UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

MACHINE LEARNING-BASED CLASSIFIER OF GAIT PROFILES OF INDIVIDUALS WITH CEREBRAL PALSY TREATED IN THE ORITEL NETWORK

Rosario María Ulloa Barrientos

Thesis
for the degree of Master of Science in
Engineering with a major in Electrical Engineering.

Advisor Name
Esteban Javier Pino Quiroga

Date
September 01, 2025

Concepción (Chile)

Abstract

This thesis aimed to develop a gait pattern classifier for individuals with cerebral palsy (CP), using data collected from multiple motion analysis laboratories within the Organización Internacional de Teletones (ORITEL) network. To achieve this, the multicenter database *Movement Analysis Network* was created and standardized, comprising 156 studies from 8 laboratories, ensuring consistency and comparability across different motion capture systems.

Eight machine learning models were evaluated, considering class balancing strategies. The multilayer perceptron (MLP) with class weighting achieved the best performance, reaching an accuracy of 82% and F1-scores of 0.81 for diparesis, 0.63 for hemiparesis, and 0.98 for the control group. External validation on 63 additional trials confirmed the model's generalization capability, although performance for hemiparesis remained lower due to class imbalance.

Feature importance analysis and statistical tests (Median, Mann–Whitney U, Cliff's delta, and Bonferroni correction), showed that spatiotemporal parameters such as double support, swing phase, and stance phase, together with specific kinematic features such as minimum angle knee flexion/extesion (acLKFE), were key discriminators between hemiparesis and diparesis. These findings aligned with clinical literature and reinforced the interpretability of the model.

An exploratory analysis of treatment associations was also conducted. For hemiparesis, the most frequent treatments were orthoses, orthopedic surgery, and physical therapy, while for diparesis the most common were botulinum toxin, orthopedic surgery, and physical therapy. Although limited by incomplete treatment information, this analysis provided preliminary insights and a basis for future studies on treatment effectiveness.

In conclusion, this research successfully achieved its objectives: building a standardized multicenter database, developing a robust gait classifier, identifying clinically relevant features, and exploring preliminary associations with treatments. The database *Movement Analysis Network* represents a substantial contribution to multicenter gait research and has the potential to become a key tool for supporting clinical decision-making in the management of cerebral palsy.

Resumen

Esta tesis tuvo como propósito desarrollar un clasificador de patrones de marcha en individuos con parálisis cerebral (PC), utilizando datos obtenidos de distintos laboratorios de análisis de movimiento pertenecientes a la red Organización Internacional de Teletones (ORITEL). Para ello, se creó y estandarizó la base de datos multicéntrica *Movement Analysis Network*, que reunió 156 estudios provenientes de 8 centros, asegurando la comparabilidad y homogeneidad de la información entre diferentes sistemas de captura de movimiento.

Se evaluaron ocho modelos de aprendizaje automático, considerando estrategias de balanceo de clases. El perceptrón multicapa (MLP) con ponderación de clases alcanzó el mejor desempeño, con una exactitud del 82% y F1-scores de 0.81 para diparesia, 0.63 para hemiparesia y 0.98 para sujetos sin patología. La validación externa en 63 ensayos adicionales confirmó la capacidad de generalización del modelo, aunque se observó un menor rendimiento en hemiparesia debido al desbalance de la base de datos.

El análisis de importancia de variables y pruebas estadísticas (Media, Mann–Whitney U, delta de Cliff y corrección de Bonferroni), mostró que parámetros espaciotemporales como doble apoyo, fase de oscilación y fase de apoyo, junto con variables cinemáticas específicas como el mínimo del ángulo de flexión/extensión de la rodilla (acLKFE), fueron determinantes para diferenciar hemiparesia de diparesia. Estos resultados coincidieron con la literatura y reforzaron la interpretabilidad clínica del clasificador.

Asimismo, se realizó un análisis exploratorio de la asociación entre patrones de marcha y tratamientos. En hemiparesia, los tratamientos más frecuentes fueron órtesis, cirugía ortopédica y fisioterapia, mientras que en diparesia destacaron la toxina botulínica, la cirugía ortopédica y la fisioterapia. Aunque esta parte del estudio se vio limitada por la falta de información completa, permitió delinear tendencias iniciales y sentar bases para futuros estudios sobre la efectividad terapéutica.

En conclusión, esta investigación cumplió con los objetivos planteados: construir una base de datos estandarizada, desarrollar un clasificador robusto de patrones de marcha, identificar variables clínicas relevantes y explorar asociaciones preliminares con tratamientos. La base de datos *Movement Analysis Network* constituye un aporte sustancial a la investigación multicéntrica y tiene el potencial de convertirse en una herramienta clave para apoyar la toma de decisiones clínicas en el manejo de la parálisis cerebral.

Acknowledgments

I want to thank my family for their endless support and understanding during this process. To my sister Isidora, who—even from far away—was always close, cheering me on and giving me her unconditional support. To my dogs, Merlín and Princesa, whose love and silliness made even the hardest days easier to handle.

A very special thanks to Ignacio, for his infinite patience, for standing by my side every single day, and for keeping me grounded when my head was spinning in a thousand directions. To my lab friends, for filling this process with laughs, stories, and good company (sometimes too much food, too). And to my friends Josefa and Bernardita, for always being there to listen to my rants and celebrate the little victories with me.

I also want to thank my advisor, Professor Esteban Pino, for his guidance and for never losing patience with me. To Professor Manuela Galli and Lucia Donno, for their super valuable feedback that always helped push this project forward. To Verónica Rattin, Mónica Morante, and Claudio Rozbaczylo, for always being ready to lend me a hand whenever I needed it. And finally, to all the professionals from the ORITEL network, part of the *Movement Analysis Network* project, for their huge help and willingness during the data collection.

The authors thank the National Center for Artificial Intelligence CENIA FB210017, Basal ANID, for their support.

I acknowledge the financial support from ANID BECAS/Magíster Nacional 2024.

Table of Contents

Tables Index

Figures Index

Abreviaciones

CP Cerebral Palsy

EMG Electromyography

ML Machine Learning

MLP Multilayer Perceptron

ORITEL Organización Internacional de Teletones

PCA Principal Component Analysis

Polimi Politécnico de Milán

Chapter 1: Introduction.

1.1 General Introduction

Gait is one of the cyclical movements that individuals perform daily, a process that has been studied for many years due to its complexity. It consists of a series of events and has a unique component that distinguishes each individual, known as the concept of gait signature [1]. Gait analysis is of great relevance in the fields of medicine, rehabilitation, and biomechanics. By assessing how a person walks you can provide valuable information about their health and functional capacity [2]. Motion capture systems and gait analysis techniques have become essential tools for the diagnosis and monitoring of a wide variety of medical conditions, such as neuromuscular disorders, orthopedic injuries, and physical disabilities [3].

Gait laboratories play a crucial role in the study and analysis of human movement patterns during locomotion. These laboratories utilize motion capture systems and various techniques to gather precise and detailed data on individuals' gait [3]. It is important to note that such laboratories are present in rehabilitation centers in many countries, such as in the Teletón Network.

The evaluation conducted in gait laboratories provides detailed information about gait, obtaining both the kinematics and kinetics of the individual. This enables analysis for the subsequent diagnosis and treatment proposal by specialists. One of the multiple benefits of these assessment is that they enable the tracking of the patient's progress over time [4].

Among the pathologies associated with gait, Cerebral Palsy (CP) proves to be one of the most recurrent [5]. This neuromotor condition, affecting a significant number of individuals, entails a variety of manifestations in movements and coordination during gait [6]. Understanding gait patterns in individuals with CP is essential for diagnosis and clinical assessment [7].

1.2 Problem Statement

The lack of uniformity and standardization in gait laboratories is a recurring issue that hinders the comparison and exchange of information among different research centers and rehabilitation clinics. Additionally, there is variability in the way gait examinations are conducted and treatments are proposed for individuals in centers such as Teletón, where CP is among the most treated pathologies.

1.3 Motivation

The motivation arises from the *Movement Analysis Network* project, an initiative of Politecnico di Milano, Organización Internacional de Teletones (ORITEL), and Universidad de Concepción. The focus of this project is to promote collaboration, research, and continuous improvement in the field of motion analysis. This project enables professionals in this sector to learn from each other and advance their research more effectively. Thus, it aims to better characterize gait patterns in patients with neuro-musculoskeletal conditions affecting gait, facilitating improved treatment selection.

The *Movement Analysis Network* aims to establish a data network with accumulated experiences, allowing the sharing of best practices, methodologies, and results from each laboratory. In the future, a shared database is expected to be used with different techniques to support and analyze motion analysis laboratories. Furthermore, it aims to significantly contribute to human capital development, facilitating the training of professionals in the field.

1.4 Goals

1.4.1 Main Goal

Develop a machine learning–based gait pattern classifier for individuals with cerebral palsy, using a standardized gait database created from the motion analysis laboratories of the ORITEL Network, while identifying the main features and exploring associations with treatment options to support medical decision-making.

1.4.2 Specific Goals

- Standardize the collected data to ensure consistency and comparability across different laboratories.
- Develop a gait pattern classifier using Machine Learning (ML) techniques, specifically designed for patients with CP.

- Analyze which specific clinical variables, such as the severity level of CP or the age of individuals, are responsible for the classification of gait profiles.
- Associate the classification of gait profiles in individuals with CP with different treatments and effectiveness to support medical decision-making.

1.5 Hypothesis

The classification of gait patterns in patients with CP based on assessments in motion analysis laboratories will support the selection of an appropriate treatment in at least 80% of cases, enabling personalized interventions.

1.6 Scope and Limitations

Data collection requires ethics committee approval in each country; each center is responsible for obtaining this approval and send the data.
The database contains information on patients with various pathologies; however, for the purposes of this study, only data from patients with CP will be used.
The database is intended to expand over time, facilitating a range of studies aimed at advancing knowledge and care for patients with gait-related conditions, including CP.

1.7 Structure of the Document

This document is structured as follows:

- Chapter 1: Introduction. This chapter provides an overview of the context, motivation, and objectives of the study.
- Chapter 2: State of the Art. This chapter reviews the current state of research on gait analysis, focusing on CP and the use of ML techniques for gait pattern classification.
- Chapter 3: Methodology. This chapter describes the methodology used to develop the gait pattern classifier, including data collection, preprocessing, and classification techniques.

- Chapter 4: Database: Movement Analysis Network. This chapter presents the database used in the study, detailing the data collection process, standardization, and the characteristics of the dataset.
- Chapter 5: Results. This chapter presents the outcomes of gait pattern classification, the analysis of the most relevant features identified by the best-performing classifier, and the association of the resulting profiles with suggested treatments.
- Chapter 6: Discussion and Conclusion. This chapter discusses the implications of the results, the limitations of the study, and potential future work. Summarizes the main findings of the study and their significance in the context of gait analysis for patients with CP.

1.8 Work Plan

The work plan was divided into several stages in order to achieve the main goal of developing a gait pattern classifier for patients with CP. The stages are as follows:

- Data Collection: Collect data from motion analysis laboratories, ensuring that it is in the required format and consistent across different centers.
- Data Standardization: Standardize the collected data to ensure consistency and comparability across different laboratories.
- Gait Pattern Classifier: Develop and train a machine learning model to classify gait patterns in patients with CP using motion analysis data, and evaluate its performance through appropriate statistical metrics, including F1-score, accuracy, precision, and recall.
- Analysis of Clinical Variables: Analyze the relationship between selected clinical variables and the classification of gait profiles, identifying the features that most influence the results. Support the analysis with statistical tests to explore correlations and differences between groups.
- Treatment association with the classes model: Identify the treatments most frequently recommended by specialists for each gait pattern class in the dataset, and complement this with a literature review to discuss evidence on their potential benefits.

Capítulo 1: Introducción.

1.1 Introducción General

La marcha es uno de los movimientos cíclicos que las personas realizan a diario, un proceso que ha sido estudiado durante muchos años debido a su complejidad. Consiste en una serie de eventos y posee un componente único que distingue a cada individuo, conocido como el concepto de firma de la marcha [1]. El análisis de la marcha tiene una gran relevancia en los campos de la medicina, la rehabilitación y la biomecánica. Evaluar la forma en que una persona camina puede proporcionar información valiosa sobre su estado de salud y su capacidad funcional [2]. Los sistemas de captura de movimiento y las técnicas de análisis de la marcha se han convertido en herramientas esenciales para el diagnóstico y el seguimiento de una amplia variedad de condiciones médicas, como trastornos neuromusculares, lesiones ortopédicas y discapacidades físicas [3].

Los laboratorios de marcha juegan un papel crucial en el estudio y análisis de los patrones de movimiento humano durante la locomoción. Estos laboratorios utilizan sistemas de captura de movimiento y diversas técnicas para recopilar datos precisos y detallados sobre la marcha de los individuos [3]. Es importante destacar que este tipo de laboratorios están presentes en centros de rehabilitación de muchos países, como en la Red Teletón.

La evaluación realizada en los laboratorios de marcha proporciona información detallada sobre la marcha, obteniendo tanto la cinemática como la cinética del individuo. Esto permite realizar análisis para el posterior diagnóstico y propuesta de tratamiento por parte de especialistas. Uno de los múltiples beneficios de estas evaluaciones es que permiten el seguimiento del progreso del paciente en el tiempo [4].

Entre las patologías asociadas a la marcha, la parálisis cerebral resulta ser una de las más recurrentes [5]. Esta condición neuromotora, que afecta a un número significativo de individuos, conlleva una variedad de manifestaciones en los movimientos y la coordinación durante la marcha [6]. Comprender los patrones de marcha en personas con parálisis cerebral es esencial para el diagnóstico y la evaluación clínica [7].

1.2 Planteamiento del Problema

La falta de uniformidad y estandarización en los laboratorios de marcha es un problema recurrente que dificulta la comparación y el intercambio de información entre distintos centros de investigación y clínicas de rehabilitación. Además, existe variabilidad en la forma en que se realizan los exámenes de marcha y se proponen los tratamientos en centros como Teletón, donde la parálisis cerebral se encuentra entre las patologías más tratadas.

1.3 Motivación

La motivación surge del proyecto *Movement Analysis Network*, una iniciativa del Politécnico de Milán, ORITEL y la Universidad de Concepción. El enfoque de este proyecto es promover la colaboración, la investigación y la mejora continua en el ámbito del análisis de movimiento. Este proyecto permite que los profesionales de este sector aprendan unos de otros y avancen en sus investigaciones de manera más efectiva. Así, se busca caracterizar mejor los patrones de marcha en pacientes con condiciones neuromusculoesqueléticas que afectan la marcha, facilitando una mejor selección de tratamientos.

Movement Analysis Network tiene como objetivo establecer una red de datos con experiencias acumuladas, permitiendo el intercambio de buenas prácticas, metodologías y resultados de cada laboratorio. En el futuro, se espera contar con una base de datos compartida en la que se utilicen diferentes técnicas para apoyar y analizar los laboratorios de análisis de movimiento. Además, se busca contribuir significativamente al desarrollo de capital humano, facilitando la formación de profesionales en el área.

1.4 Objetivos

1.4.1 Objetivo General

Desarrollar un clasificador de patrones de marcha basado en ML en individuos con parálisis cerebral, utilizando una base de datos estandarizada de marcha creada a partir de los laboratorios de análisis de movimiento de la Red ORITEL, identificando las principales características y explorando asociaciones con opciones de tratamiento que apoyen la toma de decisiones médicas.

1.4.2 Objetivos Específicos

- Estandarizar los datos recopilados para asegurar consistencia y comparabilidad entre distintos laboratorios.
- Desarrollar un clasificador de patrones de marcha utilizando técnicas de ML, específicamente diseñado para pacientes con parálisis cerebral.
- Analizar qué variables clínicas específicas, como el nivel de severidad de la parálisis cerebral o la edad de los individuos, son responsables de la clasificación de los perfiles de marcha.
- Asociar la clasificación de perfiles de marcha en individuos con parálisis cerebral con diferentes tratamientos y su efectividad, para apoyar la toma de decisiones médicas.

1.5 Hipótesis

La clasificación de patrones de marcha en pacientes con parálisis cerebral, basada en evaluaciones realizadas en laboratorios de análisis de movimiento, apoyará la selección de un tratamiento adecuado en al menos el 80% de los casos, permitiendo intervenciones personalizadas.

1.6 Alcance y Limitaciones

La recopilación de datos requiere la aprobación de los comités de ética en cada país; cada centro es responsable de obtener esta aprobación y enviar los datos.
La base de datos contiene información de pacientes con diversas patologías; sin embargo, para los fines de este estudio, solo se utilizarán los datos de pacientes con parálisis cerebral.
Se espera que la base de datos se expanda con el tiempo, facilitando una serie de estudios orientados a avanzar en el conocimiento y la atención de pacientes con condiciones relacionadas con la marcha, incluyendo la parálisis cerebral.

1.7 Estructura del Documento

Este documento está estructurado de la siguiente manera:

- Capítulo 1: Introducción. Este capítulo proporciona una visión general del contexto, la motivación y los objetivos del estudio.
- Capítulo 2: Estado del Arte. Este capítulo revisa el estado actual de la investigación sobre análisis de la marcha, con énfasis en la parálisis cerebral y el uso de técnicas de ML para la clasificación de patrones de marcha.
- Capítulo 3: Metodología. Este capítulo describe la metodología utilizada para desarrollar el clasificador de patrones de marcha, incluyendo la recopilación de datos, el preprocesamiento y las técnicas de clasificación.
- Capítulo 4: Base de Datos: Movement Analysis Network. Este capítulo presenta la base de datos utilizada en el estudio, detallando el proceso de recopilación, la estandarización y las características del conjunto de datos.
- Capítulo 5: Resultados. Este capítulo presenta los resultados de la clasificación de patrones de marcha, el análisis de las variables más relevantes según el mejor clasificador y la asociación de los perfiles obtenidos con los tratamientos sugeridos.
- Capítulo 6: Discusión y Conclusión. Este capítulo analiza las implicancias de los resultados, las limitaciones del estudio y el trabajo futuro. Resume los principales hallazgos del estudio y su relevancia en el contexto del análisis de la marcha en pacientes con parálisis cerebral.

1.8 Plan de Trabajo

El plan de trabajo se dividió en varias etapas con el fin de alcanzar el objetivo general de desarrollar un clasificador de patrones de marcha para pacientes con parálisis cerebral. Las etapas son las siguientes:

- Recolección de Datos: Recopilar datos de los laboratorios de análisis de movimiento, asegurando que estén en el formato solicitado y sean consistentes entre distintos centros.
- Estandarización de Datos: Estandarizar los datos recopilados para asegurar consistencia y comparabilidad entre los laboratorios.
- Clasificador de Patrones de Marcha: Desarrollar y entrenar un modelo de aprendizaje automático para clasificar los patrones de marcha en pacientes con parálisis cerebral a partir de datos de análisis de movimiento, y evaluar su desempeño mediante métricas estadísticas apropiadas, incluyendo F1-score, exactitud, precisión y exhaustividad.

- Análisis de Variables Clínicas: Analizar la relación entre variables clínicas seleccionadas y la clasificación de perfiles de marcha, identificando las características que más influyen en los resultados. Apoyar el análisis con pruebas estadísticas para explorar correlaciones y diferencias entre grupos.

- Asociación de Tratamientos con las Clases del Modelo: Identificar los tratamientos más frecuentemente recomendados por especialistas para cada clase de patrón de marcha en el conjunto de datos, y complementar esto con una revisión de la literatura para discutir la evidencia sobre sus posibles beneficios.

Chapter 2: State of the Art

2.1 Cerebral Palsy and Gait

CP encompasses a group of permanent disorders affecting movement and posture development, attributed to non-progressive disturbances occurring in the fetal or infant brain during development. These disorders result in limitations in activity and movement, which change with age and development, making them dynamic [8]. Studying this pathology is complex due to its broad spectrum of motor impairments, making generalization of results and the application of standardized approaches challenging.

CP is one of the most common motor disabilities in childhood [9]. In Europe, the estimated incidence of CP is approximately 2.08 cases per 1000 births, highlighting the significant presence of this condition in the population [8, 10]. A study conducted in different regions of China observed an increase in the prevalence of CP in children and adolescents spanning the period from 1988 to 2020 [11]. These data underscore the importance of understanding and addressing CP, a condition that affects a considerable proportion of the population, with significant implications for healthcare and rehabilitation services.

Cerebral palsy is commonly defined based on three main characteristics: the type of motor dysfunction (spastic, dyskinetic, ataxic and mixed) [8, 12], the distribution of motor impairment (hemiparesis, diparesis, tetraparesis), and the severity level (mild, moderate, or severe), typically categorized using the Gross Motor Function Classification System (GMFCS) levels I through V [13].

Type of motor dysfunction classification is detailed below:

1. Spastic CP: characterized by muscle stiffness and spasms, which can affect one or more limbs. Spastic CP is the most common classification, accounting for 80% of cases [8].
2. Dyskinetic CP: involves involuntary and uncontrolled movements [8].
3. Ataxic CP: characterized by difficulties in balance and coordination [8].
4. Mixed CP: Combines characteristics of more than one type of CP, potentially involving the aforementioned types [8].

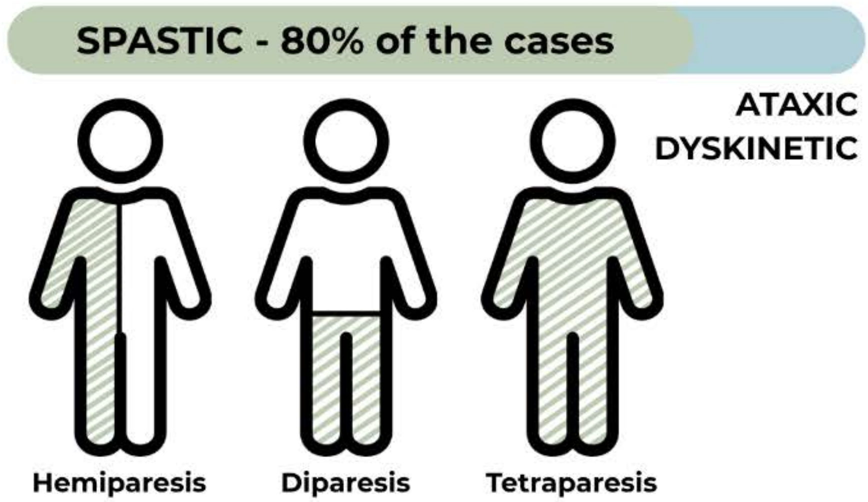

Fig. 2.1: Cerebral Palsy Classification: 80% of cases fall under spastic CP, further divided into Hemiparesis, Diparesis, and Tetraparesis.

The distribution of motor impairment can be identified, such as hemiparesis, diparesis, and tetraparesis [14] [15] [7]. In Fig. 2.1, a basic scheme of the CP classification is presented, indicating the body parts affected according to the category type.

Hemiparesis: One of the lateral halves of the body is affected, either the left or the right [16]. In this category, different types of gait profiles can be identified, which are described below:

- Type I — Drop Foot: Increased knee flexion at initial contact, increased hip flexion, and increased lordosis [15].
- Type II Includes all deviations present in Type I plus increased plantarflexion and knee hyperextension [15].
    1. True Equinus.
    2. Equinovarus foot and knee recurvatum.
- Type III — Jumping Equinus: Encompasses all deviations from Types I and II along with reduced movement [15].
- Type IV — Three-dimensional Impaired Gait: Corresponds to a combination of all the above deviations plus reduced hip movement and increased lordosis [15].

Diparesis: A more pronounced involvement in the lower half rather than the upper half of the body can be observed in this type of spastic CP [16]. The common gait patterns for this classification are presented below.

- True Equinus: The ankle is in plantar flexion throughout the entire stance phase ("toe walking") [7].
- Jump Gait: Equinus at the ankle (partially at the end of the stance phase), flexion of the knee and hip (especially at the beginning of the stance phase), anterior pelvic tilt, and increased lumbar lordosis [7].
- Apparent Equinus: The ankle has a normal range of motion, but the knee and hip are excessively flexed, and the heel lifts off the ground during walking [7].
- Squat Gait: The ankle is excessively dorsiflexed throughout the entire stance phase, and the knee and hip are excessively flexed [7].

Tetraparesis: All four limbs of the body are affected, therefore, there is a greater involvement of the body compared to hemiparesis and diparesis.

For the specific purpose of this study, only data related to hemiparesis and diparesis associated with spastic CP will be analyzed. As the analysis will focus exclusively on gait, the patterns of tetraparesis may be confused with those of diparesis, since both conditions similarly affect both legs.

Because specific gait patterns can be identified, it is possible to use data obtained from gait laboratories for their classification. Accurate classification of these gait patterns for patients with CP is essential, as it plays a crucial role in designing an effective treatment plan to address the specific needs of each individual affected by CP. This personalized approach significantly contributes to improving the quality of life and functionality of patients throughout their development.

The severity of CP refers to how much the motor difficulties affect a person's daily life. To evaluate this in a standardized way, the Gross Motor Function Classification System (GMFCS) is commonly used. This system classifies motor function into five levels, from I to V. Level I includes individuals who can walk without help, while Level V refers to those with very limited mobility who require assistance for almost all physical activities. The classification is based on what the person can do on their own, such as sitting, standing, or moving around, and it helps health professionals plan treatments, monitor progress, and provide the right support [17].

CP is characterized by affecting motor function, making the study of gait in individuals with this condition crucial for clinical understanding, treatment planning, research innovation, and enhancing the quality of life for individuals with CP.

2.1.1 Evidence-Based Rehabilitation Strategies for Hemiparesis and Diparesis

Evidence-Based Foundations for Hemiparesis treatments: A study conducted by Faccioli [18] provides evidence-based recommendations for the management and motor rehabilitation of children and adolescents with CP. The authors highlight that both physical and occupational therapy are effective in promoting muscle use, improving strength, and enhancing motor control. In cases where lower limb function is affected, interventions such as ankle–foot orthoses (AFOs) are recommended to support mobility and alignment. For individuals with spastic CP, a common subtype of botulinum toxin injections are advised to reduce spasticity and facilitate movement. Additionally, functional electrical stimulation is suggested as part of the therapeutic protocol to further improve neuromuscular activation.

These findings are also supported by Rana [19], who emphasize the role of multidisciplinary approaches, particularly physical and occupational therapies, in improving muscle coordination and motor abilities in hemiparetic children. The review further confirms the clinical utility of botulinum toxin and orthotic devices like AFOs for managing lower limb impairments, and acknowledges electrostimulation as a supportive strategy to enhance muscle activity and prevent disuse.

In addition, Gonzalez [20] conducted a systematic review focusing specifically on physical therapy interventions in children with CP. Their findings reinforce the effectiveness of physical therapy in enhancing motor function, postural control, and muscle strength, particularly in individuals with spastic presentations such as hemiparesis. The study outlines that tailored physical therapy programs often including neurodevelopmental techniques, strength training, and functional mobility exercises can significantly improve the quality of movement and participation in daily activities.

Evidence-Based Foundations for Diparesis treatments: The clinical practice guidelines published by Al Shami [21] support the use of several evidence based interventions for individuals with spastic diparesis, the most frequent form of bilateral CP. Among the recommended treatments, physical and occupational therapy are emphasized as key strategies to improve motor function and promote autonomy. In children with lower limb spasticity, orthopedic surgery is indicated to correct musculoskeletal deformities and improve gait function. For carefully selected candidates, selective dorsal rhizotomy (SDR) is proposed as a neurosurgical alternative to reduce spasticity and improve movement coordination.

This is reinforced by Shamsoddini [22], who review multiple evidence-based approaches for managing spasticity in children with CP. Their findings support the combined use of botulinum toxin injections, orthoses, and strengthening programs as effective interventions to reduce hypertonia, enhance functional mobility, and avoid secondary complications. These treatments, when implemented as part of an individualized care plan, contribute significantly to improving gait patterns and overall motor performance in children with spastic diparesis.

The use of hinged ankle-foot orthoses (AFOs) can improve gait function in children with spastic diparesis cerebral palsy by enhancing ankle dorsiflexion during stance and swing phases, thereby facilitating foot clearance and promoting a more efficient gait pattern. Proper adjustment of the dorsiflexion angle in the AFO has been shown to influence gait parameters such as stride length, cadence, and stability, supporting its role as an effective intervention for improving functional mobility in this population [23].

Additionally, Rana [19] corroborate the use of orthopedic surgery and selective dorsal rhizotomy as effective interventions for managing spastic diparesis, particularly in cases with severe muscle tightness and postural deformities. Their review highlights that a combination of pharmacological (botulinum toxin), surgical, and physical rehabilitation strategies leads to better outcomes in bilateral CP presentations.

2.2 Motion Analysis in Gait Assessment

Gait is a complex motor activity that has been extensively studied in the fields of rehabilitation and sports science. Currently, the most commonly used techniques for gait assessment include inertial measurement units (IMUs), which are favored for their low cost and ease of use [24], as well as technologies typically found in specialized gait laboratories, such as optoelectronic motion capture systems. These lab-based systems offer high precision and are widely regarded as the gold standard in clinical practice [25] [26] [27]. However, they are often expensive and require trained personnel to operate, making them less accessible for routine evaluations.

In addition, complementary tools such as force platforms, Electromyography (EMG), and 2D or 3D video recordings are frequently used to enrich both IMU-based and laboratory analyses, offering a more comprehensive understanding of gait dynamics [28]. Despite these advancements, the interpretation of gait data remains highly dependent on clinical expertise. This reliance introduces subjectivity, inter-observer variability, and time-consuming procedures, which ultimately hinder the standardization and scalability of assessments across multiple rehabilitation centers [29].

Given that gait has been evaluated for decades, a considerable amount of data is now available, offering the opportunity to automate certain aspects of the analysis and support clinical decision-making. In response to these challenges, there is a growing interest in developing intelligent systems capable of automatically interpreting or classifying gait patterns from sensor data. These systems are designed to offer objective, consistent, and scalable insights, which are especially useful in healthcare settings with high demand or limited resources [30].

Such limitations have paved the way for data-driven approaches, particularly those based on machine learning, which are increasingly being explored to enhance gait assessment through automated feature extraction, classification, and clinical interpretation.

2.3 Motion Capture Data in Biomechanics

In recent years, inertial measurement units (IMUs) have become very popular in gait analysis due to their low cost, portability, and ease of use outside the lab. As a result, many public datasets have been created using IMUs, focused on applications like activity recognition, fall detection, and pathological gait classification. Well-known examples include the MHEALTH [31] and PAMAP2 [32] datasets, as well as others targeting specific populations such as older adults or people with neurodegenerative diseases.

These datasets have played an important role in developing machine learning models for gait analysis. However, IMUs come with some limitations, they don't offer highly accurate estimates of joint angles, ground reaction forces, or full-body 3D kinematics.

To address this, several datasets have been built using optoelectronic motion capture systems, like those commonly found in gait labs. These systems use infrared cameras and reflective markers to track movement in three dimensions with high precision. Although more expensive and requiring trained personnel, the data they provide is extremely valuable for both clinical and biomechanical research.

Some notable examples include the dataset by Van Criekinge et al. [33], which combines optoelectronic motion capture with force platforms and EMG from both healthy participants and stroke survivors, and the dataset by Santos et al. [34], which synchronizes optical motion capture with IMUs. Other recent datasets also include sagittal video recordings and optical tracking at different walking speeds, enabling more comprehensive analysis and training of robust classification models.

Regarding motion datasets acquired using optoelectronic camera systems, specifically from individuals with CP, there are a few notable examples. One of them is a dataset available on Figshare, which includes 451 gait trials from children with cerebral palsy, all collected in the same gait laboratory using a Vicon motion capture system [35]. These data were used to train and validate gait classification models based on joint motion patterns.

Another example is a dataset published by KU Leuven, which includes kinematic and kinetic data of children with CP walking on a treadmill under three different conditions: baseline walking, walking with verbal feedback, and walking with virtual reality feedback on hip extension. These trials were also recorded using a Vicon system and are intended to study the immediate effects of feedback strategies on gait patterns [36].

However, there is currently a lack of publicly available datasets that combine gait data from multiple motion analysis laboratories. This highlights an existing gap, especially for developing models that generalize across clinical settings. In this context, the ongoing effort to build a multicenter dataset through the Oritel Network using real data from diverse rehabilitation centers in Latin America offers a valuable opportunity. As this dataset continues to grow over time, it may support future clinical applications and create new opportunities for developing gait pattern classifiers using machine learning techniques, with the potential to directly improve patient care for individuals with cerebral palsy.

2.4 NoSQL Databases Management

In recent years, NoSQL databases have gained significant traction as alternatives to traditional relational database management systems, especially in scenarios that demand high scalability, availability, and flexible data models. Unlike relational databases that enforce strict schemas and rely on structured query language (SQL), NoSQL systems are schema-less or schema-flexible, enabling storage and retrieval of semi-structured or unstructured data at scale [37].

NoSQL databases are generally categorized into four major types [38, 39]:

- **Document-based databases:** These store data in document formats such as JSON or BSON. They are suitable for applications where data structures evolve frequently. Examples include MongoDB and CouchDB [40, 41].
- **Key-value stores:** These are the simplest form of NoSQL databases, where data is stored as a collection of key-value pairs. This model is efficient for high-performance lookups and is often used in caching or session management. Examples include Redis and Amazon DynamoDB [42,

43].

- **Column-family stores:** Instead of storing data in rows, these databases store data in columns, making them ideal for large-scale analytical processing. Examples include Apache Cassandra and HBase [44, 45].
- **Graph databases:** These are optimized for representing and traversing relationships between data entities. They are particularly useful in recommendation systems, fraud detection, and social networks. Examples include Neo4j and Amazon Neptune [46, 47].

Among these, **MongoDB** has emerged as one of the most widely adopted document oriented databases. According to Ouyang [48], MongoDB offers robust horizontal scalability via sharding, replication through replica sets, and support for flexible data models. It enables developers to store complex, nested documents without the need for rigid schemas, which simplifies the development cycle.

MongoDB also supports indexing, aggregation pipelines, and secondary reads, making it well-suited for a variety of workloads, from real-time analytics to content management systems [49]. It supports tunable consistency levels through snapshot isolation, providing a balance between performance and consistency depending on the application's needs.

However, as pointed out by Ouyang [48], MongoDB's flexibility can also be a drawback in systems that require strong transactional guarantees. MongoDB prioritizes scalability and flexibility, but in some configurations it may sacrifice immediate consistency, which can be problematic for applications that cannot tolerate temporary inconsistencies. Additionally, poorly designed schemas or indexing strategies can lead to performance bottlenecks in large scale deployments.

NoSQL databases offer a wide range of solutions tailored to specific types of data and access patterns. The decision to adopt a particular NoSQL database, whether document-based like MongoDB, key-value like Redis, or columnar like Cassandra, depends largely on the specific requirements of the application, such as the type of data being handled, consistency needs, expected query patterns, and scalability constraints. Therefore, choosing the appropriate NoSQL model is a matter of aligning the database's strengths with the system's priorities.

2.5 Machine Learning for Gait Pattern Classification: Current State

In recent years, gait pattern classification has become a highly active research area, with applications ranging from the detection of foot abnormalities and gait anomalies to the classification of patterns in individuals with Parkinson's disease, gait phase recognition, and the identification of pathological gait profiles in various populations, including individuals with CP [50, 51, 52, 7].

The application of ML in gait related tasks has included diverse objectives such as step detection, gait event segmentation, pathology identification, asymmetry detection, and neurological gait studies [53, 54, 55, 56]. These tasks rely on data from various modalities, including inertial sensors (IMUs), optoelectronic motion capture systems, electromyography (EMG), and even 2D/3D video recordings.

Several studies have demonstrated the potential of ML models across clinical and research settings. One study trained support vector machines (SVMs) using inertial measurement unit (IMU) data and spatiotemporal gait parameters to classify orthopedic conditions such as knee or hip osteoarthritis and spinal stenosis, achieving an accuracy of 82% in distinguishing pathological from normal gait [57]. Another investigation developed a machine learning pipeline to classify artificially impaired gait patterns caused by joint-specific restrictions, also using IMUs, and reported classification accuracies exceeding 91% when employing SVM, Random Forest, and XGBoost algorithms [58].

More recently, deep learning approaches including convolutional neural networks (CNNs), long short-term memory (LSTM) models, and attention-based architectures, have shown promise in modeling the temporal and spatial dynamics of gait. A 2025 systematic review highlights the growing use of CNNs, SVMs, and ensemble methods in Parkinson's gait analysis, reporting classification accuracies exceeding 99% when combining multimodal data sources [59]. However, the review also underscores persistent challenges such as the limited size and diversity of existing datasets and concerns related to model interpretability in clinical applications.

This ability to detect subtle and nonlinear gait deviations has generated growing interest among researchers and clinicians alike. Beyond improving classification performance, ML methods are being investigated as tools for developing personalized clinical interventions. For instance, a recent study applied Random Forest and Decision Tree models to predict gait recovery in individuals with spinal cord injuries at discharge from a rehabilitation center, offering decision-support capabilities for personalized treatment strategies [60]. Another investigation involving post-stroke individuals used a gait mat system to extract 39 spatiotemporal features, which were then classified using ML algorithms such as RF, SVM., and KNN. The study reported classification accuracies above 85%, demonstrating the potential of these techniques to identify gait biomarkers associated with lesion location and to inform targeted therapies [61]. Furthermore, a 2024 systematic review highlighted that AI-assisted gait analysis tools can improve clinical outcomes, including stride length, cadence, and walking distance, while also enhancing patient adherence to rehabilitation programs. These findings underscore the value of ML-based approaches in optimizing clinical decision-making and treatment personalization [62].

Despite the progress in the field, important challenges still need to be addressed. Many studies rely on small or relatively homogeneous datasets, which can limit how well the models perform when applied to real world clinical populations. Additionally, the black-box nature of many deep learning approaches raises concerns about their interpretability and the level of trust clinicians can place in them. Overcoming these limitations is essential to ensure that ML-based gait analysis tools can be safely and effectively integrated into routine clinical practice.

2.6 Machine Learning for Gait Pattern Classifiers in Cerebral Palsy

CP is one of the most common gait-related neurological disorders seen in pediatric rehabilitation centers such as Teletón. However, the wide variability in motor impairments and movement patterns associated with CP makes it difficult to develop gait classification systems that are both accurate and generalizable. A recent systematic review found that while some models like Random Forest, have achieved classification accuracies as high as 94%, most studies vary significantly in sample size and experimental design, limiting their clinical relevance [63]. This has led to growing interest in ML approaches that can support clinical decision-making.

To address these challenges, researchers have explored a variety of strategies, such as aligning clinical assessments with data-driven predictions, linking biomechanical gait patterns to orthotic recommendations, analyzing gait indices across clinical subgroups, and applying pattern recognition techniques to both 2D and 3D motion data [7, 64, 65, 66, 67]. Despite these efforts, the diversity of gait patterns among individuals with CP continues to limit model generalization, emphasizing the need for more flexible and scalable solutions.

Recent studies have applied ML techniques to classify gait patterns in individuals with CP, using both traditional models and deep learning architectures. For example, one study showed promising results in classifying diparetic gait patterns, although it was constrained by a small and clinically homogeneous dataset [7]. Another used accelerometer data from wearable sensors to classify physical activity in children with CP, comparing multiple ML models in real-world environments [68].

Al-Sowi [69] compiled a national dataset of 305 children with CP in Jordan, featuring 316 clinical features derived from standardized assessments. They tested five ML classifiers (K-Star, MLP, Naïve Bayes, Random Tree, and SVM) on two classification tasks: CP type (six classes) and GMFCS level (five levels). Using 10-fold cross-validation repeated 10 times, the Multilayer Perceptron (MLP) model achieved 84% accuracy for CP type and 53% for GMFCS level, demonstrating how structured clinical data can improve classification outcomes.

Recent work has also focused on selecting more clinically meaningful features for classification, particularly kinematic variables related to the ankle and knee [1]. While earlier models often relied on pelvic features [7, 64], recent trends show a shift toward variables that align better with rehabilitation goals and orthotic planning.

Building on these efforts, Arias-Valdivia [70] developed a deep learning model to distinguish between hemiplegia and diplegia in children with CP, using postural control data from force platforms. Their model reached an accuracy of 76.4%, offering a non invasive and efficient alternative for CP subtype identification. Similarly, Slijepcevic [71] explored functional gait disorder classification using ground reaction force data from 279 patients, applying PCA and LDA to establish a solid baseline for multiclass classification.

In summary, while ML-based classification of gait patterns in CP continues to evolve and shows strong potential, more work is needed. Future studies should aim to use larger and more diverse datasets, better represent the clinical spectrum of CP, and improve model interpretability to enhance their clinical applicability and support personalized rehabilitation strategies.

2.7 Key clinical features differentiating Hemiparesis and Diparesis in Cerebral Palsy

At a glance, hemiparesis can be distinguished from diparesis by the symmetry of gait. Hemiparesis affects one side of the body, creating noticeable differences between the affected and unaffected limbs, while diparesis involves both sides, often with greater impairment in the lower limbs [72].

- **Joint Range of Motion:** In hemiparesis, the affected limb usually shows reduced range of motion, leading to compensatory movements on the opposite side. In diparesis, both lower limbs have reduced range of motion in a more uniform pattern. This may result in a longer stance phase on both sides in diparesis, while in hemiparesis the stance phase may be shorter on the affected side. Similarly, stride length in hemiparesis is often shorter on the affected side, whereas in diparesis it is shorter but similar on both sides [73, 74, 75].

- **Balance and Coordination:** Hemiparetic individuals may have balance issues due to unilateral weakness, while those with diparesis experience overall stability and coordination challenges due to bilateral lower limb involvement [76, 77]. Double support time increases when instability is present [78]. Because of muscle weakness and poor coordination, walking speed is often slower in diparesis, while in hemiparesis it can be normal or slightly reduced, depending on severity [79, 80].

In kinematic terms, the main difference lies in asymmetry versus symmetry of the impairments. Hemiparesis produces clear side to side differences in joint angles, range of motion, and angular velocities, often leading to compensations such as increased pelvic tilt, obliquity, or rotation, as well as exaggerated hip abduction or external rotation on the unaffected side to help foot clearance. These asymmetries are also evident at the ankle and knee, where the affected limb often shows reduced dorsiflexion during swing, increased plantarflexion at initial contact, and limited knee flexion ("stiff knee gait"), while the unaffected limb may compensate with greater dorsiflexion and knee flexion [73, 81]. Diparesis, on the other hand, shows bilaterally reduced joint movement with fewer side to side differences, including symmetrical limitations in ankle dorsiflexion, push off, and knee flexion, resulting in a more rigid and less variable gait pattern [82, 83].

Chapter 3: Methodology

The proposed methodology to carry out this study will be structured in several phases with the aim of comprehensively addressing the classification of gait profiles in individuals with CP. In the first stage, the focus will be on creation of the database. Second, the classifier was developed and trained, encompassing feature preprocessing, model selection, and hyperparameter optimization. Third, model accuracy and generalization were assessed through an internal evaluation. Fourth, an external validation was conducted on an independent cohort that was not used during training or internal testing, providing an unbiased estimate of performance. Finally, the classifier's predictions were associated with the most frequently recommended treatments for each CP subtype (hemiparesis and diparesis), and a feature importance analysis was performed to identify the variables that most strongly contributed to the model's decisions.

3.1 Creation of the Database

The database was compiled using previously acquired data from motion analysis laboratories at various Teletón institutes, which are part of the ORITEL network, across the countries of Mexico, El Salvador, Colombia, Uruguay, and Chile. Each laboratory was responsible for the selection and anonymization of data. These dataset includes kinematic and spatiotemporal parameters, in addition to providing relevant details about the patient's diagnosis and treatment, further details can be found in Table 3.1. Optionally, some studies also provide kinetic and EMG data.

It is important to mention that in order to the creation of the Database, it was necessary to obtain approval from the ethics committee to use the data. This approval ensures that the data usage complies with ethical standards and protects the rights and privacy of the participants. In the case of Chile, the approval was granted by the *Comité de Ética Científico del Servicio de Salud Concepción* for the data from Teletón Santiago and Concepción centers, with the study identified by the code 24–01–01 and approved on April 23rd of 2024. In case of the other countries, the approval was obtained by each center independently, and the data were sent to the database after anonymization.

Prior to data collection, a survey was conducted with the 8 participating gait laboratories to gather basic information about each. This survey provided the initial basis for initiating the common database. Table 3.2 shows some of the responses obtained.

Table 3.1: Data record of each study from *Movement Analysis Network* database.

Data Record	Detail information
Acquisition Information	Laboratory Name Acquisition System (Vicon or BTS GaitLab) Year Acquisition
Anthropometric Measurements	Weight Height Pelvic Width Pelvic Depth Lower Limb Length Knee Diameter Ankle Diameter
Medical history	Study Objective (pre-post surgery, orthotic modification, etc.) Clinical (rehabilitation or evaluation) Diagnostic Treatment Comments
Data	Kinematic Kinetic EMG

The gait laboratories of the participating institutes use different acquisition systems, including Vicon Systems and BTS GaitLab, both of which are systems based on optoelectronic cameras, as shown in Table 3.2. Therefore, being two different systems, the studies are stored in different file formats, because of that, it was necessary to standardize the data from both systems to ensure that they can be compared and analyzed uniformly.

- **Vicon**
  Regarding the Vicon system, studies are exported in C3D format, which stores detailed information about the kinematics and kinetics of the movement. It also contains all the necessary information for reading, visualization, and motion analysis [84]. These C3D files are compatible with publicly available software, such as Python and Matlab, as long as the corresponding libraries are added [85].

Table 3.2: Overview of participating laboratories and their equipment within the *Movement Analysis Network*.

Country	City	System type	Start year	Number of force platforms	Number of optoelectronic cameras
México (3)	Saltillo	BTS	2021	6	8
	Guadalajara	BTS	2012	6	8
	Tlalnepantla	BTS	2000	6	8
El Salvador (1)	La libertad	BTS	2004	2	8
Colombia (1)	Bogota	BTS	2001	8	8
Uruguay (1)	Montevideo	VICON	2011	2	8
Chile (2)	Santiago	BTS	2002	4	8
	Concepción	VICON	2008	3	8

- **BTS GaitLab**

  Unlike Vicon, BTS GaitLab has the capability to export data in TDF and/or EMT files. Considering the difficulty in opening TDF files, the decision was made to work with EMT files.

  EMT data corresponds to a text file in table format, allowing visualization on various platforms such as Excel [86]. This file indicates the unit of measurement, the number of cycles in the study, and the data for graphing kinetic or kinematic information, as applicable.

Regarding the spatiotemporal parameters, both systems were required to provide the data in .xlsx format, as this information is only available in gait reports, which also include subject-specific details. To address this, a Python script was developed and supplied to extract only the relevant information from the reports and store it in .xlsx files, thereby streamlining the standardization process and ensuring data anonymization.

The database *Movement Analysis Network* included the following pathologies: CP, spina bifida, congenital muscular dystrophies, stroke, degenerative neurological disorders, neuromuscular diseases, amputations, idiopathic toe walking, arthrogryposis multiplex congenita, spinal cord injuries, traumatic brain injury, as well as other less common gait-related conditions.

3.1.1 Data Standardization

After collecting the data from the participating laboratories, a standardization process was performed to ensure that all information followed a consistent format. This step was essential to enable accurate analysis and reliable classification of gait patterns. The entire procedure was carried out in MATLAB and included the following stages:

For C3D-formatted data:

- **Data Cleaning:** Outliers and erroneous data were removed to improve accuracy and ensure the reliability of the analysis.
- **Selection of the Most Representative Gait Cycle:** Since C3D files typically contain the entire gait study rather than a pre-selected cycle, an automated method was used to identify and extract the most representative gait cycle. This ensures that the classifier operates on the most relevant portion of each trial. Example in Fig.3.1 shows the standardization of the hip angle data from a C3D file, where the most representative gait cycle is highlighted.
- **Data Normalization:** All variables were normalized to a common scale to facilitate meaningful comparisons and improve model performance.
- **Sampling Frequency Adjustment:** When needed, the sampling frequency was resampled to 100 Hz, the standard in gait analysis, to ensure consistency across datasets.
- **Variable Renaming and Label Replacement:** To ensure consistency across datasets from different sources, all variable names were standardized using a common naming convention. The complete list of standardized labels can be found in Appendix A, specifically in Tables A.2, A.4, and A.1.

For EMT-formatted data:

- **Data Cleaning:** As with the C3D files, any outliers or anomalies were removed to ensure data integrity.
- **Unit Normalization:** Measurement units were standardized to ensure that all variables are expressed consistently, which is critical for reliable comparisons.
- **Variable Renaming and Label Replacement:** A unified naming convention was also applied to the EMT data, following the same scheme used for the C3D format. The full list of standardized labels is available in Appendix A, in Tables A.2, A.4, and A.1.

After the standardization process, the data was uploaded to the *Movement Analysis Network*, a database created using MongoDB. This NoSQL system stores the standardized data in JSON format,

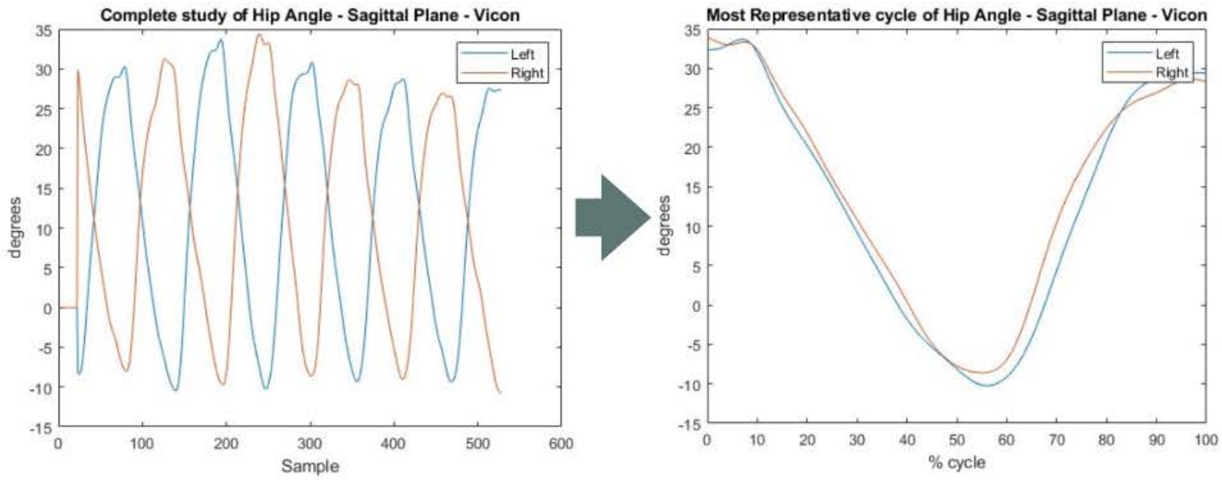

Fig. 3.1: C3D Hip Angle: gait-cycle standardization (0–100%).
Right: full gait cycles for both limbs; **Left:** clinician-selected representative stride.

allowing for flexible and scalable access. The database simplifies data retrieval and supports analysis and classification tasks across different laboratories.

It is worth noting that this ongoing database will continue to expand over time. In the future, there are plans to leverage Artificial Intelligence tools to increasing the information obtained, automate analysis processes and decision-making, and uncover hidden patterns and insights that could contribute to significant advancements in various fields of study.

3.1.2 Data Selection for Gait Pattern Classification

After the standardization process and the creation of the *Movement Analysis Network* database, the data used to develop the gait pattern classifier for CP were selected. Only subjects diagnosed with CP were included, following the specific inclusion and exclusion criteria described in Section 3.1.3. In addition, data from subjects without gait pathologies, provided by Politécnico de Milán (Polimi), were incorporated as the control group. These control data were not integrated into the *Movement Analysis Network* database itself but were considered for the analysis and the development of the classifier. Moreover, the control group data underwent the same standardization process to ensure consistency in format with the rest of the dataset.

3.1.3 Criteria for inclusion and exclusion

- **Inclusion**:

1. All patients, regardless of gender, who have been evaluated in the gait laboratories must be older than 7 years, as they have reached gait maturity.
2. Subjects who have the capacity for independent or self-assisted gait with technical aids, i.e., without the need for assistance from third parties.
3. Individuals must have at least 4 gait trials recorded during the study, with each trial containing 2 gait cycles (one per limb).

- **Exclusion**:

  1. Subjects with history of orthopedic or neurosurgical surgeries prior to their first gait study in the Teletón gait laboratory.
  2. Individuals who have undergone botulinum toxin injections within 6 months prior to their first gait study in Teletón gait laboratory.
  3. All patients whose first evaluation was conducted at Teletón before the year 2019.

3.2 Classifier Development and Training with Machine Learning Techniques

For the development of the gait profile supervised classifier of 3 classes, it was essential to work with a well-structured and standardized database. From the *Movement Analysis Network* database, data from subjects diagnosed with CP were selected. Additionally, data from a control group of healthy subjects, provided by Polimi, were included for comparison. The database selected was processed and analyzed using Python, included information about each subject, along with four to five trials per study containing kinematic and spatiotemporal parameters. In total, data from 95 studies were used: 72 corresponded to individuals with CP, and 23 to healthy individuals. Among the 72 individuals diagnosed with CP, 30 presented with hemiparesis and 42 with diparesis. It is important to note that all participants with cerebral palsy had spastic CP, with the exception of a single case, and all were classified within GMFCS (Gross Motor Function Classification System) levels I to III, indicating mild to moderate functional limitations. In order to increase the dataset size, all trials from each study were considered, making sure that trials from the same study were assigned to the same group, either training or testing. This resulted in a total of 413 trials. The data distribution is presented in Table 3.3.

1. **Standardization of the diagnosis label:** Since all subject information was entered by the professional responsible for exporting the data, it was necessary to standardize the diagnosis labels. Specifically, the labels were normalized as follows: *"hemiparesis"* for cases of CP with hemiparesis presentation, *"diparesis"* for subjects diagnosed with CP with diparesis involvement,

Table 3.3: Distribution of the selected dataset used for classification, including the number of studies, sex distribution, and total trials across cerebral palsy gait patterns and healthy controls.

Classes	Number of studies	Female studies (n)	Male studies (n)	Number of trials
Cerebral palsy with hemiparesis	30	13	17	121
Cerebral palsy with diparesis	42	13	27	177
Healthy individuals	23	11	12	115
Total	95	38	57	413

and *"sin patología"* for healthy individuals. After this standardization, the dataset was manually reviewed to ensure that each subject's diagnosis was correctly assigned.

2. **Input data selection:** The input data for the classifier was selected based on the available information in the database and included both numerical and categorical variables. The numerical features consisted of statistical descriptors (mean, standard deviation, kurtosis, skewness, minimum, and maximum) for each kinematic variable. In addition, the following spatiotemporal parameters were included: mean support phase, mean swing phase, mean double support, mean support duration, mean swing duration, mean stride duration, mean step length, mean stride length, cadence, step width, and average speed. Several anthropometric measurements were also considered, including: weight, height, pelvis width, knee diameter, ankle diameter, lower limb length, and pelvis depth. On the other hand, the categorical features included the diagnosis label, the marker protocol used, the condition under which the exam was performed, the age of the subject, and the acquisition system.
   For the categorical variables, the fuction *LabelEncoder* from the *sklearn* library was used to convert them into numerical values, allowing them to be processed by the classifier. The complete list of input variables can be found in Appendix A, specifically in the Tables A.1, A.2, and A.3.

3. **Balance the dataset:** The table 3.3 shows that the dataset is unbalanced, with a higher number of trials for individuals with diparesis compared to those with hemiparesis and healthy individuals. To address this issue, two strategies were explored. First, the *SMOTE* algorithm from *imblearn* was applied to balance the dataset by generating synthetic samples for the minor-

ity classes, ensuring that each class had an equal number of samples. The resulting balanced dataset consisted of 177 trials for each class. Second, the *sample weights* approach was tested, which does not generate new data but instead assigns higher importance to underrepresented classes during training, allowing the model to better account for their lower frequency.

4. **PCA for reduction of dimensions:** Another key step to prevent data leakage was the application of the Principal Component Analysis (PCA) algorithm from *sklearn* library, which reduces the dimensionality of the dataset while preserving most of its variance. This reduction helps minimize the risk of overfitting and enhances the classifier's performance. By applying PCA, a smaller set of features was obtained, still containing the most relevant information for the classification task.

5. **Data Splitting:** The dataset was split into training and testing sets, with 80% of the data used for training and 20% for testing. To ensure that all trials from the same study were assigned to the same partition, either training or testing, and to maintain class balance across folds, the *StratifiedGroupKFold* function from the *sklearn* library was used. This cross-validation strategy combines stratification, to preserve the class distribution and grouping, to avoid splitting data from the same study across folds. This approach helps prevent data leakage and ensures that the classifier is evaluated on truly unseen data, improving the reliability of the model's performance estimates.

6. **Optuna for hyperparameter optimization:** The *Optuna* [87] library was used to optimize the hyperparameters of the classifier. This library allows for efficient and automated hyperparameter tuning, improving the model's performance. This step involved Bayesian search to find the best combination of hyperparameters for each classifier. Instead of testing all combinations or choosing them randomly, Optuna learns from previous results to suggest better options at each step. It uses a method called Tree-structured Parzen Estimator (TPE) [88], which helps pick the most promising values based on past performance, making the search faster and smarter. For each classifier, 200 optimization trials were executed to explore the hyperparameter space.

7. **Classifier validation and testing**: The model's generalization ability was evaluated using several traditional machine learning algorithms, including Support Vector Machine (SVM) [89], decision tree [90], logistic regression [91], random forest [92], Extreme Gradient Boosting (XGBoost) Classifier [93], gradient boosting classifier [94], Stochastic gradient descent (SGD)

Classifier [95], and MLP [96]. A 10-fold cross-validation procedure was first applied on the training set to select the best performing model. The selected classifier was then evaluated on an independent test set (20% of the data) to assess its generalization ability. Performance was assessed using several evaluation metrics, including accuracy, precision, recall, F1-score, and the confusion matrix. These metrics offer a comprehensive understanding of the model's effectiveness and its ability to distinguish between different gait patterns. In particular, the F1-score uses the *weighted* average (`average='weighted'`), which is recommended for imbalanced datasets such as the one used in this study.

8. **Hyperparameter Tunning:** For hyperparameter tuning, multiple combinations of parameters were explored and progressively refined based on the model's performance. The search space was adjusted and narrowed according to the characteristics and behavior of each algorithm. The best hyperparameters were selected based on the results obtained through cross-validation. Additionally, an analysis of the top 10 trials for each model was conducted to assess the relative importance of different hyperparameters, helping to identify which ones had the greatest impact on performance.

9. **External validation:** To further evaluate the robustness of the classifier, an external validation was conducted using a separate test set that was not involved in training or hyperparameter tuning. This allowed for an unbiased assessment of the model's performance on truly unseen data. The external set included 63 trials from individuals with CP, of which 18 corresponded to hemiparesis and 45 to diparesis. The number of trials per subject in this set varied (2, 3, or 4 trials), unlike the more uniform distribution in the training data. No healthy control subjects were included in this evaluation, as the goal was to specifically test the model's ability to distinguish between different CP presentations.

Figure 3.2 presents a schematic of the evaluation workflow for the ML-based classifier. The dataset is split 80/20 (train/test); ten-fold cross-validation on the training set, guided by Optuna, tunes hyperparameters and identifies the best performing model, which is then refit and evaluated on the held out test set.

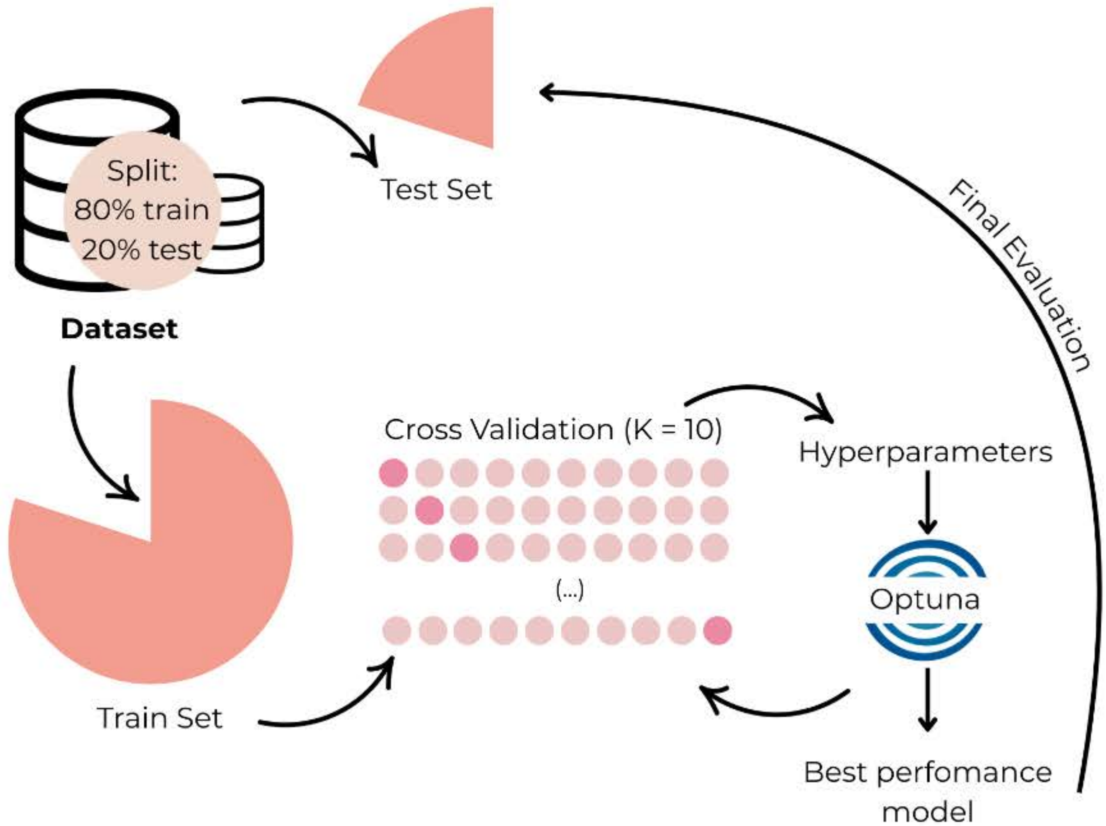

Fig. 3.2: Schematic of the evaluation workflow for the ML-based classifier.

3.3 Important Features for Classification Analysis

To identify the most relevant parameters for distinguishing hemiparesis from diparesis, Principal Component Analysis (PCA) was applied to the complete dataset, which included kinematic, spatiotemporal, and anthropometric variables. This approach reduced dimensionality while retaining the variables with the greatest contribution to the variance, highlighting the features most strongly associated with the classification task.

The variables with the highest PCA loadings were selected and compared between classes (hemiparesis or diparesis). For each variable, class medians were calculated, and differences were assessed using the Mann–Whitney U test. Effect sizes were estimated with Cliff's delta, and p-values were adjusted for multiple comparisons using the Bonferroni method. Additionally, a bibliographic review was conducted to evaluate whether the identified features have been previously reported as relevant for differentiating these gait patterns.

3.3.1 Median

The median is a measure of central tendency representing the middle value of an ordered dataset [97]. For a dataset $X = \{x_1, x_2, \ldots, x_n\}$ sorted in ascending order, it can be computed as shown in Equation 3.3.1:

$$\text{Median}(X) = \begin{cases} x_{\frac{n+1}{2}}, & \text{if } n \text{ is odd,} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & \text{if } n \text{ is even.} \end{cases} \tag{3.3.1}$$

In this study, the median for each feature was calculated separately for the hemiparesis and diparesis classes, providing a robust estimate of the central tendency that is less sensitive to outliers than the mean.

3.3.2 Mann–Whitney U

The Mann–Whitney U test is a non-parametric statistical method used to assess whether two independent samples come from the same distribution [98]. It is particularly useful when the assumption of normality is not satisfied, as it relies on rank comparisons rather than raw data.

Given two independent samples of sizes n_1 and n_2, the rank sums R_1 and R_2 are computed after jointly ranking all observations. The test statistics U_1 and U_2 are defined as:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1, \tag{3.3.2}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2, \tag{3.3.3}$$

and the final test statistic U is taken as:

$$U = \min(U_1, U_2). \tag{3.3.4}$$

Under the null hypothesis H_0 (that both groups follow the same distribution), the expected mean and variance of U are:

$$\mu_U = \frac{n_1 n_2}{2}, \quad \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}. \tag{3.3.5}$$

For sufficiently large samples, U is approximated by a normal distribution, and the standardized statistic is computed as:

$$Z = \frac{U - \mu_U}{\sigma_U}. \tag{3.3.6}$$

The corresponding two-sided p-value, denoted as p_{raw}, was then derived from the standard normal distribution. If $p_{\text{raw}} < 0.05$, the null hypothesis was rejected, indicating a statistically significant difference between the hemiparesis and diparesis groups. In this study, the Mann–Whitney U test was applied independently to each feature, and the resulting p_{raw} values were subsequently corrected for multiple testing using the Bonferroni method.

3.3.3 Cliff's Delta

Cliff's delta is a non-parametric effect size measure that quantifies the difference between two independent samples [99]. Given two samples X and Y of sizes n_x and n_y:

$$\delta = \frac{\#(x_i > y_j) - \#(x_i < y_j)}{n_x \cdot n_y} \tag{3.3.7}$$

where $\#(x_i > y_j)$ is the number of all possible pairs (x_i, y_j) such that x_i from class X is greater than y_j from class Y, and $\#(x_i < y_j)$ is the number of pairs where x_i is less than y_j. The value of δ ranges from -1 (all values in X are smaller than in Y) to $+1$ (all values in X are greater than in Y), with 0 indicating complete overlap between the distributions.

3.3.4 Bonferroni Correction

To control the family-wise error rate due to multiple comparisons, the Bonferroni correction was applied [100]. Given m hypotheses tested and their corresponding raw p-values $p_1, p_2, \ldots, p_m$, the Bonferroni-adjusted p-value for each hypothesis is computed as:

$$p_{\text{bonferroni}} = \min(1, m \cdot p_i) \tag{3.3.8}$$

where:

- p_i is the raw p-value of the i-th hypothesis,

- m is the total number of hypotheses tested,
- $p_{\text{bonferroni}}$ is the Bonferroni-adjusted p-value, truncated at 1.

The null hypothesis for the i-th test is rejected if $p_{\text{bonferroni}} \leq \alpha$, where α is the desired family-wise error rate (e.g., 0.05). In this study, m corresponded to the number of features tested, and α was set to 0.05.

3.4 Association of the classifier with treatments:

For the selection of the most appropriate treatment for each gait pattern (hemiparesis and diparesis), the most frequently recommended treatments found in the database used to train the classifier were considered. This information was further complemented by a literature review focused on identifying the most effective evidence, based treatments for each condition. As a result, the treatments were systematically associated with each predicted gait pattern by the classifier.

Chapter 4: Database: Movement Analysis Network

This dataset was created as part of the *Movement Analysis Network* project, a collaborative initiative between Polimi, ORITEL, and Universidad de Concepción. One of the main objectives of this initiative is to establish a comprehensive and historical database of motion analysis data collected from several Teletón rehabilitation centers across the ORITEL network.

4.1 Clinical Scope and Contributing Centers

To define the clinical scope of the database, a survey was conducted among the participating Teletón centers to identify the most commonly treated conditions. Based on the results, the following pathologies were included in the database: cerebral palsy (CP), spina bifida, congenital muscular dystrophies, stroke, degenerative neurological disorders, neuromuscular diseases, amputations, idiopathic toe walking, arthrogryposis multiplex congenita, spinal cord injuries, traumatic brain injury, as well as other less common gait-related conditions. These categories reflect the real-world distribution of diagnoses treated across the ORITEL network and ensure that the database captures a representative spectrum of clinical presentations.

As of the time of this study, the database contains 156 gait studies from individuals with a wide range of clinical diagnoses, including, but not limited to cerebral palsy, neuromuscular disorders, orthopedic impairments, and typically developing controls. The database is designed to be continuously updated, enabling future analyses, such as evaluating treatment outcomes across multiple rehabilitation centers.

The data was contributed by eight motion analysis laboratories from the following Teletón centers:

- **Mexico:** Saltillo, Guadalajara, Tlalnepantla.
- **El Salvador:** La Libertad.
- **Colombia:** Bogotá.
- **Uruguay:** Montevideo.
- **Chile:** Santiago, Concepción.

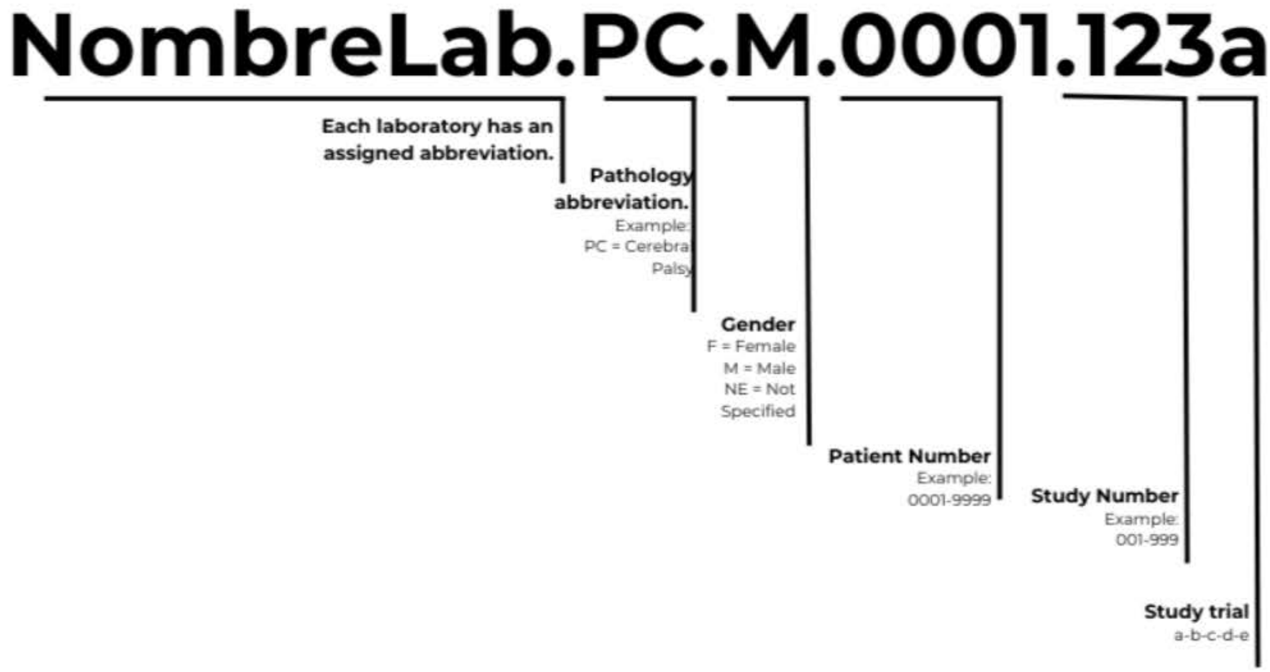

Fig. 4.1: Diagram of the overall data organization in the *Movement Analysis Network* database, illustrating the hierarchical structure of study codes and associated files.

Each study includes between two and five gait trials. However, for the purposes of this research, only studies with at least four trials and a confirmed diagnosis of CP were included in the final analysis.

4.2 Data Acquisition and Anonymization

All data were anonymized prior to submission to the central repository. Each laboratory was responsible for removing any personally identifiable information. The final dataset was stored in JSON format and uploaded to a MongoDB NoSQL database, enabling scalable, flexible access and facilitating

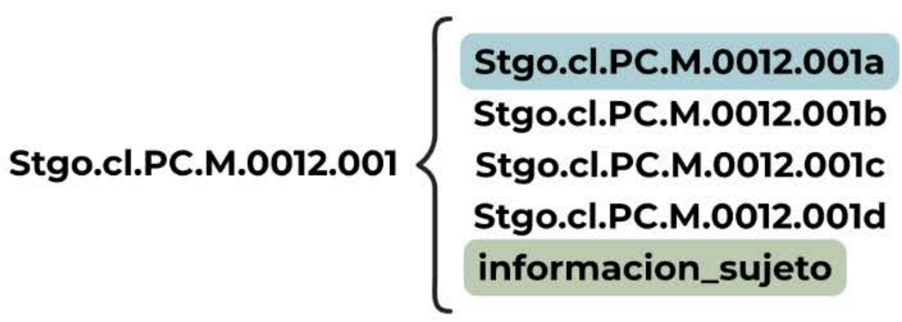

Fig. 4.2: Example of the hierarchical structure of a single subject study within the *Movement Analysis Network* database.

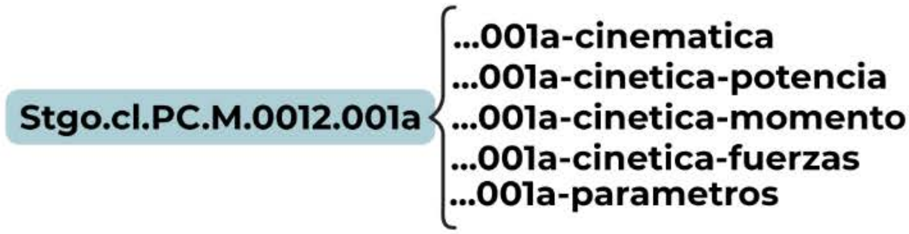

Fig. 4.3: Example of the hierarchical structure of a trial within a study.

Fig. 4.4: Example of the subject information associated with a study.

downstream integration with machine learning pipelines.

Data were acquired using either BTS or Vicon motion capture systems, depending on the center. Due to differences in acquisition protocols and file formats, a standardization process was necessary. Gait data included two representative gait cycles per trial one for each lower limb, containing kinematic (joint angles), optional kinetic (joint moments, power, and ground reaction forces), and spatiotemporal parameters.

Spatiotemporal parameters were extracted from BTS gait reports using a custom Python script provided to each center. This approach ensured that parameter extraction remained consistent and secure without compromising data anonymity. The extracted parameters were compiled in structured XLSX files, which were later linked to each corresponding trial entry.

4.3 Cycle Selection Criteria

The selection of the representative gait cycle was performed manually by the clinical expert at each laboratory. In BTS systems, only the selected cycle is typically exported, while Vicon exports the entire recording of each trial. Nevertheless, within the C3D files, Vicon also stores the samples that indicated the start and end of the clinician selected gait cycle. During preprocessing, only the designated segment was extracted for inclusion in the database, ensuring consistency across all centers.

It is important to consider that in people's gait, there are differences in multiple variables such

Table 4.1: Detailed list of kinematic and spatiotemporal variables included in the dataset, with their anatomical planes, measurement units, and file formats.

Contents	Variables	Plane	Measurement unit	Type of file
Kinematic	Ankle angle	Sagittal	degrees	EMT/C3D
	Ankle angle	Transverse	degrees	EMT/C3D
	Knee angle	Sagittal	degrees	EMT/C3D
	Hip angle	Sagittal	degrees	EMT/C3D
	Hip angle	Frontal	degrees	EMT/C3D
	Hip angle	Transverse	degrees	EMT/C3D
	Pelvis angle	Sagittal	degrees	EMT/C3D
	Pelvis angle	Frontal	degrees	EMT/C3D
	Pelvis angle	Transverse	degrees	EMT/C3D
Spatiotemporal parameters	Stance phase	Right/Left	%	XLSX/C3D
	Swing phase	Right/Left	%	XLSX/C3D
	Double support	Right/Left	%	XLSX/C3D
	Stance duration	Right/Left	s	XLSX/C3D
	Swing duration	Right/Left	s	XLSX/C3D
	Stride duration	Right/Left	s	XLSX/C3D
	Step length	Right/Left	m	XLSX/C3D
	Stride length	Right/Left	m	XLSX/C3D
	Cadence	Overall values	steps/min	XLSX/C3D
	Step width	Overall values	m	XLSX/C3D
	Average speed	Overall values	m/s	XLSX/C3D

as walking speed, step width, stride length, among others. Therefore, even if all centers use a sampling rate of 100 Hz, which is the case, the duration of a gait cycle will vary between individuals. Some cycles will be shorter and contain fewer samples, while others will be longer and contain more samples.

4.4 Data Standardization and Organization

To unify the dataset structure across all centers and motion capture systems, a standardization process was conducted. Two MATLAB scripts were developed, one for C3D files (used by Vicon systems)

Table 4.2: Detailed list of kinetic variables included in the dataset, with their anatomical planes, measurement units, and file formats.

Contents	Variables	Plane	Measurement unit	Type of file
Kinetic	Ankle moment	Sagittal	Nm	EMT/C3D
	Knee moment	Sagittal	Nm	EMT/C3D
	Knee moment	Frontal	Nm	EMT/C3D
	Hip moment	Sagittal	Nm	EMT/C3D
	Hip moment	Frontal	Nm	EMT/C3D
Kinetic	Ankle power	Sagittal	W	EMT/C3D
	Knee power	Sagittal	W	EMT/C3D
	Knee power	Frontal	W	EMT/C3D
	Hip power	Sagittal	W	EMT/C3D
	Hip power	Frontal	W	EMT/C3D
Kinetic	Force Ground Reaction Vertical	Sagittal	N	EMT/C3D
	Force Ground Reaction Antero/Posterior	Sagittal	N	EMT/C3D
	Force Ground Reaction Medio/Lateral	Sagittal	N	EMT/C3D

and another for EMT files (used by BTS systems). As part of this process, a standardized naming scheme was implemented to label all study files, as illustrated in Figure 4.1. Each filename encodes the center name, pathology, subject gender, subject identifier, study ID, and trial number. This naming convention facilitates efficient file management and ensures that every entry is uniquely identifiable across the entire database. These scripts also reorganized and cleaned the data to conform to a unified structure, as shown in Figures 4.2, 4.3, and 4.4.

Each study entry is required to include kinematic data, spatiotemporal parameters, and a folder containing the subject's metadata. Kinetic variables are optional and depend on the availability of data from the respective center. In the Table 4.1 and Table 4.2 provides a detailed overview of the dataset contents, including the specific variables, anatomical planes, units of measurement, and corresponding file formats.

Kinematic data consist of joint angles measured across various anatomical planes. When available, kinetic data include joint moments, joint powers and ground reaction forces. Additionally, the spatiotemporal parameters encompass metrics such as stance phase, swing phase, double support, stance

Table 4.3: Distribution of the *Movement Analysis Network* database, including the number of female and male participants and the breakdown by number of trials per study.

Teletón center	Number of studies	Female studies (n)	Male studies (n)	Studies 5 trials	Studies 4 trials	Studies 3 trials	Studies 2 trials
Santiago - Chile	43	17	26	0	43	0	0
Concepción - Chile	13	7	6	0	1	2	10
Bogota - Colombia	29	16	13	0	29	0	0
Guadalajara - Mexico	15	3	12	0	15	0	0
Saltillo - Mexico	15	6	9	0	15	0	0
Tlalnepantla - Mexico	18	7	11	0	2	0	16
La Libertad - El Salvador	8	3	5	0	8	0	0
Montevideo - Uruguay	15	7	8	12	3	0	0
Total	156	66	90	12	116	2	26

duration, swing duration, stride duration, step length, stride length, cadence, step width and mean walking speed

4.5 Database Distribution

The following table shows the distribution of the database, including the number of studies per center, the number of studies involving female and male participants, as well as the number of studies with 5, 4, 3, and 2 trials, Table 4.3 summarizes this information, showing the number of studies conducted at each Teletón center.

Since the dataset includes both complete studies, containing kinematic (KIN), kinetic (KKT), and spatiotemporal parameters (STP), and partial studies, which include only KIN and STP, Table 4.4 presents the distribution of studies by pathology, along with the number of complete and partial records for each. In the Figure 4.5 shows the distribution of studies by pathology, highlighting the prevalence of cerebral palsy (CP) as the most common condition in the database, followed by other pathologies such as spina bifida and congenital muscular dystrophies.

The age distribution of the data is presented in Table 4.5 and Figure 4.6. The subjects in the database range in age from 7 to 64 years, with a mean age of 15 and median age of 11. The majority of participants are under 18 years old, indicating a strong focus on children and adolescents. Figure 4.6 provides a breakdown of age distribution by center, allowing for a clearer view of how the age ranges

Table 4.4: Distribution of studies in the *Movement Analysis Network* database by pathology type and data completeness. Full data correspond to studies including kinematic (KIN), kinetic (KKT), and spatiotemporal (STP) parameters, whereas partial data include only kinematic and spatiotemporal parameters.

Pathologies	Pathology code	Full Data (KIN + KKT + STP)	Partial Data (KIN + STP only)
Cerebral Palsy	PC	70	33
Spina Bifida	EB	1	0
Congenital Muscular Dystrophies	DMC	3	0
Cerebrovascular Disease	EVC	1	0
Degenerative Disorders	TD	3	0
Neuromuscular Diseases	ENM	3	0
Amputees	AMP	1	0
Idiopathic Toe Walking	MIPP	1	0
Congenital Arthrogryposis	AMC	2	0
Acquired Spinal Cord Injuries	LMA	4	0
Traumatic Brain Injury	TEC	1	0
Other Pathologies	OTROS	32	1
	Total	122	34

vary across different locations.

4.6 NoSQL Database

The database is stored using MongoDB, a NoSQL system that offers flexible and scalable data management. This choice was made because the dataset is diverse and continues to grow some studies include kinematic and kinetic data, while others do not. A rigid structure, like the one required by

Table 4.5: Age distribution of subjects included in the *Movement Analysis Network* database.

Age range	Age in years
Minimum age	7
Maximum age	64
Mean age	15
Median age	11

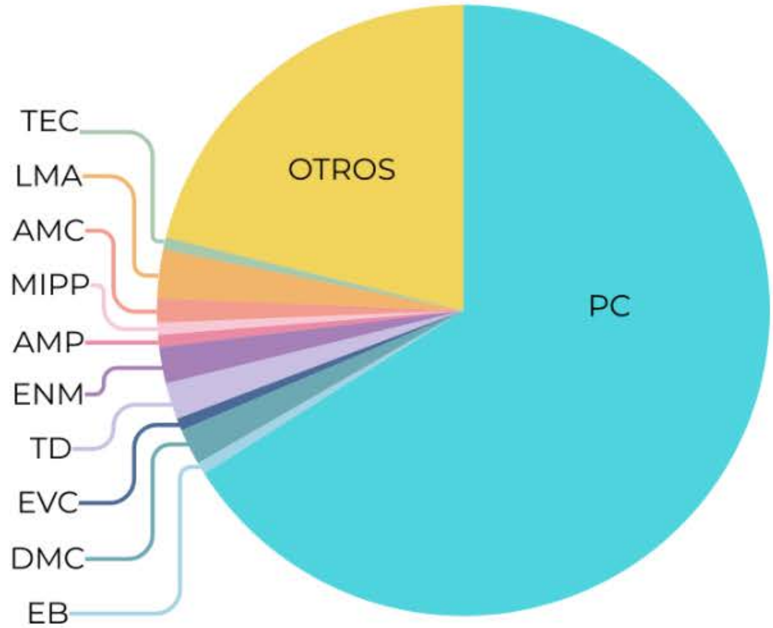

Fig. 4.5: Distribution of studies in the *Movement Analysis Network* database according to pathology type.

traditional relational databases, would make it difficult to accommodate these variations. In contrast, NoSQL flexible schema allows each study to be stored as an individual document with only the relevant fields. This makes it easier to handle different types of data and adapt to future changes. Each document contains key information such as spatiotemporal parameters, subject metadata, and kinematic or kinetic data when available.

4.7 Conclusion

This chapter described the creation and structure of the *Movement Analysis Network* database, developed as part of a collaborative effort between multiple Teletón centers across the ORITEL network. The goal of this initiative was to gather and standardize clinical gait data from different countries, ensuring a diverse and representative sample of pathologies, especially those most commonly treated in rehabilitation centers, such as CP.

Thanks to the collaboration between centers and the use of unified processing scripts, it was possible to standardize data collected using different motion capture systems (BTS and Vicon) and formats. This makes the database consistent and ready for large scale analyses or future integration with machine learning tools.

Given the variability in data content some studies include full kinematic, kinetic, and spatiotem-

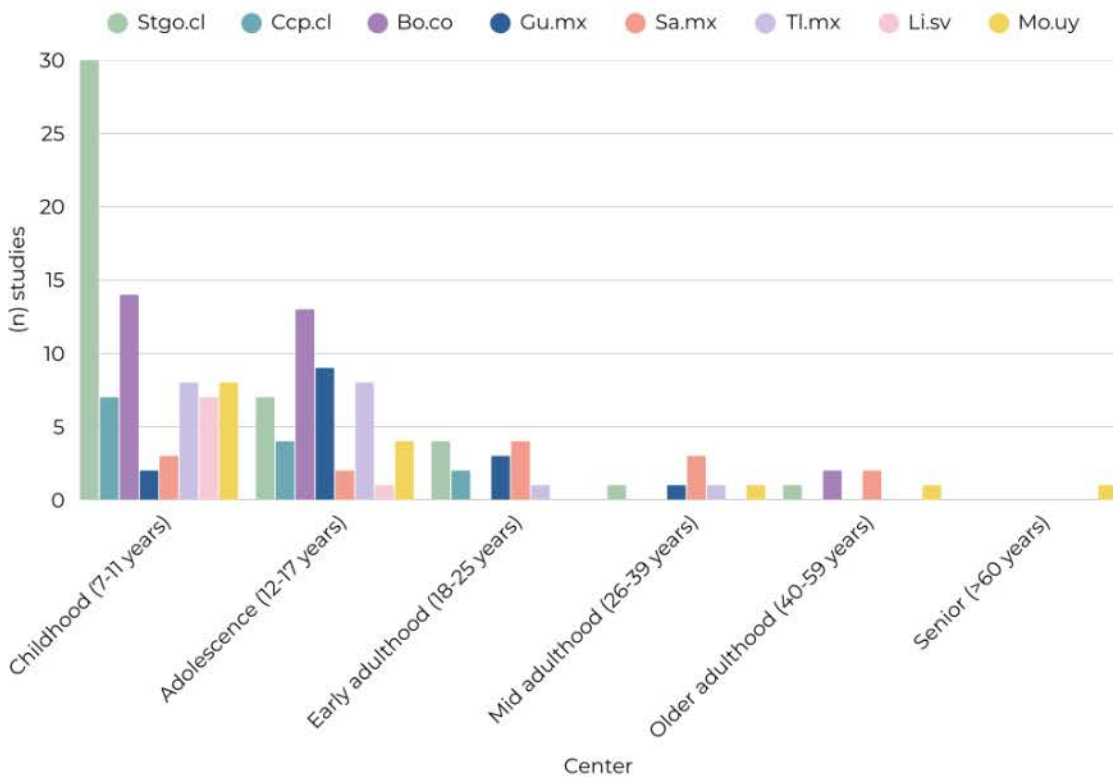

Fig. 4.6: Age distribution of subjects in the *Movement Analysis Network* database, grouped by participating center.

poral data, while others contain only a subset, a flexible storage solution was essential. MongoDB, a NoSQL database, was chosen for this reason. Its document based structure allows each study to store only the available information without needing a fixed schema, making it easy to expand and adapt the database over time.

Currently, the database includes 156 studies, most of which involve pediatric subjects, reflecting the population typically served by Teletón centers. The database includes details such as gender, age, number of trials, and type of pathology, providing a rich foundation for future research.

In summary, the *Movement Analysis Network* database offers a solid and scalable resource for analyzing gait patterns across different clinical conditions and contexts. Moreover, it is designed to grow over time as new studies are added by participating centers, ensuring that the dataset remains relevant, up to date, and increasingly valuable for research and clinical applications.

Chapter 5: Results

This chapter presented the results of gait-pattern classification using ML methods. First, model performance was reported; next, feature importance was analyzed and statistical comparisons between classes were performed; finally, the association between treatments and gait patterns was discussed.

5.1 Machine Learning Model Performance

This section presents the results of ML classifiers evaluated in the methodology. The models were assessed using four key performance metrics: weighted F1-score, accuracy, precision, and recall. Together, these metrics offer a comprehensive view of each model's ability to correctly classify gait patterns across the three classes: diparesis, hemiparesis, and no pathology.

Table 5.1 shows the comparative results of the eight models evaluated using Smote to handle class imbalance. The MLP achieved the highest overall performance, with a weighted F1-score of 0.8600 and an accuracy of 85.85%, followed closely by the SGD classifier and the Random Forest model. The remaining classifiers, including SVM, Gradient Boosting, Logistic Regression, XGBoost, and Decision Tree, exhibited slightly lower performance across all metrics.

Table 5.2 presents the comparative results of the eight models evaluated using sample weights to address class imbalance. In this approach, no synthetic data were generated; instead, the learning algorithm was adjusted to give higher importance to underrepresented classes. The MLP achieved the highest overall performance, with a weighted F1-score of 0.8899 and an accuracy of 88.35%. Notably, a performance gap begins to emerge at this point, as the next best model, the Random Forest classifier reached a lower F1-score of 0.8369. This suggests that the MLP not only outperforms the other models but does so by a more considerable margin compared to previous configurations.

The MLP model, which achieved the best performance in both configurations, performed even better when using sample weights for class balancing. Therefore, this configuration was selected for further analysis. In the Figure 5.1, the confusion matrix obtained through 10-fold cross-validation on the training set is presented.

The performance of the MLP classifier model using sample weights is summarized in Figure 5.2 and Table 5.3.

Table 5.1: Performance comparison of ML models trained with SMOTE balancing, evaluated using F1-score, accuracy, precision, and recall.

Model	F1-score	Accuracy	Precision	Recall
MLP	0.8600	0.8585	0.7996	0.7929
SGD	0.8387	0.8415	0.7692	0.7551
Random Forest	0.8352	0.8288	0.8093	0.7840
SVM	0.8253	0.8306	0.8024	0.7499
Gradient Boosting	0.8163	0.8219	0.7629	0.7432
Logistic Regression	0.8133	0.8071	0.7368	0.7264
XGBoost	0.8066	0.8046	0.7966	0.7402
Decision Tree	0.8011	0.8025	0.7269	0.7203

Table 5.2: Performance comparison of ML models trained using Sample Weight balancing, evaluated with F1-score, accuracy, precision, and recall.

Model	F1-score	Accuracy	Precision	Recall
MLP	0.8899	0.8835	0.8574	0.8521
Random Forest	0.8369	0.8303	0.7837	0.8030
Decision Tree	0.8220	0.8156	0.7782	0.7704
Logistic Regression	0.8108	0.8041	0.7742	0.7686
Gradient Boosting	0.8103	0.8093	0.7482	0.7413
SVM	0.8199	0.8452	0.7084	0.7674
XGBoost	0.8077	0.8013	0.7388	0.7555
SGD	0.8066	0.7978	0.7651	0.7618

Table 5.3: Classification performance of the best model (MLP with Sample Weights) on the test dataset.

	Precision	Recall	F1-score	Support
Diparesis	0.80	0.82	0.81	34
Hemiparesis	0.67	0.60	0.63	20
No Pathology	0.96	1.00	0.98	25
Accuracy			0.82	79

Where the Figure 5.2 shows the confusion matrix corresponding to the evaluation on the test set. Table 5.3 reports the detailed performance metrics on the test set, with an overall accuracy of 82%. Diparesis was recognized with an F1-score of 0.81, Hemiparesis with 0.63, and *No pathology* with the

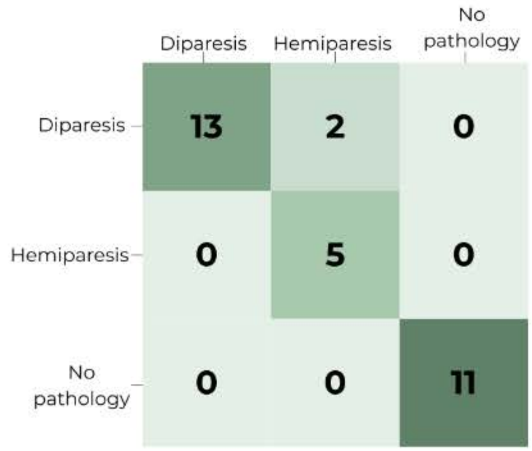

Fig. 5.1: Confusion matrix for the best performing MLP (sample weights), obtained via 10-fold cross-validation on the training set.

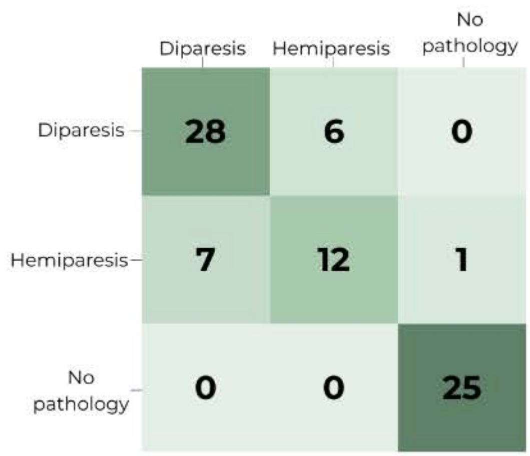

Fig. 5.2: Confusion matrix for the best performing MLP (sample weights) on the held-out test set.

highest performance (F1-score = 0.98).

Table 5.4 presents the optimal hyperparameters identified for the MLP model (Sample Weights) through the Optuna optimization process. As described in the methodology, this optimization involved 200 trials combined with 10-fold cross-validation to ensure a robust and generalizable selection of hyperparameters. For the top three most influential hyperparameters, the learning rate was explored within the range $[10^{-5}, 10^{-2}]$, the number of Principal Component Analysis (PCA) between [1, 16], and the dropout rate between [0.1, 0.5]. These parameters were iteratively adjusted across trials to determine the best combination that maximized the model's performance.

Figure 5.3 shows the relative importance of each hyperparameter for the MLP model (Sample Weights). In this context, importance is derived from the average change in the objective function across all trials when a hyperparameter is perturbed. This visualization provides insight into which

Table 5.4: Optimal hyperparameter configuration of the best-performing MLP model with sample weights.

Hyperparameter	Value
Number of layers	3
Number of units l0	128
Number of units l1	240
Number of units l2	176
Dropout	0.1715
Learning rate	$1.651{\cdot}10^{-4}$
Batch size	128
Activation	gelu
Optimizer	adamw
Weight decay	$1.8926{\cdot}10^{-6}$
Scaler	Minmax
Number components PCA	9

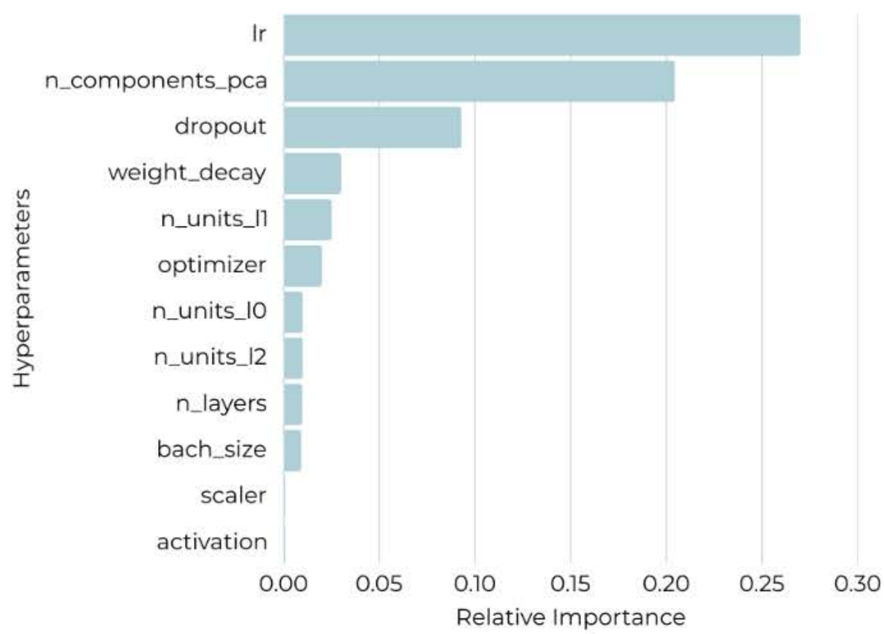

Fig. 5.3: Relative importance of the hyperparameters for the MLP model with sample weights.

hyperparameters have the greatest impact on the model's performance.

For further evaluation, the MLP model (using sample weights) was tested on an external dataset consisting of 63 trials: 45 diparesis cases and 18 hemiparesis cases. These studies included subjects with 2, 3, and 4 trials each, and none of them had been previously. The results of this evaluation are summarized in the confusion matrix shown in Figure 5.4 and a performance summary in Table 5.5.

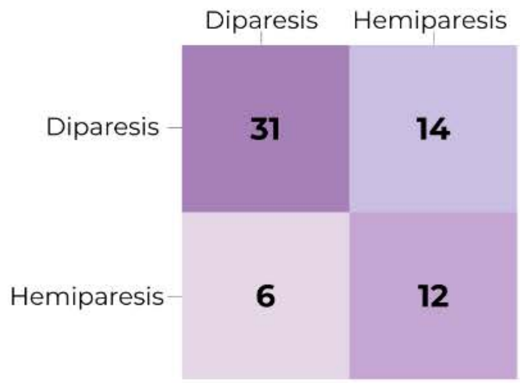

Fig. 5.4: Confusion matrix of the MLP (Sample Weights) model evaluated on an external test set.

Table 5.5: Performance of the MLP (Sample Weights) evaluated on external data

	Precision	Recall	F1-score	Support
Diparesis	0.84	0.69	0.76	45
Hemiparesis	0.46	0.67	0.55	18
Accuracy			0.68	63

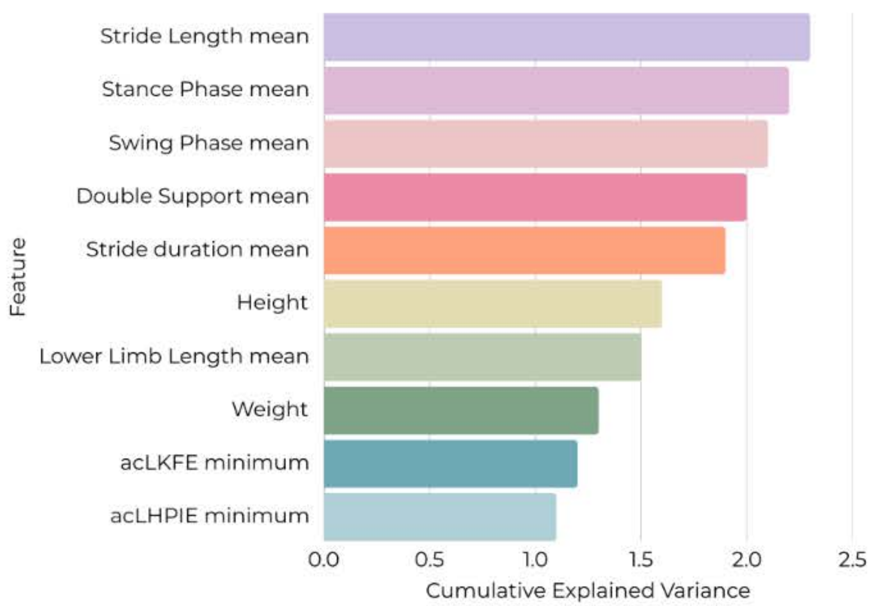

Fig. 5.5: Top 10 features with the highest importance in the classification task, as determined by the trained MLP model.

5.2 Important Features for Classification

The Figure 5.5 illustrates the features with the highest importance scores. The analysis of the most significant features contributing to the classification of gait patterns in patients with CP.

Table 5.6: Median Values and Statistical Comparison of top 10 Features in Hemiparesis and Diparesis.

Feature	Median Hemiparesis	Median Diparesis	p_{raw}	Cliff's delta	$p_{\text{bonferroni}}$
Stride Lenght mean	0.8954	0.8590	$2.3422 \cdot 10^{-2}$	0.1623	$2.3422 \cdot 10^{-1}$
Stance Phase mean	60.9550	59.3000	$1.2740 \cdot 10^{-3}$	0.2301	$1.2740 \cdot 10^{-2}$
Swing Phase mean	39.0450	40.7000	$7.8384 \cdot 10^{-4}$	-0.2399	$7.8384 \cdot 10^{-3}$
Double Support mean	10.1500	15.0000	$3.1729 \cdot 10^{-13}$	-0.5204	$3.1729 \cdot 10^{-12}$
Stride Duration mean	1.0580	1.0300	$2.0167 \cdot 10^{-1}$	0.0912	1.0000
Height	142.0000	133.0000	$1.4404 \cdot 10^{-3}$	0.2256	$1.4404 \cdot 10^{-2}$
Lower Limb Lenght mean	74.0000	71.5000	$3.1709 \cdot 10^{-4}$	0.2550	$3.1709 \cdot 10^{-3}$
Weight	39.0000	33.6000	$3.1372 \cdot 10^{-1}$	0.0714	1.0000
acLKFE minimum	4.4000	17.4000	$9.2206 \cdot 10^{-10}$	-0.4336	$9.2205 \cdot 10^{-9}$
acLHPIE minimum	-8.8500	-3.5000	$1.5228 \cdot 10^{-2}$	-0.1719	$1.5228 \cdot 10^{-1}$

Table 5.6 summarizes the statistical comparison of the ten most relevant features between the hemiparesis and diparesis classes. For each feature, the table reports the median value for each class, the raw p-value from the Mann–Whitney U test (p_{raw}), the effect size estimated by Cliff's delta, and the p-value was adjusted for multiple comparisons using the Bonferroni correction ($p_{\text{bonferroni}}$).

5.3 Classifier association with treatment selection

To associate the classifier's predictions with treatment selection, the suggested treatments provided by the ORITEL centers were analyzed and compared. These treatments were filled in by the professionals responsible for data export at each center. The goal was to identify the most commonly recommended treatments for the hemiparesis and diparesis classes.
In addition, a literature review was conducted, along with meetings with healthcare professionals and specialists in the field, to better understand the effectiveness of each treatment.

Suggested treatments for hemiparesis:

- Botulinum toxin
- Orthoses (Ankle–foot orthosis)
- Physical therapy

- Orthopedic surgery
- Occupational therapy
- Exercise and Strengthening Program

Suggested treatments for diparesis:

- Botulinum toxin
- Orthoses
- Physical therapy
- Orthopedic surgery
- Selective dorsal rhizotomy
- Exercise and Strengthening Program

Table 5.7: Distribution of hemiparesis and diparesis CP cases and number of cases with documented treatments

Pathology	Number of studies	Number of studies with treatments
Hemiparesis CP	34	28
Diparesis CP	42	24
Total	76	52

Table 5.8: Frequency of recommended treatments for hemiparesis

Recommended treatment	Frequency in Database
Orthoses	13
Orthopedic Surgery	11
Physical Therapy	8
Occupational Therapy	7
Botulinum toxin	3
Exercise and Strengthening Program	3

In Table 5.8, the frequency of recommended treatments for hemiparesis is presented, while Table 5.9 shows the corresponding data for diparesis. These tables summarize the number of cases in

Table 5.9: Frequency of recommended treatments for diparesis

Recommended treatment	Frequency in Database
Botulinum toxin	9
Orthopedic surgery	5
Physical Therapy	5
Exercise and Strengthening Program	4
Selective dorsal rhizotomy	3
Orthoses	1

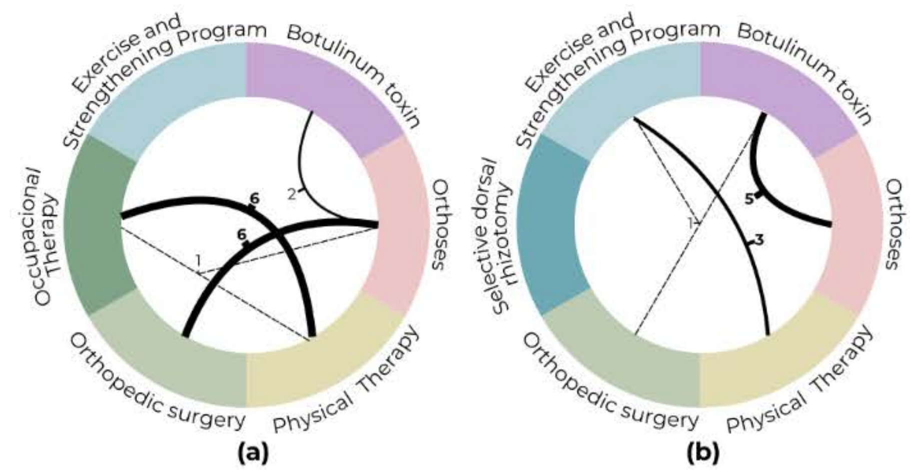

Fig. 5.6: Chord diagrams illustrating the frequency of treatment co-occurrence across the included studies. Thicker lines represent more frequently reported combinations. Solid lines indicate associations between two treatments, while dashed lines indicate associations among three treatments. Panel **(a)** corresponds to studies on hemiparesis, and panel **(b)** to studies on diparesis.

the database in which each treatment was suggested by specialists. It is important to note that the total number of cases with documented treatments is lower than the overall number of cases in the database (Table 5.7), as treatment information was not available for all cases.

Chapter 6: Discussion and Conclusion

6.1 Discussion

6.1.1 Creation of the Gait Database

The creation of the database was a long and demanding process that involved collecting gait data from several laboratories, preparing materials, organizing meetings, and providing training sessions. This required strong coordination between centers and added extra work for the clinical professionals involved, showing the dedication and teamwork of everyone who participated. Thanks to these efforts, the data-collection campaign was successful, resulting in a repository of **156** studies from **8** laboratories, including subjects with CP and other gait-related conditions. The studies vary in length, with some including 5 trials, others 4, 3, or 2 trials per study. This diversity adds value to the database, as it reflects the real-world differences in data collection across centers.

After gathering the data, a standardization process was carried out to ensure all datasets followed the same format, regardless of the gait capture system used. The final step was uploading the standardized data to a NoSQL database. While the database is already a valuable resource, there are still areas for improvement. Not all records include treatment information, and the process for adding new data could be made easier. Improving the documentation would also help other researchers use the database more effectively.

Despite these challenges, the database now provides a strong foundation for reproducible analyses, multi-center comparisons, and the development of clinically meaningful models. It also sets the stage for future growth, with opportunities for richer data annotations, easier data submissions, and broader collaboration. Overall, this database represents a significant step forward for ORITEL and for gait research in the region, and it will continue to improve as more data and features areted a substantive advance for ORITEL and for gait research in the region.

6.1.2 Performance and Generalization of the MLP Classifier

For the development of the ML model, a comparative analysis was conducted between 8 different models, testing two different approaches to balance the dataset, smote and sample weights. The MLP with sample weights achieved the best performance, with an accuracy of 82% and class-wise F1-scores of 0.81 (diparesis), 0.63 (hemiparesis), and 0.98 (no pathology), as shown in Figure 5.3. This performance surpassed that of the other models in both accuracy and F1 score, demonstrating the MLP's effectiveness in handling imbalanced data. Based on the literature review, these results are within the expected range. While no directly comparable study was found, similar outcomes have been reported in works classifying the severity of CP and in subcategories of gait patterns such as hemiparesis and diparesis. It is also important to highlight that this model was trained nd tested on a dataset of 413 trials collected from multiple gait laboratories, selected according to the inclusion and exclusion criteria described in Section 3.1.3 of the methodology. According to the state of the art, this represents a relatively large dataset, which may have contributed to the model's strong performance.

The use of sample weights allowed the MLP model to assign greater importance to the minority class, which is essential when working with imbalanced datasets. This strategy helped the model to better learn the features of the minority class, resulting in improved performance metrics. Class weighting rebalanced the loss while preserving class priors and avoiding synthetic samples that may distort boundaries, particularly with label-encoded categorical features. It also respected the *StratifiedGroupKFold* grouping and strengthened the minority signal without inflating the data, thereby reducing overfitting and aligning train and held-out distributions.

The optimal configuration was achieved with a three-layer architecture: 128 neurons in the first layer, 240 in the second, and 176 in the third, using the Gelu activation function and a learning rate of $1.651 \cdot 10^{-4} (Table\, 5.4).Theanalysisofhyperparameterimportancerevealedthatthemostinfluentialfactorsu$

As shown in Figure 5.2, the confusion matrix showed that the model had difficulty telling apart the different gait patterns, especially between diparesis and hemiparesis. To better understand how well the model could generalize, it was tested on an external dataset of 63 trials that had not been previously used. Most of these trials were from diparesis cases, and the dataset included subjects with 2, 3, 4, or 5 trials per study. This helped to check how robust the model was when the number of trials per subject varied.

As shown in Table 5.3, the F1 score for diparesis during training was 0.81 and for hemiparesis was 0.63. On the external evaluation, presented in the Table 5.5, the F1 score for diparesis dropped to 0.76 and for hemiparesis to 0.55. This decrease was expected, since the external dataset had a different composition and fewer hemiparesis samples. The results suggest that the model was better at classifying diparesis, likely because there was more diparesis data available in both training and testing. Increasing the number of hemiparesis samples could help improve the model's performance for that class.

Overall, these results demonstrate that the proposed MLP model, trained on a relatively large and heterogeneous dataset, is capable of achieving strong performance in the classification of gait patterns in CP, particularly in diparesis cases. However, the imbalance between classes remains a limiting factor, and future work should focus on expanding the dataset, especially for hemiparesis, as well as exploring advanced balancing strategies or domain adaptation techniques to further enhance the model's generalization capacity.

6.1.3 Feature Importance and Statistical Analysis

The feature importance analysis indicated that the ten most relevant features for the MLP classifier, were stride length mean, stance phase mean, swing phase mean, double support mean, stride duration mean, height, lower limb length mean, weight, acLKFE minimum, and acLHPIE minimum (Figure 5.5). These variables align with previous literature, which consistently emphasizes the role of spatiotemporal and kinematic parameters as central descriptors of gait abnormalities in CP.

The Mann–Whitney U test was applied to compute the raw p-values (P_{raw}), which represent the probability of observing the differences between distributions by chance. A $P_{raw} < 0.05$ indicated that the observed differences were unlikely to be due to random variation, and therefore the distributions could be considered significantly different. As shown in Table 5.6, most variables presented statistically significant differences between hemiparesis and diparesis. In particular, double support and acLKFE minimum reached very low P_{raw} values, reinforcing their role as key discriminators between the two groups. These findings suggest that the classifier may have relied heavily on these features to capture clinically meaningful gait differences.

In particular, double support exhibited the largest effect size (Cliff's $\delta = -0.5204$), strongly associated with diparesis. This finding is consistent with clinical descriptions, as children with diparesis often present with greater instability and require longer periods of double support compared to those with hemiparesis. Similarly, swing and stance phases showed significant differences, with diparesis displaying a longer swing phase and a shorter stance phase. These temporal adaptations may reflect compensatory strategies related to bilateral motor impairments, whereas hemiparesis is typically characterized by asymmetric loading and reduced stance on the affected limb.

The kinematic features acLKFE minimum and acLHPIE minimum also contributed to the separation between groups. The acLKFE minimum, in particular, presented a medium-to-large negative effect size ($\delta = -0.4336$), suggesting that knee flexion minima were more pronounced in diparesis. This result supports the literature describing excessive knee flexion during gait in diparetic patients, often linked to crouch gait patterns. In contrast, acLHPIE minimum differences were weaker and did not remain significant after Bonferroni correction, indicating that hip internal-external rotation contributed less consistently to group discrimination.

Anthropometric variables such as height and lower limb length also differed significantly, with children in the diparesis group showing smaller median values. These results are in line with previous studies reporting growth delays and reduced limb length in more severely affected CP subtypes. However, variables such as stride duration and weight did not reach statistical significance, suggesting that their contribution to classification.

After applying the Bonferroni correction, the most robust differences between hemiparesis and diparesis were observed for double support, acLKFE minimum, lower limb length, swing phase, stance phase, and height ($p_{\text{bonf}} < 0.05$). This reinforces the interpretation that a combination of spatiotemporal features and lower-limb kinematics are key discriminators between these two CP subtypes.

Overall, the statistical analysis not only confirmed the relevance of the features identified, but also highlighted clinically meaningful distinctions between hemiparesis and diparesis. The results suggest that the classifier relied on physiologically interpretable parameters, particularly double support, swing phase, and stance phase, when distinguishing gait patterns. These findings strengthen the validity of the model and point to the potential of combining spatiotemporal and kinematic descriptors for automated gait classification in CP.

6.1.4 Treatment association with gait patterns classifier

When analyzing the association of treatments with hemiparesis and diparesis in CP, several limitations of the dataset needed to be considered. First, not all studies included information about treatments, which restricted the scope of the analysis. In addition, because the treatment assigned to each patient remained the same across all of their trials, the variability of treatment-related data was inherently limited. Another important factor was that rehabilitation centers often reported treatment recommendations rather than verified evidence of completed interventions. As a result, it was not possible to evaluate the actual effectiveness of treatments unless multiple studies of the same patient, together with their full medical history over time, were available.

Despite these constraints, an exploratory frequency analysis was carried out to identify how often specific treatments were mentioned for hemiparesis and diparesis. This approach provided an initial overview of potential treatment patterns, which could be expanded and validated with larger datasets in future work. To complement these findings, a review of the literature was performed in order to compare the treatments most frequently reported in the database with the evidence of their effectiveness in the management of gait impairments. While this analysis did not allow for definitive conclusions, it provided a starting point for future studies aimed at linking treatment types to clinical outcomes in CP.

For hemiparesis, the treatments proposed included botulinum toxin, orthoses (especially ankle-foot orthoses), physical therapy, orthopedic surgery, occupational therapy, and exercise or strengthening programs. In contrast, for diparesis, the main treatments recommended were botulinum toxin, orthoses (mainly ankle-foot orthoses), physical therapy, orthopedic surgery, selective dorsal rhizotomy, and exercise or strengthening programs.

According to the frequency analysis presented in the Table 5.8, the three most common treatments for hemiparesis were orthoses, orthopedic surgery, and physical therapy. For diparesis, the three most common were botulinum toxin, orthopedic surgery, and physical therapy, as shown in the Table 5.9. These results suggested that while there was some overlap between the two conditions, there were also differences in treatment priorities, particularly with the more frequent use of botulinum toxin in diparesis and occupational therapy in hemiparesis.

Furthermore, the analysis of treatment combinations provided additional insight into how interventions were used together. Chord diagrams (Figure 5.6) illustrated these associations, where thicker lines indicated combinations that appeared more frequently in the studies. For hemiparesis, occupational therapy was often recommended alongside physical therapy, while orthopedic surgery was commonly paired with orthoses. For diparesis, the most frequent combination was botulinum toxin with the use of orthoses.

Overall, this exploratory analysis highlighted preliminary patterns in the association between treatments and CP subtypes. Although limited by the available data, the findings suggested that certain treatments were consistently emphasized for hemiparesis and diparesis, both individually and in combination. These results underscored the importance of expanding datasets and conducting longitudinal studies to better understand how treatment strategies influence clinical outcomes in CP.

6.2 Conclusions

The objectives defined at the beginning of this work were successfully achieved.

First, the creation and standardization of a multi-center gait database were accomplished. Data from **156** studies collected across **8** laboratories were harmonized and stored in a unified structure, ensuring comparability and reproducibility. This process established a valuable resource for future clinical and research applications, meeting the goal of data standardization across centers.

Second, a gait pattern classifier based on machine learning was developed. Among the models evaluated, the MLP with class weighting achieved the best performance, reaching an overall accuracy of **82%** and strong F1-scores across classes. The classifier generalized well to external data, particularly for diparesis, demonstrating its robustness and fulfilling the goal of building an effective model for CP gait classification. In doing so, the study confirmed the initial hypothesis, showing that the classification of gait patterns based on motion analysis laboratory assessments can support treatment selection in at least **80%** of cases, thus enabling more personalized interventions.

Third, clinical variables relevant to classification were analyzed. Through PCA-based feature importance and statistical testing (Mann–Whitney U, Cliff's delta, and Bonferroni correction), spatiotemporal and kinematic parameters such as double support, swing phase, stance phase, and acLKFE were identified as key discriminators between hemiparesis and diparesis. These results confirmed the role of clinically interpretable features in the classifier's decision process, meeting the objective of linking classification to meaningful variables.

Finally, an exploratory analysis of treatment associations was performed. Although the dataset presented limitations (incomplete treatment annotations and lack of longitudinal tracking), frequency analysis revealed preliminary treatment patterns for hemiparesis and diparesis, supported by the literature. Orthoses, orthopedic surgery, and physical therapy were most frequent in hemiparesis, while botulinum toxin, orthopedic surgery, and physical therapy predominated in diparesis. This partially fulfilled the goal of linking gait classification with treatments and provided a foundation for future studies.

In summary, the study achieved its main goal of developing a ML-based gait classifier for CP using multi-center data. The integration of a standardized database, robust classification models, statistical validation of key features, and exploratory treatment associations represent a significant contribution to gait analysis research.

6.3 Future Perspectives

The *Movement Analysis Network* database established in this work represents a first step toward a collaborative and standardized resource for gait research. However, there are several directions in which this work can be extended to further increase its clinical and scientific impact.

First, it is expected that the database will continue to grow over time as new studies are incorporated. The addition of new cases will not only improve the representativeness of the data but also allow longitudinal analyses of gait patterns, enabling the study of kinematic changes across time. To maximize its usability, one pending task is to complete and refine the current records. In particular, adding structured information about treatment recommendations provided by rehabilitation centers, as well as recording the history of treatments previously applied to each subject, would enable future studies on the effectiveness of interventions for CP and other gait-related conditions.

Another important direction is to make the database fully accessible for professionals within ORITEL, fostering multi-center collaborations and supporting diverse clinical and research projects. Providing clear documentation and user-friendly access tools will be essential to ensure that the database is widely adopted and integrated into routine clinical research workflows.

In addition, there is a need to enrich diagnostic information. Future work should aim to systematically record and standardize the three key elements of CP diagnosis: severity level, gait pattern, and CP subtype. Having these dimensions clearly defined in the database would open the door to comparative studies examining their interrelations, their relative contribution to gait characterization, and their usefulness in guiding treatment decisions.

Overall, the *Movement Analysis Network* has the potential to become a cornerstone for gait analysis in CP and related conditions. By expanding its size, improving clinical annotations, and facilitating collaborative access, this resource could support a wide range of studies aimed at advancing personalized treatment strategies and deepening our understanding of gait impairments.

6.4 Scientific Articles and Contributions

The paper entitled "Data Standardization for Gait Analysis in the ORITEL Network" [101] was disseminated through various academic venues. It was presented at the Congreso Anual de Ingeniería Biomédica (CAIB 2024) and at the 3D Analysis of Human Movement, Rehabilitation, Sports Medicine and Biomechanics Conference (3DAHM 2024), which also led to a conference publication in IEEE Xplore. In addition, the work was showcased in poster format at the Escuela de Verano en Inteligencia Computacional (EVIC 2024).

Furthermore, the article titled “Multicenter Standardized Gait Database from the ORITEL Network for Machine Learning-Based Classification of Cerebral Palsy Patterns” has been submitted to the journal *IEEE Journal of Biomedical and Health Informatics* and is currently under review. This paper details the development of the machine learning classifier and its performance evaluation, contributing to the scientific community’s understanding of gait analysis in cerebral palsy.

Capítulo 6: Discusión y Conclusión

6.1 Discusión

6.1.1 Creación de la Base de Datos de Marcha

La creación de la base de datos fue un proceso largo y exigente que involucró la recopilación de datos de marcha desde varios laboratorios, la preparación de materiales, la organización de reuniones y la realización de sesiones de capacitación. Esto requirió una fuerte coordinación entre centros y añadió carga adicional de trabajo a los profesionales clínicos involucrados, lo que reflejó la dedicación y el trabajo en equipo de todos los participantes. Gracias a estos esfuerzos, la campaña de recolección de datos fue exitosa, resultando en un repositorio de **156** estudios provenientes de **8** laboratorios, incluyendo sujetos con parálisis cerebral y otras condiciones relacionadas con la marcha. Los estudios variaron en su extensión, con algunos incluyendo 5 ensayos, otros 4, 3 o 2 ensayos por estudio. Esta diversidad aportó valor a la base de datos, al reflejar las diferencias reales en la recolección de datos entre los distintos centros.

Posteriormente, se llevó a cabo un proceso de estandarización para asegurar que todos los conjuntos de datos siguieran el mismo formato, independientemente del sistema de captura de marcha utilizado. El paso final consistió en cargar los datos estandarizados en una base de datos NoSQL. Si bien la base de datos ya constituía un recurso valioso, aún existían áreas de mejora. No todos los registros incluían información sobre tratamientos, y el proceso para añadir nuevos datos podía simplificarse. Asimismo, mejorar la documentación ayudaría a que otros investigadores pudieran utilizar la base de datos de manera más efectiva.

A pesar de estos desafíos, la base de datos proporcionó una base sólida para análisis reproducibles, comparaciones multicentro y el desarrollo de modelos clínicamente relevantes. También sentó las bases para un crecimiento futuro, con oportunidades de contar con anotaciones más completas, procesos de envío de datos más simples y una mayor colaboración. En conjunto, esta base de datos representó un avance significativo para ORITEL y para la investigación en marcha en la región.

6.1.2 Rendimiento y Generalización del Clasificador MLP

Para el desarrollo del modelo de ML, se realizó un análisis comparativo entre 8 modelos distintos, probando dos enfoques diferentes para balancear el conjunto de datos: *smote* y ponderación de muestras. El MLP con ponderación de clases alcanzó el mejor rendimiento, con una exactitud del 82% y valores de F1 por clase de 0.81 (diparesia), 0.63 (hemiparesia) y 0.98 (sin patología), como se muestra en la Figura 5.3. Este desempeño superó al de los demás modelos tanto en exactitud como en F1, demostrando la efectividad del MLP en el manejo de datos desbalanceados. De acuerdo con la revisión de literatura, estos resultados se ubicaron dentro del rango esperado. Aunque no se encontraron estudios directamente comparables, se reportaron resultados similares en trabajos que clasificaban la severidad de la parálisis cerebral y en subcategorías de patrones de marcha como hemiparesia y diparesia. Es importante destacar que este modelo fue entrenado y probado con un conjunto de 413 ensayos recolectados de múltiples laboratorios de marcha, seleccionados según los criterios de inclusión y exclusión descritos en la Sección 3.1.3 de la metodología. Según el estado del arte, este conjunto representó un tamaño relativamente grande, lo que pudo haber contribuido al sólido desempeño del modelo.

El uso de ponderación de clases permitió que el modelo MLP otorgara mayor importancia a la clase minoritaria, lo cual fue esencial al trabajar con conjuntos de datos desbalanceados. Esta estrategia ayudó al modelo a aprender mejor las características de la clase minoritaria, resultando en métricas de desempeño superiores. La ponderación reequilibró la función de pérdida preservando las proporciones de clase y evitando muestras sintéticas que pudieran distorsionar los límites, particularmente en variables categóricas codificadas. También respetó el agrupamiento definido por *StratifiedGroupKFold* y fortaleció la señal de la clase minoritaria sin inflar los datos, reduciendo así el sobreajuste y alineando las distribuciones de entrenamiento y prueba externa.

La configuración óptima se logró con una arquitectura de tres capas: 128 neuronas en la primera capa, 240 en la segunda y 176 en la tercera, utilizando la función de activación Gelu y una tasa de aprendizaje de $1.651 \cdot 10^{-4}$ (Tabla 5.4). El análisis de importancia de hiperparámetros reveló que los factores más influyentes fueron la tasa de aprendizaje, el número de componentes de PCA y la tasa de *dropout* (Figura 5.3).

Como se muestra en la Figura 5.2, la matriz de confusión indicó que el modelo tuvo dificultad para distinguir entre los diferentes patrones de marcha, especialmente entre diparesia y hemiparesia. Para evaluar la capacidad de generalización del modelo, este fue probado con un conjunto externo de 63 ensayos que no había sido utilizado previamente. La mayoría de estos ensayos correspondieron a casos de diparesia, e incluyeron sujetos con 2, 3, 4 o 5 ensayos por estudio, lo que permitió verificar la robustez del modelo frente a la variación en el número de ensayos por sujeto.

Como se observa en la Tabla 5.3, el puntaje F1 para diparesia durante el entrenamiento fue de 0.81 y para hemiparesia de 0.63. En la evaluación externa, presentada en la Tabla 5.5, el F1 para diparesia disminuyó a 0.76 y para hemiparesia a 0.55. Esta disminución era esperada, dado que el conjunto externo tenía una composición distinta y menos muestras de hemiparesia. Los resultados sugirieron que el modelo fue más eficaz en la clasificación de diparesia, probablemente debido a la mayor disponibilidad de datos de esta clase en entrenamiento y prueba. Aumentar el número de muestras de hemiparesia podría mejorar el rendimiento del modelo en dicha categoría.

En conjunto, estos resultados demostraron que el modelo MLP propuesto, entrenado en un conjunto relativamente grande y heterogéneo, fue capaz de alcanzar un buen desempeño en la clasificación de patrones de marcha en parálisis cerebral, particularmente en casos de diparesia. Sin embargo, el desbalance de clases siguió siendo un factor limitante, y se recomendó que trabajos futuros se enfocaran en expandir el conjunto de datos, especialmente para hemiparesia, además de explorar estrategias avanzadas de balanceo o técnicas de adaptación de dominio para mejorar aún más la capacidad de generalización del modelo.

6.1.3 Importancia de Características y Análisis Estadístico

El análisis de importancia de características indicó que las diez variables más relevantes para el clasificador MLP fueron: longitud de zancada promedio, fase de apoyo promedio, fase de balanceo promedio, doble apoyo promedio, duración de zancada promedio, estatura, longitud promedio de las extremidades inferiores, peso, acLKFE mínimo y acLHPIE mínimo (Figura 5.5). Estas variables coincidieron con la literatura previa, que resaltó consistentemente el rol de los parámetros espaciotemporales y cinemáticos como descriptores centrales de las anormalidades de marcha en la parálisis cerebral.

La prueba U de Mann–Whitney fue aplicada para calcular los valores p crudos (P_{raw}), que representaron la probabilidad de observar las diferencias entre distribuciones por azar. Un $P_{raw} < 0.05$ indicó que las diferencias observadas difícilmente se debieron a la variación aleatoria, y por tanto las distribuciones podían considerarse significativamente distintas. Como se muestra en la Tabla 5.6, la mayoría de las variables presentaron diferencias estadísticamente significativas entre hemiparesia y diparesia. En particular, el doble apoyo y el acLKFE mínimo alcanzaron valores P_{raw} muy bajos, reforzando su rol como discriminadores clave entre los dos grupos. Estos hallazgos sugirieron que el clasificador pudo haberse basado fuertemente en estas características para capturar diferencias clínicamente relevantes en la marcha.

El doble apoyo mostró el mayor tamaño del efecto (Cliff's $\delta = -0.5204$), fuertemente asociado con diparesia. Este hallazgo fue consistente con descripciones clínicas, ya que los niños con diparesia suelen presentar mayor inestabilidad y requieren períodos más prolongados de doble apoyo en comparación con aquellos con hemiparesia. De manera similar, las fases de balanceo y apoyo mostraron diferencias significativas: la diparesia presentó una fase de balanceo más larga y una fase de apoyo más corta. Estas adaptaciones temporales pueden reflejar estrategias compensatorias relacionadas con las limitaciones motoras bilaterales, mientras que la hemiparesia suele caracterizarse por una carga asimétrica y una reducción del apoyo en la extremidad afectada.

Las características cinemáticas acLKFE mínimo y acLHPIE mínimo también contribuyeron a la separación entre grupos. En particular, el acLKFE mínimo presentó un tamaño del efecto medio a grande ($\delta = -0.4336$), sugiriendo que los mínimos de flexión de rodilla fueron más pronunciados en diparesia. Este resultado respaldó la literatura que describe una flexión excesiva de rodilla durante la marcha en pacientes diparéticos, frecuentemente asociada a patrones de marcha en flexión. En contraste, las diferencias en acLHPIE mínimo fueron más débiles y no se mantuvieron significativas tras la corrección de Bonferroni, lo que indicó que la rotación interna-externa de cadera aportó menos de manera consistente a la discriminación entre grupos.

Las variables antropométricas como la estatura y la longitud de las extremidades inferiores también mostraron diferencias significativas, siendo menores en el grupo diparético. Estos resultados fueron coherentes con estudios previos que reportaron retrasos en el crecimiento y reducción de la longitud de las extremidades en subtipos más severos de parálisis cerebral. Sin embargo, variables como la duración de la zancada y el peso no alcanzaron significancia estadística, sugiriendo una contribución menor a la clasificación.

Tras aplicar la corrección de Bonferroni, las diferencias más robustas entre hemiparesia y diparesia se observaron en doble apoyo, acLKFE mínimo, longitud de extremidades inferiores, fase de balanceo, fase de apoyo y estatura ($p_{\text{bonf}} < 0.05$). Esto reforzó la interpretación de que una combinación de parámetros espaciotemporales y cinemática de extremidades inferiores son claves en la discriminación de estos dos subtipos de parálisis cerebral.

En conjunto, el análisis estadístico no solo confirmó la relevancia de las características identificadas, sino que también destacó diferencias clínicamente significativas entre hemiparesia y diparesia. Los resultados sugirieron que el clasificador se apoyó en parámetros fisiológicamente interpretables, particularmente el doble apoyo, la fase de balanceo y la fase de apoyo, al distinguir los patrones de marcha. Estos hallazgos fortalecieron la validez del modelo y señalaron el potencial de combinar descriptores espaciotemporales y cinemáticos para la clasificación automatizada de la marcha en parálisis cerebral.

6.1.4 Asociación de Tratamientos con el Clasificador de Patrones de Marcha

Al analizar la asociación de tratamientos con hemiparesia y diparesia en parálisis cerebral, fue necesario considerar varias limitaciones del conjunto de datos. En primer lugar, no todos los estudios incluyeron información sobre tratamientos, lo que restringió el alcance del análisis. Además, debido a que el tratamiento asignado a cada paciente se mantuvo constante en todos sus ensayos, la variabilidad de la información relacionada con tratamientos fue inherentemente limitada. Otro factor importante fue que los centros de rehabilitación a menudo reportaron recomendaciones de tratamiento en lugar de evidencia verificada de intervenciones completadas. Como resultado, no fue posible evaluar la efectividad real de los tratamientos a menos que existieran múltiples estudios de un mismo paciente junto con su historial médico completo a lo largo del tiempo.

A pesar de estas restricciones, se realizó un análisis exploratorio de frecuencias para identificar qué tratamientos fueron mencionados con mayor frecuencia en hemiparesia y diparesia. Este enfoque proporcionó una primera visión de los patrones potenciales de tratamiento, que podrían ampliarse y validarse con conjuntos de datos más grandes en el futuro. Para complementar estos hallazgos, se llevó a cabo una revisión de la literatura con el fin de comparar los tratamientos más reportados en la base de datos con la evidencia existente sobre su efectividad en el manejo de alteraciones de la marcha. Aunque este análisis no permitió conclusiones definitivas, ofreció un punto de partida para futuros estudios orientados a vincular los tipos de tratamiento con resultados clínicos en parálisis cerebral.

Para hemiparesia, los tratamientos propuestos incluyeron toxina botulínica, órtesis (especialmente órtesis tobillo-pie), fisioterapia, cirugía ortopédica, terapia ocupacional y programas de ejercicio o fortalecimiento. En contraste, para diparesia, los principales tratamientos recomendados fueron toxina botulínica, órtesis (principalmente tobillo-pie), fisioterapia, cirugía ortopédica, rizotomía dorsal selectiva y programas de ejercicio o fortalecimiento.

Según el análisis de frecuencias presentado en la Tabla 5.8, los tres tratamientos más comunes en hemiparesia fueron órtesis, cirugía ortopédica y fisioterapia. Para diparesia, los tres más comunes fueron toxina botulínica, cirugía ortopédica y fisioterapia, como se muestra en la Tabla 5.9. Estos resultados sugirieron que, aunque existió cierta superposición entre ambas condiciones, también hubo diferencias en las prioridades terapéuticas, particularmente con el uso más frecuente de toxina botulínica en diparesia y de terapia ocupacional en hemiparesia.

Asimismo, el análisis de combinaciones de tratamientos entregó información adicional sobre cómo se emplearon las intervenciones en conjunto. Los diagramas de cuerdas (Figura 5.6) ilustraron estas asociaciones, donde líneas más gruesas indicaron combinaciones que aparecieron con mayor frecuencia en los estudios. Para hemiparesia, la terapia ocupacional se recomendó a menudo junto con fisioterapia, mientras que la cirugía ortopédica se combinó frecuentemente con órtesis. Para diparesia, la combinación más habitual fue toxina botulínica con el uso de órtesis.

En conjunto, este análisis exploratorio destacó patrones preliminares en la asociación entre tratamientos y subtipos de parálisis cerebral. Aunque limitado por los datos disponibles, los hallazgos sugirieron que ciertos tratamientos fueron consistentemente priorizados para hemiparesia y diparesia, tanto de forma individual como en combinación. Estos resultados subrayaron la importancia de ampliar los conjuntos de datos y realizar estudios longitudinales para comprender mejor cómo las estrategias de tratamiento influyen en los resultados clínicos en parálisis cerebral.

6.2 Conclusiones

Los objetivos definidos al inicio de este trabajo fueron alcanzados de manera satisfactoria.

En primer lugar, se logró la creación y estandarización de una base de datos multicéntrica de marcha. Se recopilaron datos de **156** estudios provenientes de **8** laboratorios, los cuales fueron armonizados y almacenados en una estructura unificada que garantiza su comparabilidad y reproducibilidad. Este proceso estableció un recurso valioso para futuras aplicaciones clínicas y de investigación, cumpliendo con el objetivo de estandarización de datos entre centros.

En segundo lugar, se desarrolló un clasificador de patrones de marcha basado en técnicas de aprendizaje automático. Entre los modelos evaluados, el MLP con ponderación de clases alcanzó el mejor desempeño, logrando una exactitud global de **82%** y altos valores de F1-score en las distintas clases. El clasificador mostró una buena capacidad de generalización al ser probado con datos externos, particularmente en casos de diparesia, lo que demuestra su robustez y cumple con el objetivo de construir un modelo efectivo para la clasificación de la marcha en parálisis cerebral. De esta manera, se confirmó la hipótesis inicial, ya que la clasificación de patrones de marcha basada en evaluaciones de laboratorios de análisis de movimiento puede apoyar la selección de un tratamiento adecuado en al menos el **80%** de los casos, permitiendo intervenciones más personalizadas.

En tercer lugar, se analizaron las variables clínicas relevantes para la clasificación. A través de técnicas de reducción de dimensionalidad (PCA) y pruebas estadísticas (Mann–Whitney U, delta de Cliff y corrección de Bonferroni), se identificaron parámetros espaciotemporales y cinemáticos, como el doble apoyo, las fases de balanceo y apoyo, y el mínimo de acLKFE, como discriminadores clave entre hemiparesia y diparesia. Estos resultados confirmaron el rol de características clínicamente interpretables en el proceso de decisión del clasificador, cumpliendo el objetivo de vincular la clasificación con variables significativas.

Finalmente, se realizó un análisis exploratorio de la asociación entre los tratamientos y los perfiles de marcha. Aunque el conjunto de datos presentó limitaciones (información incompleta sobre tratamientos y ausencia de seguimiento longitudinal), el análisis de frecuencias permitió identificar patrones preliminares para hemiparesia y diparesia, respaldados por la literatura. En hemiparesia se destacaron las órtesis, la cirugía ortopédica y la fisioterapia, mientras que en diparesia fueron más frecuentes la toxina botulínica, la cirugía ortopédica y la fisioterapia. Este análisis cumplió parcialmente el objetivo de vincular la clasificación de la marcha con los tratamientos y sentó una base para futuros estudios.

En resumen, el estudio alcanzó su objetivo principal de desarrollar un clasificador de marcha basado en aprendizaje automático para parálisis cerebral, utilizando datos multicéntricos. La integración de una base de datos estandarizada, la construcción de modelos robustos de clasificación, la validación estadística de características clave y el análisis exploratorio de tratamientos representan una contribución significativa al estudio de la marcha.

6.3 Perspectivas Futuras

La base de datos *Movement Analysis Network* establecida en este trabajo representa un primer paso hacia un recurso colaborativo y estandarizado para la investigación de la marcha. Sin embargo, existen varias direcciones en las que este trabajo puede ampliarse para incrementar aún más su impacto clínico y científico.

En primer lugar, se espera que la base de datos continúe creciendo con el tiempo a medida que se incorporen nuevos estudios. La adición de nuevos casos no solo mejorará la representatividad de los datos, sino que también permitirá realizar análisis longitudinales de los patrones de marcha, posibilitando el estudio de cambios cinemáticos a lo largo del tiempo. Para maximizar su utilidad, una tarea pendiente es completar y refinar los registros actuales. En particular, añadir información estructurada sobre las recomendaciones de tratamiento entregadas por los centros de rehabilitación, así como registrar el historial de tratamientos previamente aplicados a cada sujeto, permitiría futuros estudios sobre la efectividad de las intervenciones para la parálisis cerebral y otras condiciones relacionadas con la marcha.

Otra dirección importante es hacer que la base de datos sea totalmente accesible para los profesionales dentro de ORITEL, fomentando colaboraciones multicentro y apoyando diversos proyectos clínicos y de investigación. Proporcionar documentación clara y herramientas de acceso amigables será esencial para garantizar que la base de datos sea ampliamente adoptada e integrada en los flujos de trabajo rutinarios de la investigación clínica.

Además, existe la necesidad de enriquecer la información diagnóstica. El trabajo futuro debería orientarse a registrar y estandarizar de manera sistemática los tres elementos clave del diagnóstico de parálisis cerebral: el nivel de severidad, el patrón de marcha y el subtipo. Contar con estas dimensiones claramente definidas en la base de datos abriría la puerta a estudios comparativos que examinen sus interrelaciones, su contribución relativa a la caracterización de la marcha y su utilidad en la orientación de decisiones de tratamiento.

En general, la *Movement Analysis Network*tiene el potencial de convertirse en una referencia fundamental para el análisis de la marcha en parálisis cerebral y condiciones relacionadas. Al ampliar su tamaño, mejorar las anotaciones clínicas y facilitar el acceso colaborativo, este recurso podría apoyar una amplia gama de estudios destinados a avanzar en estrategias de tratamiento personalizadas y a profundizar nuestra comprensión de las alteraciones de la marcha.

6.4 Artículos Científicos y Contribuciones

El artículo titulado "Data Standardization for Gait Analysis in the ORITEL Network" [101] fue difundido en diversas instancias académicas. Se presentó en el Congreso Anual de Ingeniería Biomédica (CAIB 2024) y en la 3D Analysis of Human Movement, Rehabilitation, Sports Medicine and Biomechanics Conference (3DAHM 2024), lo que además dio lugar a una publicación de conferencia en IEEE Xplore. Adicionalmente, el trabajo fue expuesto en formato póster en la Escuela de Verano en Inteligencia Computacional (EVIC 2024).

Asimismo, el artículo titulado "Multicenter Standardized Gait Database from the ORITEL Network for Machine Learning-Based Classification of Cerebral Palsy Patterns" ha sido enviado a la revista *IEEE Journal of Biomedical and Health Informatics* y se encuentra actualmente en proceso de revisión. Este trabajo detalla el desarrollo del clasificador basado en aprendizaje automático y su evaluación de desempeño, contribuyendo a la comprensión de la comunidad científica respecto al análisis de la marcha en parálisis cerebral.

Bibliography

[1] F. Horst, D. Slijepcevic, Simak, B. Horsak, S. Wolfgang, and M. Zeppelzauer, "Modeling biological individuality using machine learning: A study on human gait," *Computational and Structural Biotechnology Journal*, vol. 21, pp. 3414–3423, June 2023.

[2] L. M. Silva and N. Stergiou, "Chapter 7 - the basics of gait analysis," in *Biomechanics and Gait Analysis*, N. Stergiou, Ed. Academic Press, 2020, pp. 225–250. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128133729000075

[3] I. Klöpfer-Krämer, A. Brand, H. Wackerle, J. Müßig, I. Kröger, and P. Augat, "Gait analysis – available platforms for outcome assessment," *Injury*, vol. 51, no. 2, pp. 90–96, May 2020.

[4] S. Mihcin, S. Ciklacandir, M. Kocak, and A. Tosun, "Wearable motion capture system evaluation for biomechanical studies for hip joints," *Journal of Biomechanical Engineering*, vol. 143, no. 4, February 2021.

[5] R. A. States, J. J. Krzak, Y. Salem, E. M. Godwin, A. W. Bodkin, and M. L. McMulkin, "Instrumented gait analysis for management of gait disorders in children with cerebral palsy: A scoping review," *Gait & Posture*, vol. 90, no. 3, pp. 1–8, October 2021.

[6] H. K. Graham, P. Thomason, K. Willoughby, T. Hastings-Ison, R. Van Stralen, B. Dala-Ali, P. Wong, and E. Rutz, "Musculoskeletal pathology in cerebral palsy: A classification system and reliability study," *Children*, vol. 8, no. 3, p. 252, Mar. 2021.

[7] D. Slijepcevic, M. Zeppelzauer, F. Unglaube, A. Kranzl, C. Breiteneder, and B. Horsak, "Explainable machine learning in human gait analysis: A study on children with cerebral palsy," *IEEE Access*, vol. 11, no. 3, pp. 65 906–65 923, July 2023.

[8] M. Sadowska, B. Sarecka-Hujar, and I. Kopyta, "Cerebral palsy: Current opinions on definition, epidemiology, risk factors, classification and treatment options," *Neuropsychiatr Dis Treat*, vol. 16, pp. 1505–1518, June 2020.

[9] A. L. Schwabe, "Comprehensive care in cerebral palsy," *Physical Medicine and Rehabilitation Clinics of North America*, vol. 31, no. 1, pp. 1–13, February 2020.

[10] M. D. Peterson and E. A. Hurvitz, "Cerebral palsy grows up," *Mayo Clinic Proceedings*, vol. 96, no. 6, pp. 1404–1406, June 2021.

[11] S. Yang, J. Xia, J. Gao, and L. Wang, "Increasing prevalence of cerebral palsy among children and adolescents in china 1988–2020: A systematic review and meta-analysis," *Journal of Rehabilitation Medicine*, vol. 53, no. 5, May 2021.

[12] M. Kamate and M. Detroja, "Which is the most common physiologic type of cerebral palsy?" *Neurology India*, vol. 70, no. 3, pp. 1048–1051, July 2022.

[13] S. M. Reid, J. B. Carlin, and D. S. Reddihough, "Using the Gross Motor Function Classification System to describe patterns of motor severity in cerebral palsy," *Developmental Medicine & Child Neurology*, vol. 53, pp. 1007–1012, 2011.

[14] V. Skoutelis, A. D. Kanellopoulos, V. A. Kontogeorgakos, A. Dinopoulos, and P. J. Papagelopoulos, "The orthopaedic aspect of spastic cerebral palsy," *Journal of Orthopaedics*, vol. 22, pp. 553–558, November 2020.

[15] E. L. Dugan and J. S. Shilt, "The role of motion analysis in surgical planning for gait abnormalities in cerebral palsy," *Phys Med Rehabil Clin N Am*, vol. 31, no. 1, pp. 107–115, February 2020.

[16] S. Bhattacharjee, "A review on the types, risk factors, diagnosis, and treatment of cerebral palsy," *Indian Journal of Physical Medicine & Rehabilitation*, vol. 31, no. 4, p. 97, 2020.

[17] K. Merhy, M. de Oliveira, G. Bella, and C. Maurer-Morelli, "Epidemiological and functional profile of children with cerebral palsy assisted at the unicamp clinical hospital," *Pediatric Health, Medicine and Therapy*, vol. 16, pp. 47–59, March 2025.

[18] S. Faccioli, E. Pagliano, A. Ferrari, C. Maghini, M. F. Siani, G. Sgherri, G. Cappetta, G. Borelli, G. M. Farella, M. Foscan, M. Viganò, S. Sghedoni, S. Perazza, and S. Sassi, "Evidence-based management and motor rehabilitation of cerebral palsy children and adolescents: A systematic review," *Frontiers in Neurology*, vol. 14, p. 1171224, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fneur.2023.1171224/full

[19] M. Rana, J. Upadhyay, A. Rana, S. Durgapal, and A. Jantwal, "A systematic review on etiology, epidemiology, and treatment of cerebral palsy," *International Journal of Nutrition, Pharmacology, Neurological Diseases*, vol. 7, no. 4, pp. 76–83, 2017. [Online]. Available: https://doi.org/10.4103/ijnpnd.ijnpnd_26_17

[20] N. A. Gonzalez, R. R. Sanivarapu, U. Osman, L. A. Kumar, A. Sadagopan, A. Mahmoud, M. Begg, M. Tarhuni, M. N. Fotso, and S. Khan, "Physical therapy interventions in children with cerebral palsy: A systematic review," *Cureus*, vol. 15, no. 8, p. e43846, aug 2023.

[21] A. Al Shami, "A comprehensive review of rehabilitation strategies post-orthopedic intervention in pediatric cerebral palsy," *Journal of Musculoskeletal Surgery and Research*, vol. 9, pp. 42–48, 2025.

[22] A. Shamsoddini, S. Amirsalari, M.-T. Hollisaz, A. Rahimnia, and A. Khatibi-Aghda, "Management of spasticity in children with cerebral palsy," *Iranian Journal of Pediatrics*, vol. 24, no. 4, pp. 345–351, August 2014.

[23] I. S. Son, D. H. Lee, S. H. Hong, K. H. Lee, and G. H. Lee, "Comparison of gait ability of a child with cerebral palsy according to the difference of dorsiflexion angle of hinged ankle-foot orthosis: A case report," *American Journal of Case Reports*, vol. 20, pp. 1454–1459, 2019.

[24] S. Lin, K. Evans, D. Hartley, S. Morrison, S. McDonald, M. Veidt, and G. Wang, "A review of gait analysis using gyroscopes and inertial measurement units," *Sensors*, vol. 25, no. 11, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/25/11/3481

[25] L. Molteni and G. Andreoni, "Comparing the accuracy of markerless motion analysis and optoelectronic system for measuring gait kinematics of lower limb," *Bioengineering (Basel)*, vol. 12, no. 4, p. 424, April 2025.

[26] S. Scataglini, E. Abts, C. Van Bocxlaer, M. Van den Bussche, S. Meletani, and S. Truijen, "Accuracy, validity, and reliability of markerless camera-based 3d motion capture systems versus marker-based 3d motion capture systems in gait analysis: A systematic review and meta-analysis," *Sensors*, vol. 24, no. 11, p. 3686, 2024.

[27] A. Bartoszek, A. Struzik, S. Jaroszczuk, M. Wozniewski, and B. Pietraszewski, "Comparison of the optoelectronic bts smart system and imu-based myomotion system for the assessment of gait variables," *Acta of Bioengineering and Biomechanics*, vol. 24, no. 1, pp. 103–116, February 2022.

[28] M. Abdullah, A. A. Hulleck, R. Katmah *et al.*, "Multibody dynamics-based musculoskeletal modeling for gait analysis: a systematic review," *Journal of NeuroEngineering and Rehabilitation*, vol. 21, p. 178, 2024. [Online]. Available: https://doi.org/10.1186/s12984-024-01458-y

[29] A. A. Hulleck, D. Mohan, N. Abdallah, M. El-Rich, and K. Khalaf, "Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies," *Frontiers in Medical Technology*, vol. 4, 12 2022.

[30] D. S. Y. Vun, R. Bowers, and A. McGarry, "Vision-based motion capture for the gait analysis of neurodegenerative diseases: A review," *Gait Posture*, vol. 112, pp. 95–107, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0966636224001280

[31] Gaurav, "Mobile health (mhealth) dataset," https://www.kaggle.com/datasets/gaurav2022/mobile-health, 2022, accessed: 2025-06-16.

[32] A. Calev, "Time-series models — pamap2 dataset," https://www.kaggle.com/code/avrahamcalev/time-series-models-pamap2-dataset, 2023, accessed: 2025-06-16.

[33] T. Van Criekinge, W. Saeys, S. Truijen, L. Vereeck, L. H. Sloot, and A. Hallemans, "A full-body motion capture gait dataset of 138 able-bodied adults across the life span and 50 stroke survivors," *Scientific Data*, vol. 10, no. 1, p. 852, December 2023.

[34] G. Santos, M. Wanderley, T. Tavares, and A. Rocha, "A multi-sensor human gait dataset captured through an optical system and inertial measurement units," *Scientific Data*, vol. 9, no. 1, p. 545, September 2022.

[35] T. D. Laet, A. Nieuwenhuys, E. Papageorgiou, and K. Desloovere, "3D gait analysis data of children with CP used for gait classification," 4 2017. [Online]. Available: https://figshare.com/articles/dataset/3D_gait_analysis_data_of_children_with_CP_used_for_gait_classification/4877432

[36] T. De Mulder, H. Adams, T. Dewit, G. Molenaers, A. Van Campenhout, and K. Desloovere, "Dataset on the effects of verbal and virtual reality feedback on gait in children with cerebral palsy," https://rdr.kuleuven.be/dataset.xhtml?persistentId=doi:10.48804/JUNNWW, 2024, associated with: De Mulder et al., 2024, Children, DOI: 10.3390/children11050524. Accessed: 2025-06-16.

[37] A. Moniruzzaman and S. A. Hossain, "Nosql database: New era of databases for big data analytics—classification, characteristics and comparison," *arXiv preprint arXiv:1307.0191*, 2013.

[38] P. J. Sadalage and M. Fowler, *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*. Addison-Wesley, 2012.

[39] J. Han, E. Haihong, G. Le, and J. Du, "Survey on nosql database," *IEEE Pervasive Computing*, vol. 10, no. 4, pp. 10–20, 2011.

[40] MongoDB Inc. Mongodb. [Online]. Available: https://www.mongodb.com/

[41] Apache couchdb. [Online]. Available: https://couchdb.apache.org/

[42] Redis. [Online]. Available: https://redis.io/

[43] Amazon dynamodb. [Online]. Available: https://aws.amazon.com/es/dynamodb/

[44] Apache cassandra. [Online]. Available: https://cassandra.apache.org/_/index.html

[45] Apache hbase. [Online]. Available: https://hbase.apache.org/

[46] Neo4j: The world's leading graph database. [Online]. Available: https://neo4j.com/

[47] Amazon neptune. [Online]. Available: https://aws.amazon.com/es/neptune/

[48] M. Sharma and N. K. Trivedi, "A comprehensive review of mongo db: Features, advantages, and limitations," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–5.

[49] Mongodb manual. [Online]. Available: https://www.mongodb.com/docs/manual/

[50] U. Sunarya, Y. Sun Hariyani, J. Cho, T.; Roh, J. Hyeong, I. Sohn, S. Kim, and C. Park, "Feature analysis of smart shoe sensors for classification of gait patterns," *Sensors*, vol. 20, no. 21, p. 6253, November 2020.

[51] E. Balaji, D. Brindha, and R. Balakrishnan, "Supervised machine learning based gait classification system for early detection and stage classification of parkinson's disease," *Applied Soft Computing*, vol. 94, p. 106494, September 2020.

[52] X. Zhang, H. Zhang, J. Hu, J. Zheng, X. Wang, J. Deng, Z. Wan, H. Wang, and Y. Wang, "Gait pattern identification and phase estimation in continuous multilocomotion mode based on inertial measurement units," *IEEE Sensors Journal*, vol. 22, no. 17, pp. 16 952–16 962, May 2022.

[53] P. Khera and N. Kumara, "Role of machine learning in gait analysis: a review," *Journal of Medical Engineering & Technology*, vol. 44, no. 8, pp. 441–467, October 2020.

[54] E. Balaji, D. Brindha, and R. Balakrishnan, "Supervised machine learning based gait classification system for early detection and stage classification of parkinson's disease," *Applied Soft Computing*, vol. 94, p. 106494, September 2020.

[55] C. Fricke, J. Alizadeh, N. Zakhary, T. B. Woost, M. Bogdan, and C. Joseph, "Evaluation of three machine learning algorithms for the automatic classification of emg patterns in gait disorders," *Front. Neurol.*, May 2021.

[56] B. Chen, C. Chen, J. Hu, Z. Sayeed, J. Qi, H. Darwiche, B. Little, S. Lou, M. Darwish, and C. Foote, “Computer vision and machine learning-based gait pattern recognition for flat fall prediction,” *Sensors*, p. 7960, October 2022.

[57] C. Dammeyer, C. Nüesch, R. M. S. Visscher, Y. K. Kim, P. Ismailidis, M. Wittauer, K. Stoffel, Y. Acklin, C. Egloff, C. Netzer, and A. Mündermann, “Classification of inertial sensor-based gait patterns of orthopaedic conditions using machine learning: A pilot study,” *Journal of Orthopaedic Research*, vol. 42, no. 7, pp. 1463–1472, July 2024, epub 2024 Feb 11. [Online]. Available: https://doi.org/10.1002/jor.25797

[58] S. Hwang, J. Kim, S. Yang, H.-J. Moon, K.-H. Cho, I. Youn, J.-K. Sung, and S. Han, “Machine learning based abnormal gait classification with imu considering joint impairment,” *Sensors*, vol. 24, p. 5571, 08 2024.

[59] H. Rabie and M. A. Akhloufi, “A review of machine learning and deep learning for Parkinson’s disease detection,” *Discover Artificial Intelligence*, vol. 5, no. 1, p. 24, 2025. [Online]. Available: https://doi.org/10.1007/s44163-025-00241-9

[60] Anonymous, “Predicting walking ability at discharge after spinal cord injury using machine learning,” *Journal of NeuroEngineering and Rehabilitation*, vol. 21, no. 1, p. 25, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/38847729/

[61] ——, “Predicting brain lesion location from gait features using machine learning in stroke patients,” *Journal of NeuroEngineering and Rehabilitation*, vol. 18, no. 1, p. 89, 2021. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12142055/

[62] ——, “Ai-assisted gait analysis in physical therapy: A systematic review of tools and rehabilitation outcomes,” *Preprint on ResearchGate*, 2024. [Online]. Available: https://www.researchgate.net/publication/392894946

[63] A. Nahar, S. Paul, and M. J. Saikia, “A systematic review on machine learning approaches in cerebral palsy research,” *PeerJ*, vol. 12, p. e18270, Oct. 2024.

[64] H. Darbandi, M. Baniasad, S. Baghdadi, A. Khandan, A. Vafaee, and F. Farahmand, “Automatic classification of gait patterns in children with cerebral palsy using fuzzy clustering method,” *Clinical Biomechanics*, vol. 73, pp. 189–194, March 2020.

[65] J. Choisne, N. Fourrier, G. Handsfield, N. Signal, D. Taylor, N. Wilson, S. Stott, and T. Besier, “An unsupervised data-driven model to classify gait patterns in children with cerebral palsy,” *Journal of Clinical Medicine*, vol. 9, no. 5, p. 1432, March 2020.

[66] S. Tsitlakidis, M. Schwarze, F. Westhauser, K. Heubisch, A. Horsch, S. Hagmann, S. Wolf, and M. Götze, "Gait indices for characterization of patients with unilateral cerebral palsy," *Journal of Clinical Medicine*, vol. 9, no. 12, p. 3888, October 2020.

[67] A. G. Melanda, J. R. Davids, A. C. Pauleto, A. R. Pelegrinelli, A. E. Kuntze Ferreira, L. A. Knaut, P. R. G. Lucareli, and S. M. Smaili, "Reliability and validity of the gait classification system in children with cerebral palsy (gcs-cp)," *Gait & Posture*, vol. 98, pp. 355–361, October 2022.

[68] M. Ahmadi, M. O'Neil, E. Baque, R. Boyd, and S. Trost, "Machine learning to quantify physical activity in children with cerebral palsy: Comparison of group, group-personalized, and fully-personalized activity classification models," *Sensors*, vol. 20, p. 3976, July 2020.

[69] A. M. Al-Sowi, N. AlMasri, B. Hammo, and F. A.-Z. Al-Qwaqzeh, "Cerebral palsy classification based on multi-feature analysis using machine learning," *Informatics in Medicine Unlocked*, vol. 37, p. 101197, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914823000394

[70] J. T. A. Valdivia, V. G. Rojas, and C. A. Astudillo, "Deep learning-based classification of hemiplegia and diplegia in cerebral palsy using postural control analysis," *Scientific Reports*, vol. 15, no. 1, p. 8811, 2025. [Online]. Available: https://doi.org/10.1038/s41598-025-93166-3

[71] D. Slijepcevic, M. Zeppelzauer, A. Gorgas, C. Schwab, M. Schüller, A. Baca, C. Breiteneder, and B. Horsak, "Automatic classification of functional gait disorders," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1653–1661, 2018. [Online]. Available: https://doi.org/10.1109/JBHI.2017.2785682

[72] A. Kocak, F. Yarar, and U. Cavlak, "Effects of dual task on gait velocity and cadence in cerebral palsied children with spastic hemiparesis or diparesis," *Acta Neurologica Belgica*, vol. 121, no. 1, pp. 175–179, 2021. [Online]. Available: https://doi.org/10.1007/s13760-020-01380-9

[73] T. Y. Choi, D. Park, D. Shim, J. O. Choi, J. Hong, Y. Ahn, E. S. Park, and D. W. Rha, "Gait adaptation is different between the affected and unaffected legs in children with spastic hemiplegic cerebral palsy while walking on a changing slope," *Children*, vol. 9, no. 5, p. 593, 2022.

[74] T. Y. Choi, D. Park, D. Shim, J.-o. Choi, J. Hong, Y. Ahn, E. S. Park, and D.-w. Rha, "Gait adaptation is different between the affected and unaffected legs in children with spastic hemiplegic cerebral palsy while walking on a changing slope," *Children*, vol. 9, no. 5, 2022. [Online]. Available: https://www.mdpi.com/2227-9067/9/5/593

[75] C. J. Kim and S. M. Son, “Comparison of spatiotemporal gait parameters between children with normal development and children with diplegic cerebral palsy,” *Journal of Physical Therapy Science*, vol. 26, no. 9, pp. 1317–1319, 2014.

[76] A. Szopa and M. Domagalska-Szopa, “Postural stability in children with cerebral palsy,” *Journal of Clinical Medicine*, vol. 13, no. 17, 2024. [Online]. Available: https://www.mdpi.com/2077-0383/13/17/5263

[77] M. A. Elshafey, M. S. Abdrabo, and R. K. Elnaggar, “Effects of a core stability exercise program on balance and coordination in children with cerebellar ataxic cerebral palsy,” *Journal of Musculoskeletal and Neuronal Interactions*, vol. 22, no. 2, pp. 172–178, 2022.

[78] K. Sharifmoradi, M. Kamali, and A. a. Tahmasebi, “Dynamic balance during gait in children with spastic diplegic cerebral palsy versus normal children,” *Physical Treatments - Specific Physical Therapy*, vol. 8, no. 1, 2018. [Online]. Available: http://ptj.uswr.ac.ir/article-1-357-en.html

[79] X. Wang and Y. Wang, “Gait analysis of children with spastic hemiplegic cerebral palsy,” *Neural Regeneration Research*, vol. 7, no. 20, pp. 1578–1584, 2012.

[80] S. Chakraborty, A. Nandy, and T. Kesar, “Gait deficits and dynamic stability in children and adolescents with cerebral palsy: A systematic review and meta-analysis,” *Clinical Biomechanics*, 10 2019.

[81] J. Zhou, E. E. Butler, and J. Rose, “Kinematic and kinetic gait patterns in children with spastic cerebral palsy,” *Frontiers in Human Neuroscience*, vol. 11, p. 103, 2017.

[82] M. Domagalska-Szopa and A. Szopa, “Gait pattern differences among children with bilateral cerebral palsy,” *Frontiers in Neurology*, vol. 10, p. 183, 2019. [Online]. Available: https://doi.org/10.3389/fneur.2019.00183

[83] K. B. Bumbard, H. Herrington, C.-H. Goh, and A. Ibrahim, “Incorporation of torsion springs in a knee exoskeleton for stance phase correction of crouch gait,” *Applied Sciences*, vol. 12, no. 14, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/14/7034

[84] The 3d standard. [Online]. Available: https://www.c3d.org/

[85] *BEYOND MOTION*, Vicon Motion Systems Limited, 2020.

[86] *SmartAnalyzer Handbook*, BTS Bioengineering., 2009.

[87] Optuna: A hyperparameter optimization framework. [Online]. Available: https://optuna.org/

[88] S. Tao, P. Peng, Y. Li, H. Sun, Q. Li, and H. Wang, "Supervised contrastive representation learning with tree-structured parzen estimator bayesian optimization for imbalanced tabular data," *Expert Systems with Applications*, vol. 237, p. 121294, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423017967

[89] Support Vector Machines — scikit-learn documentation. [Online]. Available: https://scikit-learn.org/stable/modules/svm.html

[90] sklearn.tree.DecisionTreeClassifier — scikit-learn documentation. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html

[91] sklearn.linear_model.LogisticRegression — scikit-learn documentation. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

[92] sklearn.ensemble.randomforestclassifier — scikit-learn documentation. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[93] Xgboost python api — xgboost 2.0.3 documentation. [Online]. Available: https://xgboost.readthedocs.io/en/stable/python/python_api.html

[94] sklearn.ensemble.gradientboostingclassifier — scikit-learn documentation. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html

[95] sklearn.linear_model.sgdclassifier — scikit-learn documentation. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html

[96] sklearn.neural_network.mlpclassifier — scikit-learn documentation. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

[97] J. A. Rice, *Mathematical Statistics and Data Analysis*, 3rd ed. Thomson Brooks/Cole, 2007.

[98] R. Wall Emerson, "Mann-whitney u test and t-test," *Journal of Visual Impairment & Blindness*, vol. 117, no. 1, pp. 99–100, 2023, (Original work published 2023). [Online]. Available: https://doi.org/10.1177/0145482X221150592

[99] M. R. Hess and J. D. Kromrey, “Robust confidence intervals for effect sizes: A comparative study of cohen’sd and cliff’s delta under non-normality and heterogeneous variances,” in *annual meeting of the American Educational Research Association*, vol. 1. Citeseer, 2004.

[100] R. E. Blakesley, S. Mazumdar, M. A. Dew, P. R. Houck, G. Tang, I. Reynolds, C. F., and M. A. Butters, “Comparisons of methods for multiple hypothesis testing in neuropsychological research,” *Neuropsychology*, vol. 23, no. 2, pp. 255–264, Mar. 2009.

[101] R. Ulloa, E. J. Pino, M. Morante, L. Llorente, and M. Galli, “Data standardization for gait analysis in the oritel network,” in *2024 International Symposium on 3D Analysis of Human Movement (3DAHM)*, 2024, pp. 1–5.

Appendix A: Database Structure and Measurement Units

Table A.1 presents the detailed structure of the *Subject's information* in the created database, including measurement units where applicable.

Tables A.2,A.3, andA.4 present the nomenclature defined for the standardization of variable names. It should be noted that all records in the database include kinematic and spatiotemporal parameters, whereas kinetic data are optional. Consequently, the information described in Table A.4 is not available for all entries in the *Movement Analysis Network* database.

Table A.1: Structure of the subject information stored in the *Movement Analysis Network* database

Name of the structure	Subfile names	File contents
edad	—	Age (Years)
informacion_adquisicion	condicion_examen	Examination conditions
	sistema_adquisicion	Acquisition system
	modelo_sistema	System model
	protocolo_marcadores	Marker protocol
	fecha_adquisicion	Month and year of acquisition
medidas_antropometricas	peso	weight in kilograms
	talla	height in centimeters
	ancho_pelvis	pelvic width in centimeters
	profundidad_pelvis	pelvic depth in centimeters
	longitud_extremidad_inferior	lower limb length in centimeters
	diametro_rodilla	knee diameter in centimeters
	diametro_tobillo	ankle diameter in centimeters
antecedentes_medicos	objetivo	The objective for which the study was conducted
	diagnostico	Diagnosis of the study subject
	tratamiento	Indicated treatment according to the study findings
	comentarios	Any subject-provided comments relevant to the study
nombre_archivos	cinematica	Name of the kinematics file (if applicable; N/A if not)
	cinetica	Names of the kinetic files (if applicable; N/A if not)
	parametros	Parameter file name (if applicable; N/A if not)
	emg	EMG file name (if applicable; N/A if not)

Table A.2: Nomenclature of the kinematic variables of the participating laboratories

Name of the variable	Plane	Nomenclature
Angle cycle Right Ankle Flexion/Extension	Sagittal	acRAFE
Angle cycle Right Ankle Internal/External	Transverse	acRAIE
Angle cycle Right Knee Flexion/Extension	Sagittal	acRKFE
Angle cycle Right Hip Flexion/Extension	Sagittal	acRHPFE
Angle cycle Right Hip Abduction/Adduction	Frontal	acRHPAA
Angle cycle Right Hip Internal/External	Transverse	acRHPIE
Angle cycle Right Pelvic tilt	Sagittal	acRPTILT
Angle cycle Right Pelvic obliquity	Frontal	acRPOBLI
Angle cycle Right Pelvic rotation	Transverse	acRPROT
Angle cycle Left Ankle Flexion/Extension	Sagittal	acLAFE
Angle cycle Left Ankle Internal/External	Transverse	acLAIE
Angle cycle Left Knee Flexion/Extension	Sagittal	acLKFE
Angle cycle Left Hip Flexion/Extension	Sagittal	acLHPFE
Angle cycle Left Hip Abduction/Adduction	Frontal	acLHPAA
Angle cycle Left Hip Internal/External	Transverse	acLHPIE
Angle cycle Left Pelvic tilt	Sagittal	acLPTILT
Angle cycle Left Pelvic obliquity	Frontal	acLPOBLI
Angle cycle Left Pelvic rotation	Transverse	acLPROT

Table A.3: Spatiotemporal Parameters Mapping Table

Spatiotemporal Parameter	Nomenclature
Support phase	`Fase_de_Apoyo`
Swing phase	`Fase_de_Balanceo`
Double support phase	`Doble_Apoyo`
Support duration	`Duracion_Apoyo`
Swing duration	`Duracion_Balanceo`
Stride duration	`Duracion_Zancada`
Step length	`Largo_del_Paso`
Stride length	`Largo_Zancada`
Cadence	`Cadencia`
Step width	`Ancho_del_Paso`
Average speed	`Vel_Prom`

Table A.4: Nomenclature of the kinetic variables of the participating laboratories

Name of the variable	Plane	Nomenclature
Torque cycle Right Ankle Flexion/Extension	Sagittal	tcRAFE
Torque cycle Right Knee Flexion/Extension	Sagittal	tcRKFE
Torque cycle Right Hip Flexion/Extension	Sagittal	tcRHPFE
Torque cycle Right Hip Abduction/Adduction	Frontal	tcRHPAA
Torque cycle Right Knee Abduction/Adduction (Varo/Valgo)	Frontal	tcRKAA
Torque cycle Left Ankle Flexion/Extension	Sagittal	tcLAFE
Torque cycle Left Knee Flexion/Extension	Sagittal	tcLKFE
Torque cycle Left Hip Flexion/Extension	Sagittal	tcLHPFE
Torque cycle Left Hip Abduction/Adduction	Frontal	tcLHPAA
Torque cycle Left Knee Abduction/Adduction (Varo/Valgo)	Frontal	tcLKAA
Power cycle Right Ankle Flexion/Extension	Sagittal	pcRAFE
Power cycle Right Knee Flexion/Extension	Sagittal	pcRKFE
Power cycle Right Hip Flexion/Extension	Sagittal	pcRHPFE
Power cycle Right Knee Abduction/Adduction	Frontal	pcRKAA
Power cycle Right Hip Abduction/Adduction	Frontal	pcRHPAA
Power cycle Left Ankle Flexion/Extension	Sagittal	pcLAFE
Power cycle Left Knee Flexion/Extension	Sagittal	pcLKFE
Power cycle Left Hip Flexion/Extension	Sagittal	pcLHPFE
Power cycle Left Knee Abduction/Adduction	Frontal	pcLKAA
Power cycle Left Hip Abduction/Adduction	Frontal	pcLHPAA
Force cycle Right Ground Reaction Vertical	Sagittal	fcRGRVE
Force cycle Right Ground Reaction Antero/Posterior	Sagittal	fcRGRAP
Force cycle Right Ground Reaction Medio/Lateral	Sagittal	fcRGRML
Force cycle Left Ground Reaction Vertical	Sagittal	fcLGRVE
Force cycle Left Ground Reaction Antero/Posterior	Sagittal	fcLGRAP
Force cycle Left Ground Reaction Medio/Lateral	Sagittal	fcGRML