la colaboración con algunas de las empresas proveedoras de los sistemas de análisis y captura de movimiento disponibles en el laboratorio *WeCareLab* (por ejemplo, STT, desarrolladora del software *Clinical3DMA* para la captura y análisis de datos de movimiento, con la que ya se ha establecido un contacto previo durante la realización del presente TFM). De esta forma, en futuros esfuerzos se podrían incluir los modelos y/o componentes desarrollados en el presente trabajo en dicho *software*, consiguiendo así una transferencia real y práctica a la sociedad en general, y al ámbito clínico en particular.

6.4. DEDICACIÓN Y COSTES

En el Anexo F se incluye una valoración en términos de costes (tiempo y dinero) del trabajo realizado. Se describe la distribución del tiempo invertido durante el desarrollo de este TFM, detallando las horas invertidas en las diferentes fases del proyecto, así como los gastos de personal de *software* y de *hardware* generados durante el desarrollo del TFM.

Bibliografía

- [1] C. Cifuentes, F. Martínez y E. Romero, «Análisis teórico y computacional de la marcha normal y patológica: una revisión», *Revista Med*, vol. 18, n.º 2, pág. 182, 2010. DOI: [10.18359/rmed.1311](https://doi.org/10.18359/rmed.1311).
- [2] M. B. A. Larrieu, «Estudio de la marcha mediante analisis simultáneo de marcadores opticos y EMG», *XXVI SEMINARIO INGENIERIA BIOMEDICA*, 2017.
- [3] F. Pareja, *Más de cien escalas en neurología*. Madrid: Grupo Aula Médica, 2008, ISBN: 978-84-7885-461-5.
- [4] M. E. Tinetti, «Performance-Oriented Assessment of Mobility Problems in Elderly Patients», *Journal of the American Geriatrics Society*, vol. 34, n.º 2, págs. 119-126, 1986. DOI: [10.1111/j.1532-5415.1986.tb05480.x](https://doi.org/10.1111/j.1532-5415.1986.tb05480.x).
- [5] I. González y col., «An Ambulatory System for Gait Monitoring Based on Wireless Sensorized Insoles», *Sensors*, vol. 15, n.º 7, págs. 16 589-16 613, jul. de 2015. DOI: [10.3390/s150716589](https://doi.org/10.3390/s150716589).
- [6] A. V. Voegeli, *Lecciones básicas de biomecánica del aparato locomotor*. Springer Science & Business Media, 2000, ISBN: 9788407001981. dirección: <https://books.google.es/books?id=SQljKjpZwY8C>.
- [7] I. Tolosa Guzman, *Introducción al análisis de la marcha*. Editorial Universidad del Rosario, mayo de 2013. DOI: <https://doi.org/10.2307/j.ctvm203z2>.
- [8] M. Nordin y V. H. Frankel, *Biomecánica basica del sistema musculoesqueletico 3 Ed.* McGraw-Hill Interamericana, 2004, ISBN: 9788448606350. dirección: <https://books.google.es/books?id=bM4fAAAACAAJ>.
- [9] C. Hernández, A. B. Meseguer Hernández y M. Carles Pérez, *Manual teórico-práctico de valoración y abordaje fisioterápico: Extremidad superior e inferior marcha y postura*. Editum. Ediciones de la Universidad de Murcia, 2012, ISBN: 9788415463436.
- [10] J. Perry y col., *Análisis de la Marcha. Función normal y patológica*. Editorial Base, 2015, ISBN: 978-84-15706-48-9.
- [11] J. Daza Lesmes, *Evaluación clínico-funcional del movimiento corporal humano*. Editorial Médica Panamericana, 2007, ISBN: 9789589181614.
- [12] T. Stöckel y col., «The mental representation of the human gait in young and older adults», *Frontiers in Psychology*, vol. 6, jul. de 2015. DOI: [10.3389/fpsyg.2015.00943](https://doi.org/10.3389/fpsyg.2015.00943).
- [13] D. E. Andisco y J. C. Muñoz, *Conceptos de biomecánica. Con aplicaciones para ciencias de la salud y el deporte*. Impresiones Buenos Aires Editorial, 2006, ISBN: 978-987-22305-3-1.

- [14] A. V. Moreno, E. G. Gutiérrez y J. C. P. Moreno, «Consideraciones para el análisis de la marcha humana. Técnicas de videogrametría, electromiografía y dinamometría.», *Revista Ingeniería Biomédica*, 2008. DOI: [10.24050/19099762.N3.2008.39](https://doi.org/10.24050/19099762.N3.2008.39).
- [15] J. Cámara, «Análisis de la marcha: sus fases y variables espacio-temporales», *Entramado*, vol. 7, n.º 1, págs. 160-173, 2011, ISSN: 1900-3803.
- [16] P. P. Soriano y S. L. Belloch, *Biomecánica básica aplicada a la actividad física y el deporte*. 2015, ISBN: 978-84-9910-180-4.
- [17] M. P. Murray, A. B. Drought y R. C. Kory, «Walking Patterns of Normal Men», *The Journal of bone and joint surgery. American volume*, vol. 46, págs. 335-60, 1964.
- [18] A. H. Copado y E. S. Cogolludo, *Análisis y biomecánica del movimiento (Valoración de la condición física e intervención en accidentes)*, Editex, ed. Editex, 2019, ISBN: 9788413211091.
- [19] M. Latarjet y A. R. Liard, *Anatomía humana. Tomo 1*. Ed. Médica Panamericana, 2004, ISBN: 9789500613682.
- [20] M. W. Whittle, *An Introduction to Gait Analysis*. Butterworth-Heinemann Ltd, 2006, ISBN: 978-0750688833.
- [21] R. H. Fitzgerald, H. Kaufer y A. L. Malkani, *Ortopedia. Tomo 1*. Ed. Médica Panamericana, 2004, ISBN: 978-8479037949.
- [22] À. M. Vived, *Fundamentos de fisiología de la actividad física y el deporte*. 2005.
- [23] C. A. Fukuchi, R. K. Fukuchi y M. Duarte, «A public dataset of overground and treadmill walking kinematics and kinetics in healthy individuals», *PeerJ*, vol. 6, abr. de 2018. DOI: [10.7717/peerj.4640](https://doi.org/10.7717/peerj.4640).
- [24] D. Neumann, *Kinesiology of the musculoskeletal system : foundations for rehabilitation*. St. Louis, Mo: Mosby/Elsevier, 2010, ISBN: 9780323039895.
- [25] J. Dicharry, «Kinematics and Kinetics of Gait: From Lab to Clinic», *Clinics in Sports Medicine*, vol. 29, n.º 3, págs. 347-364, jul. de 2010. DOI: [10.1016/j.csm.2010.03.013](https://doi.org/10.1016/j.csm.2010.03.013).
- [26] R. W. Baker, *Measuring Walking*. Mac Keith Press, 23 de abr. de 2013, 246 págs., ISBN: 1908316667.
- [27] O. Pinzone, M. H. Schwartz y R. Baker, «Comprehensive non-dimensional normalization of gait data», *Gait & Posture*, vol. 44, págs. 68-73, feb. de 2016. DOI: [10.1016/j.gaitpost.2015.11.013](https://doi.org/10.1016/j.gaitpost.2015.11.013).
- [28] A. L. Hof, «Scaling gait data to body size», *Gait & Posture*, vol. 4, n.º 3, págs. 222-223, mayo de 1996. DOI: [10.1016/0966-6362\(95\)01057-2](https://doi.org/10.1016/0966-6362(95)01057-2).
- [29] E. Chao y col., «Normative data of knee joint motion and ground reaction forces in adult level walking», *Journal of Biomechanics*, vol. 16, n.º 3, págs. 219-233, ene. de 1983. DOI: [10.1016/0021-9290\(83\)90129-x](https://doi.org/10.1016/0021-9290(83)90129-x).
- [30] T. M. Owings y M. D. Grabiner, «Variability of step kinematics in young and older adults», *Gait & Posture*, vol. 20, n.º 1, págs. 26-29, ago. de 2004. DOI: [10.1016/s0966-6362\(03\)00088-2](https://doi.org/10.1016/s0966-6362(03)00088-2).
- [31] E. B. Titianova y col., «Gait Characteristics and Functional Ambulation Profile in Patients with Chronic Unilateral Stroke», *American Journal of Physical Medicine & Rehabilitation*, vol. 82, n.º 10, págs. 778-786, oct. de 2003. DOI: [10.1097/01.phm.0000087490.74582.e0](https://doi.org/10.1097/01.phm.0000087490.74582.e0).

- [32] V. Macellari, C. Giacomozzi y R. Saggini, «Spatial-temporal parameters of gait: reference data and a statistical method for normality assessment», *Gait & Posture*, vol. 10, n.º 2, págs. 171-181, oct. de 1999. DOI: [10.1016/s0966-6362\(99\)00021-1](https://doi.org/10.1016/s0966-6362(99)00021-1).
- [33] D. A. Winter, «Kinematic and kinetic patterns in human gait: Variability and compensating effects», *Human Movement Science*, vol. 3, n.º 1-2, págs. 51-76, mar. de 1984. DOI: [10.1016/0167-9457\(84\)90005-8](https://doi.org/10.1016/0167-9457(84)90005-8).
- [34] H. Gill y J. O'Connor, «Heelstrike and the pathomechanics of osteoarthritis: a pilot gait study», *Journal of Biomechanics*, vol. 36, n.º 11, págs. 1625-1631, nov. de 2003. DOI: [10.1016/s0021-9290\(03\)00189-1](https://doi.org/10.1016/s0021-9290(03)00189-1).
- [35] S. A. Gard, «Use of Quantitative Gait Analysis for the Evaluation of Prosthetic Walking Performance», *JPO Journal of Prosthetics and Orthotics*, vol. 18, n.º Proceedings, P93-P104, ene. de 2006. DOI: [10.1097/00008526-200601001-00011](https://doi.org/10.1097/00008526-200601001-00011).
- [36] H. W. Wallmann, «Introduction to Observational Gait Analysis», *Home Health Care Management & Practice*, vol. 22, n.º 1, págs. 66-68, ago. de 2009. DOI: [10.1177/1084822309347344](https://doi.org/10.1177/1084822309347344).
- [37] L. M. Gutiérrez-Robledo y C. D'Hyver, *Geriatría*. México: El Manual Moderno, 2014, ISBN: 9786074484038.
- [38] J. C. Paz y M. P. West, *Acute Care Handbook for Physical Therapists*. Elsevier, 2014. DOI: <https://doi.org/10.1016/C2011-0-05707-1>.
- [39] P. A. Soler y L. R. Mañas, *Tratado de medicina geriátrica : fundamentos de la atención sanitaria a los mayores*. Barcelona etc: Elsevier, 2020, ISBN: 9788491132981.
- [40] C. Roqueta y col., «Experiencia en la evaluación del riesgo de caídas. Comparación entre el test de Tinetti y el Timed Up & Go», *Revista Española de Geriatría y Gerontología*, vol. 42, n.º 6, págs. 319-327, nov. de 2007. DOI: [10.1016/s0211-139x\(07\)73570-9](https://doi.org/10.1016/s0211-139x(07)73570-9).
- [41] S. Mathias, U. S. Nayak y B. Isaacs, «Balance in elderly patients: the "get-up and go" test.», *Archives of physical medicine and rehabilitation*, vol. 67, págs. 387-9, 6 1986.
- [42] D. Podsiadlo y S. Richardson, «The Timed "Up & Go": A Test of Basic Functional Mobility for Frail Elderly Persons», *Journal of the American Geriatrics Society*, vol. 39, n.º 2, págs. 142-148, feb. de 1991. DOI: [10.1111/j.1532-5415.1991.tb01616.x](https://doi.org/10.1111/j.1532-5415.1991.tb01616.x).
- [43] J. Daly y col., «Development and testing of the Gait Assessment and Intervention Tool (G.A.I.T.): A measure of coordinated gait components», *Journal of Neuroscience Methods*, vol. 178, n.º 2, págs. 334-339, abr. de 2009. DOI: [10.1016/j.jneumeth.2008.12.016](https://doi.org/10.1016/j.jneumeth.2008.12.016).
- [44] M. D. G. Fogeda y col., «Spanish Cross-cultural Adaptation of the Gait Assessment and Intervention Tool», *PM&R*, vol. 11, n.º 9, págs. 954-962, mar. de 2019. DOI: [10.1002/pmrj.12054](https://doi.org/10.1002/pmrj.12054).
- [45] J. Saengsuwan y P. Sirasaporn, «Validity and Reliability of the Thai Version of the Gait Assessment and Intervention Tool (G.A.I.T.)», *Stroke Research and Treatment*, vol. 2020, págs. 1-6, ago. de 2020. DOI: [10.1155/2020/1710534](https://doi.org/10.1155/2020/1710534).
- [46] L. Wolfson y col., «Gait Assessment in the Elderly: A Gait Abnormality Rating Scale and Its Relation to Falls», *Journal of Gerontology*, vol. 45, n.º 1, págs. M12-M19, ene. de 1990. DOI: [10.1093/geronj/45.1.m12](https://doi.org/10.1093/geronj/45.1.m12).
- [47] M. Cameron y L. Monroe, *Physical rehabilitation for the physical therapist assistant*. St. Louis, Mo: Elsevier/Saunders, 2011, ISBN: 9780323290937.

- [48] J. M. VanSwearingen y col., «The Modified Gait Abnormality Rating Scale for Recognizing the Risk of Recurrent Falls in Community-Dwelling Elderly Adults», *Physical Therapy*, vol. 76, n.º 9, págs. 994-1002, sep. de 1996. DOI: [10.1093/ptj/76.9.994](https://doi.org/10.1093/ptj/76.9.994).
- [49] A. A. Rodríguez y col., «Gait training efficacy using a home-based practice model in chronic hemiplegia», *Archives of Physical Medicine and Rehabilitation*, vol. 77, n.º 8, págs. 801-805, ago. de 1996. DOI: [10.1016/s0003-9993\(96\)90260-9](https://doi.org/10.1016/s0003-9993(96)90260-9).
- [50] C. Estrada-Barranco, R. Cano-de-la-Cuerda y F. Molina-Rueda, «Construct validity of the Wisconsin Gait Scale in acute, subacute and chronic stroke», *Gait & Posture*, vol. 68, págs. 363-368, feb. de 2019. DOI: [10.1016/j.gaitpost.2018.12.020](https://doi.org/10.1016/j.gaitpost.2018.12.020).
- [51] V. Kumar, K. Arya y S. Pandian, «Rivermead Visual Gait Assessment: A Reliable and Valid Post-Stroke Gait Measure for Developing Countries», oct. de 2016.
- [52] A. Hochsprung y col., «Validez clínica de las variables cuantitativas de la marcha en pacientes con esclerosis múltiple. Comparativa entre el test de los 25 pies y el sistema electrónico de marcha GAITRite», *Revista de Neurología*, vol. 59, n.º 01, pág. 8, 2014. DOI: [10.33588/rn.5901.2013413](https://doi.org/10.33588/rn.5901.2013413).
- [53] M. G. del Río y M. M. Ruiz, «Escala de movilidad de 12 Ítems para esclerosis múltiple: análisis mediante diagnósticos de enfermería», *Revista Científica de la Sociedad Española de Enfermería Neurológica*, vol. 51, págs. 23-26, ene. de 2020. DOI: [10.1016/j.sedene.2019.05.001](https://doi.org/10.1016/j.sedene.2019.05.001).
- [54] R. N. Boyd y H. K. Graham, «Objective measurement of clinical findings in the use of botulinum toxin type A for the management of children with cerebral palsy», *European Journal of Neurology*, vol. 6, s23-s35, nov. de 1999. DOI: [10.1111/j.1468-1331.1999.tb00031.x](https://doi.org/10.1111/j.1468-1331.1999.tb00031.x).
- [55] T. Herman y col., «The Dynamic Gait Index in healthy older adults: The role of stair climbing, fear of falling and gender», *Gait & Posture*, vol. 29, n.º 2, págs. 237-241, feb. de 2009. DOI: [10.1016/j.gaitpost.2008.08.013](https://doi.org/10.1016/j.gaitpost.2008.08.013).
- [56] S. H. An y col., «Validity of the Original and Short Versions of the Dynamic Gait Index in Predicting Falls in Stroke Survivors», *Rehabilitation Nursing*, vol. 42, n.º 6, págs. 325-332, nov. de 2017. DOI: [10.1002/rnj.280](https://doi.org/10.1002/rnj.280).
- [57] H. Read y col., *Journal of Pediatric Orthopedics*, vol. 23, n.º 3, págs. 296-301, mayo de 2003. DOI: [10.1097/00004694-200305000-00005](https://doi.org/10.1097/00004694-200305000-00005).
- [58] A. Ong, S. Hillman y J. Robb, «Reliability and validity of the Edinburgh Visual Gait Score for cerebral palsy when used by inexperienced observers», *Gait & Posture*, vol. 28, n.º 2, págs. 323-326, ago. de 2008. DOI: [10.1016/j.gaitpost.2008.01.008](https://doi.org/10.1016/j.gaitpost.2008.01.008).
- [59] B. Toro, C. J. Nester y P. C. Farren, «The Development and Validity of the Salford Gait Tool: An Observation-Based Clinical Gait Assessment Tool», *Archives of Physical Medicine and Rehabilitation*, vol. 88, n.º 3, págs. 321-327, mar. de 2007. DOI: [10.1016/j.apmr.2006.12.028](https://doi.org/10.1016/j.apmr.2006.12.028).
- [60] C. R. Guevara y L. H. Lugo, «Validez y confiabilidad de la Escala de Tinetti para población colombiana», *Revista Colombiana de Reumatología*, vol. 19, n.º 4, págs. 218-233, dic. de 2012. DOI: [10.1016/s0121-8123\(12\)70017-8](https://doi.org/10.1016/s0121-8123(12)70017-8).
- [61] D. M. Haro, «Laboratorio de análisis de marcha y movimiento», *Revista Médica Clínica Las Condes*, vol. 25, n.º 2, págs. 237-247, mar. de 2014. DOI: [10.1016/s0716-8640\(14\)70034-3](https://doi.org/10.1016/s0716-8640(14)70034-3).

- [62] B. M., R. R y A. R., «Comparison of two Motion Capture Systems by means of joint Trajectories of human gait», Tesis doct., mayo de 2016. DOI: [10.17488/rmib.37.2.2](https://doi.org/10.17488/rmib.37.2.2).
- [63] *Análisis cinemático y cinético de la marcha humana*, vol. 17, 1, 2003.
- [64] T. Renkawitz y col., «Leg length and offset differences above 5 mm after total hip arthroplasty are associated with altered gait kinematics», *Gait & Posture*, vol. 49, págs. 196-201, sep. de 2016. DOI: [10.1016/j.gaitpost.2016.07.011](https://doi.org/10.1016/j.gaitpost.2016.07.011).
- [65] Steffen y col., «Gait analysis with IMU: Gaining new orientation information of the lower leg», en *Proceedings of the International Conference on Biomedical Electronics and Devices*, SCITEPRESS - Science, 2014. DOI: [10.5220/0004787701270133](https://doi.org/10.5220/0004787701270133).
- [66] R. Martínez Mendez y col., «Uso de sensores inerciales en la medición y evaluación de movimiento humano para aplicaciones en la salud», *Ideas en ciencia*, vol. 37, págs. 61-75, jun. de 2012.
- [67] N. F. Ribeiro y C. P. Santos, «Inertial measurement units: A brief state of the art on gait analysis», en *2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG)*, IEEE, 2017. DOI: [10.1109/enbeng.2017.7889458](https://doi.org/10.1109/enbeng.2017.7889458).
- [68] A. Muro-de-la-Herran, B. Garcia-Zapirain y A. Mendez-Zorrilla, «Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications», *Sensors*, vol. 14, n.º 2, págs. 3362-3394, feb. de 2014. DOI: [10.3390/s140203362](https://doi.org/10.3390/s140203362).
- [69] M. Paulich y col., «Xsens MTw Awinda: Miniature Wireless Inertial-Magnetic Motion Tracker for Highly Accurate 3D Kinematic Applications», 2018. DOI: [10.13140/RG.2.2.23576.49929](https://doi.org/10.13140/RG.2.2.23576.49929).
- [70] M. L. Romero, M. T. A. Carrere y L. L. Alcázar, «Actividad electromiográfica normal en la marcha humana», *Biomechanica*, vol. IV, n.º 7, dic. de 1996. DOI: [10.5821/sibb.v4i7.1580](https://doi.org/10.5821/sibb.v4i7.1580).
- [71] V. Agostini y col., «Surface Electromyography Applied to Gait Analysis: How to Improve Its Impact in Clinics?», *Frontiers in Neurology*, vol. 11, sep. de 2020. DOI: [10.3389/fneur.2020.00994](https://doi.org/10.3389/fneur.2020.00994).
- [72] J. Perry, «TWO Chapter One The Contribution of Dynamic Electromyography to Gait Analysis», 2005.
- [73] I. M. Rull y R. C. M. Marrero, *Biomecánica clínica de las patologías del aparato locomotor*. Elsevier Masson, 2016, ISBN: 9788445816806.
- [74] W. Tao y col., «Gait Analysis Using Wearable Sensors», *Sensors*, vol. 12, n.º 2, págs. 2255-2283, feb. de 2012. DOI: [10.3390/s120202255](https://doi.org/10.3390/s120202255).
- [75] I. G. Diaz, «The role of gait variability in ageing, frailty and the onset of cognitive decline. Exploring quantitative techniques for gait analysis.», Universidad de Castilla-La Mancha, 2018.
- [76] S. Karia y col., «Plantar pressure distribution and gait stability: Normal VS high heel», en *2016 2nd IEEE International Symposium on Robotics and Manufacturing Automation (ROMA)*, IEEE, sep. de 2016. DOI: [10.1109/roma.2016.7847822](https://doi.org/10.1109/roma.2016.7847822).
- [77] E. Surer y A. Kose, «Methods and Technologies for Gait Analysis», en *Computer Analysis of Human Behavior*, Springer London, 2011, págs. 105-123. DOI: [10.1007/978-0-85729-994-9_5](https://doi.org/10.1007/978-0-85729-994-9_5).

- [78] C. He y col., «Dynamic stability and spatiotemporal parameters during turning in healthy young adults», *BioMedical Engineering*, vol. 17, n.º 1, sep. de 2018. DOI: [10.1186/s12938-018-0558-5](https://doi.org/10.1186/s12938-018-0558-5).
- [79] P. Khera y N. Kumar, «Role of machine learning in gait analysis: a review», *Journal of Medical Engineering & Technology*, vol. 44, n.º 8, págs. 441-467, oct. de 2020. DOI: [10.1080/03091902.2020.1822940](https://doi.org/10.1080/03091902.2020.1822940).
- [80] S. M. Alfayeed y B. S. Saini, «Human Gait Analysis Using Machine Learning: A Review», en *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, IEEE, mar. de 2021. DOI: [10.1109/iccike51210.2021.9410678](https://doi.org/10.1109/iccike51210.2021.9410678).
- [81] R. Williamson y B. Andrews, «Gait event detection for FES using accelerometers and supervised machine learning», *IEEE Transactions on Rehabilitation Engineering*, vol. 8, n.º 3, págs. 312-319, 2000. DOI: [10.1109/86.867873](https://doi.org/10.1109/86.867873).
- [82] M. Lempereur y col., «A new deep learning-based method for the detection of gait events in children with gait disorders: Proof-of-concept and concurrent validity», *Journal of Biomechanics*, vol. 98, pág. 109 490, ene. de 2020. DOI: [10.1016/j.jbiomech.2019.109490](https://doi.org/10.1016/j.jbiomech.2019.109490).
- [83] M. Zago y col., «Machine-Learning Based Determination of Gait Events from Foot-Mounted Inertial Units», *Sensors*, vol. 21, n.º 3, pág. 839, ene. de 2021. DOI: [10.3390/s21030839](https://doi.org/10.3390/s21030839).
- [84] J. D. Farah, N. Baddour y E. D. Lemaire, «Gait phase detection from thigh kinematics using machine learning techniques», en *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, mayo de 2017. DOI: [10.1109/memea.2017.7985886](https://doi.org/10.1109/memea.2017.7985886).
- [85] J. Taborri y col., «A Novel HMM Distributed Classifier for the Detection of Gait Phases by Means of a Wearable Inertial Sensor Network», *Sensors*, vol. 14, n.º 9, págs. 16 212-16 234, sep. de 2014. DOI: [10.3390/s140916212](https://doi.org/10.3390/s140916212).
- [86] F. M. Rueda y col., «Convolutional Neural Networks for Human Activity Recognition Using Body-Worn Sensors», *Informatics*, vol. 5, n.º 2, pág. 26, mayo de 2018. DOI: [10.3390/informatics5020026](https://doi.org/10.3390/informatics5020026).
- [87] W.-M. Lee, *Python machine learning*. Indianapolis, IN: Wiley, 2019, ISBN: 9781119545637.
- [88] P. Dixon y col., «Machine learning algorithms can classify outdoor terrain types during running using accelerometry data», *Gait & Posture*, vol. 74, págs. 176-181, oct. de 2019. DOI: [10.1016/j.gaitpost.2019.09.005](https://doi.org/10.1016/j.gaitpost.2019.09.005).
- [89] B. Hu y col., «Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface- and age-related differences in walking», *Journal of Biomechanics*, vol. 71, págs. 37-42, abr. de 2018. DOI: [10.1016/j.jbiomech.2018.01.005](https://doi.org/10.1016/j.jbiomech.2018.01.005).
- [90] J. Stevenson, S. Firebaugh y H. Charles, «Biometric Identification from a Floor Based PVDF Sensor Array Using Hidden Markov Models», sep. de 2007.
- [91] X. Zeng y col., «Gait-Based Implicit Authentication Using Edge Computing and Deep Learning for Mobile Devices», *Sensors*, vol. 21, n.º 13, pág. 4592, jul. de 2021. DOI: [10.3390/s21134592](https://doi.org/10.3390/s21134592).
- [92] G. Giorgi, A. Saracino y F. Martinelli, «Using recurrent neural networks for continuous authentication through gait analysis», *Pattern Recognition Letters*, vol. 147, págs. 157-163, jul. de 2021. DOI: [10.1016/j.patrec.2021.03.010](https://doi.org/10.1016/j.patrec.2021.03.010).

- [93] C. BenAbdelkader y col., «EigenGait: Motion-Based Recognition of People Using Image Self-Similarity», en *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2001, págs. 284-294. DOI: [10.1007/3-540-45344-x_42](https://doi.org/10.1007/3-540-45344-x_42).
- [94] M. Alaqtash y col., «Automatic classification of pathological gait patterns using ground reaction forces and machine learning algorithms», en *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, ago. de 2011. DOI: [10.1109/iembs.2011.6090063](https://doi.org/10.1109/iembs.2011.6090063).
- [95] M. Khokhlova y col., «Normal and pathological gait classification LSTM model», *Artificial Intelligence in Medicine*, vol. 94, págs. 54-66, mar. de 2019. DOI: [10.1016/j.artmed.2018.12.007](https://doi.org/10.1016/j.artmed.2018.12.007).
- [96] M. Nieto-Hidalgo y col., «A vision based proposal for classification of normal and abnormal gait using RGB camera», *Journal of Biomedical Informatics*, vol. 63, págs. 82-89, oct. de 2016. DOI: [10.1016/j.jbi.2016.08.003](https://doi.org/10.1016/j.jbi.2016.08.003).
- [97] J. Luo, J. Tang y X. Xiao, «Abnormal Gait Behavior Detection for Elderly Based on Enhanced Wigner-Ville Analysis and Cloud Incremental SVM Learning», *Journal of Sensors*, vol. 2016, págs. 1-18, 2016. DOI: [10.1155/2016/5869238](https://doi.org/10.1155/2016/5869238).
- [98] C. Fricke y col., «Evaluation of Three Machine Learning Algorithms for the Automatic Classification of EMG Patterns in Gait Disorders», *Frontiers in Neurology*, vol. 12, mayo de 2021. DOI: [10.3389/fneur.2021.666458](https://doi.org/10.3389/fneur.2021.666458).
- [99] L. B. Peres y col., «Discrimination between healthy and patients with Parkinson's disease from hand resting activity using inertial measurement unit», *BioMedical Engineering OnLine*, vol. 20, n.º 1, mayo de 2021. DOI: [10.1186/s12938-021-00888-2](https://doi.org/10.1186/s12938-021-00888-2).
- [100] H. Ng y col., «Gait Classification by Support Vector Machine», en *Software Engineering and Computer Systems*, J. Mohamad Zain, W. M. b. Wan Mohd y E. El-Qawasmeh, eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, págs. 623-636, ISBN: 978-3-642-22170-5.
- [101] R. Collins, R. Gross y J. Shi, «Silhouette-based human identification from body shape and gait», en *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, IEEE, 2002. DOI: [10.1109/afgr.2002.1004181](https://doi.org/10.1109/afgr.2002.1004181).
- [102] H.-Y. Lau, K.-Y. Tong y H. Zhu, «Support vector machine for classification of walking conditions using miniature kinematic sensors», *Medical & Biological Engineering & Computing*, vol. 46, n.º 6, págs. 563-573, mar. de 2008. DOI: [10.1007/s11517-008-0327-x](https://doi.org/10.1007/s11517-008-0327-x).
- [103] R. Begg, M. Palaniswami y B. Owen, «Support Vector Machines for Automated Gait Classification», *IEEE Transactions on Biomedical Engineering*, vol. 52, n.º 5, págs. 828-838, mayo de 2005. DOI: [10.1109/tbme.2005.845241](https://doi.org/10.1109/tbme.2005.845241).
- [104] J. Kamruzzaman y R. Begg, «Support Vector Machines and Other Pattern Recognition Approaches to the Diagnosis of Cerebral Palsy Gait», *IEEE Transactions on Biomedical Engineering*, vol. 53, n.º 12, págs. 2479-2490, dic. de 2006. DOI: [10.1109/tbme.2006.883697](https://doi.org/10.1109/tbme.2006.883697).
- [105] H. Gou, L. Yan y J. Xiao, «A Gait Recognition System based on SVM and Accelerations», *MATEC Web of Conferences*, vol. 30, J. Zhou y O. Adiguzel, eds., pag. 06 001, 2015. DOI: [10.1051/mateconf/20153006001](https://doi.org/10.1051/mateconf/20153006001).
- [106] D. Szczepański, «Multilayer perceptron for gait type classification based on inertial sensors data», en *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2016, págs. 947-950.

- [107] V. B. Semwal, M. Raj y G. Nandi, «Biometric gait identification based on a multilayer perceptron», *Robotics and Autonomous Systems*, vol. 65, págs. 65-75, mar. de 2015. DOI: [10.1016/j.robot.2014.11.010](https://doi.org/10.1016/j.robot.2014.11.010).
- [108] A. S. Alharthi, S. U. Yunas y K. B. Ozanyan, «Deep Learning for Monitoring of Human Gait: A Review», *IEEE Sensors Journal*, vol. 19, n.º 21, págs. 9575-9591, nov. de 2019. DOI: [10.1109/jsen.2019.2928777](https://doi.org/10.1109/jsen.2019.2928777).
- [109] K. Açıcı y col., «A Random Forest Method to Detect Parkinson's Disease via Gait Analysis», en *Engineering Applications of Neural Networks*, Springer International Publishing, 2017, págs. 609-619. DOI: [10.1007/978-3-319-65172-9_51](https://doi.org/10.1007/978-3-319-65172-9_51).
- [110] S. Pasinetti y col., «Assisted Gait Phase Estimation Through an Embedded Depth Camera Using Modified Random Forest Algorithm Classification», *IEEE Sensors Journal*, vol. 20, n.º 6, págs. 3343-3355, 2020. DOI: [10.1109/JSEN.2019.2957667](https://doi.org/10.1109/JSEN.2019.2957667).
- [111] Z. Chen y col., «Prediction of Freezing of Gait in Parkinson's Disease Using a Random Forest Model Based on an Orthogonal Experimental Design: A Pilot Study», *Frontiers in Human Neuroscience*, vol. 15, mar. de 2021. DOI: [10.3389/fnhum.2021.636414](https://doi.org/10.3389/fnhum.2021.636414).
- [112] Q. Guo y D. Jiang, «Method for Walking Gait Identification in a Lower Extremity Exoskeleton Based on C4.5 Decision Tree Algorithm», *International Journal of Advanced Robotic Systems*, vol. 12, n.º 4, pág. 30, abr. de 2015. DOI: [10.5772/60132](https://doi.org/10.5772/60132).
- [113] M. Derlatka y M. Ichnatowski, «Decision Tree Approach to Rules Extraction for Human Gait Analysis», en *Artificial Intelligence and Soft Computing*, L. Rutkowski y col., eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, págs. 597-604, ISBN: 978-3-642-13208-7.
- [114] H. H. Manap, N. M. Tahir y R. Abdullah, «Parkinsonian Gait Motor Impairment Detection Using Decision Tree», en *2013 European Modelling Symposium*, IEEE, nov. de 2013. DOI: [10.1109/ems.2013.36](https://doi.org/10.1109/ems.2013.36).
- [115] J. Bae y M. Tomizuka, «Gait Phase Analysis based on a Hidden Markov Model», *IFAC Proceedings Volumes*, vol. 43, n.º 18, págs. 746-751, 2010. DOI: [10.3182/20100913-3-us-2015.00014](https://doi.org/10.3182/20100913-3-us-2015.00014).
- [116] A. Mannini y A. M. Sabatini, «A hidden Markov model-based technique for gait segmentation using a foot-mounted gyroscope», en *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, ago. de 2011. DOI: [10.1109/iembs.2011.6091084](https://doi.org/10.1109/iembs.2011.6091084).
- [117] G. Panahandeh y col., «Continuous Hidden Markov Model for Pedestrian Activity Classification and Gait Analysis», *IEEE Transactions on Instrumentation and Measurement*, vol. 62, n.º 5, págs. 1073-1083, mayo de 2013. DOI: [10.1109/tim.2012.2236792](https://doi.org/10.1109/tim.2012.2236792).
- [118] J. Taborri y col., «Real-time gait detection based on Hidden Markov Model: Is it possible to avoid training procedure?», en *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*, IEEE, mayo de 2015. DOI: [10.1109/memea.2015.7145188](https://doi.org/10.1109/memea.2015.7145188).
- [119] B. DeCann, A. Ross y M. Culp, «On Clustering Human Gait Patterns», en *2014 22nd International Conference on Pattern Recognition*, IEEE, ago. de 2014. DOI: [10.1109/icpr.2014.315](https://doi.org/10.1109/icpr.2014.315).
- [120] A. Rozumalski y M. H. Schwartz, «Crouch gait patterns defined using k-means cluster analysis are related to underlying clinical pathology», *Gait & Posture*, vol. 30, n.º 2, págs. 155-160, ago. de 2009. DOI: [10.1016/j.gaitpost.2009.05.010](https://doi.org/10.1016/j.gaitpost.2009.05.010).

- [121] F. Zgolli y col., «Kinematic Data Clustering for Healthy Knee Gait Characterization», en *2018 IEEE Life Sciences Conference (LSC)*, IEEE, oct. de 2018. DOI: [10.1109/lsc.2018.8572119](https://doi.org/10.1109/lsc.2018.8572119).
- [122] I. Milovanović y D. B. Popović, «Principal Component Analysis of Gait Kinematics Data in Acute and Chronic Stroke Patients», *Computational and Mathematical Methods in Medicine*, vol. 2012, págs. 1-8, 2012. DOI: [10.1155/2012/649743](https://doi.org/10.1155/2012/649743).
- [123] G. Guo y col., «Classification of Normal and Pathological Gait in Young Children Based on Foot Pressure Data», *Neuroinformatics*, vol. 15, n.º 1, págs. 13-24, sep. de 2016. DOI: [10.1007/s12021-016-9313-x](https://doi.org/10.1007/s12021-016-9313-x).
- [124] J. E. Boyd y J. J. Little, «Gait Recognition, Silhouette-Based», en *Encyclopedia of Biometrics*, Springer US, 2009, págs. 646-652. DOI: [10.1007/978-0-387-73003-5_36](https://doi.org/10.1007/978-0-387-73003-5_36).
- [125] L. Wang y col., «Silhouette analysis-based gait recognition for human identification», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, n.º 12, págs. 1505-1518, dic. de 2003. DOI: [10.1109/tpami.2003.1251144](https://doi.org/10.1109/tpami.2003.1251144).
- [126] S. Sarkar y Z. Liu, «Gait Recognition», en *Handbook of Biometrics*, A. K. Jain, P. Flynn y A. A. Ross, eds. Boston, MA: Springer US, 2008, págs. 109-129, ISBN: 978-0-387-71041-9. DOI: [10.1007/978-0-387-71041-9_6](https://doi.org/10.1007/978-0-387-71041-9_6). dirección: https://doi.org/10.1007/978-0-387-71041-9_6.
- [127] G. Nagymate y R. M. Kiss, «Application of OptiTrack motion capture systems in human movement analysis», *Recent Innovations in Mechatronics*, vol. 5, n.º 1. Sep. de 2018. DOI: [10.17667/riim.2018.1/13](https://doi.org/10.17667/riim.2018.1/13).
- [128] J. H. Canós y M. Letelier, «Metodologías Ágiles en el Desarrollo de Software», 2012.
- [129] J. Rollins. (2015). «Metodología Fundamental para la Ciencia de Datos», IBM Analytics, dirección: <https://www.ibm.com/downloads/cas/6RZMKDN8>.
- [130] G. Phillips-Wren, A. Esposito y L. C. Jain, *Advances in Data Science: Methodologies and Applications*. Springer-Verlag GmbH, 26 de ago. de 2020, 333 págs. dirección: https://www.ebook.de/de/product/39545623/advances_in_data_science_methodologies_and_applications.html.
- [131] F. Foroughi y P. Luksch, «Data Science Methodology for Cybersecurity Projects», en *Computer Science & Information Technology*, Academy & Industry Research Collaboration Center (AIRCC), feb. de 2018. DOI: [10.5121/csit.2018.80401](https://doi.org/10.5121/csit.2018.80401).
- [132] J. E. Sirvent-Belando y R. P. Garrido Chamorro, *Valoración antropométrica de la composición corporal. Cineantropometría*. San Vicente del Raspeig: Publicaciones de la Universidad de Alicante, 2009, ISBN: 978-84-9717-052-9.
- [133] L. Carmenate Milián, F. A. Moncada Chévez y B. L. Engels Waldemar, *Manual de Medidas Antropométricas*. Publicaciones SALTRA, 2014, ISBN: 978-9968-924-18-4.
- [134] P. F. Herrera y col. (). «A segmentable normative database for gait analysis based on the Helen Hayes markerset», dirección: <https://www.stt-systems.com/downloads/whitepapers/A%20segmentable%20normative%20database%20for%20gait%20analysis%20based%20on%20the%20Helen%20Hayes%20markerset.pdf>.
- [135] C. Charalambous, «Measurement of Lower Extremity Kinematics During Level Walking», *Classic Papers in Orthopaedics*, págs. 397-398, ene. de 2014. DOI: [10.1007/978-1-4471-5451-8_100](https://doi.org/10.1007/978-1-4471-5451-8_100).

- [136] A. Sweigart, *Automate the boring stuff with Python : practical programming for total beginners*. San Francisco: No Starch Press, 2015, ISBN: 9781593275990.
- [137] T. Piazza y col., «Predicting Missing Markers in Real-Time Optical Motion Capture», en *Modelling the Physiological Human*, Springer Berlin Heidelberg, 2009, págs. 125-136. DOI: [10.1007/978-3-642-10470-1_11](https://doi.org/10.1007/978-3-642-10470-1_11).
- [138] F. Crenna, G. B. Rossi y M. Berardengo, «Filtering Biomechanical Signals in Movement Analysis», *Sensors*, vol. 21, n.º 13, pág. 4580, jul. de 2021. DOI: [10.3390/s21134580](https://doi.org/10.3390/s21134580).
- [139] O. Ľupa y col., «Motion tracking and gait feature estimation for recognising Parkinson's disease using MS Kinect», *BioMedical Engineering OnLine*, vol. 14, n.º 1, oct. de 2015. DOI: [10.1186/s12938-015-0092-7](https://doi.org/10.1186/s12938-015-0092-7).
- [140] H. Lanshammar, «On practical evaluation of differentiation techniques for human gait analysis», *Journal of Biomechanics*, vol. 15, n.º 2, págs. 99-105, ene. de 1982. DOI: [10.1016/0021-9290\(82\)90041-0](https://doi.org/10.1016/0021-9290(82)90041-0).
- [141] J. Luo y col., «Properties of Savitzky–Golay digital differentiators», *Digital Signal Processing*, vol. 15, n.º 2, págs. 122-136, mar. de 2005. DOI: [10.1016/j.dsp.2004.09.008](https://doi.org/10.1016/j.dsp.2004.09.008).
- [142] W. H. Press y S. A. Teukolsky, «Savitzky-Golay Smoothing Filters», *Computers in Physics*, vol. 4, n.º 6, pág. 669, 1990. DOI: [10.1063/1.4822961](https://doi.org/10.1063/1.4822961).
- [143] C. Wang y col., «Estimation of Spatial-Temporal Gait Parameters based on the Fusion of Inertial and Film-Pressure Signals», en *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, dic. de 2018. DOI: [10.1109/bibm.2018.8621216](https://doi.org/10.1109/bibm.2018.8621216).
- [144] P. E. Caicedo-Rodríguez, C. F. Rengifo-Rodas y L. E. Rodríguez-Cheu, «A human gait temporal parameters calculation algorithm», en *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th -28th, 2016*, Springer Singapore, 2017, págs. 285-288. DOI: [10.1007/978-981-10-4086-3_72](https://doi.org/10.1007/978-981-10-4086-3_72).
- [145] C. Prakash y col., «Identification of spatio-temporal and kinematics parameters for 2-D optical gait analysis system using passive markers», en *2015 International Conference on Advances in Computer Engineering and Applications*, IEEE, mar. de 2015. DOI: [10.1109/icacea.2015.7164683](https://doi.org/10.1109/icacea.2015.7164683).
- [146] E. E. Cowley, T. L. Chevalier y N. Chockalingam, «The Effect of Heel Height on Gait and Posture», *Journal of the American Podiatric Medical Association*, vol. 99, n.º 6, págs. 512-518, nov. de 2009. DOI: [10.7547/0990512](https://doi.org/10.7547/0990512).
- [147] T. Hoenig, T. Rolvien y K. Hollander, «Footstrike patterns in runners: concepts, classifications, techniques, and implications for running-related injuries», *Deutsche Zeitschrift für Sportmedizin*, vol. 71, n.º 3, págs. 55-61, mar. de 2020. DOI: [10.5960/dzsm.2020.424](https://doi.org/10.5960/dzsm.2020.424).
- [148] B. Mariani y col., «On-Shoe Wearable Sensors for Gait and Turning Assessment of Patients With Parkinson's Disease», *IEEE Transactions on Biomedical Engineering*, vol. 60, n.º 1, págs. 155-158, ene. de 2013. DOI: [10.1109/tbme.2012.2227317](https://doi.org/10.1109/tbme.2012.2227317).
- [149] M. El-Gohary y col., «Continuous Monitoring of Turning in Patients with Movement Disability», *Sensors*, vol. 14, n.º 1, págs. 356-369, dic. de 2013. DOI: [10.3390/s140100356](https://doi.org/10.3390/s140100356).

- [150] D. Novak y col., «Toward Real-Time Automated Detection of Turns during Gait Using Wearable Inertial Measurement Units», *Sensors*, vol. 14, n.º 10, págs. 18 800-18 822, oct. de 2014. DOI: [10.3390/s141018800](https://doi.org/10.3390/s141018800).
- [151] M. Czech y S. Patel, «GaitPy: An Open-Source Python Package for Gait Analysis Using an Accelerometer on the Lower Back», *Journal of Open Source Software*, vol. 4, n.º 43, pág. 1778, nov. de 2019. DOI: [10.21105/joss.01778](https://doi.org/10.21105/joss.01778).
- [152] M. J. C. D y R. A. U. J., «Cambios asociados al envejecimiento normal en los parámetros angulares de la marcha a una velocidad controlada», *Revista médica de Chile*, vol. 144, n.º 1, págs. 74-82, ene. de 2016. DOI: [10.4067/S0034-98872016000100010](https://doi.org/10.4067/S0034-98872016000100010).
- [153] L. F. Restrepo B y J. González L., «De Pearson a Spearman», *Revista Colombiana de Ciencias Pecuarias*, vol. 20, págs. 183-192, 2007, ISSN: 0120-0690. dirección: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-06902007000200010&nrm=iso.
- [154] R. Taylor, «Interpretation of the Correlation Coefficient: A Basic Review», *Journal of Diagnostic Medical Sonography*, vol. 6, n.º 1, págs. 35-39, ene. de 1990. DOI: [10.1177/875647939000600106](https://doi.org/10.1177/875647939000600106).
- [155] M. M. Samson y col., «Differences in gait parameters at a preferred walking speed in healthy subjects due to age, height and body weight», *Aging Clinical and Experimental Research*, vol. 13, n.º 1, págs. 16-21, feb. de 2001. DOI: [10.1007/bf03351489](https://doi.org/10.1007/bf03351489).
- [156] V. Lugade, V. Lin y L.-S. Chou, «Center of mass and base of support interaction during gait», *Gait & Posture*, vol. 33, n.º 3, págs. 406-411, mar. de 2011. DOI: [10.1016/j.gaitpost.2010.12.013](https://doi.org/10.1016/j.gaitpost.2010.12.013).
- [157] J. Rosker, G. Markovic y N. Sarabon, «Effects of vertical center of mass redistribution on body sway parameters during quiet standing», *Gait & Posture*, vol. 33, n.º 3, págs. 452-456, mar. de 2011. DOI: [10.1016/j.gaitpost.2010.12.023](https://doi.org/10.1016/j.gaitpost.2010.12.023).
- [158] O. Valencia y col., «Relationship between lower limb anthropometry and temporo-spatial parameters in gait of young adults», *Retos*, n.º 33, págs. 258-260, oct. de 2017. DOI: [10.47197/retos.v0i33.58136](https://doi.org/10.47197/retos.v0i33.58136).
- [159] A. Geron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd. O'Reilly UK Ltd., 1 de oct. de 2019, 819 págs., ISBN: 1492032646.
- [160] J. Brownlee, *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python*. Machine Learning Mastery, 2020.
- [161] R. G. Mantovani y col., «An empirical study on hyperparameter tuning of decision trees», 5 de dic. de 2018. arXiv: [1812.02207](https://arxiv.org/abs/1812.02207) [cs.LG].
- [162] R. Begg y J. Kamruzzaman, «A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data», *Journal of Biomechanics*, vol. 38, n.º 3, págs. 401-408, mar. de 2005. DOI: [10.1016/j.jbiomech.2004.05.002](https://doi.org/10.1016/j.jbiomech.2004.05.002).
- [163] C.-w. Hsu, C.-c. Chang y C.-J. Lin, «A Practical Guide to Support Vector Classification Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin», nov. de 2003.
- [164] A. Mohamed, «Comparative Study of Four Supervised Machine Learning Techniques for Classification», *International Journal of Applied Science and Technology*, vol. 7, n.º 2, jun. de 2017.

- [165] E. Zanaty, «Support Vector Machines (SVMs) versus Multilayer Perception (MLP) in data classification», *Egyptian Informatics Journal*, vol. 13, n.º 3, págs. 177-183, nov. de 2012. DOI: [10.1016/j.eij.2012.08.002](https://doi.org/10.1016/j.eij.2012.08.002).
- [166] O. Beauchet y col., «Guidelines for Assessment of Gait and Reference Values for Spatiotemporal Gait Parameters in Older Adults: The Biomathics and Canadian Gait Consortiums Initiative», *Frontiers in Human Neuroscience*, vol. 11, ago. de 2017. DOI: [10.3389/fnhum.2017.00353](https://doi.org/10.3389/fnhum.2017.00353).
- [167] J. C. del Río, I. Sastre Gervás y S. Romero Yuste, «Hoja de información al paciente y consentimiento informado de casos clínicos y series de casos: propuesta de un modelo estandarizado para comunicaciones en congresos y otras publicaciones científicas», *Reumatología Clínica*, vol. 14, n.º 4, págs. 215-223, 2018, ISSN: 1699-258X. DOI: <https://doi.org/10.1016/j.reuma.2018.07.001>.

ANEXOS

Repositorios del Proyecto

En el siguiente enlace a un repositorio de *GitHub* se puede consultar el código completo del proyecto, incluyendo el *script* para la exportación de datos (presentado en la Sección 5.1.3), el código correspondiente a las tareas de preprocesamiento descritas en la Sección 5.2; *notebooks* con el entrenamiento de modelos de *Machine Learning*, tal y como se ha documentado en la Sección 5.3; y el código que implementa la herramienta final para la generación de informes y visualización de resultados (Sección 5.4).

https://github.com/SergioGonzalezVelazquez/Gait_Optical_Sensors_Tinetti

Por su parte, el *dataset* construido durante el desarrollo del trabajo es accesible desde *Kaggle*, una comunidad gratuita con recursos orientados a la ciencia de datos y el *Machine Learning*.

www.kaggle.com/serglez/gait-dataset-using-a-3d-motion-capture-system

ANEXO B

Librerías Python

En este anexo se proporciona el listado de todas las librerías de Python que se han empleado en el desarrollo del trabajo, especificando la versión exacta que se ha importado junto con un enlace al repositorio del que han sido descargadas.

Tabla B.1: Librerías *Python* utilizadas en el proyecto

Paquete	Versión	Descripción
docxtpl https://pypi.org/project/docxtpl/0.11.5/	0.11.5	Generación de documentos de <i>Microsoft Word</i> a partir de plantillas
easygui https://pypi.org/project/easygui/0.98.2/	0.98.2	Funciones para la generación de interfaces gráficas de usuario
joblib https://pypi.org/project/joblib/0.17.0/	0.17.0	Funciones de <i>pipelining</i> en <i>Python</i> . Se ha utilizado para la persistencia de modelos
keras https://pypi.org/project/keras/2.4.3/	2.4.3	<i>API</i> de <i>Deep Learning</i> de alto nivel para compilar y entrenar redes neuronales
matplotlib https://pypi.org/project/matplotlib/3.3.2/	3.3.2	Biblioteca para crear visualizaciones estáticas, animadas e interactivas
numpy https://pypi.org/project/numpy/1.19.2/	1.19.2	Librería especializada en el cálculo numérico y análisis de datos
pandas https://pypi.org/project/pandas/1.1.3/	1.1.3	Ofrece estructuras de datos y operaciones para manipular tablas numéricas
peakdetect https://pypi.org/project/peakdetect/1.1/	1.1	Funciones para la detección de picos (máximos y mínimos locales)
psutil https://pypi.org/project/psutil/5.8.0/	5.8.0	Proporciona información sobre los procesos en ejecución y el uso del sistema
pyautogui https://pypi.org/project/PyAutoGUI/0.9.52/	0.9.52	Automatización de la interfaz de usuario controlando el ratón y teclado
pyinquirer https://pypi.org/project/PyInquirer/1.0.3/	1.0.3	Construcción de interfaces de usuario por línea de comandos
scipy https://pypi.org/project/scipy/1.5.2/	1.5.2	Algoritmos matemáticos para la manipulación y filtrado de datos
seaborn https://pypi.org/project/seaborn/0.11.0/	0.11.0	Librería para la visualización de datos basada en <i>matplotlib</i> .
scikit-learn https://pypi.org/project/scikit-learn/0.23.2/	0.23.2	Librería para Aprendizaje Automático
tensorflow https://pypi.org/project/tensorflow/2.3.0/	2.3.0	Permite entrenar y ejecutar redes neuronales a gran escala
win32com https://pypi.org/project/pywin32/301/	301	Proporciona acceso a una gran parte de la <i>API</i> de <i>Windows</i> desde <i>Python</i> .

Hoja de información y consentimiento informado

En este anexo se adjunta la hoja de información y consentimiento informado elaborada durante la fase de recopilación de datos (Sección 5.1), y que, cumpliendo con lo dispuesto en el [RGPD](#), ha sido entregada y firmada por todas las personas que han colaborado en la grabación de movimientos de marcha. En este documento, se explica a los usuarios la finalidad y objetivos del estudio, se garantiza la confidencialidad con la que se tratarán los datos personales, y se informa de la posibilidad de retiro voluntario en cualquier momento.

Para su correcta redacción se ha tomado como referencia la guía publicada por la [AEMPS](#)¹, así como las directrices propuestas en otras fuentes bibliográficas [167].

¹<https://www.aemps.gob.es/investigacionClinica/medicamentos/docs/anexo8a-Ins-AEMPS-EC.pdf>

Hoja de información y consentimiento informado para la captura y tratamiento de datos de la marcha

Título del estudio	Valoración cuantitativa de la marcha mediante un sistema óptico de captura y análisis de movimiento y técnicas de Aprendizaje Automático
Autor principal	Sergio González Velázquez (sergio.gvelazquez@uclm.es)

INTRODUCCIÓN

Este documento tiene por objeto ofrecerle información sobre un estudio de investigación en el que se le invita a participar, con la finalidad de **pedir su autorización** para la captura y tratamiento de datos biomecánicos basados en un sistema de sensorización óptico. Si decide autorizar, antes debe leer detalladamente este documento y hacer todas las preguntas que precise para comprender los detalles sobre el estudio. Su decisión es completamente **voluntaria**.

OBJETIVO DEL ESTUDIO

Este trabajo de investigación pretende llevar a cabo una valoración cuantitativa de la marcha humana, aplicando técnicas de Aprendizaje Automático sobre datos sensoriales obtenidos con un sistema de captura y análisis de movimiento óptico (*Optitrack Motion Capture + Clinical3DMA*), y siguiendo el criterio de marcha en la validada Escala de Tinetti (*Performance Oriented Mobility Test, POMA*) para la evaluación del deterioro funcional.

El estudio se realiza en el marco del grupo de investigación **MAmi** de la *Universidad de Castilla-La Mancha (UCLM)*, y se presenta como Trabajo Fin de Máster en el *Máster Universitario en Ingeniería y Ciencia de Datos* impartido por la *Universidad Nacional de Educación a Distancia (UNED)*.

ACTIVIDADES DEL ESTUDIO

Para la adquisición de datos de la marcha humana se le pedirá que camine con su paso habitual a un ritmo ligero durante 5 series de 30 segundos. A continuación, si es conforme, tendrá que simular varios tipos de marchas patológicas que le serán explicadas en ese mismo momento. La grabación de estos movimientos se realiza dentro del laboratorio *WeCareLab*, ubicado en el *Instituto de Tecnologías y Sistemas de Información (ITSI)*, y tiene una duración aproximada de una hora.

Para la correcta configuración del sistema de captura y análisis de movimiento se le preguntará su edad, altura y peso, y se tomarán algunas medidas antropométricas (ancho y altura de rodilla, ancho y altura de tobillo, longitud de pie y longitud de extremidad inferior). Además, para la captura de los datos de movimiento se le colocarán marcadores reflectantes en una serie de puntos anatómicos de interés, utilizando adhesivos de doble cara y/o cintas con velcro sobre la piel o la ropa. En cualquier caso,

la participación en este estudio no producirá ninguna molestia, y no implica riesgo alguno para la salud.

PARTICIPACIÓN VOLUNTARIA

Debe saber que su participación en este estudio es totalmente voluntaria y puede decidir libremente NO participar. Asimismo, si decide participar, puede retirar el consentimiento en cualquier momento sin tener que dar explicaciones. Si durante su participación tiene alguna duda o necesita obtener más información, podrá solicitar cualquier explicación sobre cualquier aspecto del estudio y sus implicaciones.

PROTECCIÓN DE DATOS PERSONALES

La obtención, tratamiento, conservación, comunicación y cesión de sus datos se hará en cumplimiento de lo dispuesto en el Reglamento General de Protección de Datos de Carácter Personal (Reglamento UE 2016/679 del Parlamento Europeo y del Consejo, del 27 de abril de 2016) y la normativa española sobre protección de datos de carácter personal vigente. Asimismo, le informamos que, en cualquier momento, puede ejercer los derechos de acceso, rectificación, supresión, limitación y demás derechos previstos en la normativa reguladora, contactando para ello con el investigador principal de este trabajo.

En todo momento, se mantendrá la confidencialidad absoluta de sus datos, que se recogerán **anonimizados**, es decir, durante su participación en el estudio sus datos se identificarán mediante un código y los investigadores no transferirán información alguna que pueda identificarle directamente. De igual forma, si los resultados del estudio fueran susceptibles de publicación en revistas científicas, no se proporcionarán datos personales de los participantes que han colaborado en esta investigación.

CONSENTIMIENTO INFORMADO

Yo, _____ he sido informado por **Sergio González Velázquez**, autor del citado estudio de investigación, y declaro que:

1. He leído la presente hoja de información y estoy conforme con su contenido.
2. He podido hacer preguntas sobre la investigación y he recibido suficiente información.
3. Soy consciente de que durante la grabación de los datos de movimiento se tomarán imágenes de mi persona, en fotografía o vídeo.
4. He sido informado de que se mantendrá la confidencialidad de mis datos
5. Comprendo que mi participación es voluntaria y puedo retirarme del estudio en cualquier momento.

Con esto, presto mi conformidad para participar en el estudio.

Fecha:
Firma del participante:

Fecha:
Firma del investigador:

ANEXO D

Resumen del *dataset*

En este anexo se adjunta una tabla con información sobre las capturas de marcha incluidas en el conjunto de datos, cuya elaboración se ha explicado en la Sección 5.1 y es accesible desde el enlace proporcionado en el Anexo A. La Tabla D.1 etiqueta cada grabación de movimiento siguiendo la nomenclatura descrita en la Tabla 5.2 (Sección 5.1.3), indica el número de ciclos de la marcha para cada extremidad, según la segmentación explicada en la Sección 5.2.1, así como el número de giros durante la deambulaci3n, obtenidos autom1ticamente con la t3cnica presentada en la Secci3n 5.2.3. Finalmente, en las dos 3ltimas columnas se muestra el 3ndice de las zancadas filtradas en base a los criterios propuestos en la Secci3n 5.2.3.

Adem1s, tal y como se ha explicado en la Secci3n 5.2.3, se han descartado dos registros completos (*subject_07_rec_03* y *subject_10_rec_05*) por presentar una alta proporci3n de zancadas incorrectas.

Tabla D.1: Resumen del conjunto de datos

Sujeto	Captura	Etiqueta	Giros	Zancadas		Zancadas filtradas	
				Izq.	Der.	Izq.	Der.
subject_01	rec_01	norm	4	18	20	[4, 5, 11, 15, 16]	[4, 8, 12, 13, 17]
	rec_02	norm	4	18	21	[3, 4, 10, 11, 15]	[4, 8, 9, 13, 17, 18]
	rec_03	norm	4	20	19	[3, 4, 9, 13, 14, 17, 18]	[4, 8, 9, 13, 17]
	rec_04	norm	4	18	20	[3, 4, 8, 12, 13]	[4, 8, 9, 13, 17, 18]
	rec_05	norm	4	20	20	[4, 5, 9, 13, 14, 18]	[4, 8, 9, 13, 17, 18]
	rec_06	lap1	1	18	17	[9, 10]	[9]
	rec_07	lap1	1	16	18	[]	[11]
	rec_08	lap2	1	18	19	[9, 18]	[9, 18]
	rec_09	lap2	1	17	17	[11]	[]
	rec_10	lap3	1	16	16	[13, 14]	[13]
	rec_11	lap3	1	17	16	[11, 12, 13]	[11, 12]
	rec_12	lap4	1	16	17	[11]	[11]
	rec_13	lap4	1	16	17	[]	[]
	rec_14	dc	0	10	14	[]	[]
	rec_15	dc	0	11	14	[]	[]
	rec_16	pm	1	13	13	[8]	[8]
	rec_17	pm	1	13	14	[9]	[]
subject_02	rec_01	norm	3	18	19	[4, 5, 8, 9, 10, 14, 15]	[5, 8, 10, 15]
	rec_02	norm	4	18	19	[4, 9, 13, 14, 18]	[4, 5, 9, 10, 15]
	rec_03	norm	4	20	19	[5, 10, 15, 20]	[4, 9, 10, 15]
	rec_04	norm	3	20	20	[4, 5, 9, 10, 15]	[5, 10, 15, 16]
	rec_05	norm	3	20	20	[4, 9, 10, 15]	[4, 5, 10, 15, 16]
	rec_06	lap1	1	15	15	[8, 9]	[8]
	rec_07	lap1	1	15	13	[8, 9]	[6]

Tabla D.1 (continuación de la página anterior)

Sujeto	Captura	Etiqueta	Giros	Zancadas		Zancadas filtradas	
				Izq.	Der.	Izq.	Der.
	rec_08	lap2	1	15	16	[9]	[9]
	rec_09	lap2	1	16	16	[8]	[8]
	rec_10	lap3	1	16	17	[]	[12]
	rec_11	lap3	1	16	16	[9, 11]	[9, 11, 12]
	rec_12	lap4	1	17	17	[12, 13]	[13]
	rec_13	lap4	1	16	14	[]	[11]
	rec_14	dc	1	12	12	[6, 7]	[7]
	rec_15	dc	1	11	11	[7, 8]	[7]
	rec_16	pm	1	14	14	[13, 14]	[14]
	rec_17	pm	1	11	13	[10]	[11, 12]
subject_03	rec_01	norm	4	23	22	[4, 5, 9, 10, 14, 15, 19, 20]	[4, 9, 14, 19]
	rec_02	norm	4	21	21	[4, 9, 14, 18, 19]	[4, 5, 9, 10, 14, 18, 19]
	rec_03	norm	4	21	21	[4, 9, 13, 14, 18]	[4, 5, 9, 14, 19]
	rec_04	norm	4	21	22	[4, 9, 14, 19]	[4, 5, 9, 10, 14, 19]
	rec_05	norm	4	21	21	[4, 5, 13, 17, 18]	[4, 9, 13, 14, 18]
	rec_06	lap1	2	18	17	[7, 8, 14, 15]	[7, 14]
	rec_07	lap1	2	16	16	[7, 13]	[7, 14]
	rec_08	lap2	2	17	17	[7, 15]	[7]
	rec_09	lap2	2	16	18	[7, 14]	[7, 14]
	rec_10	lap3	2	17	15	[8, 15, 16]	[15]
	rec_11	lap3	2	18	17	[7, 14, 15]	[7, 14]
	rec_12	lap4	2	15	16	[6, 12]	[11]
	rec_13	lap4	2	15	15	[6]	[6, 7, 12]
	rec_14	dc	1	17	15	[11]	[11]
	rec_15	dc	1	16	17	[11]	[11]
rec_16	pm	1	23	23	[]	[13]	
rec_17	pm	1	22	22	[13]	[]	
subject_04	rec_01	norm	5	22	23	[3, 4, 7, 8, 14, 15, 18, 19]	[4, 8, 12, 16, 20]
	rec_02	norm	5	21	22	[3, 4, 7, 8, 11, 12, 15, 18, 19]	[4, 8, 12, 16, 20]
	rec_03	norm	5	19	21	[3, 4, 7, 8, 13, 14, 17, 18]	[3, 7, 11, 15, 19]
	rec_04	norm	5	20	21	[4, 8, 9, 12, 13, 16, 19, 20]	[3, 4, 8, 12, 16, 20]
	rec_05	norm	5	21	20	[3, 4, 7, 8, 11, 12, 15, 16, 20, 21]	[3, 7, 11, 15, 20]
	rec_06	lap1	2	21	20	[8, 16]	[7, 16]
	rec_07	lap1	2	21	19	[8, 17]	[8, 16, 17]
	rec_08	lap2	2	21	21	[8, 17]	[8]
	rec_09	lap2	2	21	21	[8, 16]	[]
	rec_10	lap3	2	19	18	[7, 15, 16]	[7, 15]
	rec_11	lap3	2	18	18	[7, 14, 15]	[7, 15]
	rec_12	lap4	2	18	18	[6, 7, 14, 15]	[7, 15]
	rec_13	lap4	2	17	19	[14]	[7, 15, 16]
	rec_14	dc	1	12	12	[11]	[10]
	rec_15	dc	1	11	12	[10]	[]
rec_16	pm	2	26	25	[11, 22, 23]	[10, 22]	
rec_17	pm	2	25	25	[10, 21]	[21, 22]	
subject_05	rec_01	norm	4	13	19	[3, 4, 7, 10, 11]	[4, 8, 12, 16]
	rec_02	norm	4	14	19	[5, 6]	[4, 5, 8, 12, 13, 16, 17]

Tabla D.1 (continuación de la página anterior)

Sujeto	Captura	Etiqueta	Giros	Zancadas		Zancadas filtradas		
				Izq.	Der.	Izq.	Der.	
	rec_03	norm	5	19	20	[3, 7, 11, 15, 19]	[4, 8, 9, 12, 13, 16, 17, 20]	
	rec_04	norm	5	18	20	[3, 6, 7, 10, 11, 14, 17, 18]	[4, 8, 12, 16, 20]	
	rec_05	norm	5	18	20	[3, 4, 7, 11, 14, 15, 18]	[8, 12, 20]	
	rec_06	lap1	2	14	13	[7]	[7]	
	rec_07	lap1	2	14	14	[6, 13, 14]	[13]	
	rec_08	lap2	2	14	14	[6, 12]	[6, 12]	
	rec_09	lap2	2	12	14	[6, 11]	[6, 12]	
	rec_10	lap3	2	12	12	[12]	[5, 6, 12]	
	rec_11	lap3	2	14	12	[7]	[7]	
	rec_12	lap4	2	12	11	[7]	[11]	
	rec_13	lap4	1	12	13	[6]	[6]	
	rec_14	dc	1	11	9	[9]	[7]	
	rec_15	dc	1	10	10	[6, 7]	[8]	
	rec_16	pm	3	12	13	[4, 8, 12]	[4, 5, 8, 9, 12, 13]	
	rec_17	pm	3	13	14	[4, 8, 12]	[4, 5, 8, 9, 12, 13]	
	subject_06	rec_01	norm	5	20	19	[3, 4, 7, 11, 15]	[3, 10, 14, 18]
		rec_02	norm	6	21	21	[3, 7, 14, 18, 21]	[3, 6, 10, 14, 17, 21]
rec_03		norm	6	21	20	[3, 7, 10, 14, 17, 21]	[6, 10, 13, 17]	
rec_04		norm	6	21	21	[3, 6, 10, 13, 17, 21]	[2, 6, 9, 13, 16, 20]	
rec_05		norm	6	21	21	[3, 6, 7, 10, 14, 17, 18, 21]	[2, 6, 10, 13, 17, 21]	
rec_06		lap1	3	19	19	[5, 11, 17, 18]	[5, 11, 17]	
rec_07		lap1	3	21	20	[5, 11, 17, 18]	[5, 10, 11, 17]	
rec_08		lap2	3	20	21	[4, 9, 10, 15, 16]	[10, 16]	
rec_09		lap2	4	23	22	[5, 6, 11, 12, 17, 18, 23]	[5, 11, 17]	
rec_10		lap3	4	16	15	[7, 11, 15]	[3, 7, 11]	
rec_11		lap3	4	15	15	[3, 7, 12]	[3, 7, 11, 15]	
rec_12		lap4	4	15	15	[3, 7, 10, 11, 15]	[3, 7, 11, 15]	
rec_13		lap4	4	15	15	[3, 7, 11, 15]	[3, 7, 11, 15]	
rec_14		dc	2	16	15	[7, 15]	[7, 14]	
rec_15		dc	2	16	16	[7, 14, 15]	[7, 14]	
rec_16		pm	2	19	22	[6, 14]	[8, 16]	
rec_17		pm	2	22	21	[7, 8, 15, 16]	[7, 15]	
subject_07	rec_01	norm	5	19	21	[3, 7, 14, 15, 18, 19]	[3, 4, 7, 8, 12, 16, 20]	
	rec_02	norm	5	19	21	[3, 7, 8, 11, 12, 18, 19]	[3, 4, 12, 16, 20]	
	rec_03	norm	Descartado					
	rec_04	norm	4	20	20	[3, 4, 7, 8, 12, 16, 17]	[4, 8, 12, 13, 17]	
	rec_05	norm	4	20	20	[4, 8, 12, 13, 17]	[3, 7, 8, 12, 16]	
	rec_06	lap1	2	14	14	[5, 12]	[5, 11, 12]	
	rec_07	lap1	2	16	15	[5, 6, 11, 12]	[5, 11]	
	rec_08	lap2	2	15	16	[5, 11]	[5, 6]	
	rec_09	lap2	2	16	18	[5, 11, 12]	[5, 6, 13]	
	rec_10	lap3	2	16	15	[5, 6, 13]	[5, 12]	
	rec_11	lap3	2	18	17	[5, 12]	[4, 5, 11]	
	rec_12	lap4	2	18	18	[5, 12]	[5, 12]	
	rec_13	lap4	3	17	17	[4, 10, 16, 17]	[4, 10]	
	rec_14	dc	1	11	14	[]	[13]	
	rec_15	dc	1	14	17	[10]	[]	

Tabla D.1 (continuación de la página anterior)

Sujeto	Captura	Etiqueta	Giros	Zancadas		Zancadas filtradas		
				Izq.	Der.	Izq.	Der.	
	rec_16	pm	1	23	21	[13, 14]	[]	
	rec_17	pm	1	25	25	[12]	[12, 13]	
subject_08	rec_01	norm	4	17	18	[4, 9, 13]	[4, 5, 9, 13, 14, 18]	
	rec_02	norm	4	18	18	[4, 9, 13, 14, 18]	[4, 5, 9, 10, 14, 18]	
	rec_03	norm	3	17	17	[4, 5, 9, 14]	[5, 9, 10, 14, 15]	
	rec_04	norm	4	17	18	[4, 9, 13, 14]	[4, 5, 9, 10, 14]	
	rec_05	norm	3	17	17	[4, 9, 13, 14]	[4, 5, 9, 10, 14]	
	rec_06	lap1	2	17	17	[8, 16]	[8, 15]	
	rec_07	lap1	2	18	18	[7, 13, 14]	[13]	
	rec_08	lap2	2	18	18	[6, 7, 15]	[7]	
	rec_09	lap2	2	19	20	[7, 14]	[7, 14]	
	rec_10	lap3	3	17	17	[6, 12]	[6, 12, 17]	
	rec_11	lap3	3	18	17	[5, 6, 11, 16]	[5, 11, 16]	
	rec_12	lap4	3	17	18	[5, 11, 17]	[5, 6, 11, 12, 17, 18]	
	rec_13	lap4	3	17	18	[5, 10, 15]	[5, 10, 11, 15, 16]	
	rec_14	dc	1	11	10	[8]	[7, 8]	
	rec_15	dc	1	8	9	[]	[7]	
	rec_16	pm	3	20	20	[6, 12, 13, 18, 19]	[5, 6, 12, 18]	
	rec_17	pm	3	22	21	[6, 13, 19, 20]	[5, 6, 12, 13, 19]	
subject_09	rec_01	norm	5	21	21	[0, 1, 3, 4, 5, 7, 8, 11, 12, 15, 19, 20]	[0, 1, 4, 5, 8, 12, 15, 16, 20]	
	rec_02	norm	4	17	19	[3, 4, 8, 13]	[4, 8, 9, 13, 14, 18, 19]	
	rec_03	norm	4	19	20	[1, 3, 4, 8, 12, 13, 17]	[1, 4, 8, 9, 13, 17]	
	rec_04	norm	4	18	20	[3, 7, 11, 12, 16]	[4, 8, 9, 13, 17, 18]	
	rec_05	norm	4	19	19	[4, 5, 9, 13, 17]	[4, 8, 9, 13, 17, 18]	
	rec_06	lap1	2	18	19	[5, 8, 9, 18]	[5, 8, 18]	
	rec_07	lap1	2	18	18	[1, 4, 6, 7, 13, 14]	[1, 4, 7, 13, 15]	
	rec_08	lap2	2	17	17	[7, 14, 15]	[14, 15]	
	rec_09	lap2	2	17	18	[7, 13]	[7, 13, 14]	
	rec_10	lap3	2	16	17	[8, 16]	[8, 16, 17]	
	rec_11	lap3	2	14	17	[7]	[8, 16]	
	rec_12	lap4	2	14	16	[11]	[6, 13]	
	rec_13	lap4	2	16	17	[6, 13]	[6, 7, 13, 14]	
	rec_14	dc	1	10	16	[4, 5]	[8]	
	rec_15	dc	2	10	16	[4, 5, 9]	[16]	
	rec_16	pm	2	21	21	[10, 11, 21]	[11, 20, 21]	
	rec_17	pm	2	19	21	[19]	[11, 21]	
subject_10	rec_01	norm	5	19	19	[4, 7, 8, 11, 12, 15, 19]	[4, 12, 16]	
	rec_02	norm	5	19	19	[4, 5, 8, 9, 12, 13, 16, 17, 19]	[4, 8, 16, 19]	
	rec_03	norm	5	19	20	[3, 4, 7, 8, 11, 12, 15, 16, 19]	[4, 8, 12, 20]	
	rec_04	norm	5	19	19	[3, 4, 7, 11, 12, 15, 16, 19]	[4, 8, 12, 16]	
	rec_05	norm	Descartado					
	rec_06	lap1	1	16	15	[8, 9, 10]	[8, 9]	
	rec_07	lap1	1	15	15	[9, 10]	[9]	
	rec_08	lap2	1	16	16	[10, 11, 12, 13, 14, 15]	[10, 11, 12, 13, 14, 15]	
	rec_09	lap2	1	15	16	[11]	[]	
	rec_10	lap3	1	15	14	[11, 12]	[11]	
	rec_11	lap3	1	14	13	[9, 10, 11]	[9, 10]	

Tabla D.1 (continuación de la página anterior)

Sujeto	Captura	Etiqueta	Giros	Zancadas		Zancadas filtradas	
				Izq.	Der.	Izq.	Der.
	rec_12	lap4	1	12	14	[]	[]
	rec_13	lap4	1	14	14	[]	[]
	rec_14	dc	0	9	10	[]	[]
	rec_15	dc	1	11	11	[9, 10]	[9, 10]
	rec_16	pm	1	18	20	[11]	[10]
	rec_17	pm	1	17	20	[]	[11]

ANEXO E

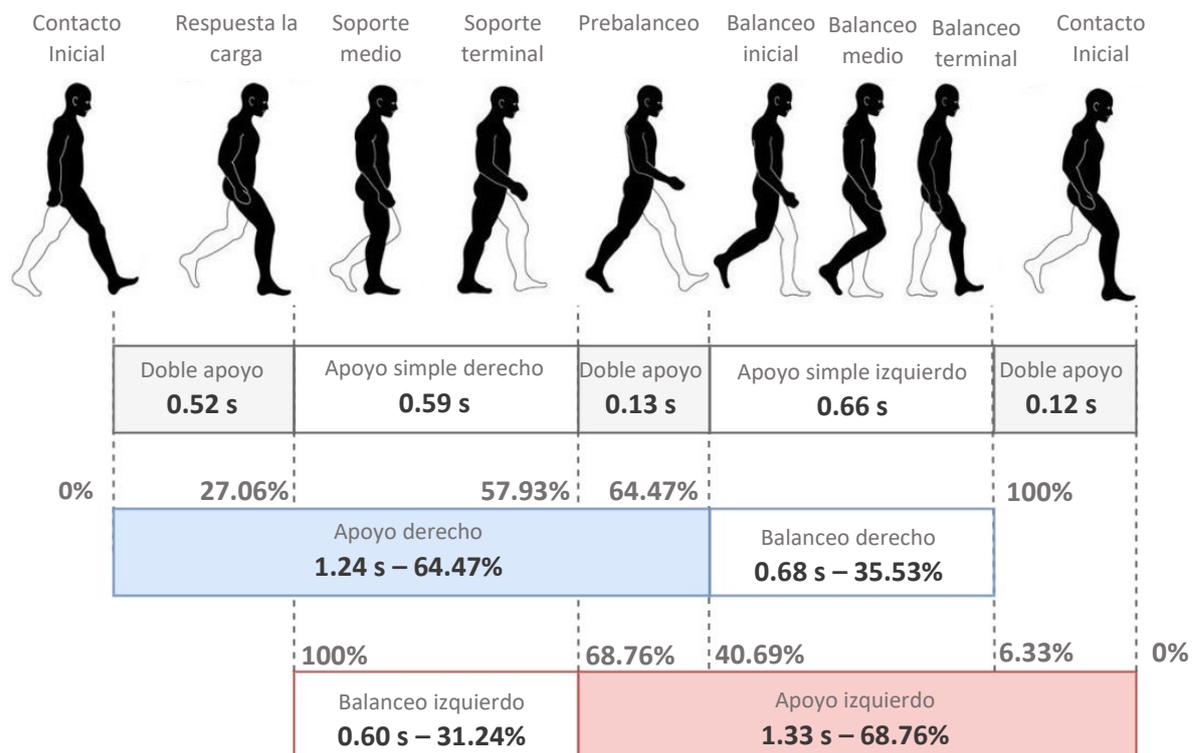
Informe biomecánico de marcha humana

En el presente anexo se proporciona un ejemplo de informe de marcha generado por la herramienta desarrollada durante el trabajo, y cuyo contenido ha sido descrito en detalle en la Sección 5.4.2. Este documento, incluye las respuestas, generadas automáticamente, a varios ítems de la escala de Tinetti, así como tablas y gráficos para presentar los parámetros espaciotemporales y cinemáticos de la marcha humana.

Características generales

72 Edad	Hombre Género	1.72 m Altura	80 kg Peso	265 cm Longitud de pie	27.51s Duración
0.7 m/s Velocidad	82.3 pasos/min Cadencia	14 Izq. Zancadas totales	13 Der. Zancadas totales	13 Izq. 12 Der. Zancadas procesadas	2 Giros

Análisis del ciclo de la marcha

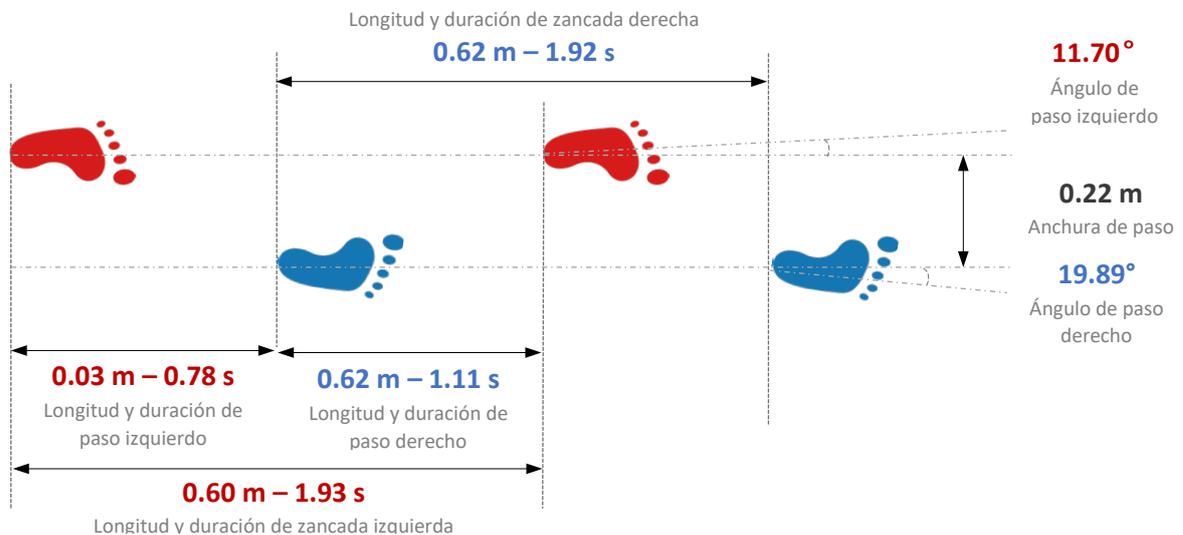


Análisis detallado

Ciclo de la marcha										
Parámetro	Fase de apoyo		Fase de balanceo		Fin de respuesta a la carga	Inicio del prebalanceo		Tiempo de doble apoyo		
Unidad	% del ciclo		% del ciclo		% del ciclo	% del ciclo		% del ciclo		
Pierna	Izq.	Der.	Izq.	Der.	Izq.	Der.	Izq.	Der.	Izq.	Der.
Media	68.76	64.47	31.24	35.53	6.33	27.06	40.69	57.93	34.40	33.60
Mediana	67.51	66.16	32.49	33.84	7.22	25.52	40.62	58.40	32.63	33.16
Desviación Estándar	4.24	3.90	4.24	3.90	2.04	4.82	2.58	3.13	5.72	6.18
Mínimo	64.42	53.71	19.07	32.63	1.11	19.43	35.29	52.13	23.89	20.57
Máximo	80.93	67.37	35.58	46.29	9.52	38.27	46.11	61.34	45.88	47.45
Valor orientativo ⁱ	62%		38%		10%		50%		25%	

Parámetros espaciotemporales

Duración y longitud de zancadas



Temporales										
	Duración de zancada segundos		Duración de paso segundos		Duración fase de apoyo segundos		Duración fase de balanceo segundos		Tiempo de doble apoyo segundos	
Unidad	Izq.	Der.	Izq.	Der.	Izq.	Der.	Izq.	Der.	Izq.	Der.
Media	1.93	1.92	0.78	1.11	1.33	1.24	0.60	0.68	0.67	0.65
Mediana	1.92	1.94	0.78	1.13	1.31	1.29	0.61	0.66	0.64	33.16
Desviación Estándar	0.17	0.10	0.03	0.09	0.16	0.11	0.08	0.06	0.16	0.14
Mínimo	1.63	1.74	0.71	0.92	1.05	0.94	0.37	0.61	0.43	0.36
Máximo	2.38	2.11	45.88	47.45	1.68	1.33	0.70	0.81	1.02	0.93
Valor orientativo ⁱⁱ	1,12 s		0,55 s		0,70 s		0,41 s		0,29 s	

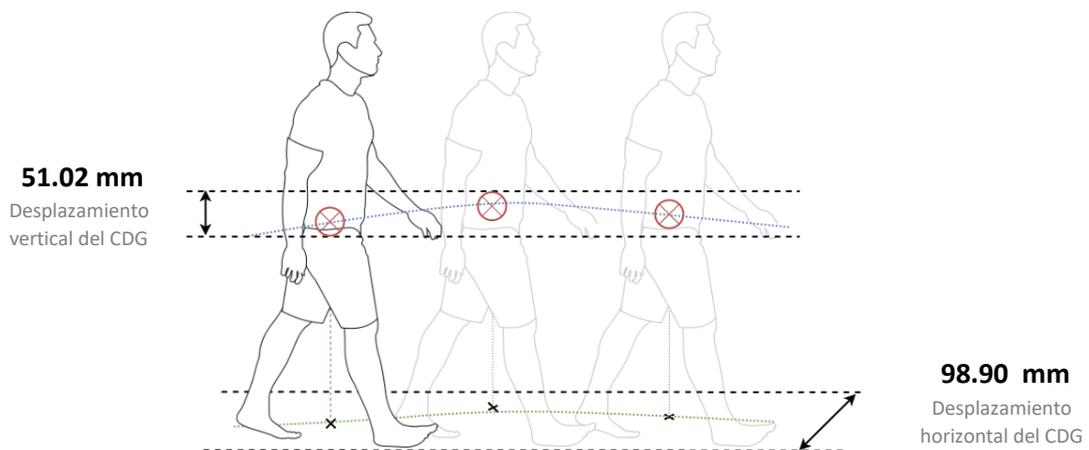
Espaciales								
	Longitud de zancada metros		Longitud de paso metros		Anchura de paso metros		Ángulo de paso grados	
Unidad	Izq.	Der.	Izq.	Der.	Izq.	Der.	Izq.	Der.
Media	0.60	0.62	0.03	0.62	0.22	0.21	11.70	19.89
Mediana	0.60	0.62	0.03	0.62	0.22	0.22	7.83	17.01
Desviación Estándar	0.06	0.05	0.03	0.04	0.02	0.03	16.81	10.41
Mínimo	0.49	0.51	0.00	0.54	0.17	0.12	2.69	12.56
Máximo	0.71	0.71	0.09	0.68	0.26	0.24	64.31	43.56
Valor orientativo ⁱⁱⁱ	0.70 a 0.82 m		0.35 a 0.41 m		0.05 a 0.10 m		5 a 8 grados	

Altura y ángulos de tobillo y pie



Espaciales								
Parámetro	Máx. altura de talón		Máx. altura del pie		Ángulo de contacto		Ángulo de despegue	
Unidad	metros		metros		grados		grados	
Pierna	Izq.	Der.	Izq.	Der.	Izq.	Der.	Izq.	Der.
Media	0.05	0.18	0.10	0.03	76.68	82.50	13.32	7.50
Mediana	0.05	0.20	0.11	0.03	75.86	82.77	14.14	7.23
Desviación Estándar	0.02	0.05	0.02	0.01	2.70	1.13	2.70	1.13
Mínimo	0.03	0.06	0.04	0.02	73.20	81.09	9.04	5.07
Máximo	0.08	0.22	0.13	0.04	80.96	84.93	16.80	8.91

Trayectoria durante la marcha



Parámetro	Desplazamiento vertical		Desplazamiento horizontal	
Unidad	milímetros		milímetros	
Pierna	Izq.	Der.	Izq.	Der.
Media	53.31	48.73	85.02	112.77
Mediana	55.42	52.00	82.38	115.72
Desviación Estándar	6.50	11.34	8.57	17.76
Mínimo	37.63	18.71	72.99	68.46
Máximo	59.76	59.72	95.89	134.86

Valoración según la Escala de Tinetti

LONGITUD Y ALTURA DE PASO	
El pie derecho no sobrepasa al izquierdo con el paso en la fase de balanceo.....0	1
El pie derecho sobrepasa al izquierdo.....1	(98.25%)
El pie derecho no se levanta completamente del suelo con el paso en la fase del balanceo.....0	1
El pie derecho se levanta completamente.....1	(96.91%)
El pie izquierdo no sobrepasa al derecho con el paso en la fase de balanceo.....0	0
El pie izquierdo sobrepasa al derecho.....1	(92.92%)
El pie izquierdo no se levanta completamente del suelo con el paso en la fase del balanceo.....0	0
El pie izquierdo se levanta completamente.....1	(93.88%)

CONTINUIDAD DE LOS PASOS	
Para o hay discontinuidad entre pasos.....0	1
Los pasos son continuos.....1	(66.89%)

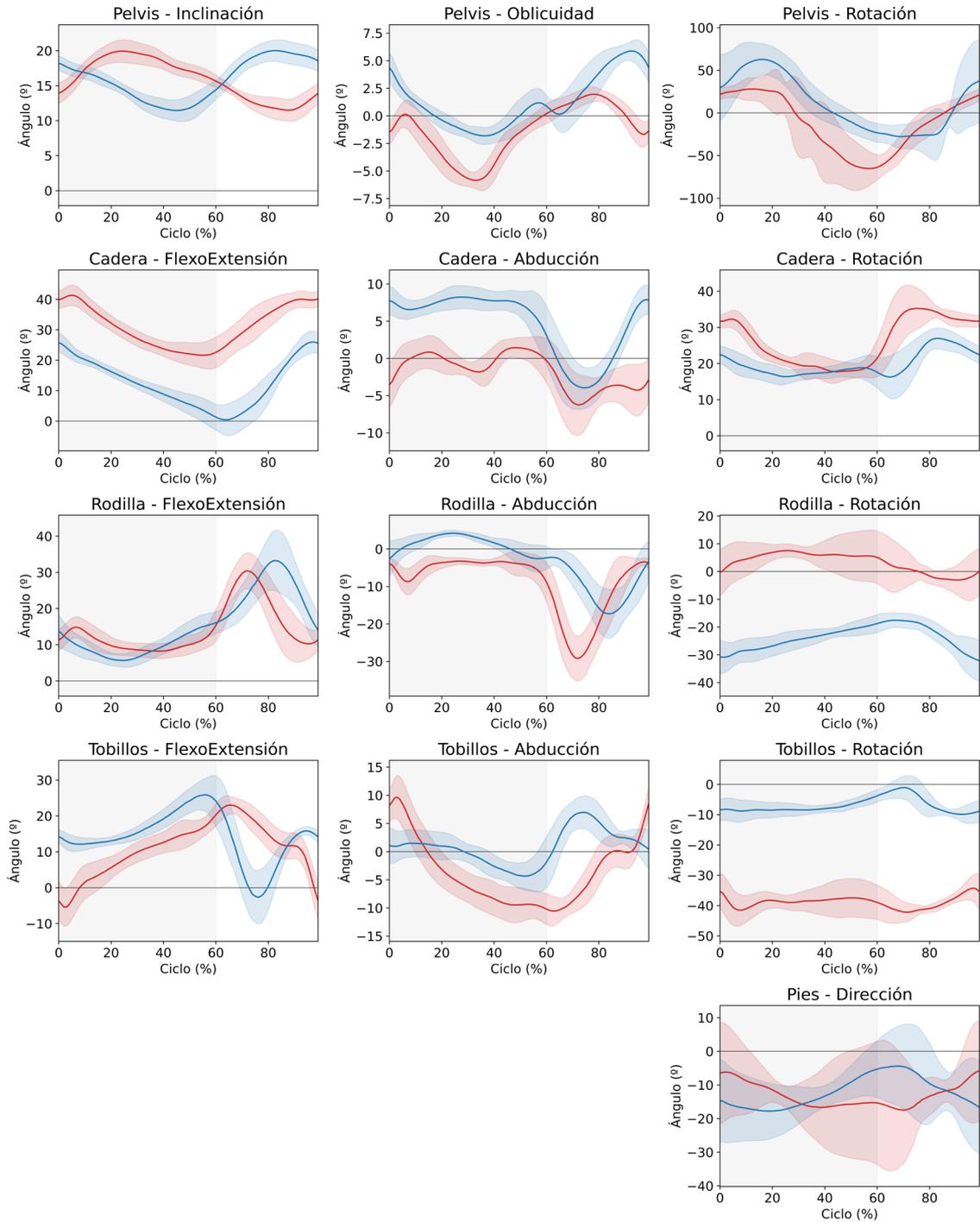
POSTURA EN LA MARCHA	
Talones separados.....0	0
Talones casi se tocan mientras camina.....1	(67.64%)

PUNTUACION GLOBAL	
	3 / 6

Respuestas generadas automáticamente en base a las características espaciotemporales y cinemáticas de la marcha. El valor porcentual incluido para cada categoría es una probabilidad que indica la certeza de la predicción.



Estudio cinemático



Análisis detallado

Pelvis						
Parámetro	Inclinación		Oblicuidad		Rotación	
	Anterior [+] / Posterior [-]		Superior [+] / Inferior [-]		Interna [+] / Externa [-]	
Pierna	Izq.	Der.	Izq.	Der.	Izq.	Der.
Media	15.8°	15.9°	-1.4°	1.4°	-11.0°	8.1°
Mínimo	11.0°	10.9°	-6.2°	-2.1°	-70.7°	-36.4°
Máximo	20.0°	20.4°	2.1°	6.3°	32.8°	78.6°
Rango articular	9.0°	9.5°	8.3°	8.5°	103.5°	115.0°
Ángulo inicial	14.2°	18.1°	-1.4°	4.6°	22.0°	19.5°
Ángulo final	13.7°	18.5°	-1.5°	4.5°	22.4°	37.1°
Ángulo de despegue	13.7°	16.1°	1.0°	-0.1°	-42.4°	-27.5°

Cadera						
Parámetro	Flexo Extensión		Aducción Abducción		Rotación	
	Flexión [+] / Extensión [-]		Aducción[+] / Abducción[-]		Interna [+] / Externa [-]	
Pierna	Izq.	Der.	Izq.	Der.	Izq.	Der.
Media	31.0°	12.3°	-1.7°	4.3°	26.5°	19.8°
Mínimo	20.2°	-1.6°	-7.6°	-5.0°	17.1°	12.7°
Máximo	42.7°	28.3°	2.7°	9.2°	38.7°	27.6°
Rango articular	22.5°	29.9°	10.2°	14.2°	21.7°	14.9°
Ángulo inicial	40.2°	26.3°	-3.5°	8.1°	31.6°	23.0°
Ángulo final	39.7°	26.0°	-3.5°	7.7°	31.5°	22.4°
Ángulo de despegue	27.9°	-0.8°	-5.2°	-1.5°	34.8°	14.5°

Rodilla						
Parámetro	Flexo Extensión		Valgo Varo		Rotación	
	Flexión [+] / Extensión [-]		Valgo [+] / Varo [-]		Interna [+] / Externa [-]	
Pierna	Izq.	Der.	Izq.	Der.	Izq.	Der.
Media	14.1°	15.7°	-8.8°	-3.1°	3.6°	-23.7°
Mínimo	7.2°	5.2°	-32.5°	-19.7°	-7.4°	-33.3°
Máximo	33.1°	34.7°	-2.1°	4.5°	12.0°	-17.3°
Rango articular	25.8°	29.5°	30.4°	24.2°	19.4°	15.9°
Ángulo inicial	11.1°	13.1°	-4.1°	-1.6°	-1.4°	-29.2°
Ángulo final	11.3°	14.6°	-3.9°	-3.9°	-0.4°	-32.3°
Ángulo de despegue	30.1°	17.1°	-27.6°	-1.9°	2.8°	-18.0°

Tobillo						
Parámetro	Flexo Extensión		Aducción Abducción		Rotación	
	Flexión Dorsal [+] / Extensión Plantar [-]		Aducción[+]/Abducción[-]		Interna [+] / Externa [-]	
Pierna	Izq.	Der.	Izq.	Der.	Izq.	Der.
Media	10.8°	14.1°	-3.8°	0.9°	-38.6°	-6.9°
Mínimo	-7.6°	-5.4°	-11.9°	-5.3°	-44.7°	-11.0°
Máximo	24.1°	27.8°	11.3°	7.8°	-31.3°	-0.2°
Rango articular	31.7°	33.3°	23.2°	13.1°	13.4°	10.8°
Ángulo inicial	-4.3°	13.9°	8.5°	1.8°	-36.4°	-7.5°
Ángulo final	-3.4°	14.1°	8.4°	0.4°	-35.6°	-8.9°
Ángulo de despegue	22.7°	16.9°	-9.4°	4.1°	-41.7°	-2.0°

Pie	Rotación	
	Interna [+] / Externa [-]	
Parámetro	Izq.	Der.
Pierna	-12.1°	-11.9°
Media	-20.8°	-21.9°
Mínimo	1.8°	-3.4°
Máximo	22.6°	18.5°
Rango articular	-8.7°	-11.3°
Ángulo inicial	-6.9°	-16.8°
Ángulo final	-14.7°	-5.5°

ⁱⁱ J. D. Lesmes, "Consideraciones del desarrollo de la marcha El ciclo de la marcha y sus divisiones," in Evaluación clínica funcional del movimiento corporal humano, Editorial Médica Panamericana, Ed. 2007, pp. 259–268.

ⁱⁱ O. Beauchet et al., "Guidelines for assessment of gait and reference values for spatiotemporal gait parameters in older adults: The biometrics and canadian gait consortiums initiative," Front. Hum. Neurosci., vol. 11, no. August 2017, doi: 10.3389/fnhum.2017.00353.

ⁱⁱⁱ R. Carles Hernández, "La marcha," in Manual Teórico-Práctico de Valoración y Abordaje Fisioterápico, 2014.

Dedicación y costes

En este anexo se incluye una valoración en términos de costes (tiempo y dinero) del trabajo realizado. Se describe la distribución del tiempo invertido durante el desarrollo de este TFM, considerando que tiene asignados 12 créditos ECTS, lo cual se corresponde con una dedicación aproximada de 300-360 horas en total.

F.1. DEDICACIÓN HORARIA

Durante el desarrollo de este Trabajo Fin de Máster (TFM), se ha realizado un recuento de las horas dedicadas a cada una de las actividades (Tabla F.1).

Tabla F.1: Horas de dedicación al TFM

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Total horas/mes
Definición del problema y anteproyecto	10	5								15
Estudio del dominio y estado del arte	5	15	5							25
Familiarización con el sistema óptico de captura	5	10	5							20
Recogida y etiquetado de datos			25	30	10					65
Preprocesamiento y preparación de datos				30	30	25				85
Aplicación y evaluación de técnicas de Aprendizaje Automático						10	25	20		55
Despliegue e integración de los modelos de <i>Machine Learning</i>							20	20	10	50
Redacción de la memoria y presentación		10	15	5			5	20	20	75
Total horas/mes	20	40	50	65	40	35	50	60	30	390

De manera adicional, se incluyen una serie de gráficas que muestran las horas dedicadas al proyecto, ordenadas por mes y por tarea (Figuras F.1 y F.2).

Distribución del trabajo por tipo de tarea

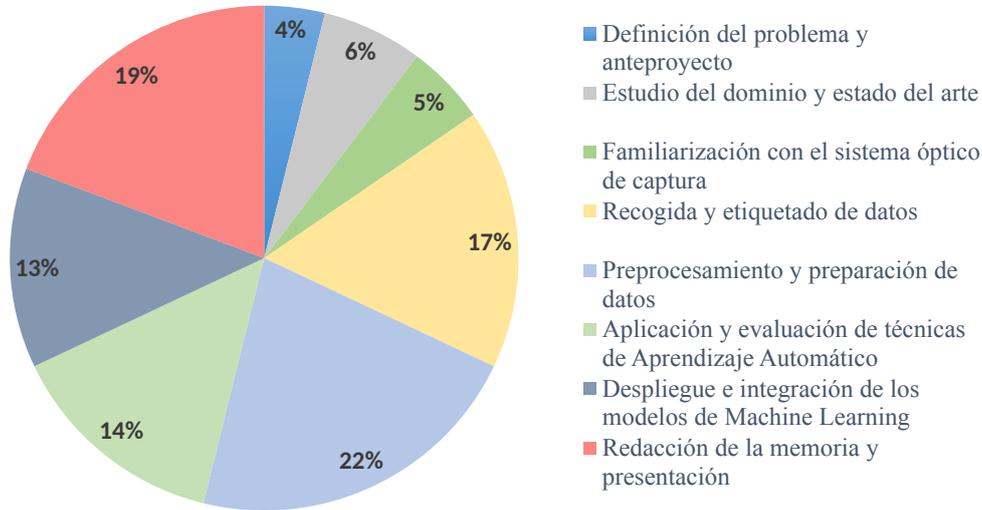


Figura F.1: Distribución del trabajo por tipo de tarea

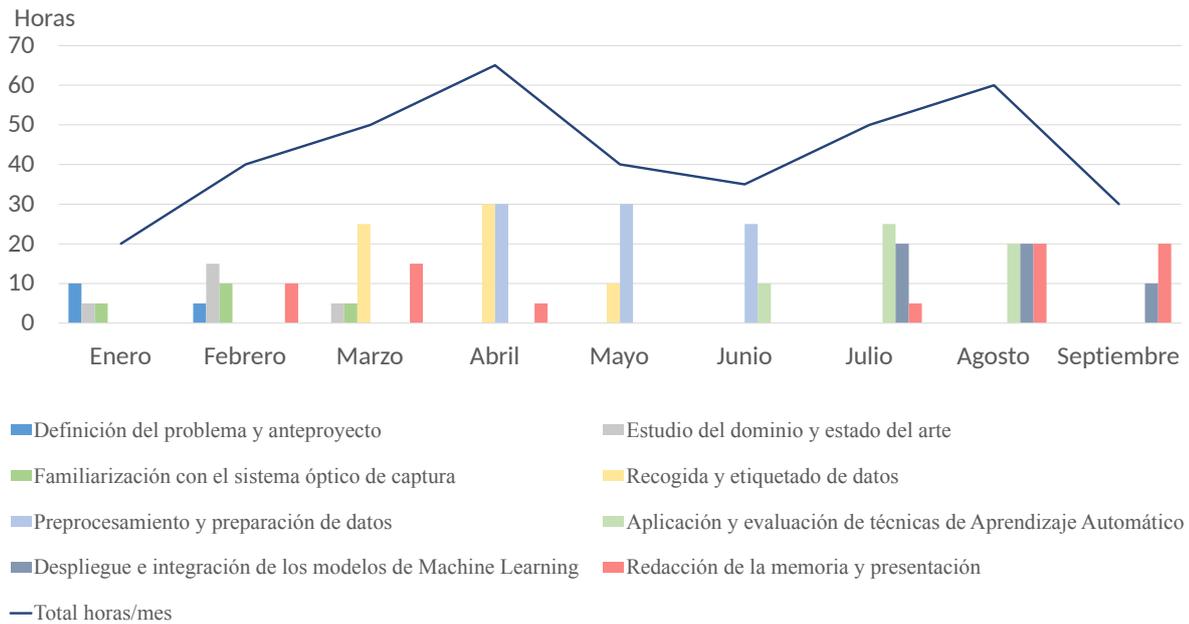


Figura F.2: Dedicación horaria y distribución de tareas por meses

F.2. PRESUPUESTO

En este apartado se detallan aspectos relacionados con el presupuesto del proyecto. En él, se incluyen los gastos de personal (Tabla F.2), de *software* y de *hardware* generados durante el desarrollo del TFM. Entre los gastos de *hardware* se ha tenido en cuenta el costo asociado a la adquisición del sistema óptico de captura de movimiento. Por su parte, en la Tabla X, se detallan las horas invertidas en el proyecto, así como el coste relacionado con la posible supuesta contratación de cada una de las categorías profesionales que estarían implicadas

en cada fase del proyecto. Los precios son los establecidos por el *Convenio Colectivo Estatal de Empresas*¹.

Tabla F.2: Coste de personal

Puesto	NºHoras	Coste/Hora	Total (€)
Analista	60	28	1.680 €
Programador	255	20	5.100 €
Responsable Documentación	75	16	1.200 €
Total	390	-	7.980 €

Tabla F.3: Coste total

Descripción	Coste (€)
Personal	7.980,00 €
Ordenador de desarrollo (<i>Acer Aspire E5-573G-56TS</i>)	579,00 €
Camáras <i>OptiTrack Primex 13W</i> (x8) + software <i>Clinical3DMA</i>	21.682,00 €
Adhesivo de velcro para los marcadores de 19 mm (x240)	50,81 €
Segmomómetro flexible	111,26 €
Cámara de vídeo <i>Anvasc Full HD 30 FPS</i>	19,99 €
Total	30.423,06 €

¹https://www.boe.es/diario_boe/txt.php?id=BOE-A-2018-3156