UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ARTIFICIAL INTELLIGENCE FOR EVALUATING GAIT IN CANCER PATIENTS UNDERGOING CHEMOTHERAPY USING IMU SENSOR DATA

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

in

Electrical and Computer Engineering

By

HAYA TCHALABI
Norman, Oklahoma
2025

ARTIFICIAL INTELLIGENCE FOR EVALUATING GAIT IN CANCER PATIENTS

UNDERGOING CHEMOTHERAPY USING IMU SENSOR DATA

A THESIS APPROVED FOR THE

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

Dr. Chad Roller, Chair

Dr. Hazem Refai, Co-Chair

Dr. Samuel Cheng

Dr. Choon Yik Tang

Dedication

This work is built on more than study.

It is shaped by people who stood beside me and made the most difficult moments easier to carry.

And above all, it is built on faith in God, who always provided more than I knew to ask for, and more than people thought was possible. Praise be to him only.

To my beloved parents, Aila & Bakour

To my supportive brother, Saad

To my dear family and friends, those around and those I hold in memory

Thank you for your support and love in all its forms

Haya

Acknowledgments

I would like to sincerely thank my advisor *Dr. Hazem Refai,* for his support and mentorship throughout this research. I am also grateful to my committee Chair *Dr. Chad Roller* for his guidance and encouragement. My thanks extend to the members of my committee, *Dr. Samuel Cheng*, and *Dr. Choon Yik Tang* for their time and effort in reviewing this thesis. Special acknowledgment to Dr. Fahd Khan for developing the foundational codes for data extraction and synchronizing the pressure mat with the IMUs for the initial extraction of gait events.

In addition, I would like to extend my appreciation to my colleagues, as well as staff members including Ms. *Denise Davis*, and the *Student Affairs'* team for always being there when I needed help.

Finally, heartfelt thanks to the *University of Oklahoma* and the *OU-Tulsa* community of students, faculty, and staff for providing an environment that allowed me to grow, learn, and complete this journey.

Table of Contents

List of Tables

List of Figures

Abstract

Gait analysis plays a crucial role in understanding motor impairments, an example of that is cancer patients undergoing chemotherapy who can be at risk of chemotherapy-induced peripheral neuropathy (CIPN). Inertial measurement units (IMUs) offer an objective and accessible means of tracking movement [1], enabling the extraction of gait features that may indicate early functional deterioration. By integrating signal processing techniques with machine learning (ML), **this thesis aims to utilize IMU gait data to identify and segment gait events, assess its potential in biometric identification of different people under two walking conditions with the use of signal processing and machine learning techniques, and quantify gait changes over multiple visits for normal walking speed, all in cancer patients undergoing chemotherapy, providing a data-driven approach to assessing their mobility impairments**.

The first stage of this work focuses on preprocessing IMU data collected from 34 individuals under two different walking speeds, using signals obtained from foot-mounted IMUs, demonstrating that each person has their own unique walking pattern from features extracted from singular value decomposition (SVD) and discrete wavelet transform (DWT) used to train a support vector machine (SVM) classifier. This classification framework establishes the foundation for understanding individualized gait patterns and their variability.

Building upon this, the later chapter shifts toward analyzing how gait evolves over time for each study subject undergoing treatment. The previously extracted (DWT) gait features are further examined using machine learning techniques for the purpose of identifying meaningful changes across multiple visits. These methods aim to track longitudinal changes and determine whether observed differences reflect inherent individual fluctuations or potential early indicators of functional decline. By integrating step-level analysis and feature-based modeling, this work

seeks to provide a more comprehensive understanding of gait deterioration before it becomes clinically apparent, to enable earlier intervention, potentially improving the patient's quality of life by addressing mobility issues before they impact a person's daily activities.

1 Introduction

Gait is a fundamental aspect of human movement, and its analysis has become a critical area of study in clinical research to assess mobility impairments associated with causations such as aging, musculoskeletal conditions, and neurological disorders like chemotherapy-induced peripheral neuropathy (CIPN). CIPN is a common side effect of neurotoxic chemotherapy agents that causes sensory deficits, balance disturbances and gait abnormalities increasing the risk of falls and injuries. One study suggests that approximately 50-90% of patients are affected by CIPN, with the condition persisting indefinitely in nearly 30-40% of cases [2]. However, despite its prevalence, CIPN remains challenging to diagnose early as clinical assessment often relies on subjective patient-reported symptoms rather than quantitative gait measures.

Inertial measurement units (IMUs) have revolutionized the process of gait analysis by providing an affordable and portable alternative to traditional motor caption systems (i.e., pressure mats and optical sensors), which require specialized facilities and may restrict natural movements. Originally developed for aerospace and navigation application, IMUs have gained prominence in biomedical research due to their ability to capture precise motion data in real world settings. By measuring acceleration, angular velocity, and magnetic field using a combination of accelerometers, gyroscopes and magnetometers, they can provide detailed insights into movement when affixed to key anatomical regions such as the feet, ankles, lumbar, and sternum. Researchers can then use these insights for continuously tracking and analyzing movement patterns and assessing gait abnormalities in clinical populations.

Previous research on gait analysis in patients with CIPN has predominantly focused on extracting traditional biomechanical parameters, such as gait velocity, stride length, and joint angles to assess mobility impairments. Studies have shown that patients undergoing chemotherapy may experience a reduction in walking speed and increased sway area indicating balance issues. While these approaches provide valuable clinical insights, they rely on predefined gait metrics that may not fully capture the complexity of gait alterations in CIPN. This method deviates from the conventional approach by shifting the focus from clinically established biomechanical measurements to a data-driven perspective using IMUs. Instead of deriving gait changes based on parameters like step length or cadence, raw IMU signals were analyzed directly using data processing techniques to uncover patterns and variations not detectable using traditional methods. Subtle changes in gait that might not yet be pronounced enough to manifest as observable clinical impairments can be detected with IMUs and serve as early indicators of deterioration. By emphasizing data-driven feature extraction over traditional clinical descriptors, the approach in this thesis aims to expand the understanding of gait abnormalities beyond current biomechanical frameworks. This begins with **segmenting gait cycles** from foot mounted IMU data to isolate distinct gait events (flat-foot, swing, heel strike) using template matching with Euclidian distances and detecting turns using angular velocity signals from an IMU placed over the lumbar area, followed by an alignment process using Euclidian distances to ensure consistency and accurate feature extraction and comparison. Once the data is structured, two mathematical transformations are applied, Singular Value Decomposition (SVD) and Discrete Wavelet Transform (DWT) which provide complimentary information about gait segments. The resulting feature matrices are then combined and used to train a Support Vector Machine (SVM) model, which successfully differentiates 34 individuals under normal walking speed and 31 under fast speed, based on their walking patterns at their first visit (baseline) with accuracies of 86.97% ,82.96%, 83.55%, 82.37% for left-foot normal pace, right-foot normal pace, left-foot fast pace, and right-foot fast pace respectively. This finding reinforces the notion that each person's gait is distinct, demonstrating the feasibility of using IMU derived gait signatures for person identification.

This classification forms the foundation upon which we build later processes, in which the study shifts towards a deeper exploration of gait evolution over time, to quantify gait changes across multiple visits and assess potential deterioration by re-examining the extracted features through a different analytical framework, which leverages logistic regression-based models, unsupervised clustering and statistical methods to highlight the specific gait events that exhibit significant shifts, quantify that shift, and determine what steps experienced it within a walk comprising of multiple passes, offering a data-driven means of identifying deviations that may not yet be clinically observable but could signal the onset of functional impairments.

1.1 Thesis Contributions

This thesis was developed with the objective of identifying patterns and variations in gait for multiple participants and over multiple visits associated with deterioration in cancer patients undergoing chemotherapy, which may not yet be clinically detectable. The use of IMU acceleration signals is a shift from traditional biomechanical metrics to a feature-based analysis utilizing Singular Value Decomposition (SVD) and Discrete Wavelet Transform (DWT) to extract features from IMU data that can be employed in training various machine learning models including Support Vector Machines (SVM) and Logistic Regression. Initially, these tools were used to establish that each person has a unique walking signature that can be used to determine how subsequent visits diverge as CIPN and treatment progress by identifying exact gait events (swing, toe off, heel strike) that experience the most change for each person.

The main contributions of this thesis are summarized below:

[1] Developed a novel gait event segmentation and alignment framework isolating key gait indicators (flat foot, swing, heel strike) and turn events using Euclidian distances, filters, and thresholds.

[2] Developed a model based on signal processing and machine learning techniques to define unique signatures per person using two IMUs placed over the feet. The achieved accuracy of the model was 86.97%, 82.96%, 83.55%, 82.37% for left-foot normal pace, right-foot normal pace, left-foot fast pace, and right-foot fast pace respectively.

[3] Applied logistic regression-based models to systematically assess how gait evolves over time for each individual throughout visits.

[4] Discovered specific gait events (toe off, swing, heel strike) and steps that exhibit strong characteristic features during the walk.

[5] Exploring the possibility of identifying gait abnormality using data driven methods instead of relying on traditional spatiotemporal parameters

1.2 Content Summary

In alignment with the outlined objectives and contributions, this thesis is structured as follows: Chapter 1 offers an introduction to the study, outlining the motivation behind it and a generalized overview of the contributions and workflow. Chapter 2 summarizes the background and provides an overview of existing research in gait analysis, in relation to segmentation, biometric identification, and CIPN as well as other diseases. Chapter 3 talks in detail about the system setup and data collection protocol. Chapter 4 explains the preprocessing steps, which involve data extraction, gait event segmentation, and segment alignment steps necessary to proceed with further processing. Chapter 5 clarifies feature extraction process using Singular Value Decomposition

(SVD) and Discrete Wavelet Transform (DWT) and demonstrates the ability of a Support Vector Machine (SVM) classifier to distinguish between individuals. Chapter 6 shifts towards gait changes between multiple visits using the previously extracted features from DWT with logistic regression, localizing which events experience change and quantifying that change. Chapter 7 provides a summary of the study's findings and discusses the conclusion and future work.

2 Gait Studies: Background and Related Work

Gait analysis is the study of human locomotion, which focuses on the biomechanical and neuromuscular aspects of walking. It involves measuring and evaluating various parameters associated with step cycles, joint movements, ground reaction forces and postural control to assess normal and pathological gait patterns [4]. The ability to measure and analyze these parameters is crucial for mobility assessment and research. Therefore, various technologies have been developed to capture gait data, ranging from optical motion capture systems and pressure sensitive walkways to wearable inertial measurement units (IMUs) which, in contrast to other technologies, provide a portable and cost-effective alternative for gait analysis, by measuring linear acceleration, angular velocity, and earth's magnetic field, enabling gait analysis outside of traditional laboratory settings [5]. A normal gait cycle consists of two primary phases, **stance and swing** as shown in

Figure 2-1. Stance phase begins when the foot makes (initial contact) with the ground and supports body weight, this is also known as **heel strike**, then it continues to the phases where the foot is **flat** and fully in contact with the ground (loading response and midstance), and it ends with **toe off** (terminal stance) where the foot pushes off to transition into the **swing** phase. Breaking gait into these distinct events allows for a more detailed assessment of movement mechanics.

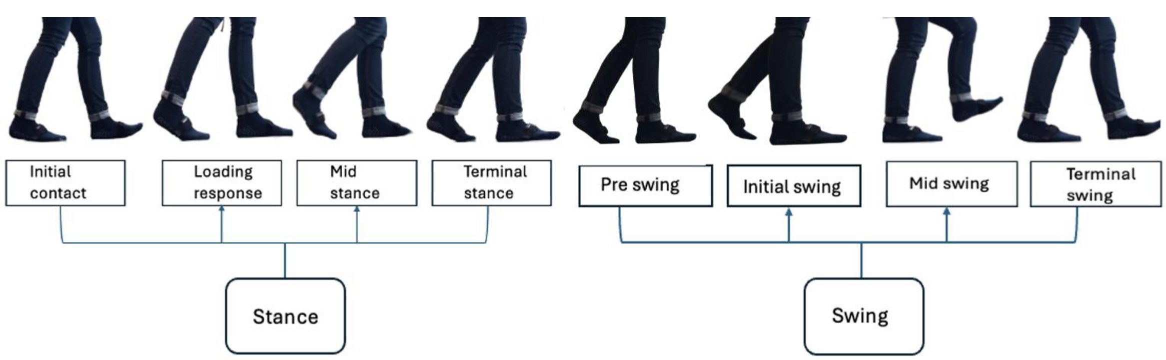


Figure 2-1. The Mechanics of Gait

Therefore, researchers have developed various **segmentation methods** to automatically identify these key events, ranging from rule-based techniques to machine learning and signal processing driven approaches. For example, in [6] the authors applied a threshold-based segmentation and double integration to compute vertical displacement for detecting key gait events. While this method is computationally simple, it's performance is often dependent on individual differences, sensor placement, and walking conditions. An alternative approach relies on plantar pressure, which provides direct insights into foot-ground interactions. In [7], the authors developed a gait segmentation method using a plantar pressure measurement system with custom-made capacitive sensors embedded in insoles, identifying gait events based on pressure shifts at different foot regions. This method is particularly effective for stance-phase detection, as it directly captures contact dynamics, but is limited in detecting swing-phase transitions, which require motion-based sensors like IMUs. To address these limitations, machine learning for segmentation has been explored in [8], where a self-adaptive threshold for heel strike detection is determined in real time via a linear regression model, improving robustness across stride frequencies. However, toe-off detection still relies on a fixed threshold, which may limit adaptability across different walking conditions. The authors in [9] investigated the use of neural networks to mitigate variability in segmentation by combining Dynamic Time Warping (DTW) and Convolutional Neural Networks (CNNs), achieving high accuracy in distinguishing stride from non-stride data and reducing mean absolute errors in event detection. In another study [10], the researchers compared four gait segmentation methods, the first one being a conventional threshold based, the second using Euclidean Dynamic Time Warping (eDTW), which aligns segments of gait signals by minimizing the Euclidean distance between them, improving robustness to variations. The third uses Probabilistic Dynamic Time Warping (pDTW), an extension of DTW that incorporates probabilistic models to refine segmentation, and the fourth is a machine learning-based approach named Hierarchical Hidden Markov Models (hHMMs), which models gait as a sequence of probabilistic states, allowing it to adaptively detect transitions between gait phases. Their findings showed that at the cost of computational complexity. The hHMM approach outperformed the others in handling irregular gait sequences. The discussed segmentation step is essential for many applications in gait analysis; by isolating distinct gait phases, researchers can extract meaningful movement patterns that contribute to several applications. One such application is **gait as a biometric marker for person identification.** Previous studies have shown that individuals exhibit distinct gait signatures that can be classified using machine learning models applied to spatiotemporal parameters or raw IMU-based features, in [11], the authors explored the feasibility of identifying users based on gait patterns recorded via accelerometers in portable devices. By extracting step level features, particularly, correlation, frequency, and data distributions, and applying machine learning models, they were able to demonstrate that gait exhibits consistent and distinctive patterns across individuals. The authors in [12] investigated the use of smartphone IMUs for this purpose, collecting data from 10 healthy subjects walking on a treadmill in controlled conditions, and applying sensor fusion to extract hip joint angles, then applying various machine learning algorithms to classify these subjects based on their hip joint patterns, they achieved an accuracy of 88.9%, demonstrating the feasibility of using hip joint angles for person identification. The study in [13] assessed the robustness of gait based identification under controlled conditions with the use of a multi-sensor dataset collected from 3D IMU signals and 3D optical sensors, using UMAP and DFT to describe characteristics of identity, the results showed that inertial sensors provided consistent performance by defining the fundamental frequency for learning identity at 6-10 Hz, stating that these components were more significant for identity learning purposes than events and other gait related biomechanical movements.

Beyond its applications in biometric identification, gait analysis plays a very crucial role in **healthcare research for mobility impairments** [14], particularly in populations affected by musculoskeletal or neurological disorders, which have been linked to increased risk of falls, loss of independence and diminished quality of life, especially in elderly populations and individuals with neurodegenerative diseases such as Parkinson's [15], multiple sclerosis [16] and peripheral neuropathy [17]. Clinical gait assessment plays a critical role in identifying abnormalities associated with these conditions and monitoring disease progression.

To quantify gait deficits, traditional gait analysis relies on spatiotemporal parameters, such as step length, stride length, step width, cadence, swing time, gait velocity, and others listed in [4]. These parameters are commonly used to assess balance, coordination, and neuromuscular function across various patient groups, allowing for early diagnosis and monitoring of disease severity.

Several studies have demonstrated the importance of spatiotemporal parameters in analyzing gait alterations across different neurological disorders. The authors in [18] have observed four significant parameters for the diagnosis of Parkinson's disease, which are step length, velocity, step width, and step width variability, feeding them into several machine learning algorithms for the purpose of PD diagnosis, as for disease severity, they have found two parameters to be relevant, which are stride width variability, and step double support time variability. In [19] the authors examined spatiotemporal parameters of gait during the 6-min walk in individuals with multiple sclerosis, and in healthy controls, recording velocity, cadence, step length, step width, step time, percentage of gait cycle in double support, and variability of step length and width. They have found that the reduction in velocity over 6 minutes is greater for patients with MS who require assistance while walking. In addition, there was an observed increase in gait variability for these patients, which indicates that the control over walking further deteriorates over the course of the 6-min walk. The authors in [20] aimed to clarify which gait parameters are affected by dementia progression. The meta-analysis revealed that during a single task, most spatiotemporal parameters of gait effectively discriminated between healthy controls and people with dementia. Whereas in dual tasks, only speed, stride length, and stride time variability discriminated between the two groups. While gait alterations in neurological disorders such as Parkinson's are well-documented, peripheral neuropathy presents a unique challenge. Several studies have investigated gait impairments in diabetic peripheral neuropathy (DPN), revealing changes such as lower velocity, shorter stride length, longer stride and stance time, higher maximum knee extension moment, smaller maximum knee joint moment, and higher maximum frontal plane ankle joint moment compared to diabetic controls without neuropathy [21, 22], in addition to lower stride frequency, gait symmetry, and gait intensity [23]. In another study focusing on gait in patients with type 2 diabetes mellitus (T2DM), it was found that gait abnormalities are common among these patients even with the absence of (DPN) [24]. The authors in [25] compared gait patients with (DPN) to healthy controls, their findings showed compensations in gait mechanisms for hip and knee joint movements, shifting the workload from distal to proximal joints, which could indicate distal muscle disfunction. In [26], the authors evaluated gait performance for patients with early (DPN), showing that abnormality in biomedical parameters which have been discussed before, such as low velocity and increased stance and double support phases, were present even in patients without clinical symptoms and signs. **Although diabetic peripheral neuropathy (DPN) and chemotherapy induced peripheral neuropathy (CIPN) arise from different underlying sources,** both conditions involve sensory and motor impairments that may lead to similar gait adaptations. However, research directly comparing the two is limited, and while gait disturbances in DPN have been more extensively studied, CIPN remains an area under exploration for gait research. The authors in [27] used a GAITRite system to obtain spatiotemporal parameters of gait from elderly participants with breast or colorectal cancer, to which statistical analysis was applied. They have found that these patients experience shorter step length and decreased velocity, with other gait parameters experiencing no notable difference. Similar findings were discovered in [3] where IMUs were used to collect gait and postural sway data from older female cancer patients with PCA being used to identify the essential domains of mobility from 15 measures. Five domains were identified and from those domains it was suggested that past fallers had a shorter stride length and slower pace. In [28], user gait data was collected from IMUs and pedometers present in their smartphones, which they were instructed to keep in their pockets, exploring the feasibility of tracking CIPN via a smartphone app. The sensor fusion algorithm from IOS was used and like the previous study they have found that patients with CIPN have a shorter step length, their findings also suggest that their sway acceleration patterns were unique while performing a walking task. While the study in [29] was focused on postural control in CIPN instead of gait, the authors have interestingly found that while CIPN symptoms experienced decline during the course of chemotherapy, postural control improved within 6 months after completion, suggesting that some aspects of motor function can recover over time.

In summary, prior studies have demonstrated the potential of IMUs for multiple purposes including biometric identification and several healthcare related fields, but there remains a vital need in exploring these areas further, particularly for monitoring the gait of patients experiencing CIPN, which remains an area under study.

3 Participants, Instrumentation Systems, and Data Collections

This section provides a comprehensive overview of the sensors and data collection systems used to capture each participant's gait data, with a brief description of the participants' metadata. The focus will be on inertial measurement units (IMUs), describing their hardware specifications, placements, and rationale for use. Pressure mats are briefly covered which was employed to mark gait events. Although its data is not the focus of subsequent analysis, by describing the type of data and the mechanisms to manage it, we lay the groundwork for the motivations and methodologies explored in the chapters that follow.

3.1 Participant Metadata

The study included 34 female cancer patients who were diagnosed with breast or gynecological cancer. All participants were able to stand and walk independently without support or assistive devices. The mean age was 63 ± 10 years. While the group primarily consisted of middle aged to older adults, only three participants were under the age of 50 (ages 32, 35 and 45) allowing for some age diversity within the dataset. Weight was also measured, and it varied across patients, ranging from 114.3 to 294.8 lbs. Several participants had weights exceeding 240 lbs, while others were under 130 lbs, reflecting a diverse range of body compositions. Participant age and weight distributions are shown in Figure 3-1

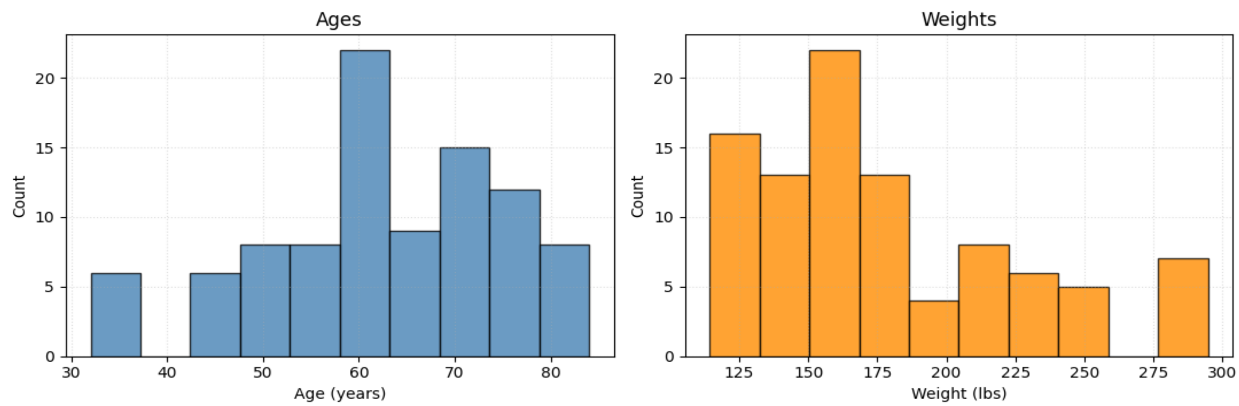


Figure 3-1. Histograms of Weights and Ages Recorded Per Visit for Each Participant

The first data collection visit occurred before the initiation of chemotherapy. A small subset of participants began the study with self-reported signs of peripheral neuropathy that is either diabetic, or chemo-induced from previous chemotherapy treatments. A separate subset appeared to develop neuropathy with the progression of chemotherapy, and a small group showed no evidence of neuropathy during the study period based on clinical evaluations. Clinical outcomes were not directly analyzed in this thesis but may serve as important links for future investigations. The participants had a number of visits ranging from one to six, during each of which they were asked to walk along a designated pathway at their own comfortable pace, reflecting their typical walking pattern. 31 of the participants were also able to perform the additional task of walking at a faster -but still self-selected- pace in their baseline visit. For each of the tasks mentioned, the participant completed 4 to 8 passes, depending on the physical abilities and well-being of the patients. This study was approved by the institutional review board (IRB) at OUHSC under protocol #11473. All participants provided consent prior to participation.

3.2 Instrumentation and System Setup

The data collection setup in this study mainly comprised of two measurement systems. The first is a pressure sensitive walkway (Strideway version 7.8, Tekscan Inc., Norwood, MA, USA) which collects foot pressure data at a sampling rate of **50 Hz**. This mat was only used as an initial reference for the identification of gait events such as swing and heel strike in the work presented in this thesis, but its data was not integrated into the final analysis.

Figure 3-2.System Setup [30,31]

The second system, which will be the main focus of this work, contains six IMUs, secured to the feet, ankles, lumbar, and sternum of each participant to capture full body gait dynamics as illustrated in Figure 3-2. Further details about IMUs are provided in Section 3.2.1.

3.2.1 Inertial Measurement Units (IMUs)

IMUs are electronic sensors that combine triaxial accelerometers, gyroscopes, and magnetometers, which capture linear acceleration, angular velocity, and earth's magnetic field respectively, making them 9-axis IMUs. They are often used in inertial navigation systems and have been widely applied in fields such as robotics, aerospace, sports science, and biomedical engineering. In the case of clinical gait analysis, IMUs are popular for their ability to provide motion data in real-world settings, without the need of a laboratory structured, controlled test, they are also easy to wear and inflict no limitations on the movement of the human body.

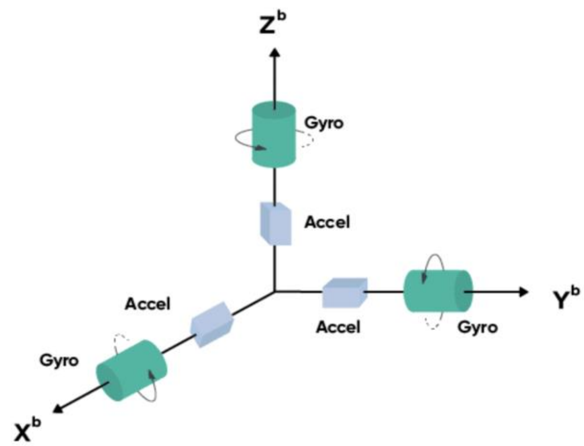


Figure 3-3. Accelerometers and Gyroscopes in the Three Axes of Movement [33]

The theoretical basis of IMUs lies in Newtonian laws of motion which state the movement of a body is uniform and linear unless an external force is acting on it, and that the force excreted on a mass will produce a proportional acceleration [35]. **Accelerometers** have many lab and industrial types, involving piezoelectric, fiber optic, capacitive, strain gage and piezoresistive, but the ones that are most commonly used in wearable devices are MEMS (micro-electro-mechanical systems) due to their compact size, low power consumption, and reliable performance in measuring human motion. These accelerometers have a reference system, with a small mass attached to it via a spring, this setup keeps track of the movement of the mass using differential capacitors and other electronic components, allowing them to measure the speeding up or slowing down of anything the device is attached to. When the accelerometer is standing still, the mass provides a certain baseline capacitance indicating no acceleration, then when movement occurs, the capacitance changes and is measured electronically and adjusted for accuracy [33]. This setup is shown in Figure 3-4.

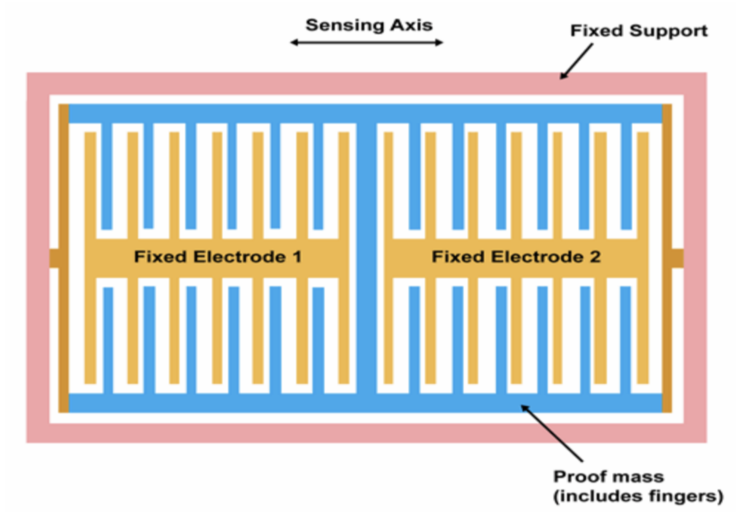


Figure 3-4. Accelerometer Principle [34]

These sensors are prone to the gravitational effects of the earth, meaning that an IMU in free fall would produce zero acceleration in the vertical axis, because it is accelerating at the same rate as earth's gravity. Conversely, when the sensor is floating "stationary in free space", it measures a vertical acceleration of $-9.81\ m/S^2$, a value that needs to be removed from the original output to ensure it is accurately representative of movement in free space [34].

Gyroscopes measure how fast an object is rotating and in which direction (Roll, Pitch, Yaw). As mentioned above regarding accelerometers, gyroscopes also have several forms including mechanical, ring laser gyroscopes (RLG), fiber-optic gyroscopes (FOG), and quartz/MEMS gyroscopes which are industrially the most popular. They rely on Coriolis effect, which occurs in rotation reference frames when an apparent deflection is experienced by an object moving withing a rotating system, causing a measurable force proportional to the object's velocity and the rotation rate, which is used to detect angular motion. In practical terms, gyroscopes have a reference frame with a resonating proof mass connected to it by mechanical springs. The first reference frame is attached to an outer reference frame, and when rotated this generates a secondary vibration perpendicular to the drive axis, known as the sense axis. Most commonly, the measurement is derived when rotation causes a change the differential capacitance between the inner and outer reference frames [33].

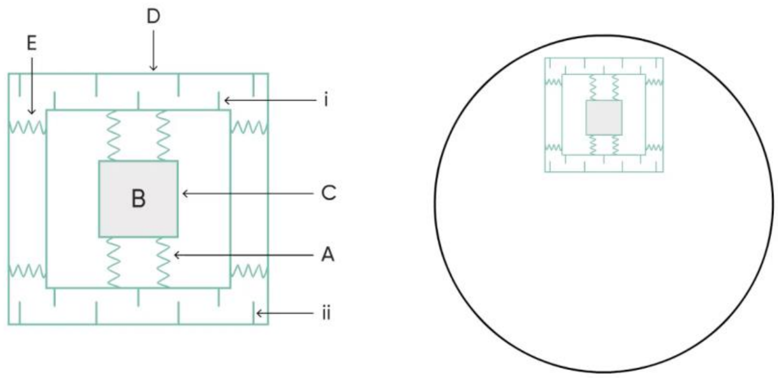


Figure 3-5. The Operation of a MEMS Coriolis Gyroscope [33]

MEMS gyroscopes tend to exhibit a drift over time, meaning small measurement errors can accumulate during prolonged motion tracking. This drift results from biases and noise within the sensor and is significantly lower in higher-end technologies like fiber-optic gyroscopes (FOGs), though at the expense of size and cost.

Magnetometers are sensors typically used in some IMUs to find their orientation with respect to the earth reference frame. They have several types, each based on a different principle of magnetic field interaction, including the Hall Effect Magnetometer, which involves generating a voltage proportional to the strength of the magnetic field perpendicularly applied to it, Magneto-Induction Magnetometers which measure how magnetized a material gets when affected by an external magnetic, and Magneto-Resistance Magnetometers which measure the ability of a certain object to alter electrical resistance when it is in the presence of an external magnetic field [33].

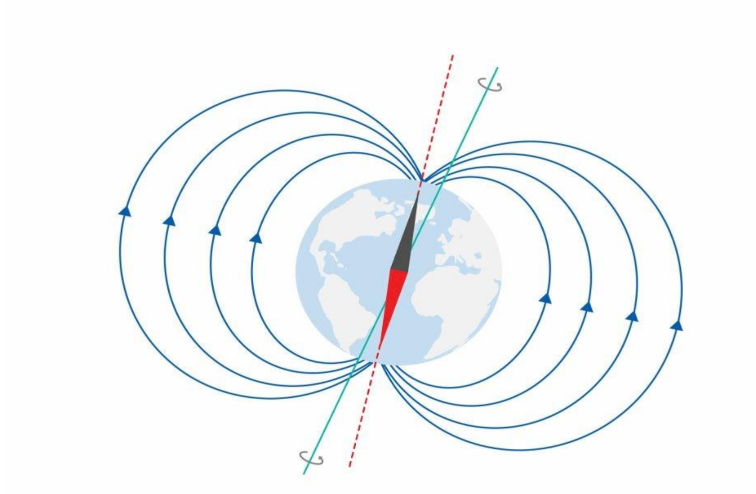


Figure 3-6. Magnetometer Principle [33]

IMUs without magnetometers are preferred in scenarios where concerns about magnetic field interference are present, or when rudimentary information about orientation serves the purpose of use. While they are beneficial in the sense of improving the stability of finding orientation changes, they need to be calibrated prior to installation in order to compensate for the interference mentioned earlier in order to avoid errors [33].

As mentioned earlier, this study utilized APDM opal IMU sensors (APDM, Inc., Portland, OR, USA) for gait data acquisition, these are compact, wireless 9-axial IMU sensors widely validated in both clinical and research settings for motion analysis, this system offers the following specifications demonstrated in Table 3-1:

Table 3-1 APDM Opal Specifications [32]

	Accelerometer	Gyroscope	Magnetometer	Barometer
Axes	3	3	3	1
Range	+/- 200 g	+/- 2000˚ /s	+/- 8 Gauss	300 - 1100 hPa

Noise Density	5 mg/√Hz for > 16 g 120 μg/√Hz for ≤ 16 g	0.025˚ /s/√Hz	2 mGauss/√Hz	1.3 Pa
Sample Rate	20 - 128 Hz	20 - 128 Hz	20 - 128 Hz	20 - 128 Hz

In this study, the highest sampling rate offered by this device was selected (**128 Hz**), in order to obtain the most nuanced information possible to ensure accurate analysis.

These sensors offer standard configured setups for motion analysis (Raw Movement Data, Gait and Balance Analysis, and Full Body Kinematics). The setup selected in this study was Gait and Balance Analysis, shown in Figure 3-7:

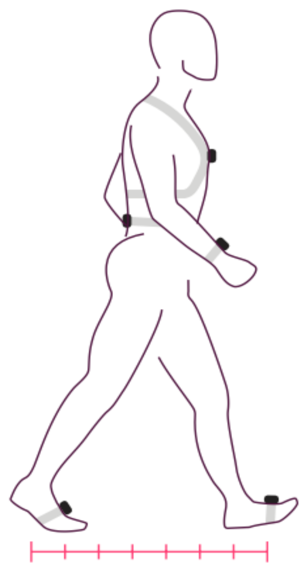


Figure 3-7. Standard APDM Opal Setup for Gait and Balance [32]

This setup was modified to match the one shown in Figure 3-2 in Section 3-2, where the wrist sensors were moved to the ankles, for the aim of collecting maximum information about the lower extremities, while maintaining the two upper sensors since this placement contains balance information.

3.3 Summary

This chapter started by describing the population from which gait data was collected (female cancer patients undergoing chemotherapy), as well as brief information of their metadata followed by the instrumentation and system setup used for collecting and processing the gait data of the study. The overall setup was discussed that comprises of IMU instruments and a pressure mat. However, the pressure mat was mainly used as a primary step in segmenting IMU data, but it was not included in further analysis. We later transitioned into detailing the functioning and theoretical principles of IMU instruments, including the type of sensors they contain (accelerometers, gyroscopes, and magnetometers), with a brief explanation of how each one works and what type of data they collect. We then briefly presented the specifications of the IMU system selected for this study (APDM Opal), The system's technical specifications were summarized, including its 9-axis sensing, high sampling rate of 128 Hz, wearable wireless design, and its ability to support a range of kinematic configurations. By establishing the characteristics and rationale for the sensors and configurations used, this chapter laid the technical foundation for the processing, segmentation, and feature extraction methods introduced in the subsequent chapters.

4 IMU Data Processing: Alignment and Outlier Removal

In this chapter, we describe the methodology used to extract, align, and prepare gait segments from raw IMU data for development of models that capture the participants' gait features. The process begins with segmenting specific gait events, namely flat foot, heel strike, and swing combined with toe-off. This enables accurate labeling of the unlabeled walking data by identifying patterns that closely resemble known event templates. Following segmentation, each extracted segment undergoes alignment to ensure consistency in timing and shape across steps as well as participants.

4.1 Data extraction

The raw gait data is originally stored in hierarchical file format (.h5), with each file corresponding to a specific walking condition for a given participant and visit. A Python script was developed to systematically extract and organize data from all relevant sensors (feet, ankles, lumbar and sternum), capturing triaxial accelerometer, gyroscope, magnetometer, and quaternion orientation signals. The script performs visit-wise organization, by matching the date of the raw data file, with the date of the corresponding CSV file containing the information. It then extracts the .h5 file into multiple separate CSV files, each of which is labeled according to the walking condition specified inside the corresponding CSV file and the position of the IMU sensors placed over the body is labeled according to the given number of the sensor. This step was necessary to ensure a uniform file structure across all participants and walking conditions before moving forward with further analysis.

4.2 Template Extraction and Matching

In this work, the segmentation method is based on matching templates with the raw data using Euclidean distances. Explicitly, representative kinematic data representing known gait events (Swing, Heelstrike) are established, referred to as templates, by synchronizing the IMU with the TekScan according to their timestamps as shown in Figure 4-1. The IMU at 128 Hz and TekScan at 50 Hz were synchronized using their Unix epoch timestamps recorded by each system. For the IMU, timestamps were directly available at the sample level. For the TekScan data, each frame had a corresponding date-time string that was parsed and converted to Unix time. Each IMU sample was then matched to the closest TekScan frame in time without temporal resampling or interpolation. This introduces a maximum timing uncertainty of $\pm 10\ ms$, corresponding to half the TekScan sampling interval.

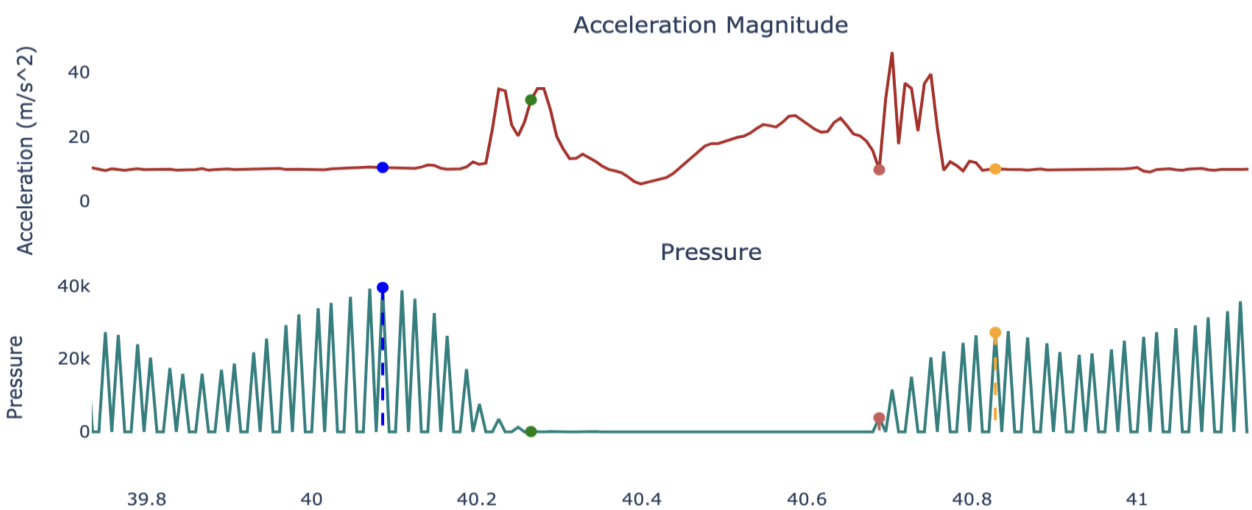


Figure 4-1. Synchronization of Pressure Mat with IMU

The labeled template segment extracted from the above method is slid across the x-axis raw unlabeled acceleration signal one point at a time. At each point, the Euclidean distance is calculated between the template and the matching-length window of the unlabeled signal. This process continues across the entire signal producing a distance profile that reflects how closely the template matches different parts of the data, as shown in Figure 4-2.

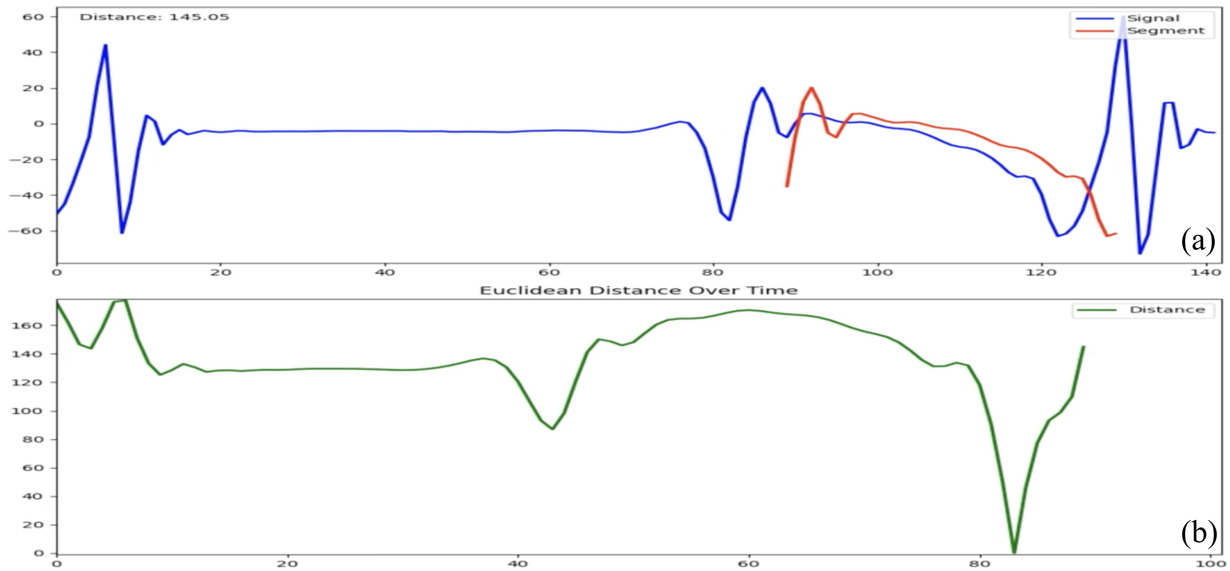


Figure 4-2.(a) Template in red is slide across unlabeled acceleration signal; (b) Calculated Euclidian distance

To detect a heel strike, the local minimum at each slice in the distance profile is found and identified as a potential heel strike event. These are points where the heel strike template matches the unlabeled data most closely. Moreover, the flat regions in the profile, are labeled as a flatfoot event since the distance profile doesn't vary significantly for a portion of datapoints. As for the swing event, a non-fixed threshold is set to identify it for each datafile. This threshold is based on the distance profile following an initial flat region, assuming the gait trial starts from a standstill position for a period of time before the subject starts walking. By analyzing a window of data that follows the flat region, the minimum distance within this window is selected as the threshold for that file. Swing events are then identified as points in the distance profile that fall on or below this threshold. The need for starting from a flat region stems from the variability in the placements of the IMU sensors across participants. Despite efforts to place the IMUs in the exact same position on the body for each participant and across sessions, slight variations in orientation are unavoidable due to the practical limitations of manual placement and physical variations among people. These differences might cause variations in the x-axis acceleration values even for the same participants but during different visits. The overall gait acceleration magnitude roughly takes on the same pattern for all participants. As a result, a single fixed threshold would be ineffective for identifying swing events across different participants and visits. Therefore, the threshold was made adaptive to account for variations in IMU orientation and to mitigate errors in swing event detection.

To summarize the process, identified events are labeled in the original data. Flat foot regions are labeled based on minimal variation in the distance profile, while heel strikes are marked according to local minima. Swing events are labeled when the distance profile falls on or below the adaptive threshold added to address acceleration variations due to participant-specific sensor placements and/or gait patterns. When labeling gait events, the order of events (labels) is closely observed to reduce mislabeling and detection errors. Flat foot tends to be the most stable and less likely to be confused with other events. Heel strike can be reliably found using the local minima in the distance profile. Once these two gait events are correctly identified, they effectively carve out the regions where swing should exist, reducing the potential of erroneously labeling swing and toe-off events and/or the chances of accidentally labeling parts of the acceleration signal that did not represent the actual event.

Since toe-off and swing occur successively and are often the events that show the most natural variation with less noise compared to heel strike and flat foot, I chose to group them into a single event segment for the rest of the analysis. Figure 4-3 presents the labeled data obtained after applying the above-described process.

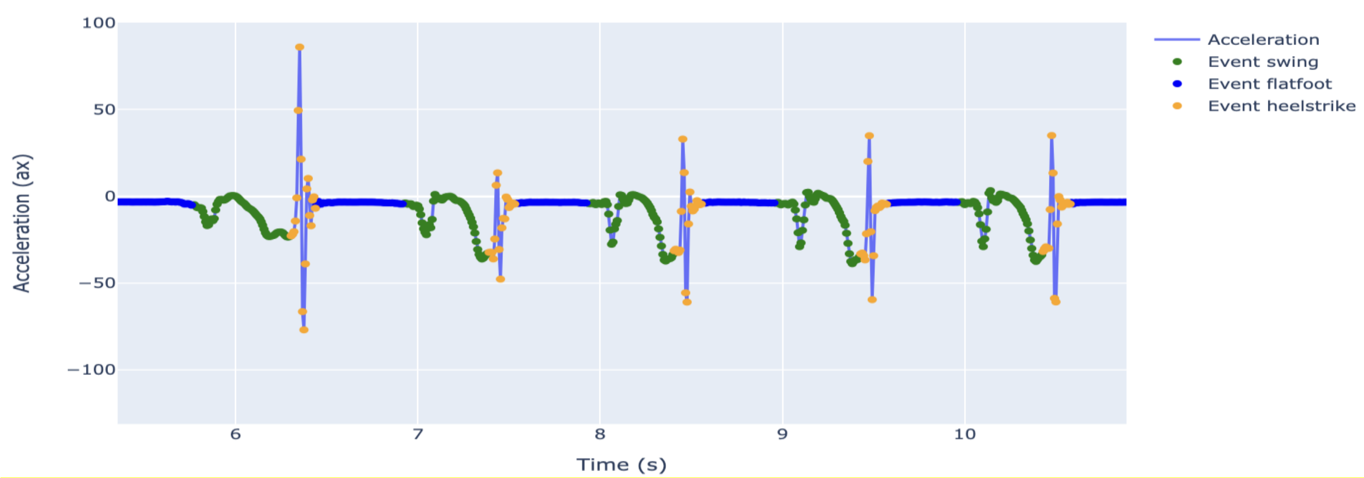


Figure 4-3. Labeled x-axis Acceleration Data

To improve the accuracy of the segmentation, several post-processing steps were applied. First, events that were too short to be meaningful were unlabeled, since these were likely due to noise or brief fluctuations that did not represent real gait activity. Additionally, transition between events, such as flat foot to swing to heel strike, were logically consistent and helped refine the accuracy of gait event identification.

The following part demonstrates the pseudocode for the explained segmentation algorithm:

Input: Labeled IMU data (acceleration: ax, ay, az), Unlabeled IMU data.

Output: Labeled IMU raw data (events: heel strike, swing, flatfoot)

- Load labeled and unlabeled IMU data.
- Define event extraction function:
  - For each labeled event (e.g., heel strike, swing):
    1. Find indices where the event is labeled.
    2. Extract corresponding segments from the signal.
- Extract heel strike and swing segments using event extraction function.
- Define Euclidean distance calculation:
  - For each labeled segment:
    1. Slide the labeled segment across the unlabeled data.
    2. Calculate Euclidean distance for each segment position.
    3. Store the resulting distance profile.
- Calculate Euclidean distances for all labeled segments (heel strike and swing).
- Define heel strike detection function:

- Identify local minima in the Euclidean distance profile between two flat regions.
- Mark these minima as heel strike events.

- Define swing detection function:
  - Identify non-flat regions in the distance profile.
  - Set a threshold based on the minimum value in the first non-flat region.
  - Mark points on or below this threshold as swing events.
- Process each file:
  - Load unlabeled IMU data.
  - Calculate Euclidean distances for heel strike and swing segments.
  - Detect heel strike events (using local minima).
  - Detect swing events (using threshold).
- Label events in the unlabeled data:
  - Initialize all labels as "unknown."
  - Label flat regions as "flatfoot."
  - Label detected heel strike and swing events.
- Post-process labels:
  - Remove short events.
  - Ensure consistency between flatfoot, heel strike, and swing labels.

4.3 Segment Alignment

Alignment of gait segments is a critical step to ensure improved results. This process was designed to enhance the comparability of the data across participants, visits and/or walking conditions. Initially, triaxial acceleration data from the IMUs was converted into a single magnitude value to provide a consistent measure of acceleration across all axes and eliminate the effect of unavoidable inconsistency of sensor placement across patients and different visits. The magnitude (M) of the acceleration vector was calculated for each time point using the following formula:

$$M = \sqrt{a_x^2 + a_y^2 + a_z^2}$$

where a_x, a_y, and a_z represent the acceleration components along the x, y, and z axes respectively. Event magnitude creates a sort of reference frame to be used to align other event segments extracted from different steps taken by the same or different participants. The longest extracted segment is designated as the reference frame segment for alignment. Each shorter segment is aligned to this reference by calculating the Euclidean distance between the segment and corresponding windows of the reference segment. The alignment that minimized the Euclidean distance is selected, ensuring that the segments are optimally aligned whilst maintaining their original shape. After alignment, all segments are padded with zeros either to their right, left, or on both sides to a target length of 90 samples which was approximately equal to the longest observed segment across the dataset (88 samples), this enables matrix operations and subsequent SVD analysis which requires all segments be the same length before they are placed into a matrix. Since the magnitude of each segment begins with an acceleration value close to 10 m/s^2 due to gravitational forces as mentioned in Chapter 3-Section 3.2.1, a mag shift was applied to correct for the gravitational force and remove large discontinuity between the zero-padded areas and the actual data. The segments were adjusted such that their minimum value became zero, ensuring a smoother transition between the segments and added zeros.

Finally, outlier removal algorithm was performed using the Interquartile Range (IQR) method. The first quartile (Q1) and third quartile (Q3) were calculated for each time point across all segments. The IQR was then determined as the difference between Q3 and Q1. Segments were considered outliers and removed if more than 15% of their samples fell outside 1.5 times the IQR from Q1 to Q3, as defined by:

Lower Bound = Q1 - 1.5 × IQR

Upper Bound = Q3 + 1.5 × IQR

Segments laying within these bounds are retained and included in the analysis. The 15% threshold retained approximately an average of 86% of the segments per patient, ensuring it was not too strict.

The alignment and outlier removal algorithm described in this chapter is an initial procedure which was applied to datasets corresponding to each individual visit's data. A more advanced alignment strategy is required to enable cross-subject comparisons for the person identification task which will be described in the next chapter. Alignment among participants forms the foundation for analyzing gait as a unique biometric signature across multiple individuals.

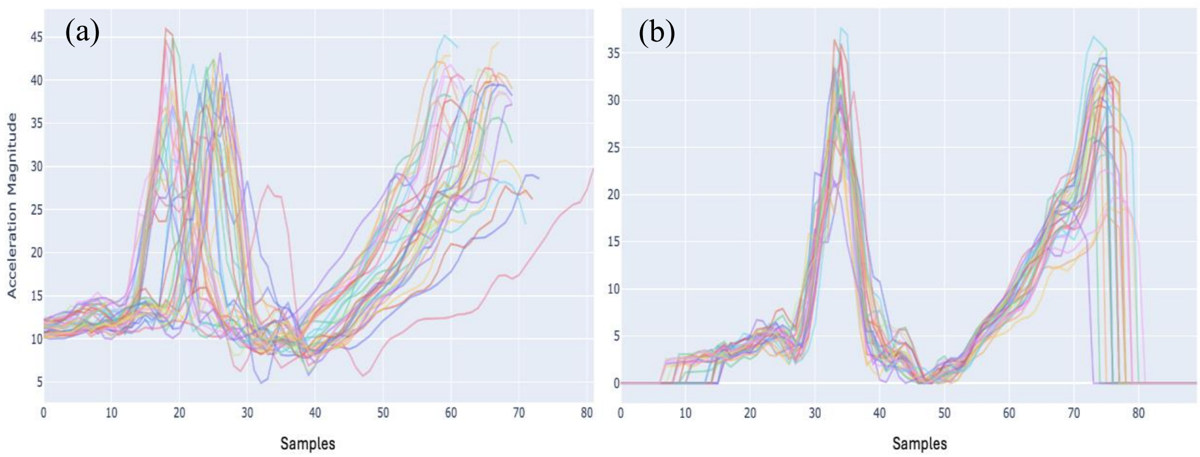


Figure 4-4. (a) unaligned swing segments (b) Aligned using minimizing Euclidean distances among swing segments capture during one visit

The pseudocode for the described alignment algorithm is as follows:

1- Locate segments based on “swing” label from the IMU data.
2- Calculate magnitude of acceleration.
3- Shift each segment so that its minimum is zero.
4- Align segments:
    - Find the longest segment.
    - Align shorter segments by sliding them over the longest to minimize Euclidian distance.
5- Pad all segments with zeros to match the target length (90 samples).
6- Apply interquartile range (IQR) to remove segments with excessive outliers.
7- Shift segments down so the minimum value for each is zero

4.4 Summary

To Summarize, this chapter briefly discussed data extraction, then detailed alignment and outlier removal algorithms that are essential for later analysis described in this thesis. Data extraction was developed to identify main gait events (Toe-off and Swing, Heel Strike, Flat Foot). First, the time indices of gait cycles captured by the pressure mat are found. The matching indices of the synchronized IMU data are then identified. As a result, IMU gait templates were extracted to represent gait events; the templates were slid over the raw IMU data and Euclidian distances were calculated to find the best match to detect the desired event. A dynamic thresholding process was applied to minimize detection errors for the swing segments. To ensure consistency across steps and participants, segments were aligned using a sliding window method that minimized Euclidean distance relative to a reference segment. Zero-padding and magnitude shifting were applied to unify segment lengths and reduce discontinuities caused by gravitational acceleration. Finally, outlier segments were identified and removed using the Interquartile Range (IQR) method to enhance data quality. This structured preprocessing ensured that the data fed into the feature extraction models was clean, consistent, and representative.

5 Modelling Gait Data for Biometric Identification

Gait analysis has been recognized as a potential biometric for person identification due to the uniqueness of an individual's walking pattern. Recent advances have focused on using wearable sensors to perform gait analysis in more realistic settings, allowing for continuous monitoring and greater flexibility.

In this chapter, I examine the ability to uniquely identify individuals walking slow or fast pace by extracting features from their gait acceleration data and using machine learning in order to classify these individuals based on data collected during their first (baseline) visits.

5.1 Cross-Participant Alignment for Subject Classification

Once swing, toe-off segments were extracted, cleaned, and aligned within each trial (intra-visit) as described in the previous chapter, some additional alignment steps were required across participants in order to enable person-based identification analysis.

A per-visit average segment was computed from all segments of each person's data. These averages were then aligned to a global reference segment, computed as the mean across all visit averages. Each individual segment was then re-aligned to its own visit's calculated aligned average segment using the custom shifting and padding routine that minimizes Euclidean distance while preserving the original shape.

The zero padding at the start and end of each segment to achieve constant segment length were replaced with values obtained from the same start and end regions of the reference segment. This was done to mitigate the sudden value jumps introduced by zero-padding—a process that might introduce high signal spectrum not existing in the original unpadded data.

Replacing the zeros aims to strike a balance between intra-visit alignment and alignment across participants, to enable fair and meaningful comparison of gait patterns across individuals using a reference averaged event pattern.

The result of this process for left-foot normal pace is demonstrated in *Figure 5-1.*

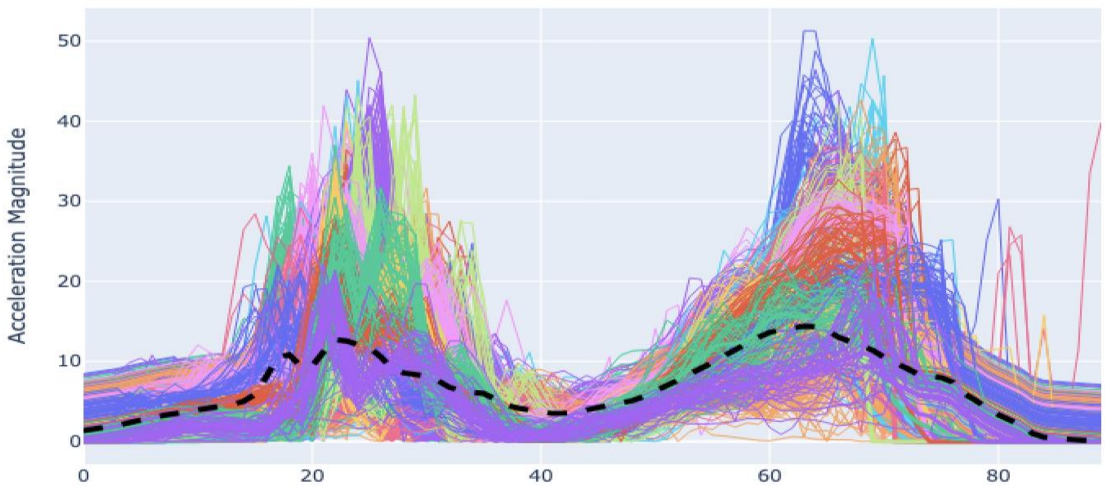


Figure 5-1. Aligned Segments Across Different Participants

5.2 Feature Extraction and Description

Singular value decomposition (SVD) and wavelet transform were both applied to the aligned magnitude acceleration as complementary techniques to extract the most informative features from the gait data. While SVD provides global energy distribution, wavelet decomposition identifies energy localized to a particular time-frequency scale. We first employed SVD to the data matrix, X, which contains aligned swing magnitude acceleration segments obtained from all participants and was decomposed as:

$$X = U\, \Sigma\, V^T$$

Where V^T contains the right singular vectors that represent the principal directions of variation in the data. By projecting each segment onto the first five rows of V^T, we created a feature matrix where each row corresponds to a gait segment and each column represents its projection onto a specific singular vector. Through this process, the aim was to create a lower dimensional space whilst ensuring that the values retained are most indicative of personal differences in gait.

Figure 5-2 shows singular values and their corresponding energy contributions, with the first one capturing approximately up to 45% and the following four capturing around 14%, 7%, 5% and 4% respectively.

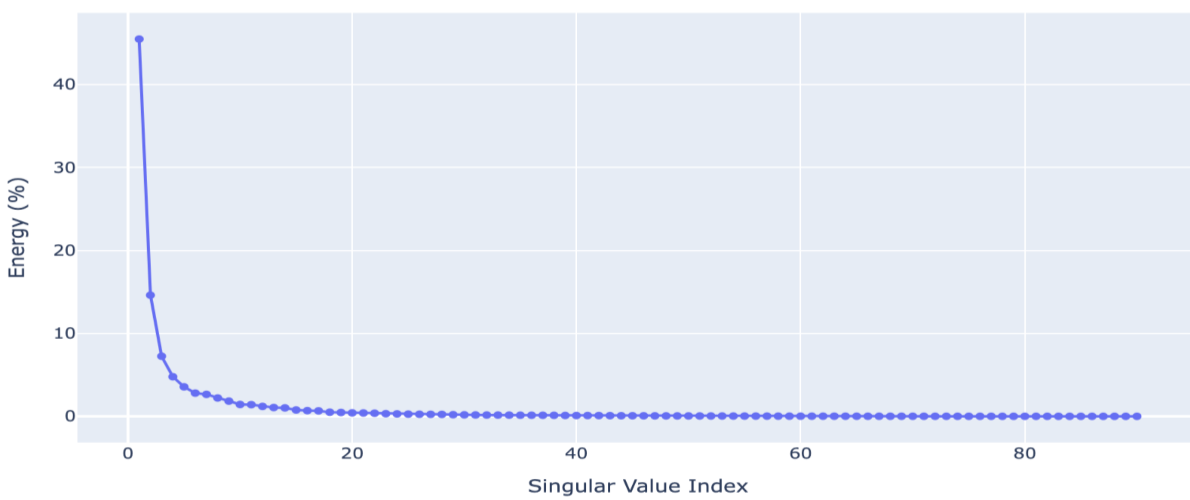


Figure 5-2. Singular Values Energy Contributions

To visualize the variability that SVD captures between individuals, the vectors formed by these five projections for different participants were plotted, as illustrated in Figure 5-3. These line plots show distinctions between participants, indicating that the projections effectively create unique patterns for each individual across the five dimensions. Participant-specific variability is evident, and the projections reveal personalized gait signatures, suggesting that even subtle differences between participants are captured in this high dimensional space.

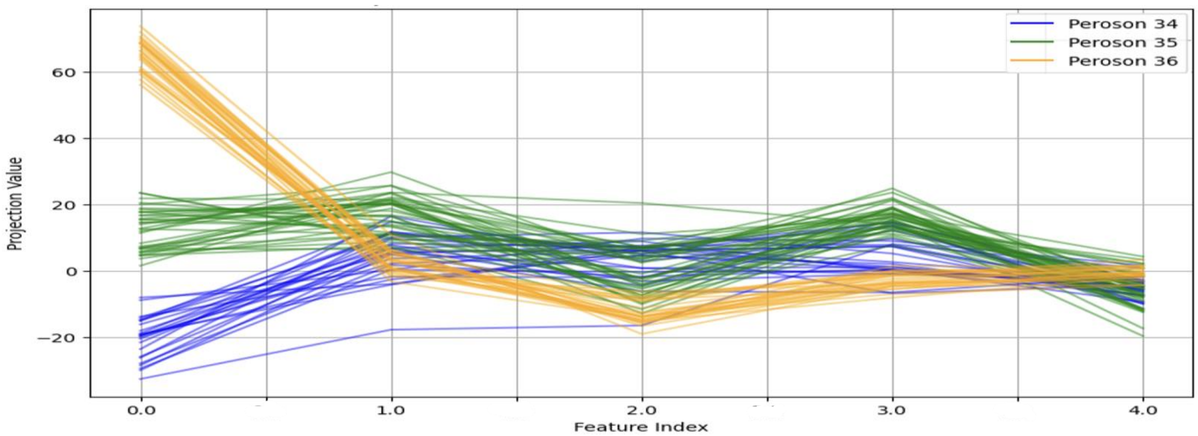


Figure 5-3. Example of SVD Projections for Three Different Participants

The second feature matrix is obtained by applying a Discrete Wavelet Transform (DWT) which captures multi resolution information in both time and frequency domain. Wavelet decomposition yields two types of coefficients: approximation coefficients (low frequency components), which represent the overall trend or slow-changing parts of the signal, and detail coefficients (high frequency components) which capture sudden changes or sharp transitions in the signal. The wavelet coefficients reflect how the frequency content of the signal changes over time, providing a richer understanding of gait patterns. More detail on this will be covered in Chapter 6. Initially, different wavelet types (Db, Haar, Symlets), and different decomposition levels (1, 2, 3) were explored experimentally. The selection of the optimal parameters was done using a grid search cross validation for multiple iterations to find the ones that yielded the highest model performance. Among the candidates, db4 at level 2 was the most consistently selected with no edge effect warnings arising and was utilized for the remainder of the analysis. Once wavelet transform is applied, the resulting coefficients are flattened and combined with the SVD projections to form the final feature matrix, which incorporates dominant spatial variations, temporal structure, and frequency components of the gait data. An example of the vectors resulting from the extraction and flattening of Discrete Wavelet Transform Coefficients can be seen in *Figure 5-4*

Figure 5-4. Example of Wavelet Coefficients Flattened into Vectors for Three Different Participants

As can be observed, these coefficients can be categorized into three subsets. The first one has the highest coefficient values (~0-25) and represents a low frequency range known as the approximation coefficients. As seen in the figure, they capture the low frequencies generated by convolving the db4 wavelet filter with our signal, and it captures the two main peaks in our segments which correspond to toe off and end of swing. Similarly, detail level 2 coefficients (~26-50) have smaller values and a higher number of coefficients, these represent the middle frequency band existed in the data; we can perceive that they are also divided into two smaller divisions, capturing the frequencies in the two peaks. Lastly, detail level 1 coefficients (~51-95) represent the highest frequency band, and has the smallest values of coefficients.

5.3 Feature Selection and Refinement

To improve the performance of the model, a refinement process was implemented to minimize redundancy in the feature matrix. This approach utilized both correlation and permutation importance, applied in tandem to ensure that only the most informative and non-redundant features were retained.

The process began by calculating the correlation across all features in the combined feature matrix (derived from both SVD and wavelet transformations). By examining pairwise correlations, pairs of features were identified with a correlation coefficient $|r| > 0.6$, indicating high mutuality. To avoid redundancy, we removed any feature correlated with another (its pair), as determined by permutation importance analysis. In addition, coefficient representative of the segment edges were removed, as they can be artificially influenced by the implemented zero padding process, when the wavelet filter is convolved with this part of the signal.

To quantify each feature's contribution to model accuracy, permutation importance evaluates how much each feature impacts the model by randomly shuffling its values and calculating accuracy reduction. Features that resulted in minimal or no accuracy loss were deemed less informative and subsequently removed. *Figure5-5* displays a bar plot of permutation importance for a subset of the top-ranked features, illustrating the diverse impact of various features on classification accuracy.

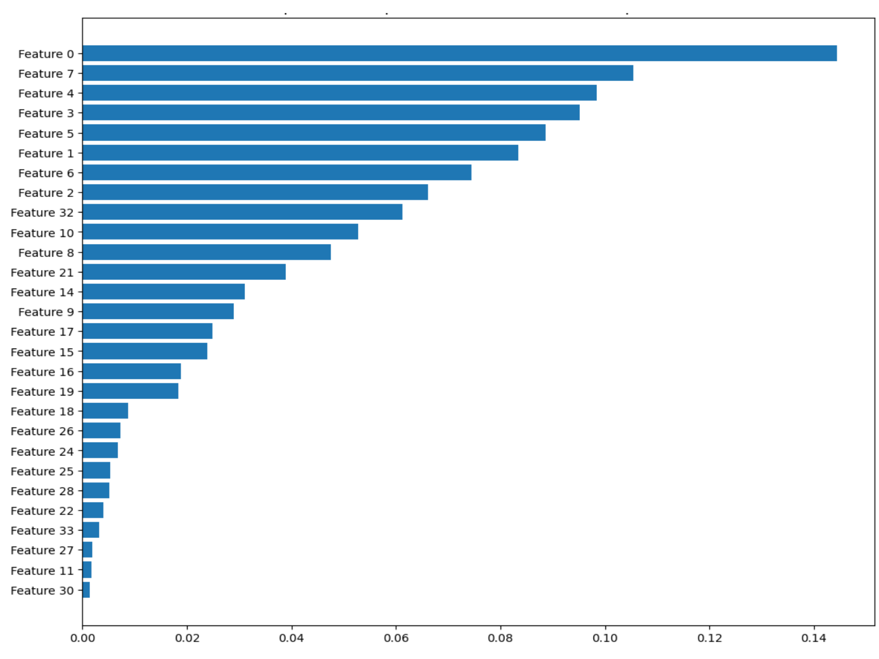


Figure 5-5. Bar Plot of Permutation Importance for Top Feature

Following the removal of redundant and low-important features, the remaining features were standardized to ensure that all feature values were scaled to unit variance. This standardization step mitigated any disproportionate influence from features with larger inherent variance, enhancing the stability and generalizability of the model. Hence, the final refined feature matrix retained only the most relevant components, balancing spatial, temporal, and frequency information extracted from the gait data. *Figure 5-6* illustrates the feature matrix post-refinement, demonstrating the reduced dimensionality and standardized values.

Figure 5-6. Line Plot of Standardized Features Post-Refinement

5.4 Classification of Individuals

For the classification stage, a Support Vector Machine (SVM) model was used to distinguish between individuals based on their extracted features. The combined feature matrix from SVD projections and wavelet coefficients is split into training and testing sets using a 70-30 random train-test split. To ensure a robust classification model and mitigate overfitting, a nested (hybrid) cross-validation approach was employed using a Support Vector Machine (SVM). Specifically, the outer loop followed the Monte Carlo cross-validation strategy (repeated random sub-sampling). The dataset was randomly split into training (70%) and testing (30%) sets over 200 independent iterations to ensure that the performance estimates were stable and minimally impacted by any particular random partitioning, as it was observed that the mean accuracy converged by this point. Adding more iterations did not significantly change the average performance. For each iteration, the dataset was randomly split into a training set (70%) and a testing set (30%). The SVM hyperparameters were then tuned inside the training set (the inner loop) by applying GridSearchCV with 5-fold cross-validation to the training set to find the best parameters, where for each hyperparameter combination in the grid, it will train on 4 folds and validate on the 5th, generating two key metrics:

- Mean test score: the average accuracy achieved across these five folds.
- Std test score: the standard deviation across the folds, indicating how consistently the model performs.

Table 5-1 illustrates one example grid search result:

Table 5-1. Cross Validation Results for One Iteration, L-foot, Normal Pace.

Combination index	mean_test_score	std_test_score
0	0.966232	0.011850
1	0.967845	0.011387
2	0.966258	0.012823
3	0.966245	0.010669
4	0.292645	0.019333
5	0.291032	0.016225
6	0.265303	0.017323
7	0.289419	0.016611
8	0.964645	0.012002
9	0.966245	0.010669
10	0.966258	0.012823

Higher mean test scores combined with lower standard deviations highlight the most reliable and well-generalized hyperparameter settings. After identifying the best hyperparameters in this inner loop, we retrained the SVM on the entire 70% training split and evaluated on the remaining 30% hold-out.

The results in Table 5-2 summarize the average accuracy, precision and recall for each foot in both conditions over all given iterations.

Table 5-2. Classification Metrics

	Normal-Lfoot	Normal-Rfoot	Fast-Lfoot	Fast-Rfoot
Accuracy	86.97%	82.96%	83.55%	82.37%
Precision	88.36%	84.35%	85.12%	84.11%
Recall	86.97%	82.96%	83.55%	82.37%

These results demonstrate satisfactory classification accuracy, particularly for the normal walking condition, while faster walking proved a bit more challenging. *Figure 5-7* shows the confusion matrix heatmaps corresponding to the above classification results, with the true labels on the y-axis and the predicted on the x-axis, the diagonal cells represent true positives while the off diagonals show false positives. The intensity of colors over the diagonal shows the high number of true predictions, while misclassification appears to have much fewer occurrences in comparison.

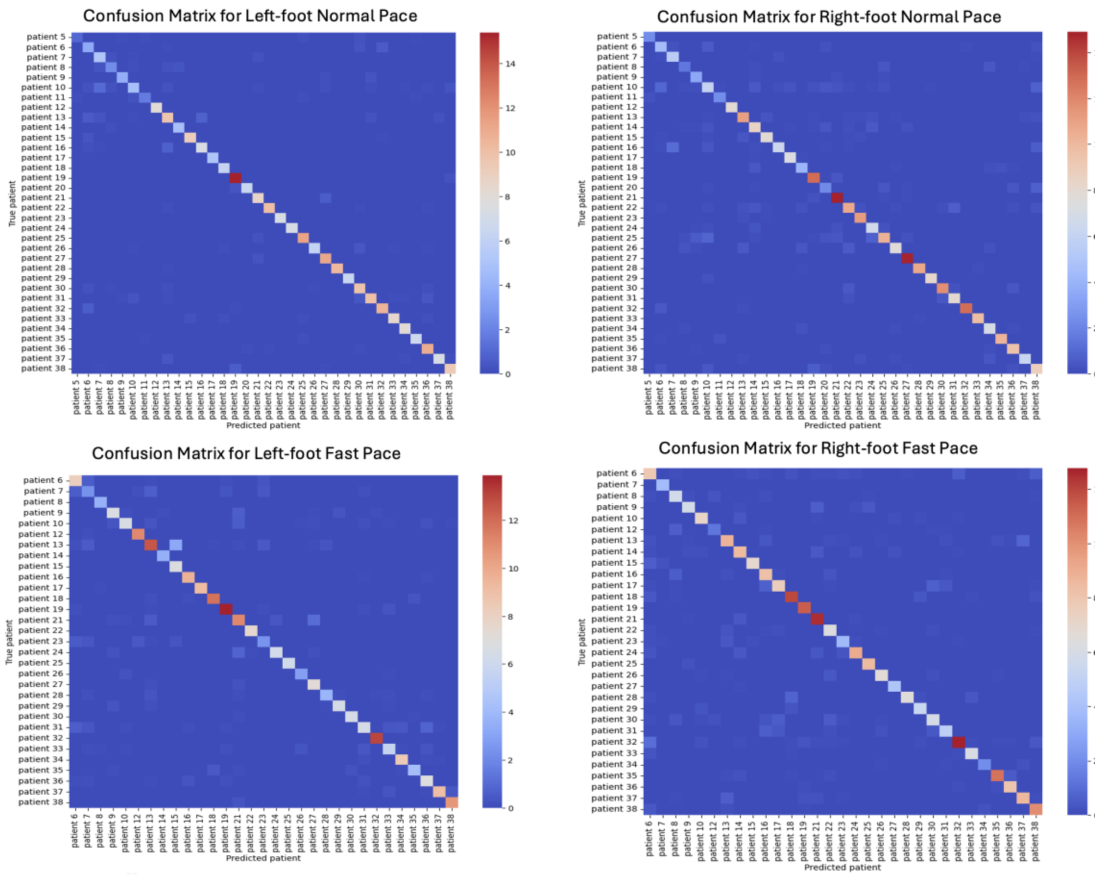


Figure 5-7. Line Plot of Standardized Features Post-Refinement

The analysis of gait patterns by integrating Singular Value Decomposition (SVD) with wavelet transform offers valuable insights into distinguishing individual characteristics in walking behaviors under various walking conditions. This Chapter focused on optimizing the use of IMU data to identify features that can differentiate between individual gait patterns effectively.

The rationale for using this combination stems from their complementary nature.

5.5 Summary

This chapter explored the identification of individuals based on gait patterns captured using foot-mounted IMU sensors. Gait event segmentation was performed to isolate meaningful portions of the walk, focusing specifically on toe-off and swing phases, which demonstrated greater variability across participants compared to flat foot and heel strike events. Following segmentation, the triaxial acceleration signals were converted to magnitude and aligned to ensure comparability. Two sets of features were extracted from the aligned segments: one using Singular Value Decomposition (SVD) to project the data onto a reduced-dimensional space, and another using Discrete Wavelet Transform (DWT) to capture time-frequency-domain characteristics. The extracted features were then combined and used to train a Support Vector Machine (SVM) model. Classification experiments were conducted separately for each foot and for two different walking speeds, demonstrating high accuracy in differentiating between the 34 individuals for normal pace, and 31 for fast pace, reaching 86.97%, 82.96%, 83.55%, 82.37% for left-foot normal pace, right-foot normal pace, left-foot fast pace, and right-foot fast pace, respectively. The success of this identification task highlights the distinctiveness of each participant's walking signature captured by an IMU system and establishes the foundation for more advanced longitudinal analyses in subsequent chapters.

6 Monitoring Gait Similarities Across Multiple Visits and Detecting Outliers

Gait is known to exhibit natural variability throughout time, even in healthy individuals. Small fluctuations in step timing, force application, or joint movements occur from day to day and are influenced by numerous factors such as fatigue, environment, and physiological state. However, for populations at risk of progressive conditions like Chemotherapy-Induced Peripheral Neuropathy (CIPN), subtle gait deviations over time could signal the early occurrence of a clinically significant impairment. Detecting such changes before they manifest into noticeable mobility decline could offer an opportunity for early intervention and improved outcomes for the patient's quality of life.

This chapter aims to quantify gait changes at the individual level across multiple visits, focusing on identifying where deviations emerge within a patient's own gait pattern relative to their baseline, and which steps experience the most pronounced changes, instead of relying on traditional spatiotemporal metrics.

The previously extracted features, based on coefficients from Discrete Wavelet Transform (DWT), form the foundation of this analysis. However, to better understand the temporal localization of changes within each stride, additional steps are applied here, which will allow for more specific interpretation of changes.

Beyond global measures of change, this chapter also investigates the more granular question of *where* changes occur within the walking sequence. analyses are fitted at the feature level, enabling the identification of specific gait events, time windows, and even individual strides that contribute most significantly to the detected changes with respect to the whole multi-pass walk.

A two-tier framework is implemented:

Tier 1 (Global Divergence): Developing a visit-level logistic regression model trained on selected features to output the probability that a new walk differs from the baseline.

Tier 2 (Localized Insights): Identifying one set of features, within participants' steps, that contribute most to the detected divergence. Partial inverse reconstruction of wavelet coefficients visualizes the affected segment in the raw signal.

Through this two-tiered approach (first at the full-feature set processing and then at the event-step processing), this chapter seeks to provide a comprehensive picture of gait evolution over time, with the goal of establishing data-driven markers for early detection of functional decline. To further support the interpretation of these changes, an unsupervised sub-analysis is introduced which clusters strides across all visits based on their overall pattern. Additionally, stride-specific outliers are flagged based on abnormal feature values, offering additional insight into the variability of the walking behavior.

To summarize the above, this chapter will address the following research pathways:

Feature explanation and selection → Global divergence (logistic regression) → Localized change (feature-by-feature logistic) → Pattern composition (k-means) → Anomaly flag.

6.1 Features for Gait Change Quantification

In previous stages of this study, gait features were extracted from aligned swing segments using Discrete Wavelet Transform (DWT) coefficients. These features provided a characterization of walking patterns at individual baseline level captured during the participants' first visit. However, to enable finer assessment of how gait evolves across multiple visits, two important extensions were introduced in this phase of the analysis: (1) the detection of turn steps, and (2) the interpretation of wavelet features by using partial reconstruction to map individual coefficients to specific time-frequency components within the gait cycle.

6.1.1 Turn Detection Using Gyroscope Data

The main motivation for incorporating this process was to retain meaningful metadata about when and where each stride occurred within the walking trial. This is an essential part of the outlier detection analysis, where the goal was to determine whether specific gait changes are most occurring in certain passes, or in specific regions within a pass (right before or after the turn).

By identifying and labelling turns, each walk can be segmented into collective individual straight-walking passes, within each, strides are labeled sequentially in order to enable the later detection of possible patterns such as:

- Whether outliers tend to occur at the start, middle, or end of a pass (possibly reflecting instability triggered by turning)
- Whether some passes contained more outliers than others (potentially indicating fatigue)
- To exclude the possibility that turning strides are associated with outliers

To achieve this, turn detection was implemented using the lumbar gyroscope signal, specifically the x-axis angular velocity. As shown in Figure 6-1 the raw gyroscope signal first undergoes two preprocessing steps, absolute value computation and low-pass filtering (LPF). Taking the absolute value emphasizes the magnitude of angular velocity regardless of direction, while low-pass filtering removes high-frequency noise, highlighting the broader rotational patterns associated with turns.

After filtering, the resulting signal displays distinct peaks corresponding to significant rotational movements. These peaks align well with turning events during gait. A threshold is then applied to the filtered signal to detect these peaks, allowing reliable identification of the time a participant initiates and completes a turn. Finally, the detected turning regions are mapped back onto the corresponding acceleration data (bottom panel of Figure 6-1).

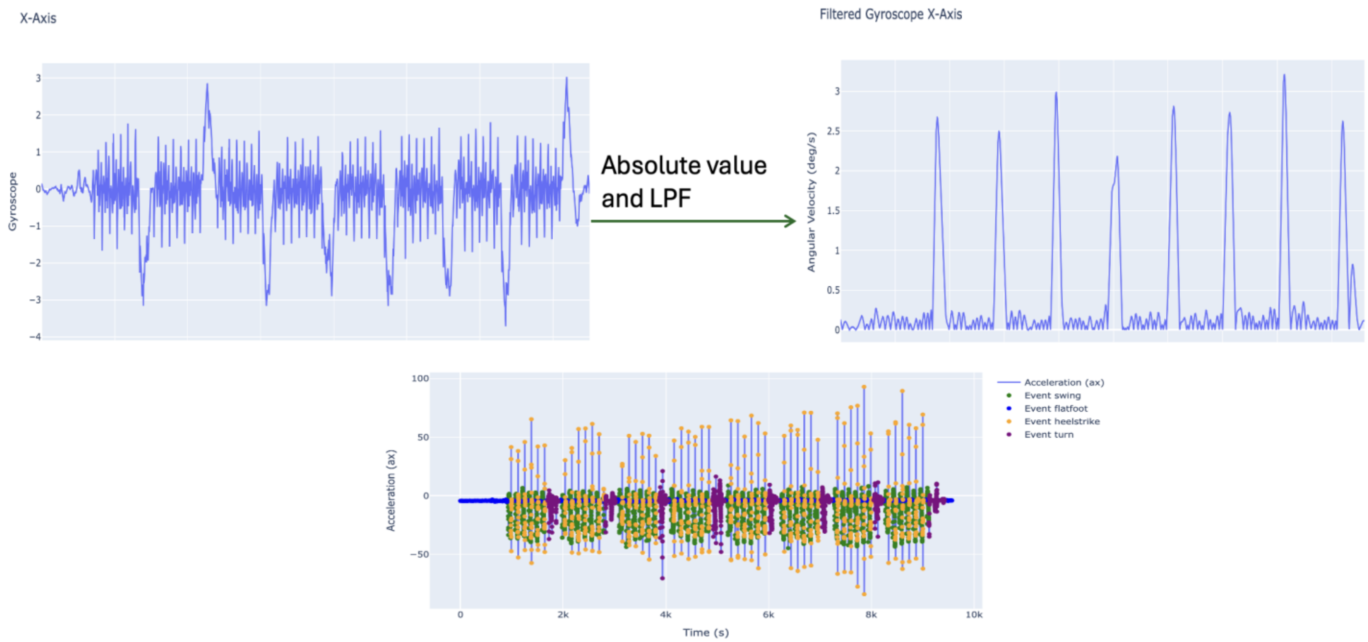


Figure 6-1. Detecting the Turns Using Lumbar Gyroscope. Top: Absolute value of the X-axis Angular Velocity After Low Pass Filtering. Bottom: Identified Turning Periods Highlighted

Once turning regions were identified, they were excluded from the bundle of strides used in divergence and outlier analysis, but their *locations* were retained. This metadata was then linked to the remaining strides, enabling the possibility of later examining stride-level outliers within a visit, not only in terms of features, but also in terms of when in the walk they occur.

6.1.2 Wavelet Feature Interpretation and Localization

The second major refinement involved enhancing the interpretability of the DWT coefficients to understand gait changes over visits in terms of the time and frequency characteristics of the signal. Discrete wavelet transform makes this possible by decomposing the signal into multiresolution components, allowing us to examine signal features in various temporal positions at multiple frequency bands. Each stage of wavelet relies on two core functions as shown in Figure 6-2: The scaling function which captures smoothness and low-frequency trends of the signal resulting in approximation coefficients, and the wavelet function which captures high frequency, localized fluctuations that correspond to detail coefficients.

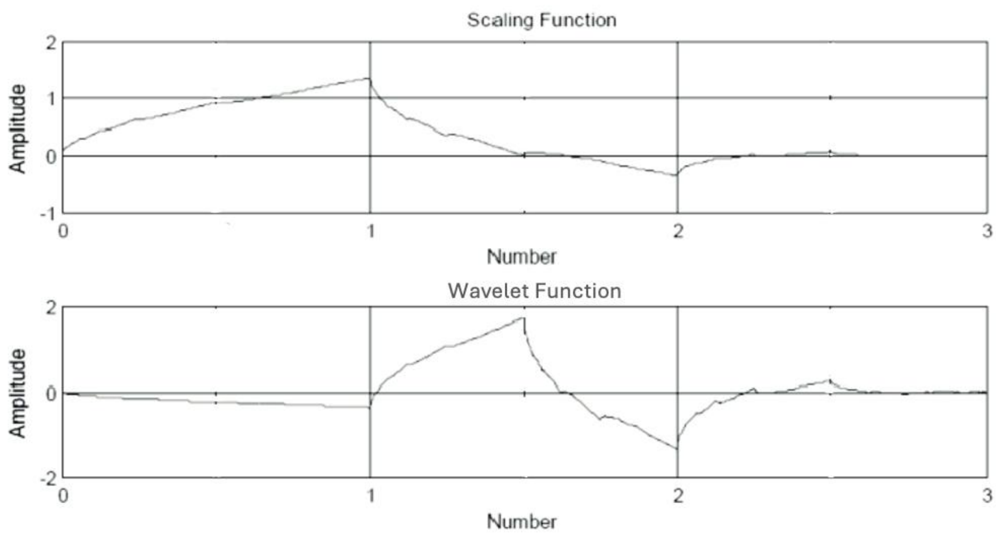


Figure 6-2. Scaling Function and Wavelet Function of a Db2(DWT) [36]

The signal is passed through a pair of filters derived from these functions and downsampled by a factor of two; each round of filtering and downsampling is referred to as a decomposition level, which reduces temporal resolution but increases that of the frequency. This process is repeated on the approximation signal only, which yields a hierarchical structure of frequency bands. The result of this process is a set of coefficients labeled as:

- CA_L : Approximation coefficients at level L
- CD_L : Detail coefficients at level L
- CD_L-1 down to CD_1

The structure of the described process is shown in Figure 6-3 adapted from MATLAB toolbox [37]:

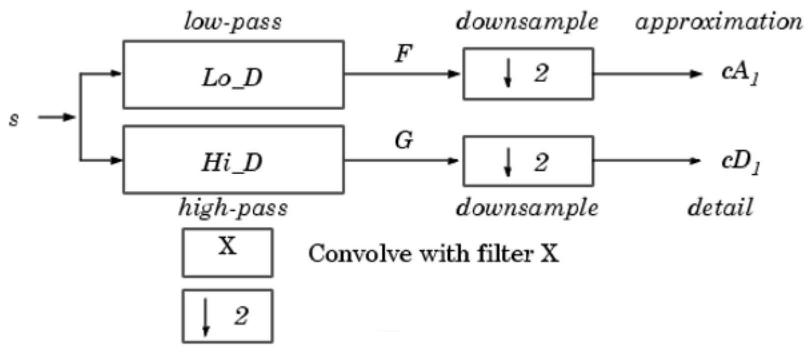


Figure 6-3. The Process of Discrete Wavelet Transform Decomposition [37]

This makes DWT suitable for capturing transient gait dynamics like toe off and heel strike events. The Db2 wavelet was selected for the purposes of the work in this chapter for its short support of (four) samples as shown in Figure 6-2, providing sharper time localization and minimal (two) vanishing moments. This means that the wavelet can represent polynomials up to degree 1 and it is sensitive to deviations from linear trends. This ensures sensitivity to localized changes while filtering out low order polynomial trends which reduces feature count [38]. For a segment length of 90 samples, dwt_max_level() which uses: $\left\lfloor log_2\left(\frac{90}{filter\ length-1}\right)\right\rfloor$, suggests a maximum possible level of 4 for a db2 wavelet. However, I selected level 2 for interpretability to avoid decomposing the undertest signal into many frequency bands. This results in one approximation and two detail coefficient sets.

To visualize and interpret the distribution and localization of energy across time and frequency in the signal, a scalogram was constructed by individually reconstructing each coefficient into a time-domain signal using inverse DWT and zeroing out all others. This allows identification of the time region where its effect was most impactful. A threshold of 10% of the peak amplitude identifies the active region of the coefficient's influence. Each coefficient's absolute value was then accumulated into a 2D grid indexed by time and wavelet level, the resulting matrix was log scaled and visualized as a time-frequency heatmap, with the x axis corresponding to gait time, and y axis corresponding to frequency bands, with approximation coefficients covering ~(0-16 HZ), and detail coefficients covering ~(16-32 HZ and 32-64 HZ). This enabled interpretable mapping between specific wavelet features and their corresponding time-frequency regions, Figure 6-4 demonstrates the process of reconstruction and its resulting heatmap.

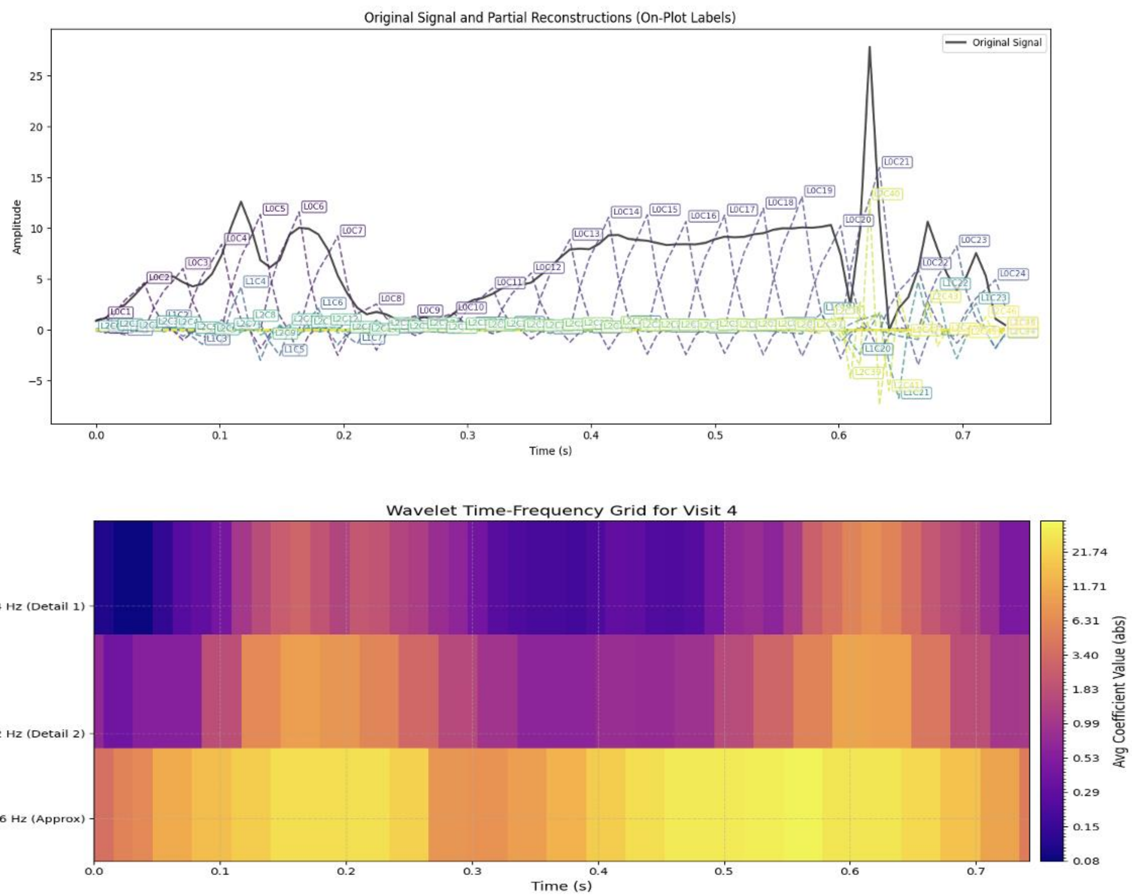


Figure 6-4. A Demonstration of Wavelet Heatmap Construction Process and its Result

The intensity of color in the heatmap represents the energy level of the coefficient. As can be seen, the low frequency band is spread out across the entire segment and contains the highest energy among the coefficients, whereas the energy of higher frequency bands is more pronounced around toe off and heel strike peaks.

Together, these additional processing steps refined the feature set to improve sensitivity in detecting localized gait changes across patient visits, supporting a more detailed and step-resolved analysis of gait evolution patterns.

6.2 Gait Change Quantification Across Visits

In the context of this study, "gait change" refers to detectable alterations in movement patterns across multiple visits. While natural variability and measurement noise are inherent challenges, the aim here is to quantify the extent to which gait patterns at later visits differ from those observed at the baseline. Building upon the feature extraction method described earlier, this section explores how these features evolve between first and subsequent visits. The probability of correctly separating a later visit from baseline serves as a measure for the degree of gait divergence, offering an individualized marker of change.

6.2.1 Feature Selection and Pruning

The analysis begins by representing each full step (toe-off+ swing+ heel strike) using discrete wavelet coefficients as shown in the previous section (note that the segments are aligned just as described in previous chapters to ensure comparability). A discrete wavelet transform (DWT) is applied to each stacked time domain signal using the Daubechies 2 (db2) wavelet as mentioned previously, with two levels of decomposition.

After transforming all gait segments, the feature space was reduced as follows:

- Energy Coverage Pruning:

  For each segment, the energy contribution of every wavelet coefficient was estimated by reconstructing the partial signal generated by that coefficient alone and summing the squared amplitude over the most active part (samples exceeding 10% of peak), the coefficients where then ranked by their **average contribution** across all segments, and the minimum set of coefficients that cumulatively captured at least 95% of total signal energy were retained.

- Edge Coefficients Removal:

  Coefficients whose partial reconstruction activated only at the very first or very last samples of the segments were discarded, those are often sensitive to boundary effects and were filtered out to avoid misleading interpretations.

These steps result in a set of features unified across all segments and visits, allowing fair comparison in the logistic model. However, because individual strides can still differ in coefficient energy even during the same visit, the feature matrix retains stride level variability. This concept will be revisited in **Section 6.2.5,** where per stride outlier detection is used to localize specific steps that deviate most from their visit level averages. Figure 6-5 shows an example of the resulting heatmap for the retained wavelet coefficients after pruning. The color intensity reflects the average absolute value of each coefficient across all segments. Figure 6-6 demonstrated the reconstruction of a representative segment using only the selected coefficients, confirming the preservation of key gait dynamics in the signal.

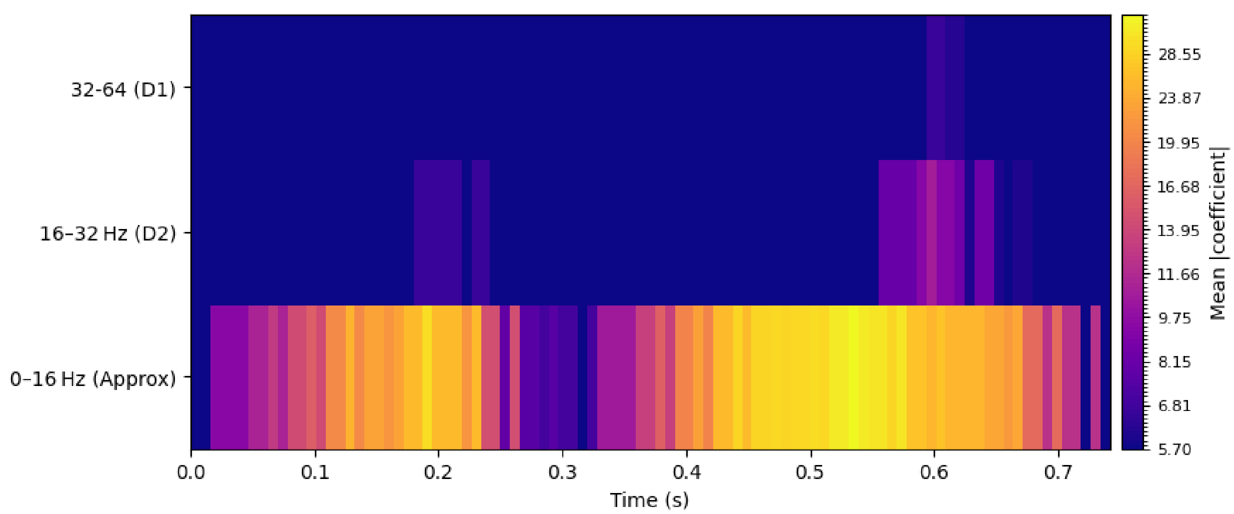


Figure 6-5. A Heatmap of Kept Coefficients

Energy captured: 94.8%

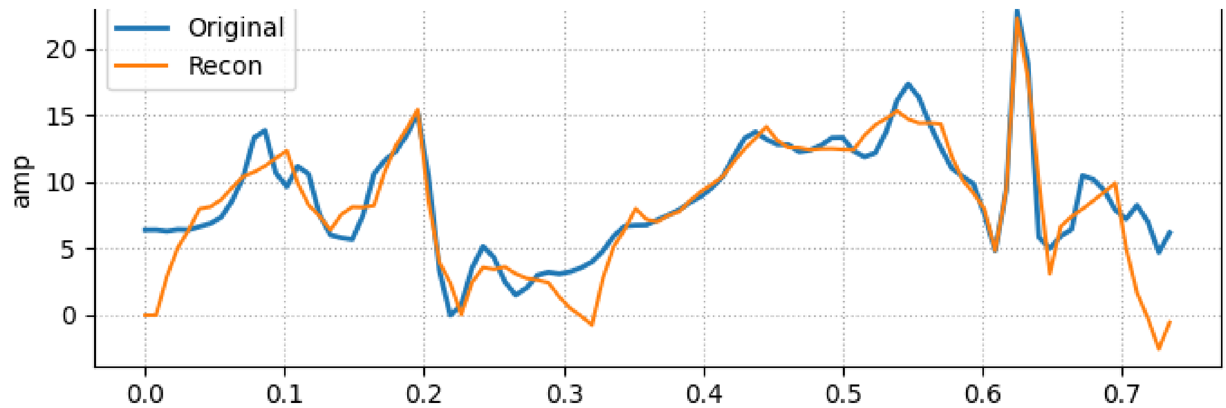


Figure 6-6. Reconstruction with Kept Coefficients

The accompanying printout reports the number of coefficients retained as follows for this example:

Initial count of coefficients: 101
Reduced count of coefficients with energy equal greater than 95% energy: 27

6.2.2 Logistic Regression for Global Divergence

To evaluate how much a patient's gait pattern diverges over time from her initial baseline, a logistic regression model with elastic net regularization was trained to classify segments as either baseline (visit 1=class 0) or non-baseline (visit ≥ 2 = class1). This model operates on the selected wavelet coefficients from the previous section and estimates the probability that a given segment matches the baseline's. The probabilities are calculated per segment, then they are averaged per visit to yield a visit level divergence score, serving as an indicator of gait evolution over chemotherapy treatment. Mathematically, logistic regression models the log-odds of the outcome as a linear combination of the input features:

$$log\left(\frac{P(Y=1|X)}{P(Y=0|X)}\right)=\beta_0+\sum_{j=1}^{d}\beta_j\,x_j$$

Where X is the input vector of selected wavelet coefficients for a single segment, $Y \in \{0,1\}$ indicates whether the segment belongs to the baseline or later visits, β_0 is the intercept term which shifts the decision boundary to adjust for baseline class probabilities, and β_j are the model weights. To improve generalization, the model uses elastic net regularization which combines both L1 (lasso) and L2 (ridge) penalties. This dual penalty ensures that the model remains interpretable by zeroing out irrelevant coefficients while also retaining enough features to avoid underfitting. However, this does not eliminate the need for the feature pruning step done earlier, since a bloated model with 100+ features would slow down the training, increase the risk of overfitting, and weaken signal to noise ratio because more noise would be added to the feature space.

To mitigate information from the test set leaking into the training set (data leakage), a leave-one-visit-out validation was used. The model was trained on all visits (including the baseline) except one and tested on the held-out visit. The baseline visit was always included in the training set so that the model learned a stable representation of what "baseline gait" looks like. The test visit was never seen during training. Prediction probabilities for the baseline visit itself were averaged across all folds to obtain an estimate of its divergence score. this average reflects the model's tendency to assign low divergence to baseline data. This maintains internal comparability even though visit 1 is never left out of training.

To handle the class imbalance, class_weight ='balance' was specified in the logistic regression, which automatically adjusts the contribution of each class in the loss function based on their frequency.

Mathematically, logistic regression minimizes the following regularized loss:

$$\mathcal{L}(\beta) = -\sum_{i=1}^{N} \omega_{y_i}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] + \lambda((1 - \alpha)\|\beta\|_2^2 + \alpha\,\|\beta\|_1$$

Where:

- p_i is the predicted probability from the sigmoid function
- ω_{y_i} is the class weight
- λ (via c= 1/ λ) controls over all regularization strength
- $\alpha \in \{0,1\}$ controls the mix between L1 and L2 penalties:

  α=1: pure L1 (lasso)

  $\alpha = 0$: pure L2 (ridge)

  In between: Elastic net

This setting ensures that each class (baseline vs non baseline) contributes equally to the loss function, even if their counts are different in the training set. Grid search with 3-fold cross validation was used to tune hyperparameters λ and α, using area under the curve (AUC) as the scoring metric.

This evaluation strategy allows for a supervised estimation of gait divergence probabilities, where high divergence indicates substantial deviation from baseline gait recorded before chemotherapy initiation. While that serves as a useful tool, it depends on binary classification. If the visit under test lies in between the two classes, or if it follows a pattern not well-represented in the training data, the predicted probability may hover around 0.5 reflecting model uncertainty. This makes the model sensitive to how well the training visits capture the diversity of possible gait changes.

To address this, **Section 2.6.4** applies unsupervised clustering to the same pruned feature set to explore the possibility that certain visits are not only more or less similar, but qualitatively different and contain underlying pattern shifts. Figure 6-7 shows an example result of the model.

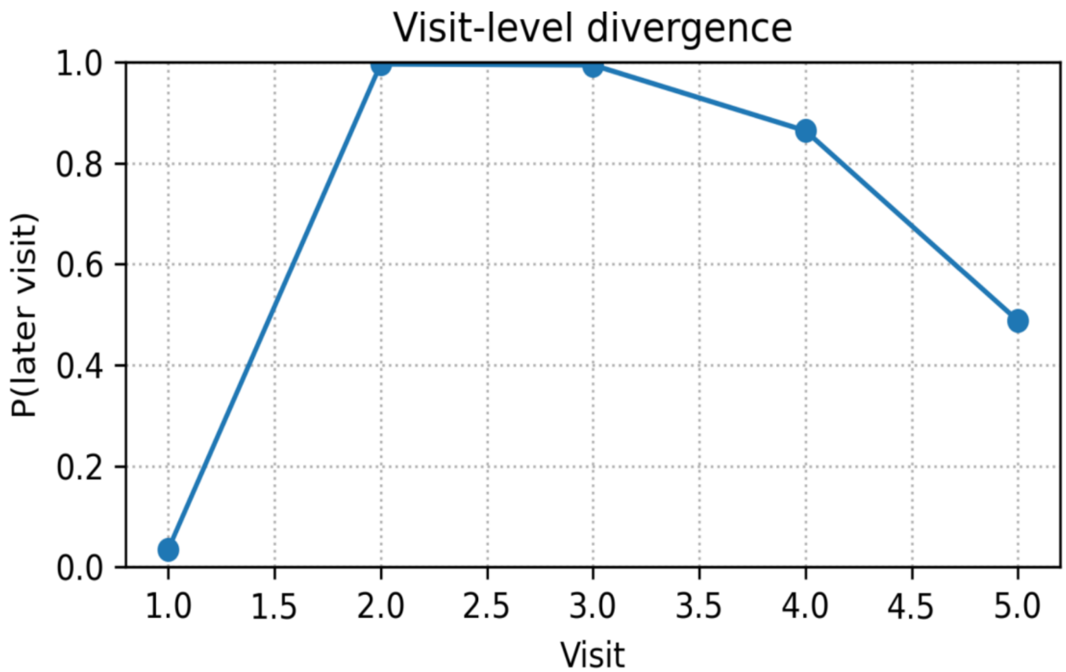


Figure 6-7. Probabilities of Divergence from the Baseline for L-foot for One Patient

6.2.3 Mapping Divergence to Gait Events for Physical Interpretation

This section transitions from the probabilistic statement of "later visits have diverged" to an interpretable explanation of why they diverged. This interpretation is done in terms of specific wavelet coefficients and the corresponding time-frequency regions of the gait signal.

After training the logistic model, the importance for each retained feature was determined by first computing the standardized mean of each feature across all strides for the baseline visit. The same was done for the visits that is being compared to. Then their difference was computed yielding a per features mean shift vector:

$$\Delta x_j = \mu_j^{(visit)} - \mu_j^{(baseline)}$$

Then the mean shift vector was multiplied with the weight learned by the elastic net, this results in a contribution defines as:

$$contribution_j = |\beta_j . \Delta x_j|$$

Where β_j is the weight learned by the logistic regression model for feature. The top features collectively driving 90% of the divergence probability were considered the most important features for the visit under evaluation.

After the initial step of determining which features are causing the most divergence from the baseline, a meaningful interpretation is needed as was explained earlier in **Section 6.1.2.** The wavelet coefficients were given meaning in terms of time and frequency, allowing to map them straight back to the signal and determine which gait event is being affected (toe off, swing, heel strike) and at which frequency band (Approx=0-16, D2= 16-32, D1=32-64) as demonstrated in Figure 6-8, which overlays a one participant example of all strides for a visit vs baseline, with color coded highlights showing the corresponding signal window.

The final step after determining the important features and their meaning in terms of the signal, is computing the mean amplitude at the defined region for each coefficient (peak to peak range), and the mean energy (sum of squares) across all strides for that visit. This helps quantify not just where changes happen but also how strongly they manifest. The detailed results are shown in a printed report accompanying each plot as exhibited below:

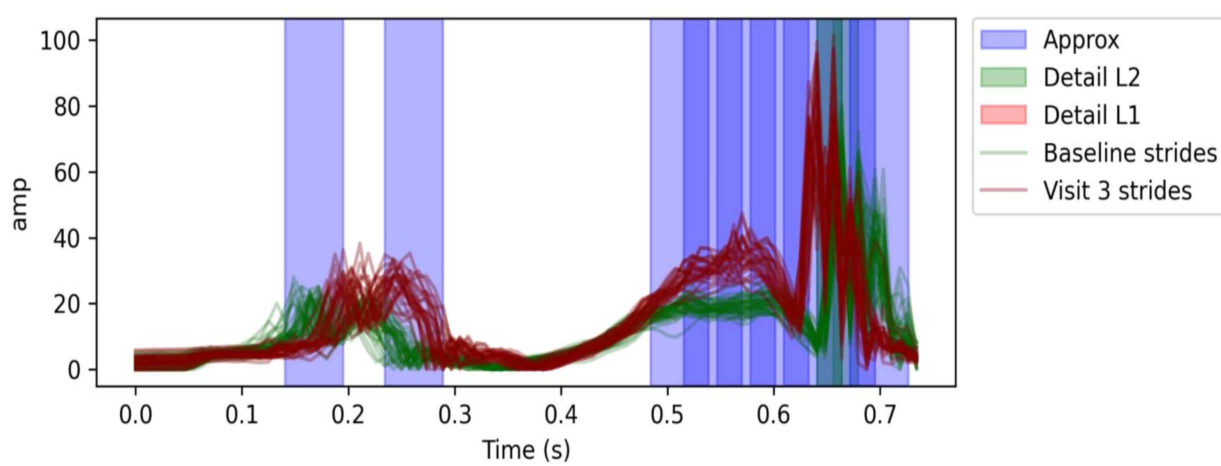


Figure 6-8. Reporting Localized Changes from Baseline and their Meaning

== Visit 3: 11 coefficients → 90 % drift report==
idx19 L0 (0.55-0.60s) (Swing): ampΔ =52.1, E= 4.58e+03,
idx18 L0 (0.52-0.57s) (Swing): ampΔ =45.8, E= 3.53e+03,
idx6 L0 (0.14-0.20s) (Toe-off): ampΔ =9.67, E =170
idx21 L0 (0.61-0.66s) (Heel-strike): ampΔ= 60.6, E =6.28e+03
idx95 L2 (0.66-0.68s) (Heel-strike): ampΔ =16, E=199
idx9 L0 (0.23-0.29s) (Swing): ampΔ= 38.6, E= 2.54e+03
Idx22 L0 (0.64-0.70s) (Heel-strike): ampΔ =76.6, E =9.95e+03
idx17 L0 (0.48-0.54s) (Swing): ampΔ =36.4, E=2.23e+03
idx20 L0 (0.58-0.63s) (Swing): ampΔ=41.3, E=2.94e+03
idx23 L0 (0.67-0.73s) (Heel-strike): ampΔ=28.2, E=1.45e+03
idx94 L2 (0.64-0.66s) (Heel-strike): ampΔ=35.3, E=929

By localizing the divergence, we move beyond a black-box classifier to an explainable gait marker framework, helping to identify early biomechanical changes across visits.

6.2.4 Unsupervised Clustering of Stride Patterns

While the previous sections established a probabilistic and interpretable framework for quantifying gait divergence over time, an additional unsupervised clustering based sub-analysis was performed to both support and further explore structures withing the gait data. The aim of this approach is to determine whether unlabeled swing segments from all visits would form naturally distinguishable groups that may correspond to different walking patterns adopted by the patient. This shows the percentages of the patterns presented in each visit, and how the patterns are shifting in an unsupervised term.

The features used for clustering were identical to those used in the logistic model up to the pruning step ($\geq 95\%\ energy\ coverage\ and\ edge\ filtering$). However, the feature set was not restricted to those with non-zero logistic regression weights, as that might introduce a supervised bias into what is intended to be a purely unsupervised analysis, suppressing possible alternative patterns of gait change that are not aligned with the logistic decision boundary.

The feature matrix was input into a K-means clustering.

The elbow method was computed in order to intelligently select the optimal number of clusters as presented in the example in Figure 6-9, wherein the total within-cluster sum of squares (inertia) is plotted against the number of clusters. To avoid selecting K manually and make the process systematic, the orthogonal distance from each point to the line connecting k=1 to k=max was computed, identifying the value of k that maximizes this distance and indicating the optimal number of clusters.

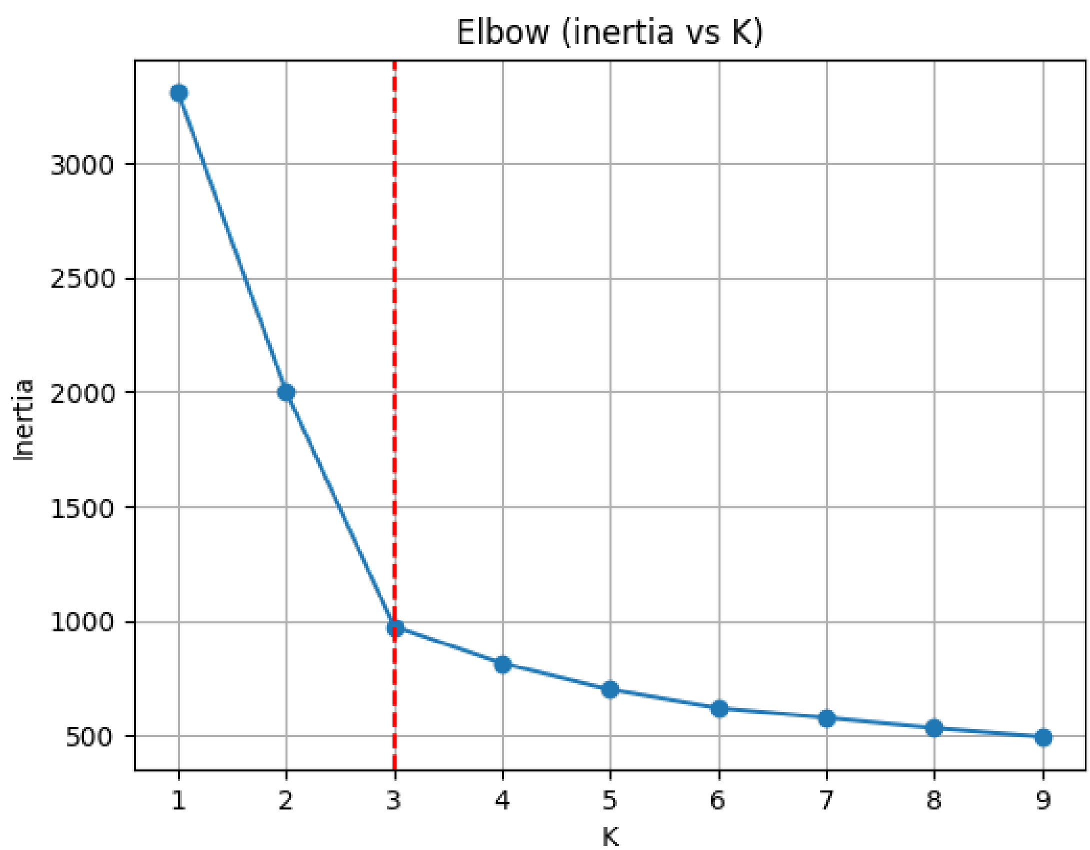


Figure 6-9. Elbow method for Systematically Selecting the Optimal Number of Clusters

Each segment was then objectively assigned a cluster label and visit-based wise distributions of these clusters were computed; an example result can be seen in Figure 6-10.

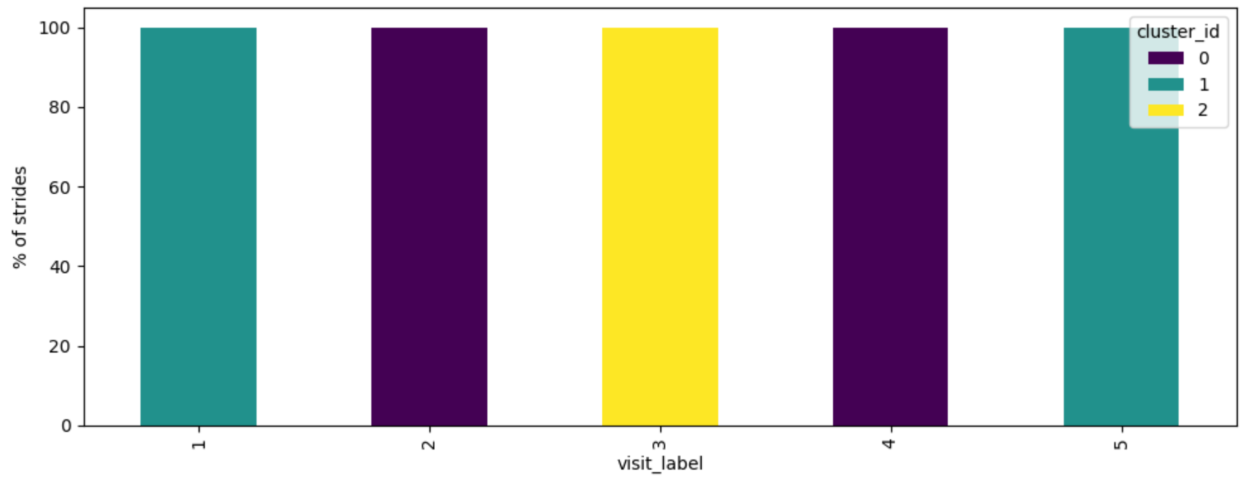


Figure 6-10. An Example of the Result of Unsupervised Clustering Pattern Formation for one Patient

By cross examining the results from this approach with the results from the logistic regression divergence probability, we could conclude supporting outcomes to the previous section as can be demonstrated in the following example table resulted from the left-foot sensor of one participant:

Table 6-1. Cross Examination of K Means Results with Drift Probability Results from Logistic Regression..

	Present Patterns from K Means	**Drift Probability from Baseline**
Visit 1	100% p1	0.035
Visit 2	100% p0	0.996
Visit 3	100% p2	0.994
Visit 4	100% p0	0.864
Visit 5	100% p1	0.488

As observed, there is correspondence between the present patterns and the probability of drift. The dominant pattern in visit 1 is pattern 1 present by 100%, which is shown in visit 5 as well, making it the most similar to visit 1. This agrees with the probability of divergence from visit 1 being the lowest for visit 5. We can also observe that visits 2,3, and 4, which have a high probability of divergence from the baseline, have formed their own separate clusters indicating different patterns. To make this cross evaluation in a more methodical way, each visit's cluster distribution was compared to that of the baseline. For every visit, after computing the percentage of strides assigned to each cluster which forms a vector of cluster proportions, I computed the following: let "b" be the cluster proportion vector for the baseline visit, and v the vector for another visit, then a similarity score can be calculated as the dot product:

$$\text{Similarity}=\sum_i \left(\frac{b_i}{100}.\frac{v_i}{100}\right)$$

This yields a score falling between 0 and 1; with higher values indicate that the visit's stride patterns have higher resemblance to the baseline, so to express the degree of change instead of similarity we compute (1-similarity).

Finally, we compare the dissimilarity scores to the logistic regression probabilities from **Section 6.2.2** by plotting them together, as shown in Figure 6-11, a strong positive correlation was observed in most cases, meaning visits that appear more divergent according to logistic regression also had stride patterns more different from the baseline, which supports the conclusion that the logistic model reflects real, structured changes in gait composition across visits.

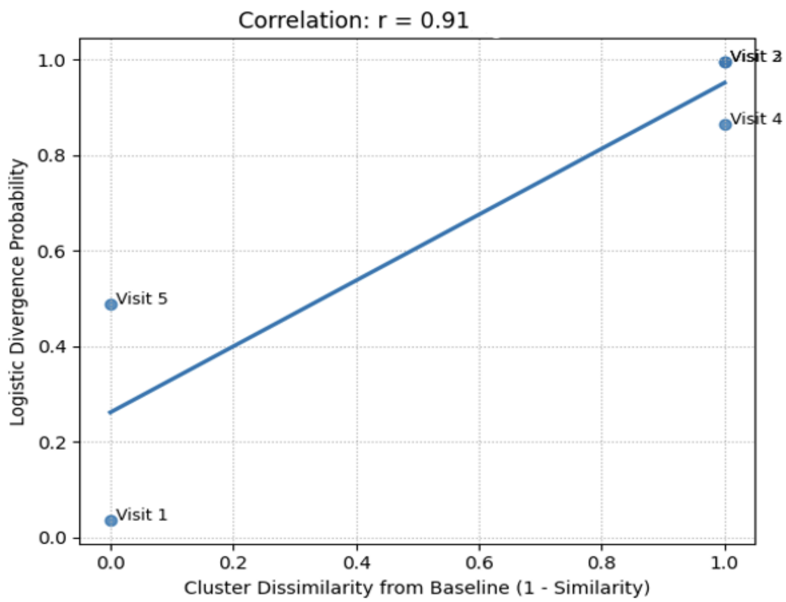


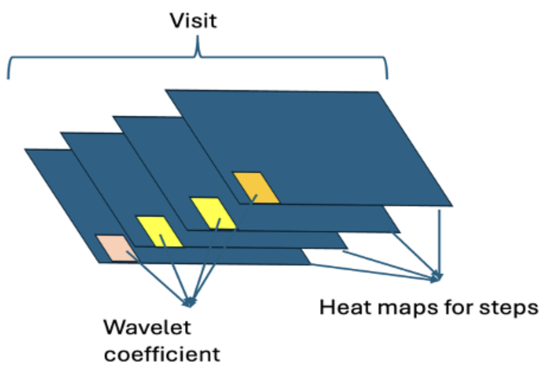
Figure 6-11. An Example of the Result of Unsupervised Clustering Pattern Formation for one Patient

The results in this approach also show how stride patterns shift across visits instead of just how different a visit is from another. Some patterns become more dominant, others disappear, or a unique pattern shows up which wasn't present in previous visits. These changes might help explain why some segments end up looking unusual later in the outlier analysis which will be presented in the next section.

6.2.5 Segment Level Outlier Detection

To explore deviation as per individual strides, a per segment outlier analysis was conducted using the same pruned wavelet features described earlier. This step was motivated by the need to identify the specific strides within a visit that show a strong departure from their mean values. These outliers can offer some insight into short-term irregularities or the emergence of new gait patterns that might correspond to underlying clinical or contextual events, or they could be due to inherent variability in human gait depending on surrounding conditions (i.e. getting ready to turn, entering or leaving the pressure mat).

This analysis begins by standardizing each feature across strides within the same visit, then a stride will be flagged as an outlier as per that feature if the feature value exceeds 2 standard deviations within the feature's distribution. Figure 6-12 shows the KDEs for the coefficients remaining after feature pruning for the previous example shown in **Section 6.2.1**.



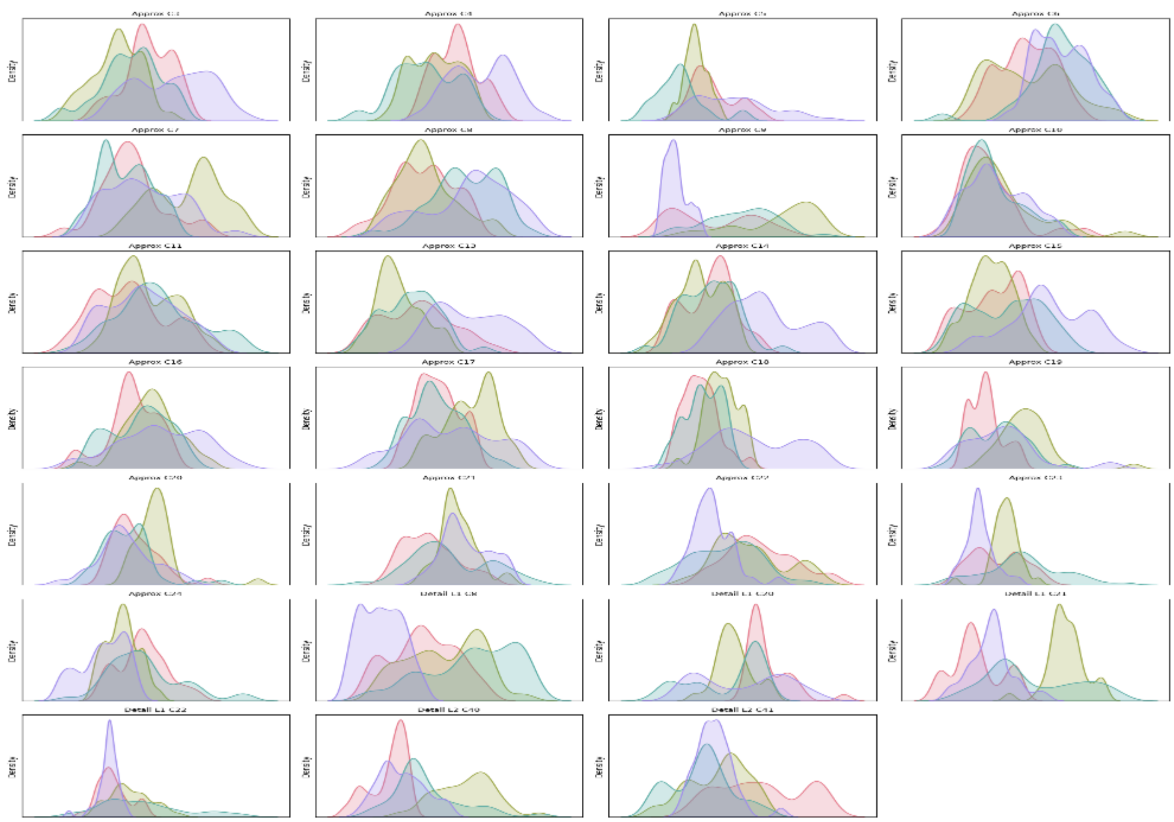


Figure 6-12. KDEs of Remaining Features for One Patient After Feature Pruning

Outlier strides are then clustered within each visit in order to explore whether certain types of irregularities tend to occur together. Each stride is represented as a binary vector, where each entry indicated whether or not a particular wavelet feature was flagged as an outlier for that stride. These binary vectors are clustered hierarchically, using Jaccard distance as the similarity measure since it is ideal for comparing binary vectors. Each leaf represents one outlier stride labeled by its pass and segment index; strides that are clustered closely share many of the same outlier coefficients. Each cluster is further summarized by identifying the wavelet bands and gait phases (toe off, swing, heel strike) that are involved most frequently. Additionally, the outlier strides are characterized by their position within the walking pass, labelled as occurring in the (beginning, middle, end) of the pass based on its segment index relative to the total number of segments in that pass. This helps identify whether anomalies occur at transition points.

And example of this grouping is shown in Figure 6-13

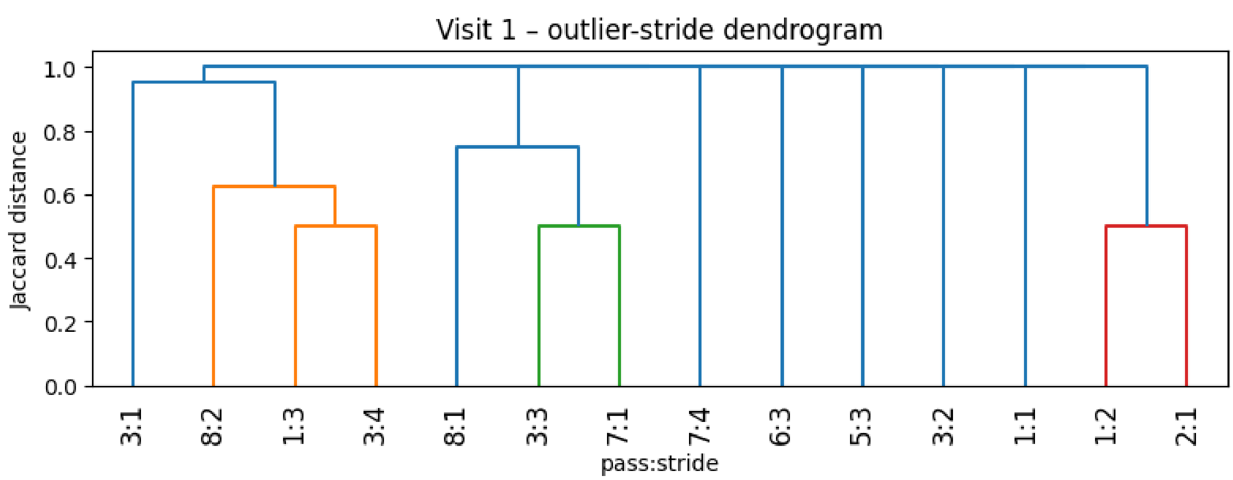


Figure 6-13. Clustering Outliers Per Feature and Per Location

Visit 1: cluster report
Cluster 1 (n=14)
overall → bands [Approx:15, D2:2, D1:1]
phases [Swing:10, Toe-off:4, Heel-strike:4]
location trend → {'begin': 8, 'end': 4, 'middle': 2}

We can observe that for this example that most outliers occur at the beginnings of passes, which could indicate that they are the outcome of transitional phases in the walk.

The results from this section start with a full stride level table listing the visit, pass, stride index, and affected coefficients. The findings are summarized by clustering based on shared coefficient anomalies and based on metadata trends showing where in the pass outlier strides tend to occur. Together, this analysis links back to the previous sections (the logistic scores and the clustering) where each one adds a different perspective on what is changing and where it is changing.

6.2.6 Discussion and Collective Results Across Patients

In the preceding sections, I detailed patient specific analyses, the goal of this section is to collectively summarize the key findings across all participants and for each foot. Due to the volume of plots, tables, and per visit metrics generated for each individual, I will present only summarized trends by grouping patients according to common patterns observed in each analysis, while the complete visualizations and some examples of metadata are provided in the **Appendix**.

- **Logistic Regression:**

  The model was run for patients with a minimum of three visits, that summed up to a total of 24 patients with a varying number of visits and a varying number of passes per visit; each patient's results were visualized as a line plot of predicted probability of divergence per visit. The left and right foot sensors were analyzed separately, all resulting graphs are presented in **Appendix A.** The results show that the most common patterns are: (1) an increase at the second visit followed by a decrease in later visits, (2) an increase in the second visit followed by a plateau in later visits, (3) fluctuating, non-monotonic divergence, (4) a linear or semi linear increase can be observed in five cases, (5) only three cases have shown minimal divergence across visits (both feet for CIP038 and right foot for CIP014, additionally her left foot experienced small divergence compared to other cases). While CIP014 was reported as one of the patients who did not develop CIPN. This result is not sufficient to create a hypothesis, and further investigation needs to be done on a larger and more versatile dataset. These findings also support the hypothesis that gait drift is not always monotonic and therefore it requires patient-specific monitoring, and that each foot experiences a different pattern of gait and drift. A notable observation which can be seen in rare cases is a later visit showing lower probability of divergence than visit 1 itself. This is due to visit 1 being used in every iteration and its results being averaged across all, which can inflate its predicted divergence. Having said that, the inclusion of visit 1 does not affect the validity of between visit comparisons, but it should be kept in mind when interpreting absolute probabilities, and the results of this model should be viewed in conjunction with per segment sub-analysis.

- **Localized Features:**

This analysis suggested that the coefficients driving 90% of the divergence are often spatially distributed rather than isolated to a specific event nor isolated to a specific patient. Examples of that are shown in **Appendix B.**

In cases where changes were more localized, they occur in toe off, heel strike, and the end of swing. Mid-swing shows the least likely divergence in amplitude or energy.

- **K-means Clustering:**

When comparing K-means cluster distributions across visits to the probabilities estimated by supervised logistic regression, a wide range of associations was observed. In some patients, as can be viewed in **Appendix C**, logistic and k-means showed strong agreements, with 33/48 cases showing strong (above 75) correlation between predicted divergence and the shifts in stride cluster composition, in 7 out of 48 cases, the agreement was moderate (between 60 and 70); The remaining 8 cases showed no correlation. This is because while the two approaches can be cross examined for exploratory purposes, they capture different aspects present in the data. K-means clusters strides based on geometric similarity, but logistic regression assigns the score based on how well a stride separates baseline from later visits. So high agreement occurs when those two aspects align. The correlation drops when top logistic coefficients are not sufficient for explaining most of the overall variance.

Beyond agreement with logistic regression, the K-means also revealed shifts in gait patterns across visits. In some patients, a cluster that dominated the first visit or early visits disappears completely by the final visit. In other cases, an initially dominant cluster disappeared in the middle but then reappeared and became dominant again toward the end. Some patients exhibited a gradual emergence of a pattern in the middle which became dominant toward the end. This indicates a possible gradual change in gait dynamics. Some profiles showed irregular shifts with no clear trend, with patterns non-monotonically disappearing and reappearing, indicating variable gait behavior. Altogether, these shifting cluster patterns suggest that gait change may manifest not only in how distinct a visit is from the baseline, but also in how internal stride patterns evolve or destabilize over time.

- **Outlier Flagging:**

Outliers appeared to be most frequently concentrated near the beginning of a walking pass across participants and visits, followed by the end, with mid-pass strides showing fewer anomalies. Additionally, outliers seem distributed across the visit rather than concentrated in a particular visit. They are probably occurring due to transition from turn to straight walking. Individual variability remains true with some exceptions. within a single gait, swing events were the most common contributors to outliers. While there exist exceptions to these patterns, they reinforce the importance of phase specific analysis in gait monitoring.

6.3 Summary

This chapter introduced a multi-tier analytical framework to characterize how individual patient's gait patterns change, due to chemotherapy treatment, and evolve across multiple visits using foot-mounted IMU data. The chapter began by introducing two key types of features: (1) turn detection from lumbar gyroscope data and (2) heatmaps of acceleration signal steps extracted from discrete wavelet transform.

Next, systematic feature selection was performed using an energy coverage pruning method which retained the minimum number of coefficients contributing $\geq 95\%$ of the signal/s total energy, while also removing edge influences coefficients. These features formed the input for all subsequent analyses.

The first major analysis used an elastic net logistic regression model to quantify how divergent the follow up visits are from the baseline (visit1). Then, feature importance scores obtained from the logistic model were examined to identify which wavelet coefficients most contributed to the calculation of the probability. Their corresponding signal contributions were localized to toe off, swing, or heel strike phases.

Unsupervised K-means clustering—a second approach—was applied to the same pruned feature space to detect underlying structures and patterns. The results were cross compared with the logistic divergence probabilities, offering a validation of the outcomes observed.

Lastly, a stride-level outlier analysis was conducted within each visit by identifying coefficients with distribution of $\geq \pm 2\ STD$.. These outlier strides were then clustered hierarchically based on

Jaccard distance of their abnormality patterns. Each cluster was characterized in terms of gait phase involvement, and position within the walking pass (beginning, middle, end).

The chapter concluded with a collective analysis that grouped patients by their shared evolution patterns for the logistic and K-means methods, laying the foundation for phenotype level clinical interpretation.

7 Conclusion and Future Work

This thesis presented a multi-level framework for analyzing gait patterns among cancer patients undergoing chemotherapy, using wearable IMUs mounted at the feet. The work staged from a gait data preprocessing, segmentation, alignment strategy, and a classification to confirm the uniqueness of gait patterns among participants. Two machine learning approaches were used to quantify both global divergence from the baseline and localized irregularities in gait cycles.

Chapters 1-4 focused on data preparation, where foot-mounted IMU signals were segmented to extract steps and gait events (swing, heel strike, flat foot) and align segments using Euclidian distance. Once the segmentation and alignment framework were established, Chapter 5 explored inter-subject classification using data from visit 1 as baseline, employing SVD and wavelet-based feature extraction methods to distinguish between individuals. This formed a foundational understanding of the discriminative qualities of raw gait acceleration data.

Chapter 6 conducted a longitudinal studying the progression and deviations of gait cycles when compared to the baseline visit. Turn detection was introduced to the segmentation framework using gyroscope data extracted from the lumbar sensor for the purpose of defining metadata for passes. Wavelet coefficients were used to construct time-frequency heat maps and to extract interpretable features with 95% cumulative energy coverage, excluding edge coefficients. These features were used to model divergence from baseline using an elastic-net logistic regression framework, yielding per-visit probabilities that show patient's gait deviation from her baseline visit. The most influential features were identified and localized to see what gait events were mostly driving the divergence and quantify divergence in terms of amplitude and energy.

To evaluate the observed divergence and explore the possibility of it reflecting a more complex reformation of gait patterns, unsupervised K-means clustering was applied to the same pruned feature space. By grouping stride segments based on common patterns, clustering revealed hidden substructures in the data in a means that aligns with divergence probabilities obtained from the logistic method.

Segment-level outliers were also detected by identifying extreme deviations relative to within feature distributions. Those were clustered to reveal recurring irregularities as per gait event and stride position within a walking pass. A collective summary was made across all participants, capturing common trends.

Taken together, the framework developed in this thesis provides a detailed and interpretable pipeline for analyzing temporal evolution of a person's gait, balancing classification with physical insights, offering a valuable tool for longitudinal gait monitoring in clinical populations.

7.1 Future Work

Several extensions of the work laid out in this thesis can further enhance its clinical relevance and analytical depth. One promising direction involves expanding the analysis to incorporate bilateral foot dynamics. While the current approach treats each foot independently, the availability of left and right foot IMUs creates the opportunity to investigate common related features. Metrics such as event time asymmetry or bilateral stance overlap could yield sensitive indicators of subtle gait degradation or compensatory adaptation, especially for populations at risk of neuropathy or balance impairments.

Another avenue could be integrating data collected by other sensors, including those placed at the lumbar, sternum, and ankle levels. These could reveal changes in trunk control, upper body stability or joint mobility, especially during turns or transitional movement. Incorporating these sensors could enable a more holistic view of gait mechanics and could support multi-sensor fusion strategies for robust monitoring in uncontrolled environments.

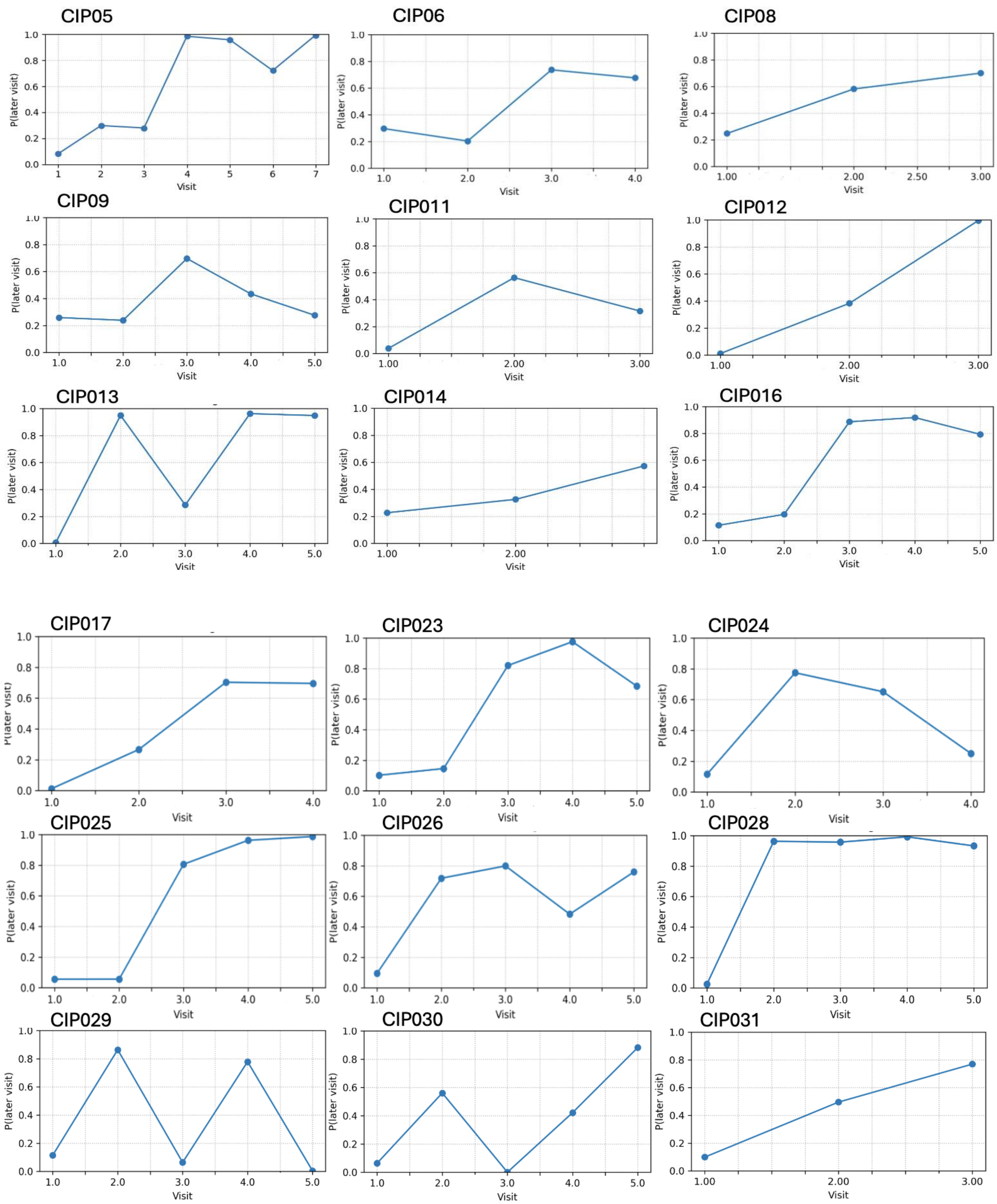
The framework could also be extended to model divergence explicitly within specific gait phases. While the work here mapped influential features to events such as toe off or swing, a more targeted approach could involve a phase-specific model. This could improve both physiological interpretability and clinical utility.

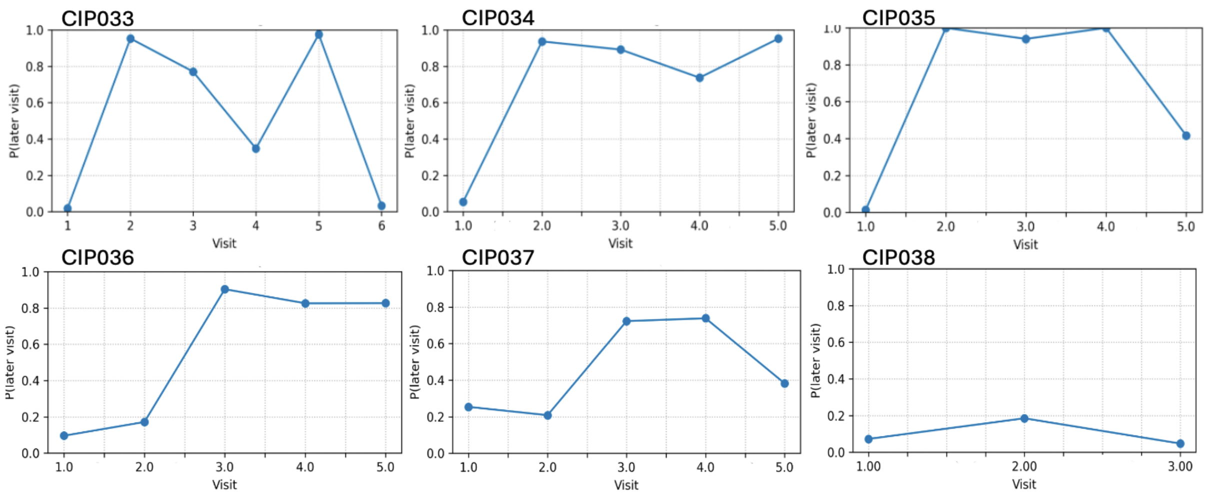
On a broader scale, future studies may aim to generalize beyond individual-level modelling, by examining whether shared trajectories emerge across groups of patients with similar clinical profiles. This could support the development of regulating gait trajectories or early warning markers which can be applied across sub-populations.

8 Appendix

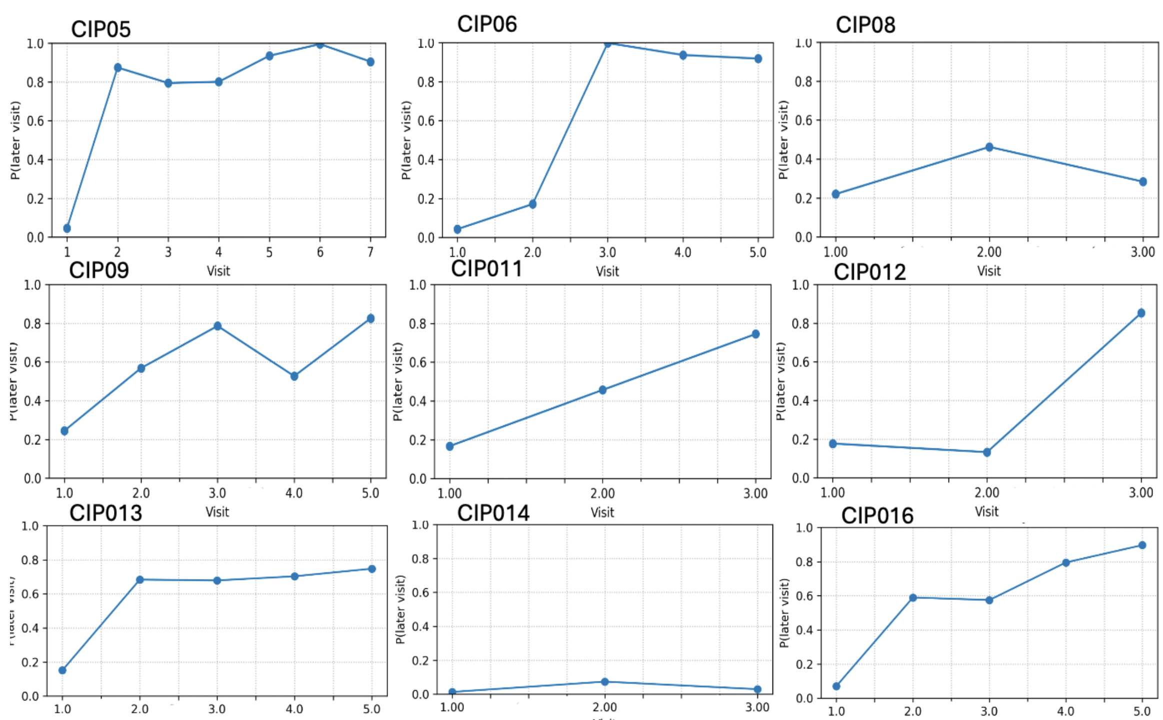
Appendix A: Global Divergence Figures for All Participants

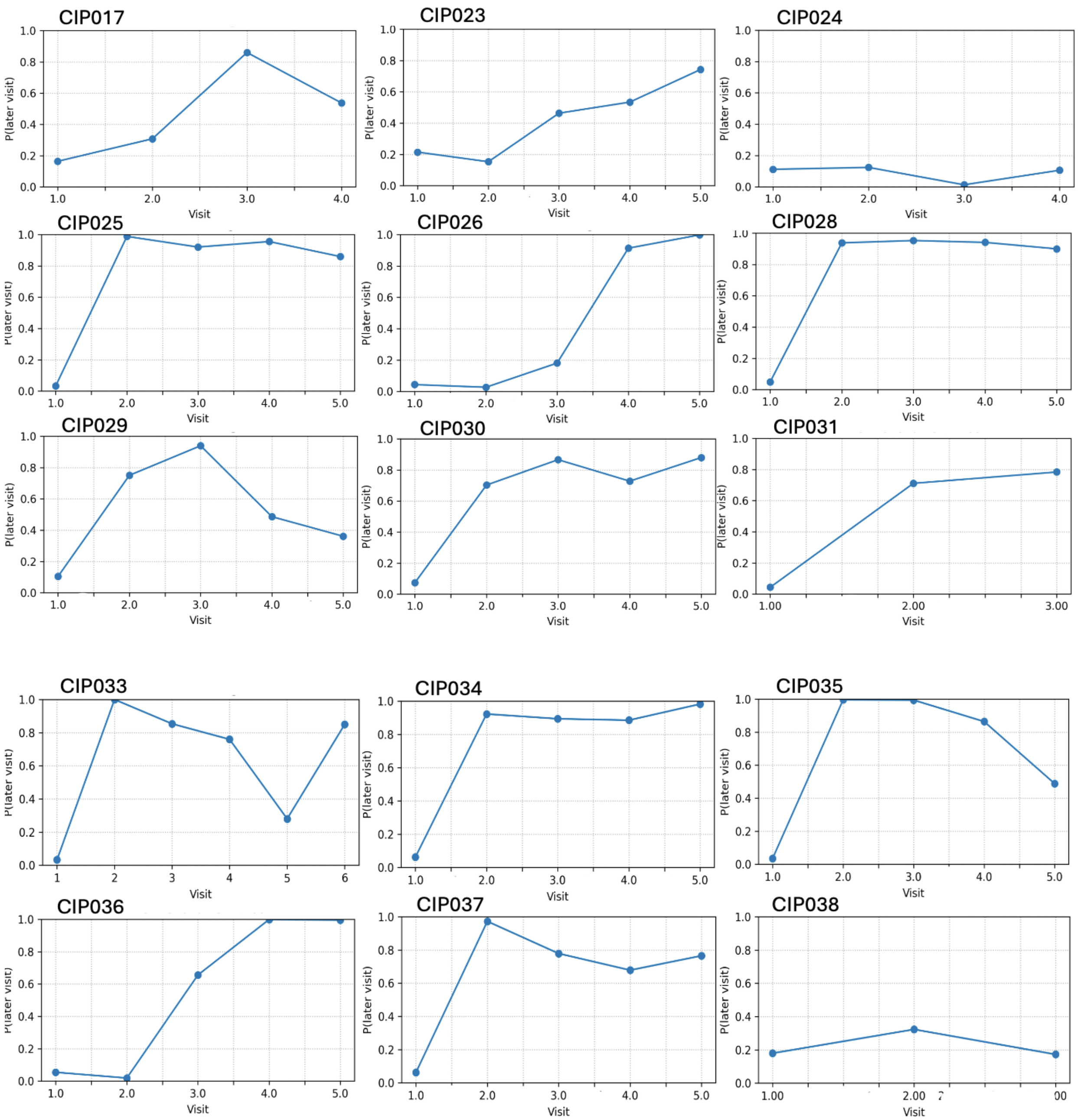
- Left Foot Sensor:





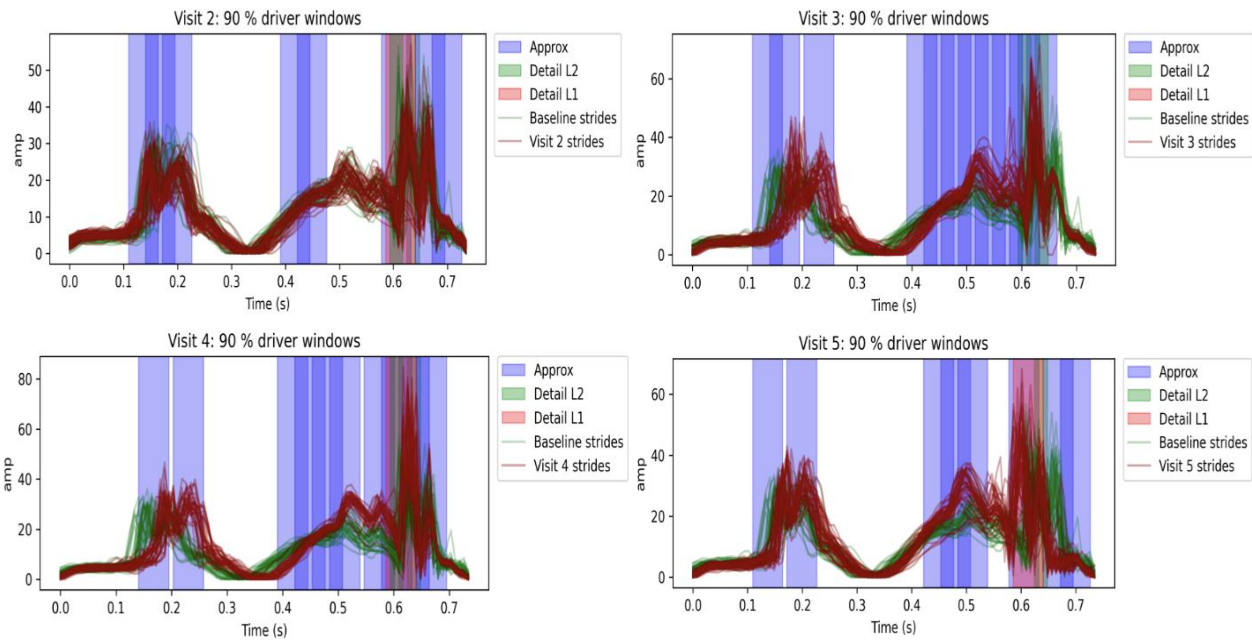
- Right Foot Sensor:



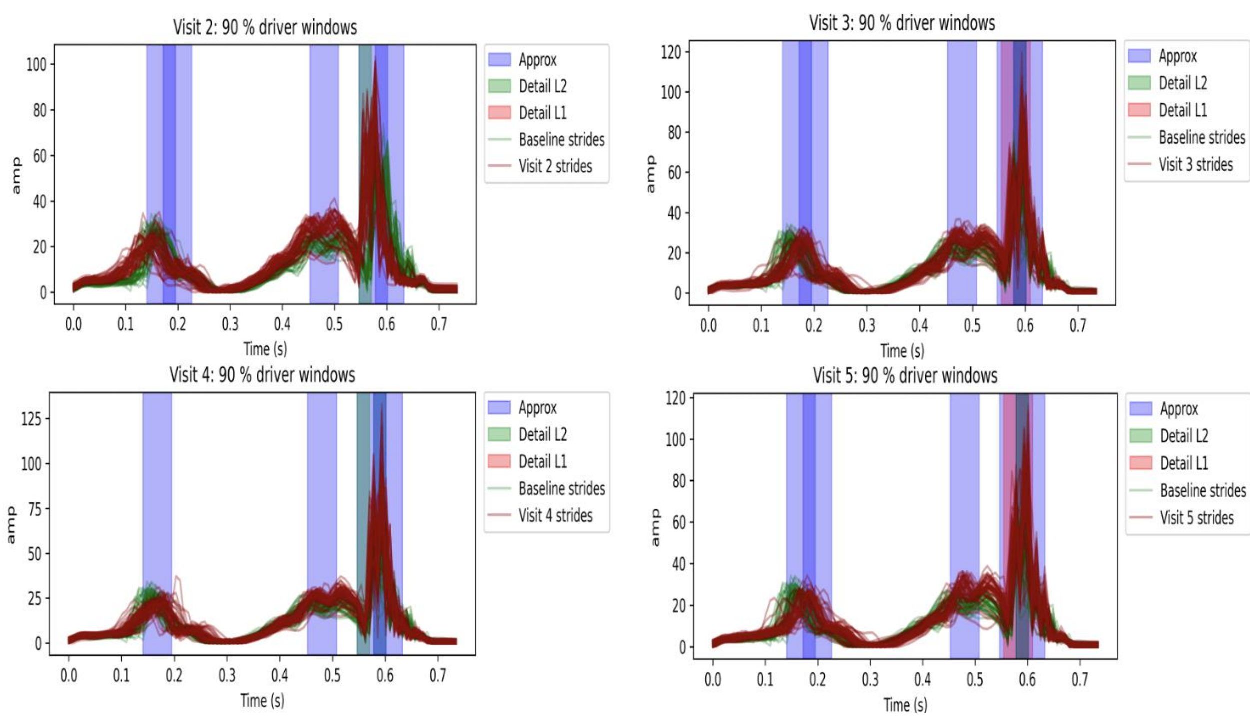

CIP017
CIP023
CIP024
CIP025
CIP026
CIP028
CIP029
CIP030
CIP031
CIP033
CIP034
CIP035
CIP036
CIP037
CIP038
P(later visit)
Visit

Appendix B: Examples of Localized Features Driving Divergence

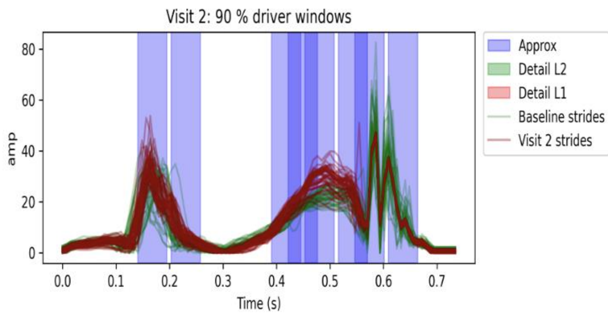
CIP023_Left_Foot:

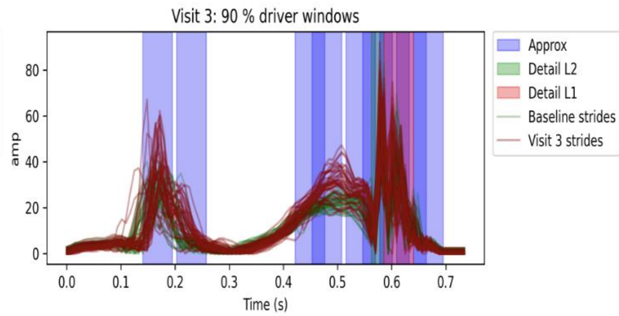


CIP030_Right_Foot:



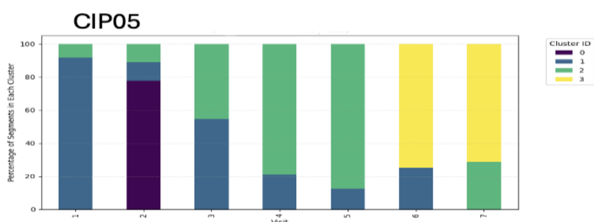
CIP031_Right_Foot:

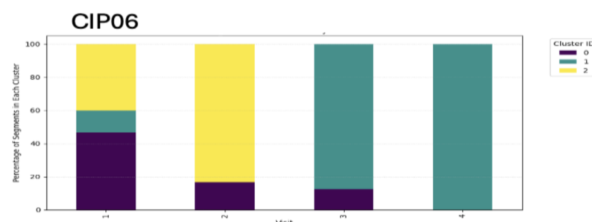


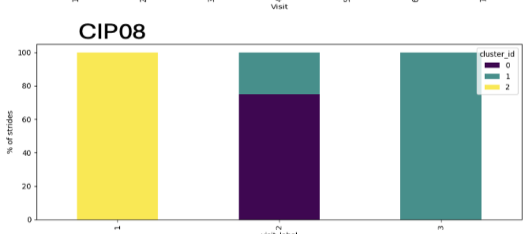


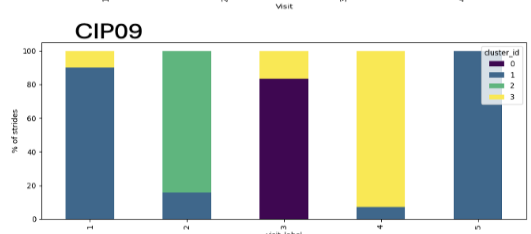
Appendix C: K-means Clustering and Correlation with the Logistic Regression Model

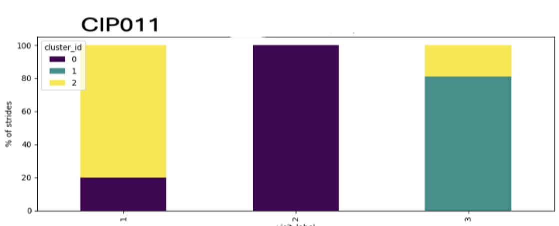
- Left Foot Sensor Clusters:

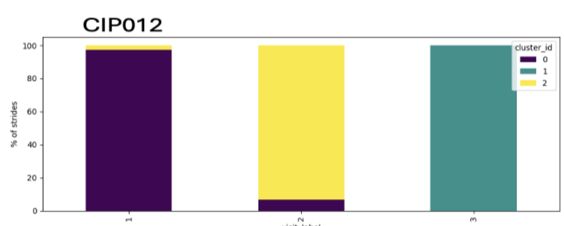


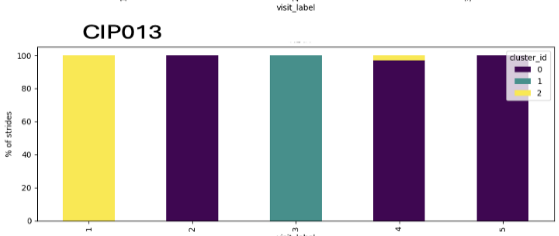


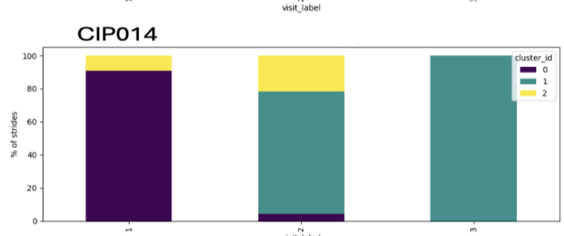




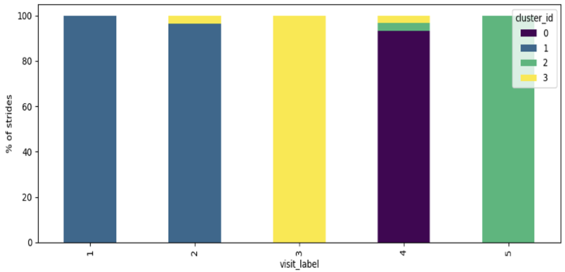




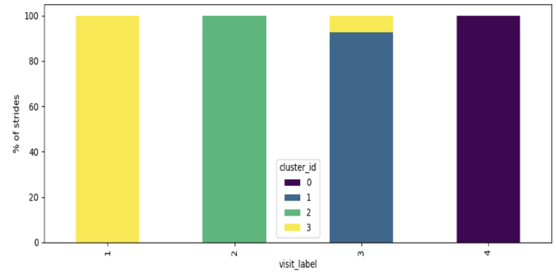




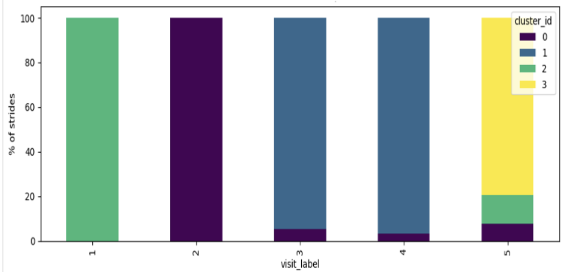
CIP016



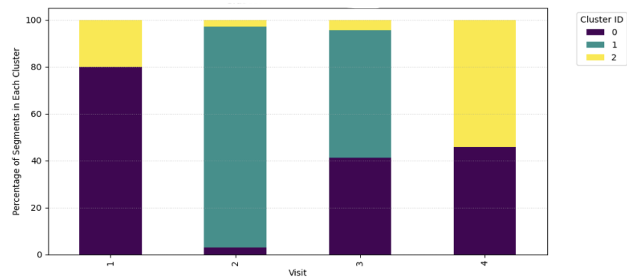
CIP017



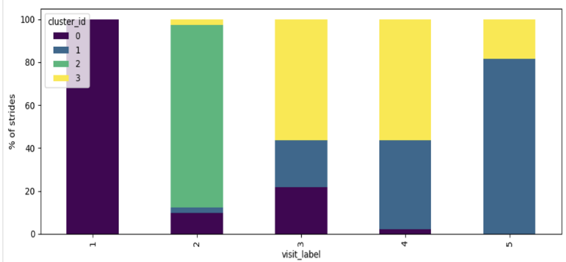
CIP023



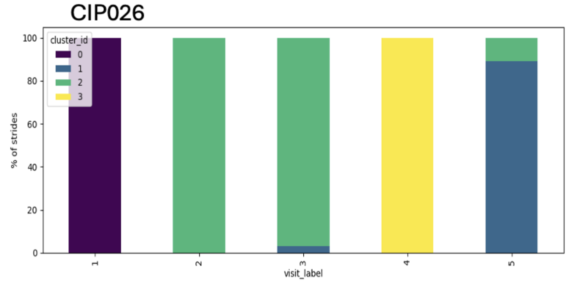
CIP024



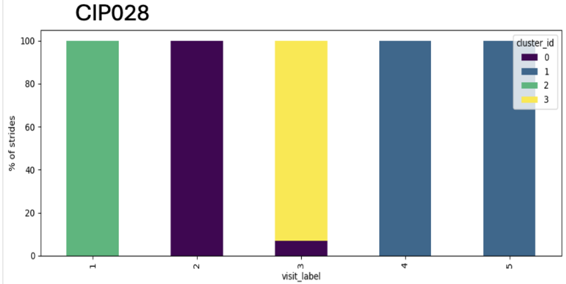
CIP025



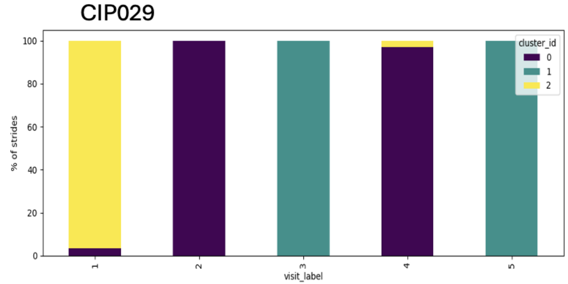
CIP026



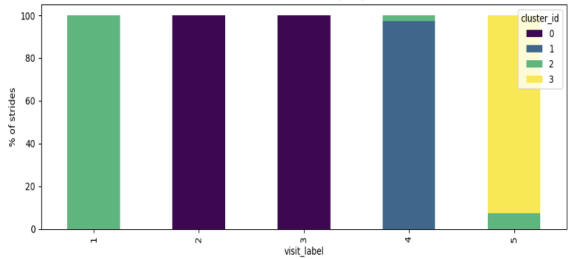
CIP028



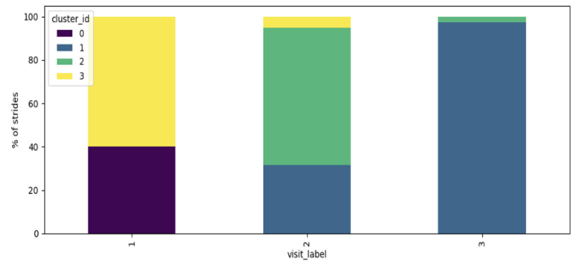
CIP029



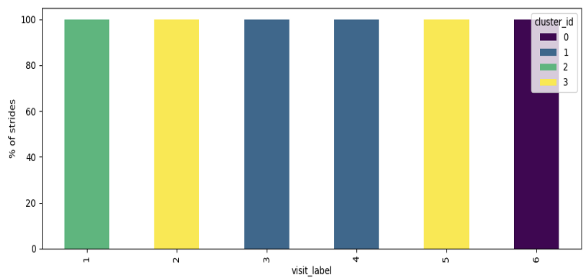
CIP030



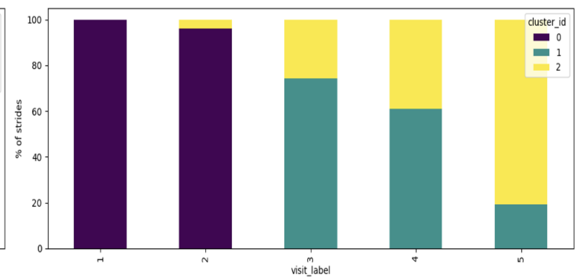
CIP031



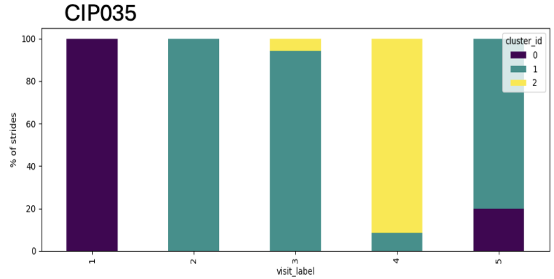
CIP033



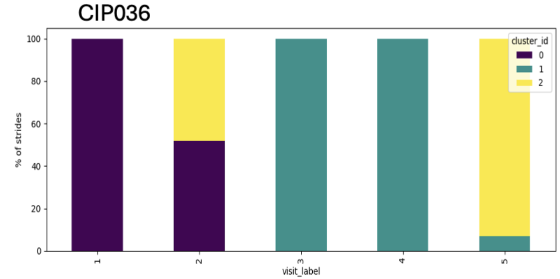
CIP034



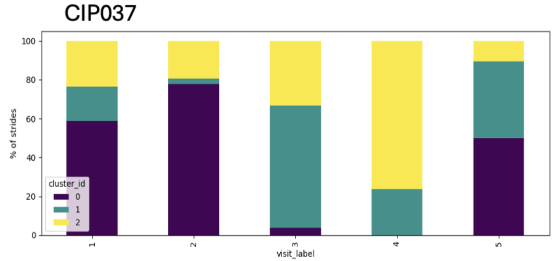
CIP035



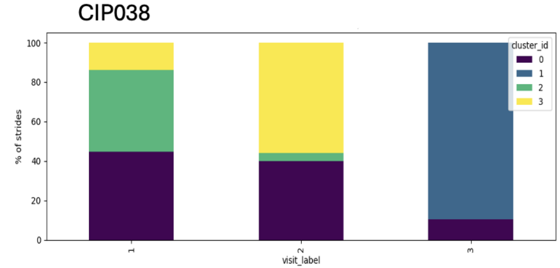
CIP036



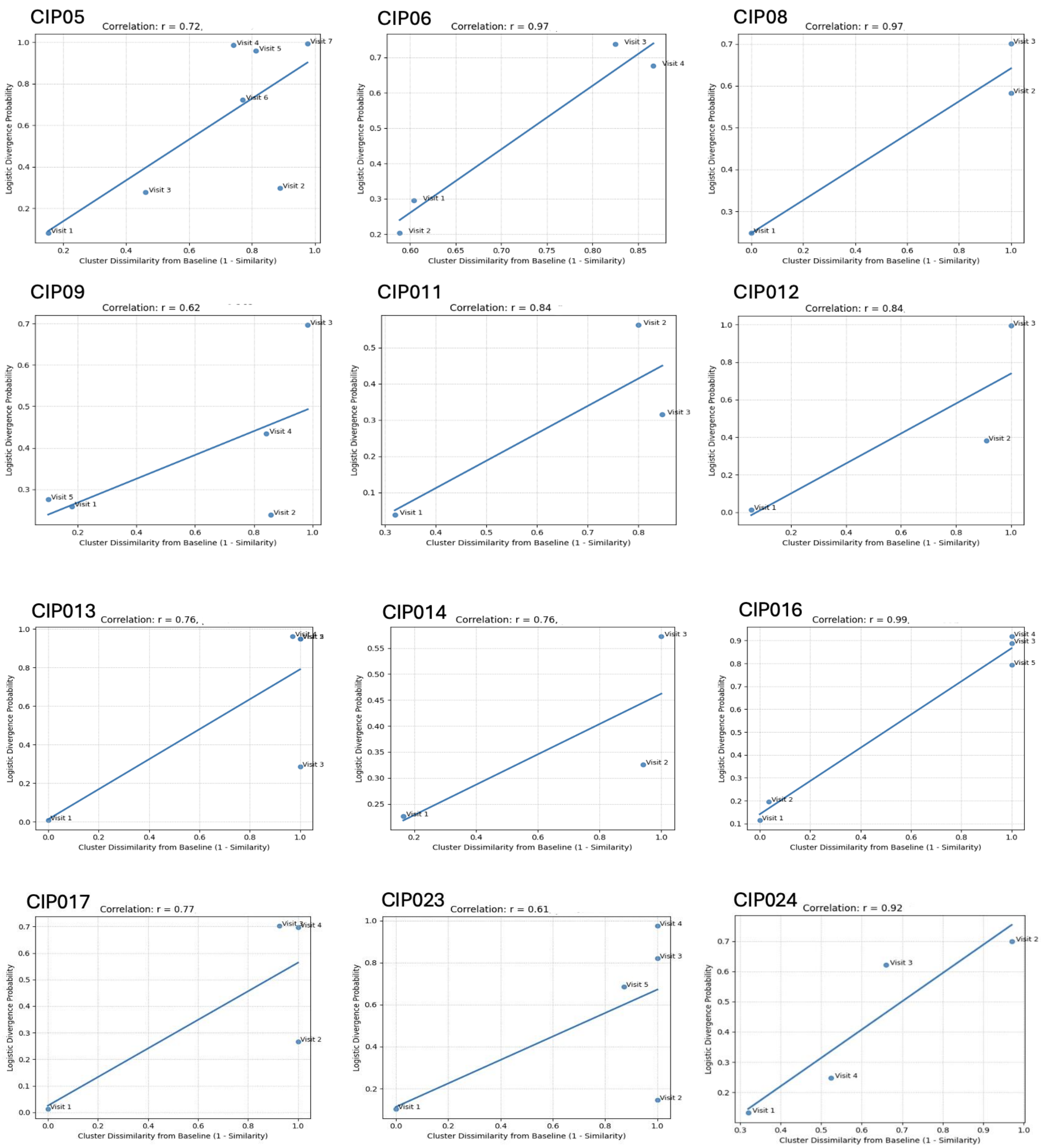
CIP037

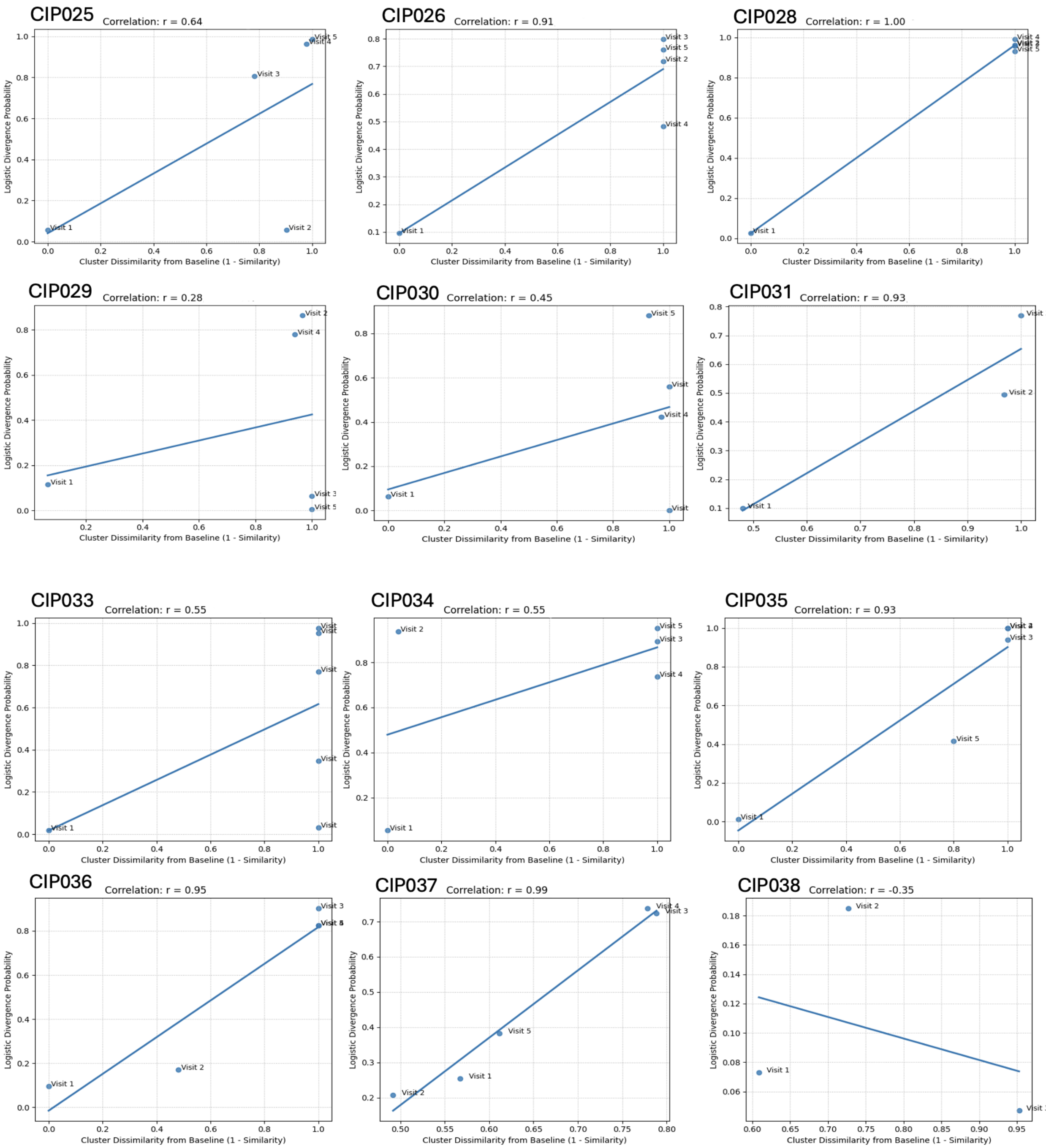


CIP038



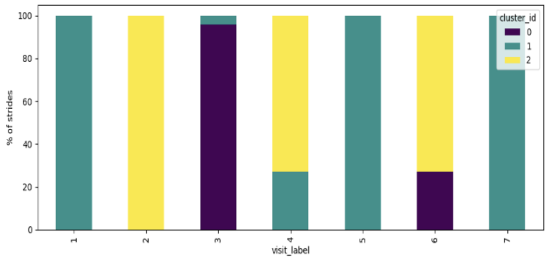
- Left Foot Sensor Cluster Correlation with Divergence Probability:



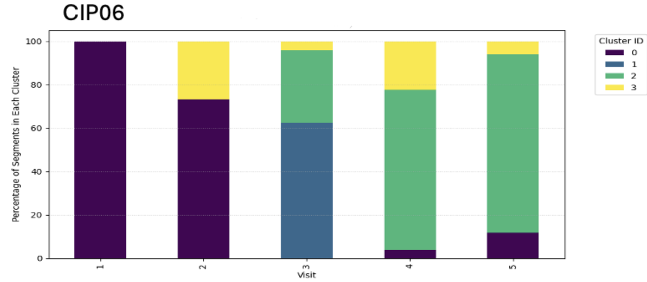

CIP025
Correlation: r = 0.64
CIP026
Correlation: r = 0.91
CIP028
Correlation: r = 1.00
CIP029
Correlation: r = 0.28
CIP030
Correlation: r = 0.45
CIP031
Correlation: r = 0.93
CIP033
Correlation: r = 0.55
CIP034
Correlation: r = 0.55
CIP035
Correlation: r = 0.93
CIP036
Correlation: r = 0.95
CIP037
Correlation: r = 0.99
CIP038
Correlation: r = -0.35
Logistic Divergence Probability
Cluster Dissimilarity from Baseline (1 - Similarity)

- Right Foot Sensor Clusters:

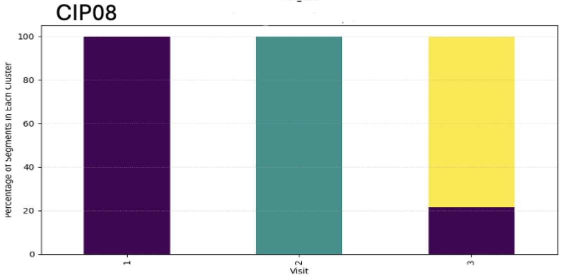
CIP05



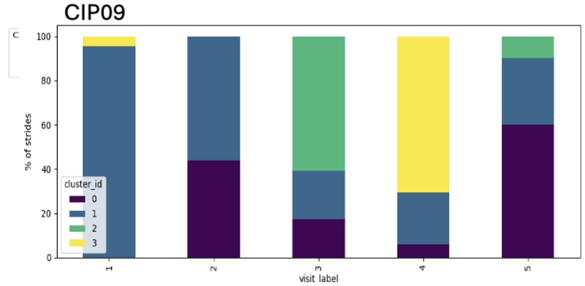
CIP06



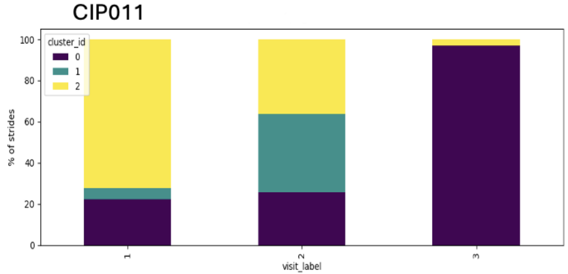
CIP08



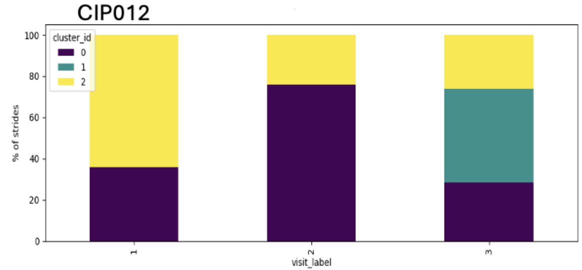
CIP09



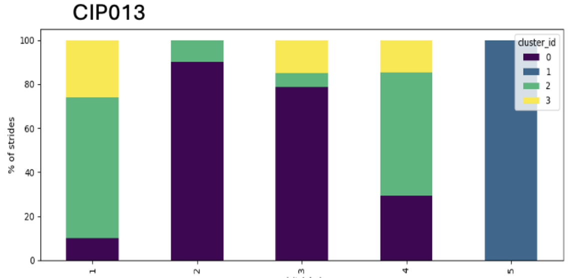
CIP011



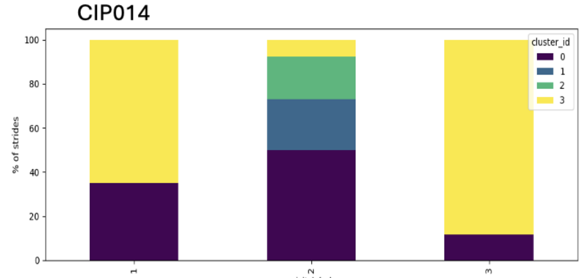
CIP012



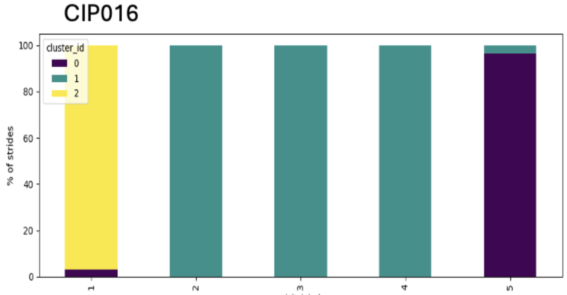
CIP013



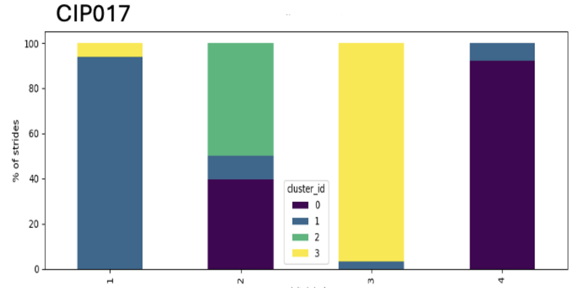
CIP014



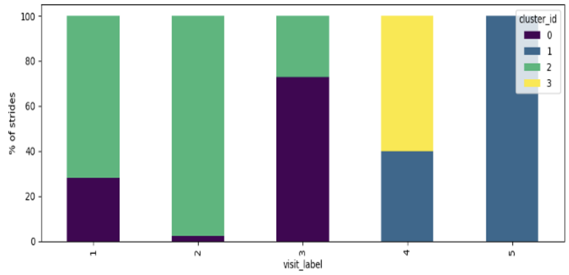
CIP016



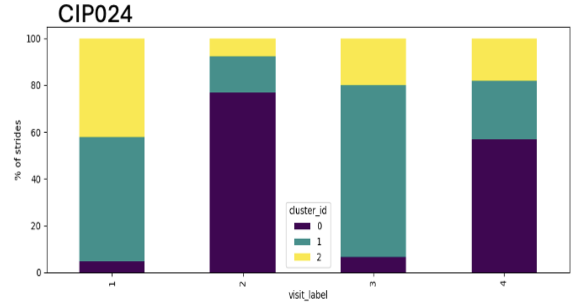
CIP017



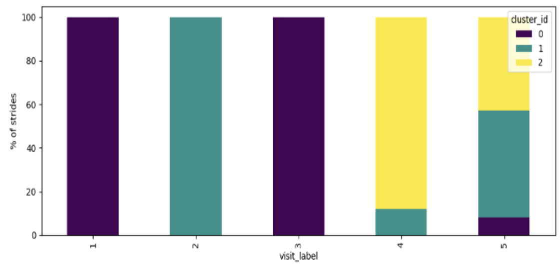
CIP023



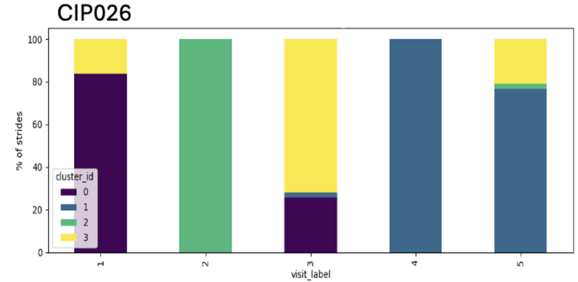
CIP024



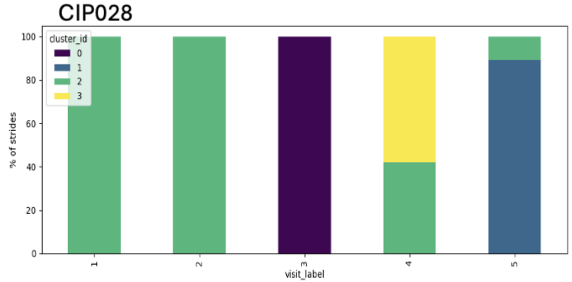
CIP025



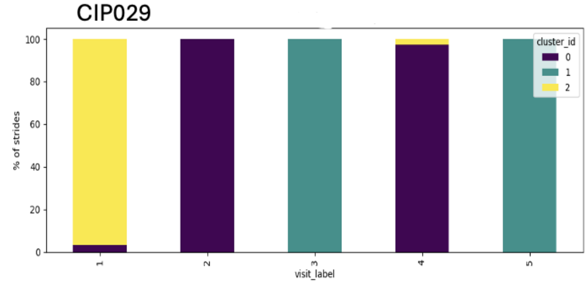
CIP026



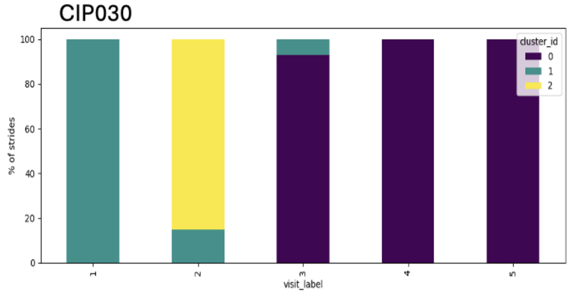
CIP028



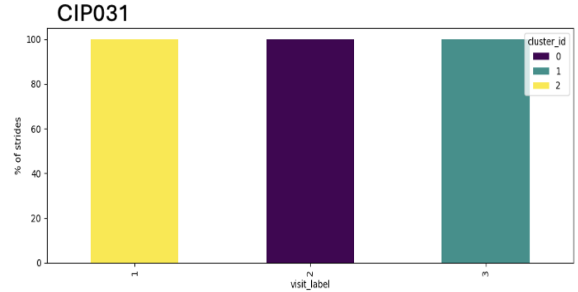
CIP029



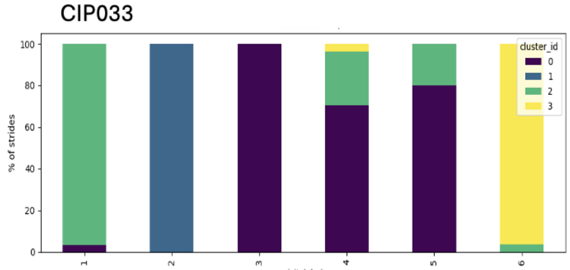
CIP030



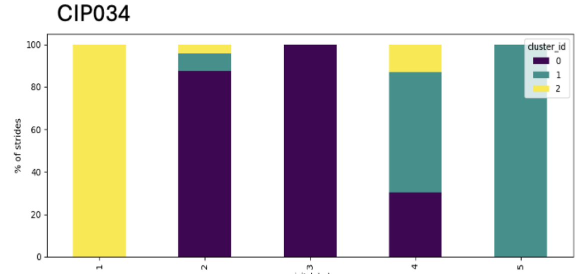
CIP031

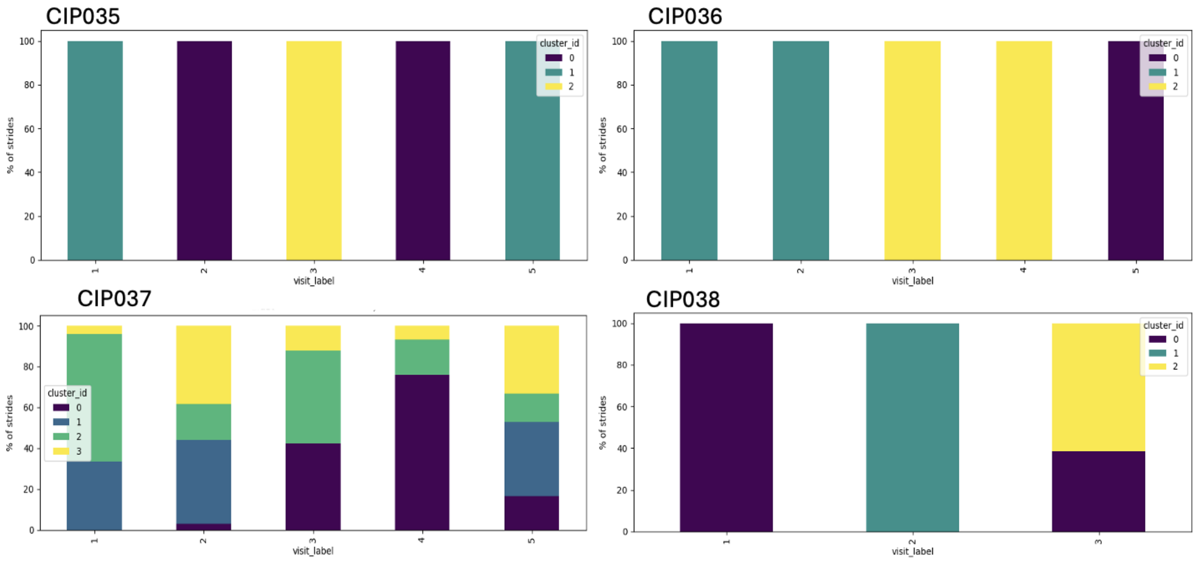


CIP033

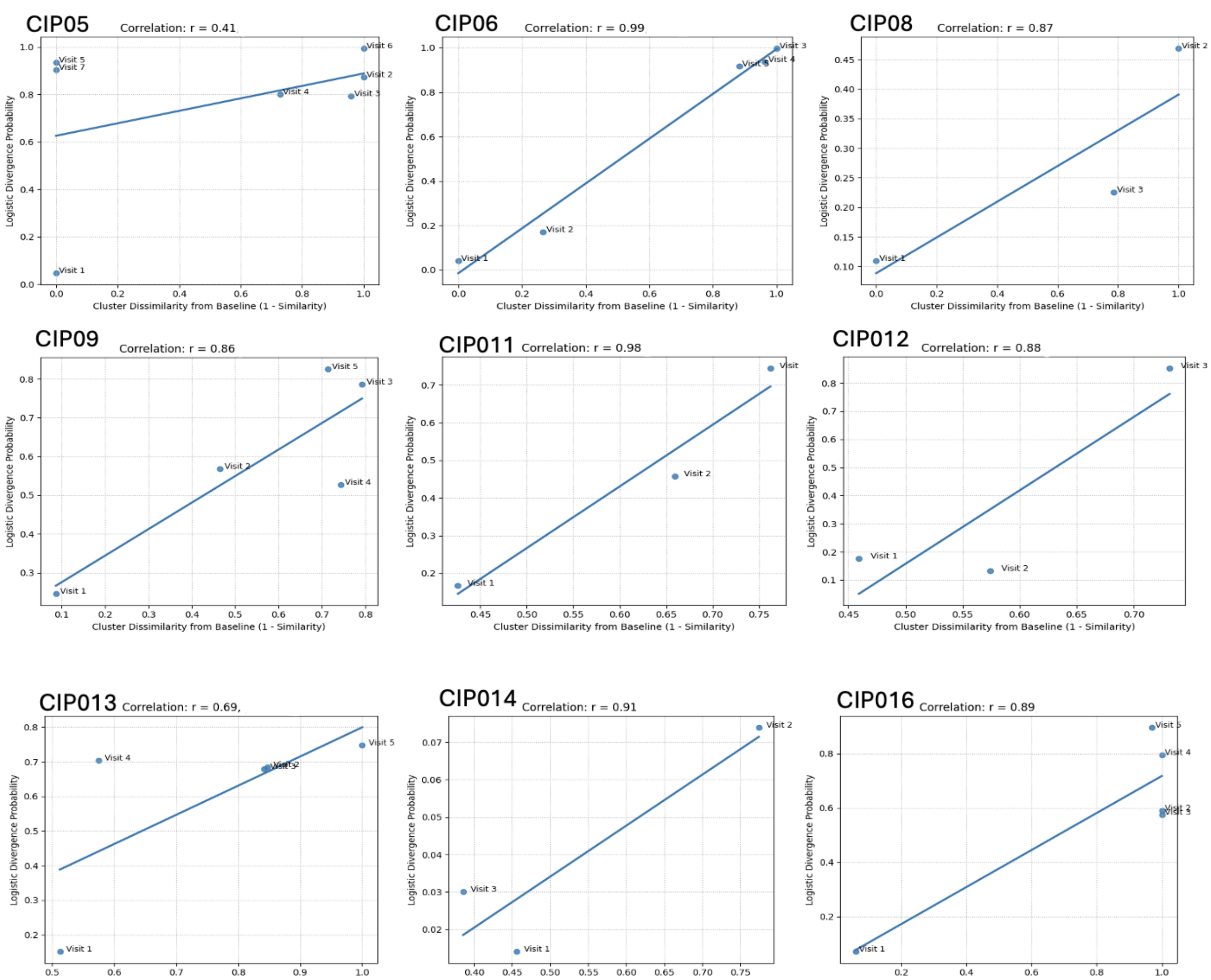


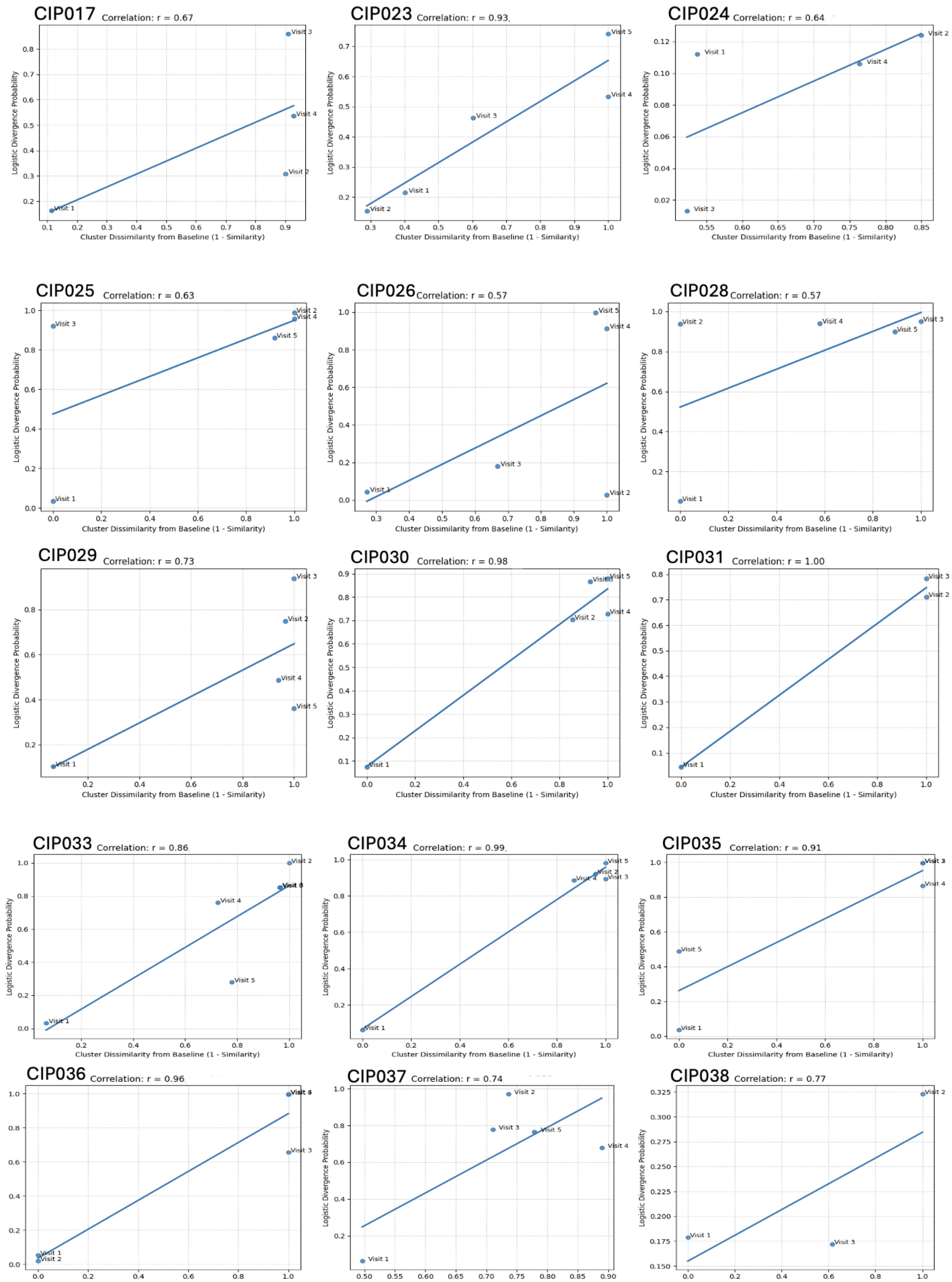
CIP034





- Right Foot Sensor Cluster Correlation with Divergence Probability:




CIP017 Correlation: r = 0.67
CIP023 Correlation: r = 0.93
CIP024 Correlation: r = 0.64
CIP025 Correlation: r = 0.63
CIP026 Correlation: r = 0.57
CIP028 Correlation: r = 0.57
CIP029 Correlation: r = 0.73
CIP030 Correlation: r = 0.98
CIP031 Correlation: r = 1.00
CIP033 Correlation: r = 0.86
CIP034 Correlation: r = 0.99
CIP035 Correlation: r = 0.91
CIP036 Correlation: r = 0.96
CIP037 Correlation: r = 0.74
CIP038 Correlation: r = 0.77
Logistic Divergence Probability
Cluster Dissimilarity from Baseline (1 - Similarity)

Appendix D: Examples of Obtained Flagged Outlier Reports

CIP06_Left_Foot:

Visit 1: cluster reports

· Cluster 1 (n=5)
overall → bands [Approx:9, D2:1]

Phases [Swing:8, Toe-off:2]
stride list → 1:1 , 1:4 , 2:4 , 5:2 , 5:3
location trend → {'end': 3, 'begin': 2}

Visit 2: cluster reports

· Cluster 1 (n=9)
overall → bands[Approx:11, D2:2, D1:2] phases[Swing:12, Toe-off:2, Heel-strike:1] stride list → 1:5 , 2:1 , 2:2 , 2:4 , 3:3 , 3:4 , 3:5 , 4:1 , 4:4
location trend → {'end': 5, 'begin': 3, 'middle': 1}

Visit 3: cluster reports

· Cluster 1 (n=9)

overall → bands[Approx:11, D2:1, D1:1] phases[Swing:10, Toe-off:2, Heel-strike:1] stride list → 1:1 , 1:2 , 1:3 , 1:5 , 2:3 , 2:4 , 3:1 , 5:1 , 5:2
location trend → {'begin': 5, 'end': 3, 'middle': 1}

Visit 4: cluster reports

· Cluster 1 (n=1)
common → bands[Approx:1] phases[Swing:1]
overall → bands[Approx:1] phases[Swing:1]
stride list → 1:1
location trend → {'begin': 1}
· Cluster 2 (n=1)
common → bands[D2:1] phases[Swing:1]
overall → bands[D2:1] phases[Swing:1]
stride list → 2:2
location trend → {'begin': 1}
· Cluster 3 (n=1)
common → bands[Approx:1] phases[Swing:1]
overall → bands[Approx:1] phases[Swing:1]
stride list → 3:1
location trend → {'begin': 1}
· Cluster 4 (n=1)
common → bands[Approx:1] phases[Swing:1]
overall → bands[Approx:1] phases[Swing:1]

stride list → 4:1
location trend → {'begin': 1}

CIP08_Left_Foot

Visit 1: cluster reports

· Cluster 1 (n=8)
overall → bands[Approx:11, D1:1] phases[Swing:7, Toe-off:3, Heel-strike:2] stride list → 1:1 , 1:3 , 1:5 , 2:1 , 2:2 , 3:1 , 3:2 , 3:3
location trend → {'begin': 5, 'end': 2, 'middle': 1}

Visit 2: cluster reports

· Cluster 1 (n=7)
overall → bands[Approx:11, D2:1] phases[Swing:7, Heel-strike:3, Toe-off:2] stride list → 1:1 , 1:3 , 2:1 , 2:3 , 2:4 , 4:2 , 4:4
location trend → {'begin': 3, 'end': 3, 'middle': 1}

Visit 3: cluster reports

· Cluster 1 (n=6)
overall → bands[Approx:12, D1:2, D2:1] phases[Swing:8, Heel-strike:4, Toe-off:3] stride list → 1:2 , 2:1 , 4:1 , 4:2 , 5:1 , 5:4
location trend → {'begin': 5, 'end': 1}

CIP037_Right_Foot

Visit 1: cluster reports

· Cluster 1 (n=14)
overall → bands[Approx:14, D2:4, D1:4] phases[Swing:13, Heel-strike:5, Toe-off:4] stride list → 1:1 , 1:3 , 1:5 , 2:1 , 2:2 , 2:5 , 3:1 , 3:3 , 4:1 , 4:2 , 4:4 , 5:3 , 6:2 , 6:3 location trend → {'begin': 7, 'end': 6, 'middle': 1}

Visit 2: cluster reports

· Cluster 1 (n=4)
overall → bands[D2:2, Approx:1, D1:1] phases[Toe-off:2, Swing:1, Heel-strike:1]
stride list → 9:1 , 9:4 , 12:2 , 12:3
location trend → {'begin': 2, 'end': 2}
· Cluster 2 (n=5)
overall → bands[Approx:11, D2:1] phases[Swing:8, Toe-off:2, Heel-strike:2]
stride list → 5:4 , 8:1 , 10:1 , 11:1 , 11:5
location trend → {'begin': 3, 'end': 2}
· Cluster 3 (n=2)
common → bands[Approx:1] phases[Swing:1]

overall → bands[Approx:1] phases[Swing:1]
stride list → 1:1 , 8:3
location trend → {'begin': 1, 'middle': 1}
· Cluster 4 (n=6)
(none common)
overall → bands[Approx:5, D2:3, D1:3] phases[Swing:8, Heel-strike:2, Toe-off:1]
stride list → 5:1 , 5:2 , 9:3 , 11:3 , 11:4 , 12:1
location trend → {'begin': 3, 'end': 2, 'middle': 1}

Visit 3: cluster reports

· Cluster 1 (n=20)
overall → bands[Approx:16, D2:5, D1:4] phases[Swing:16, Toe-off:5, Heel-strike:4] stride list → 1:1 , 1:5 , 2:1 , 2:4 , 3:1 , 3:2 , 3:3 , 3:4 , 4:1 , 4:2 , 4:3 , 5:1 , 5:2 , 5:3 , 5:4 , 6:1 , 6:4 , 6:5 , 7:2 , 8:5
location trend → {'begin': 10, 'end': 10}

Visit 4: cluster reports

· Cluster 1 (n=4)
overall → bands[Approx:6, D2:1, D1:1] phases[Toe-off:3, Swing:3, Heel-strike:2]
stride list → 2:3 , 7:1 , 8:3 , 8:4
location trend → {'end': 2, 'begin': 1, 'middle': 1}
· Cluster 2 (n=3)
overall → bands[Approx:3, D1:3, D2:1] phases[Swing:5, Heel-strike:2]

stride list → 2:4 , 5:4 , 7:3
location trend → {'end': 3}
· Cluster 3 (n=2)
common → bands[Approx:1] phases[Swing:1] overall → bands[Approx:2, D1:1] phases[Swing:3] stride list → 3:3 , 6:1

location trend → {'end': 1, 'begin': 1}
· Cluster 4 (n=5)
overall → bands[Approx:7, D2:2, D1:2] phases[Swing:8, Heel-strike:2, Toe-off:1] stride list → 1:1 , 2:1 , 4:1 , 6:2 , 8:1

location trend → {'begin': 5}

Visit 5: cluster reports

· Cluster 1 (n=20)
overall → bands[Approx:11, D1:4, D2:3] phases[Swing:13, Heel-strike:3, Toe-off:2] stride list → 1:1 , 1:5 , 2:1 , 2:2 , 2:4 , 2:5 , 3:4 , 3:5 , 4:1 , 4:3 , 4:5 , 5:1 , 5:4 , 5:5 , 6:3 , 7:1 , 7:3 , 7:4 , 8:2 , 8:4
location trend → {'end': 12, 'begin': 7, 'middle': 1}

9 References

[1] Rekant, Julie, Scott Rothenberger, and April Chambers. "Inertial measurement unit-based motion capture to replace camera-based systems for assessing gait in healthy young adults: Proceed with caution." Measurement: Sensors 23 (2022): 100396.

[2] Maihöfner, Christian, et al. "Chemotherapy-induced peripheral neuropathy (CIPN): current therapies and topical treatment option with high-concentration capsaicin." Supportive Care in Cancer 29.8 (2021): 4223-4238.

[3] Shah, Vrutangkumar V., et al. "Chemotherapy-induced peripheral neuropathy and falls in cancer survivors relate to digital balance and gait impairments." JCO precision oncology 8 (2024): e2300312.

[4] Sethi, Dimple, Sourabh Bharti, and Chandra Prakash. "A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and future work." Artificial Intelligence in Medicine 129 (2022): 102314.

[5] Trojaniello, Diana, et al. "Comparative assessment of different methods for the estimation of gait temporal parameters using a single inertial sensor: application to elderly, post-stroke, Parkinson's disease and Huntington's disease subjects." Gait & posture 42.3 (2015): 310-316.

[6] Mei, Chang, Farong Gao, and Ying Li. "A determination method for gait event based on acceleration sensors." Sensors 19.24 (2019): 5499.

[7] Aqueveque, Pablo, et al. "Gait segmentation method using a plantar pressure measurement system with custom-made capacitive sensors." Sensors 20.3 (2020): 656.

[8] Ye, Jing, et al. "An adaptive method for gait event detection of gait rehabilitation robots." Frontiers in Neurorobotics 14 (2020): 38.

[9] Wu, Jiaen, et al. "Human gait-labeling uncertainty and a hybrid model for gait segmentation." Frontiers in neuroscience 16 (2022): 976594.

[10] Haji Ghassemi, Nooshin, et al. "Segmentation of gait sequences in sensor-based movement analysis: a comparison of methods in Parkinson's disease." Sensors 18.1 (2018): 145.

[11] Mantyjarvi, J., et al. “Identifying Users of Portable Devices from Gait Pattern with Accelerometers.” Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, IEEE, 2005, pp. 973–976.

[12] Andersson, Rabé, et al. “Smartphone IMU Sensors for Human Identification through Hip Joint Angle Analysis.” Sensors 24.15 (2024): 4769.

[13] Santos, Geise, Tiago Tavares, and Anderson Rocha. "Reliability and generalization of gait biometrics using 3D inertial sensor data and 3D optical system trajectories." Scientific Reports 12.1 (2022): 8414.

[14] Hulleck, Abdul Aziz, et al. “Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies.” Frontiers in medical technology 4 (2022): 901331.

[15] Zhao, Na, et al. “Quality of life in Parkinson’s disease: A systematic review and meta-analysis of comparative studies.” CNS neuroscience & therapeutics 27.3 (2021): 270-279.

[16] Bergmann, Catherine, et al. “Multiple sclerosis and quality of life: The role of cognitive impairment on quality of life in people with multiple sclerosis.” Multiple sclerosis and related disorders 79 (2023): 104966.

[17] Engvall, Kristina, et al. “Impact of persistent peripheral neuropathy on health-related quality of life among early-stage breast cancer survivors: a population-based cross-sectional study.” Breast Cancer Research and Treatment 195.3 (2022): 379-391.

[18] Ferreira, Marta Isabel ASN, et al. “Machine learning models for Parkinson’s disease detection and stage classification based on spatial-temporal gait parameters.” Gait & Posture 98 (2022): 49-55.

[19] Socie MJ, Motl RW, Sosnoff JJ. Examination of spatiotemporal gait parameters during the 6-min walk in individuals with multiple sclerosis. Int J Rehabil Res. 2014 Dec;37(4):311-6. Doi: 10.1097/MRR.0000000000000074. PMID: 25117855.

[20] Chiaramonte, Rita, and Matteo Cioni. “Critical spatiotemporal gait parameters for individuals with dementia: A systematic review and meta-analysis.” Hong Kong Physiotherapy Journal 41.01 (2021): 1-14.

[21] Wang, Zhenchao, et al. “Gait parameters and peripheral neuropathy in patients with diabetes: A meta-analysis.” Frontiers in Endocrinology 13 (2022): 891356.

[22] Katoulis, Evangelos C., et al. “Gait abnormalities in diabetic neuropathy.” Diabetes care 20.12 (1997): 1904-1907.

[23] Hulshof, Chantal M., et al. “The association between peripheral neuropathy and daily-life gait quality characteristics in people with diabetes.” Gait & Posture 114 (2024): 152-159.

[24] Khalil, Samir H. Assaad, et al. “Impact of diabetic peripheral neuropathy on gait abnormalities in patients with type 2 diabetes mellitus.” Diabetology international 15.1 (2024): 58-66.

[25] Henderson, Adrienne D., et al. “Diabetic gait is not just slow gait: gait compensations in diabetic neuropathy.” Journal of Diabetes Research 2019.1 (2019): 4512501.

[26] Jiang X, Deng F, Rui S, Ma Y, Wang M, Deng B, Wang H, Du C, Chen B, Yang X, Boey J, Armstrong DG, Deng W, Duan X. The Evaluation of Gait and Balance for Patients with Early Diabetic Peripheral Neuropathy: A Cross-Sectional Study. Risk Manag Healthc Policy. 2022;15:543-552
https://doi.org/10.2147/RMHP.S361698

[27] Marshall, Timothy F., et al. "Chemotherapy-induced-peripheral neuropathy, gait and fall risk in older adults following cancer treatment." Journal of Cancer Research and Practice 4.4 (2017): 134-138.

[28] Chen, Ciao-Sin et al. “Chemotherapy-Induced Peripheral Neuropathy Detection via a Smartphone App: Cross-sectional Pilot Study.” JMIR mHealth and uHealth vol. 9,7 e27502. 5 Jul. 2021, doi:10.2196/27502

[29] Müller, Jana, et al. "Chemotherapy-induced peripheral neuropathy: longitudinal analysis of predictors for postural control." Scientific reports 11.1 (2021): 2398.

[30] O’Day, J., Lee, M., Seagers, K. et al. Assessing inertial measurement unit locations for freezing of gait detection and patient preference. J NeuroEngineering Rehabil 19, 20 (2022). https://doi.org/10.1186/s12984-022-00992-x

[31] Tekscan. “Webinar: An Insider’s Look Into Modular Gait Analysis Technology.” Tekscan, 2024, https://www.tekscan.com/resources/on-demand-webinar/webinar-insider%E2%80%99s-look-modular-gait-analysis-technology.

[32] Clario. “PxM Opal Tech Specs V2R.” Clario, May 2024, https://clario.com/wp-content/uploads/2024/05/24_01_PxM_OpalTechSpecsV2R_v2.pdf.

[33] Advanced Navigation. “Inertial Measurement Unit (IMU) – An Introduction.” Advanced Navigation, 2025, https://www.advancednavigation.com/tech-articles/inertial-measurement-unit-imu-an-introduction/.

[34] JOUAV. “A Complete Guide to Inertial Measurement Units (IMU).” JOUAV, 2025, https://www.jouav.com/blog/inertial-measurement-unit.html.

[35] Technaid. “IMU Working Principles.” Technaid, 2017, https://www.technaid.com/publications/imu-working-principles/.

[36] Lokesh, C., et al. “Wavelet OFDM for Power Line Communication.” Proceedings of the 2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT), IEEE, 2017.

[37] Northwestern University. “Wavelet Toolbox.” MATLAB Help, Northwestern University, https://www.ece.northwestern.edu/local-apps/matlabhelp/toolbox/wavelet/dwt.html.

[38] MathWorks. “Choose a Wavelet.” MATLAB & Simulink, MathWorks, https://www.mathworks.com/help/wavelet/gs/choose-a-wavelet.html.