

POLITECNICO DI TORINO

Master's Degree in Biomedical Engineering



Master's Degree Thesis

Data-Driven Evaluation of Optimal IMU Placement on Ice Hockey Helmets

Supervisors

Prof. Marco GAZZONI

Prof. Andrea CEREATTI

Prof. Anisoara IONESCU

Candidate

Sara INCHINGOLO

OCTOBER 2025

ACKNOWLEDGMENTS

Alla mia famiglia

Abstract

Contact sports such as ice hockey expose athletes to a considerable risk of concussions, emphasizing the importance of reliable tools for head impact monitoring. To address this challenge, a smart helmet was developed, embedding inertial measurement units (IMUs) at three distinct locations on the outer shell. Experimental tests were carried out in a controlled laboratory environment, where a pendulum impactor was used to deliver standardized impacts to a dummy’s head according to a predefined protocol. An additional IMU, placed inside the headform, served as the reference system. Each sensor recorded both linear accelerations and angular velocities along the three spatial axes, and the measurements collected from the helmet-mounted IMUs were systematically compared against the reference IMU. The objective of this study is to determine which sensor location provides signals most consistent with the ground truth, thereby identifying the most reliable placement for wearable sensors intended to capture head impact kinematics. Data processing involved segmentation and synchronization of the signals to isolate single impacts, followed by feature extraction in the time, frequency, and time–frequency domains. Features were heuristically selected to provide meaningful descriptors of the signals and to address two research questions: which helmet-mounted IMU exhibits patterns most similar to the ground truth, and whether impact direction influences measurement accuracy. A statistical analysis was conducted to compare features across sensors, while a supervised machine learning framework was employed to assess the degree of decoupling between the helmet and the headform, labeling impacts as either low or high decoupling. The results demonstrate that sensor position has a measurable effect on signal reliability. Moreover, the machine learning analysis provides additional support by quantifying the susceptibility of each sensor to decoupling, thereby offering complementary insights into overall measurement reliability.

Contents

List of Figures	VIII
List of Tables	X
1 Overview	1
1.1 Ice Hockey and the Importance of Rapid Injury Diagnosis	1
1.2 Thesis Objective and Relevance of Sensor Positioning	2
2 Introduction and State of Art	3
2.1 Definition of Traumatic Brain Injury, Mild Traumatic Brain Injury, and Concussion	3
2.2 Current Methodologies for Head Impact Assessment	4
2.2.1 Inertial Measurement Units (IMUs)	4
Physical Principles	4
Accelerometers	5
Gyroscopes	6
2.3 In-Lab Methods for Head Impact Evaluation	6
2.3.1 Real-Time Acquisition Systems	7
3 Methods	10
3.1 Experimental Setup	10
3.1.1 Impact Generation System	10
3.1.2 ATD and Headform Sensor Positioning	11
3.1.3 Sensor Specifications and Helmet IMUs Positioning	12
3.2 Protocol – Data Collection	13
3.2.1 Impact Directions	13
3.2.2 Impact Angles	14
3.2.3 Experimental Protocol	14
4 Data Processing	16
4.1 Preprocessing	17
4.1.1 Functional Calibration	17
4.1.2 Up-Sampling	18

4.1.3	Segmentation and Synchronization of Impacts	19
	Impacts Detection	19
	Extraction of Relevant Signal Segments	19
4.2	Feature Selection and Feature Extraction	20
4.2.1	Translational and Rotational Contributions to Brain Injury	20
4.2.2	Feature Selection	21
	Time-Domain Features	22
	Frequency-Domain Features	24
	Time-Frequency Domain - Wavelet Transform	26
4.2.3	Overview of Extracted Features	28
4.3	Statistical Analysis	29
4.3.1	Hypoteses and Errors	29
4.3.2	Sensor Pairs and Test Design	31
4.4	Machine Learning	32
4.4.1	ML Overview	34
4.4.2	Datasets	35
4.4.3	Train-Test Splitting	35
4.4.4	Random Forest	36
4.4.5	Support Vector Machines	37
4.4.6	Evaluation Metrics	38
5	Results and Discussions	40
5.1	Visual Inspection of Angular Velocity Signals	40
5.2	Feature Extraction Analysis	43
5.3	Quantitative Comparison and Statistical Results	46
5.3.1	Correlation between extracted features	46
5.3.2	Interpretation of the Statistical Results	47
5.3.3	Statistical Summary: Sensor Comparison and Impact Direction Influence	54
5.4	Machine Learning Results	55
5.5	Limitations	57
6	Summary and Conclusions	59
A	Impact Synchronization Analysis	61
A.1	Motivation and Objectives	61
A.2	Methodology	61
A.2.1	Synchronization among Physilog Sensors	61
A.3	Synchronization between Physilog Middle and Headform	62
A.4	Data Processing and Alignment	63
A.4.1	Trigger Detection	63

A.4.2	Drift Correction and Interpolation	63
A.4.3	Impact Segmentation	65
A.5	Results and Discussions	65
A.6	Conclusion	66
B	Principal Component Analysis for Impact Direction Verification	67
B.1	Principal Component Analysis	67
B.2	Methodology	68
B.2.1	Preprocessing and Impact Segmentation	68
B.2.2	Principal Component Analysis for Impact Direction	68
B.2.3	Impact Vector and Reference Directions	68
B.2.4	Angle Calculation	68
B.2.5	Classification by Angular Ranges	69
B.3	Results and Limitations	69
B.4	Conclusion	71
	Bibliography	72

List of Figures

2.1	Progression of TBI severity.	4
2.2	Second-order mass-spring-damper model [9].	5
2.3	Coriolis force acting on an oscillating mass inside a MEMS gyroscope [10].	6
2.4	Hybrid-III head and neck model [11]	7
2.5	Common laboratory impact configurations for helmet evaluation: (a) pendulum impactor, (b) linear impactor.	8
2.6	Examples of head impact sensor configurations.	9
3.1	Experimental Setup (pendulum impactor, sliding table, and headform assembly).	11
3.2	Sensor placement on the helmet: Physilog Top, Middle, and Bottom. . .	12
3.3	Impact Directions.	14
4.1	Overview of the analytical pipeline.	17
4.2	Feature selection overview.	22
4.3	Low vs High Decoupling.	33
4.4	Machine Learning approaches.	34
5.1	Location: Front, Pendulum arm angle: 32.	42
5.2	Location: Side, Pendulum arm angle: 50.	42
5.3	Location: Right Front-Boss, Pendulum arm angle: 70.	42
5.4	Amplitude of angular velocity signals across impact directions and angles, for the Headform reference and helmet-mounted sensors (Top, Middle, Bottom).	43
5.5	Band power in the 0–50 Hz range: distribution of spectral energy across impact directions.	45
5.6	Band power in the 50-100 Hz range: distribution of spectral energy across impact directions.	46
5.7	Band power in the 100-250 Hz range: distribution of spectral energy across impact directions.	46
5.8	Heatmap Time Domain (Side, 32°)	47
5.9	Heatmap Frequency Domain (Front Boss, 50°)	48
5.10	Heatmap Time-Frequency Domain (Front, 32°)	48

5.11	Visual summary of paired t-test results comparing the Headform reference sensor with helmet-mounted Physilog sensors (Top, Middle, Bottom) for Frequency-domain features, Front impact direction, 32°.	50
5.12	Visual summary of paired t-test results comparing the Headform reference sensor with helmet-mounted Physilog sensors (Top, Middle, Bottom) for Frequency-domain features, Front-Boss impact direction, 50°.	51
5.13	Visual summary of paired t-test results comparing the Headform reference sensor with helmet-mounted Physilog sensors (Top, Middle, Bottom) for Time-domain features, Front impact direction, 50°.	52
5.14	Visual summary of paired t-test results comparing the Headform reference sensor with helmet-mounted Physilog sensors (Top, Middle, Bottom) for Time-domain features, Side impact direction, 50°.	52
5.15	Visual summary of paired t-test results comparing the Headform reference sensor with helmet-mounted Physilog sensors (Top, Middle, Bottom) for Time-Frequency domain features, Front impact direction, 32°.	53
5.16	Final statistical synthesis: total number of features per impact condition satisfying both criteria ($p \geq 0.001$ and $d < 0.5$). The three curves represent the different helmet-mounted sensors, while the point coloration indicates Cohen's d effect size.	55
A.1	Positioning of the Physilog sensors during shaking for synchronization. . .	62
A.2	Positioning of the Physilog Middle sensor on the dummy's face during trigger events.	64
A.3	a) Alignment at the first trigger event; b) Drift observed at the second trigger event.	64
A.4	a) First impact event; b) Second impact event.	65
B.1	Principal Component Analysis.	70

List of Tables

3.1	Technical specifications of the inertial sensors used in the experiments. . .	12
3.2	Impact directions and corresponding rotational angles relative to the frontal position.	13
3.3	Impact angles and corresponding kinetic energy levels.	14
4.1	Summary of Paired t-test analysis settings	32
4.2	Dataset sizes before and after Train-Test splitting.	36
5.1	Accuracy and F1 Score for RF and SVM across datasets	56
6.1	Summary of Research Questions (RQ), Hypotheses (H), and Results . . .	60
B.1	Angular ranges for impact direction classification.	69
B.2	Macro-average F1-scores for impact classification using different sensors. .	71

Chapter 1

Overview

1.1 Ice Hockey and the Importance of Rapid Injury Diagnosis

Ice hockey is currently played in 76 countries and continues to grow in popularity worldwide. This high-intensity sport is characterized by rapid gameplay and frequent physical contact between players. Athletes can reach speeds approaching 50 km/h on skates, utilize composite sticks capable of shooting pucks at more than 160 km/h, and compete within enclosed arenas bounded by rigid plexiglass barriers [1]. These dynamic and high-speed conditions make ice hockey one of the most biomechanically traumatic team sports, especially at competitive levels, where body checks and high-impact collisions are routine. Head impacts in sport can result in a wide spectrum of concussive injuries. Although many athletes experience transient symptoms such as headache, dizziness, or confusion that resolve quickly, others may suffer from persistent impairments or long-term consequences due to cumulative trauma [2]. The timely and accurate recognition of these injuries during play has become a critical component of athlete safety, with increasing emphasis on the development of evidence-based protocols for the diagnosis and treatment of concussions. Medical personnel must be equipped with reliable tools to facilitate rapid on-field evaluation and informed decisions about athlete removal and return-to-play timelines.

In response to increasing awareness of concussion risks, several wearable technologies have been developed to monitor head impacts in real time. These systems, often integrated into helmets or other athletic equipment, are designed to capture data on the magnitude, direction, and other general information about impacts sustained during play. The recorded information can be transmitted to sideline staff, offering immediate insights into potentially injurious events and supporting timely medical evaluation [3].

In addition to helmet-based systems, another promising technology in concussion monitoring is the instrumented mouthguard. These devices are equipped with accelerometers and gyroscopes that measure linear and angular head impact kinematics with high temporal fidelity by coupling directly to the skull via the upper dentition [4]. However, a key

limitation of instrumented mouthguards is their predominant deployment in American football settings, which restricts the generalizability of mouthguard-based findings to other sports, such as ice hockey, where helmet design, impact dynamics, and playing conditions differ.

Despite their promise, these technologies are not without limitations. Challenges include the novelty of such systems, variability in biomechanical interpretation, and limited clinical validation for diagnostic purposes. Consequently, while they represent a step forward in head injury monitoring, their current role remains supportive rather than definitive in clinical decision making.

1.2 Thesis Objective and Relevance of Sensor Positioning

This thesis investigates whether a specific placement of inertial measurement units (IMUs) on an ice hockey helmet can accurately approximate head kinematics. The core objective is to identify, through a data-driven approach, the helmet-mounted sensor location that best reflects the actual motion of the head during impact events. To achieve this, data collected from multiple IMUs positioned at various locations on the helmet are systematically compared against a reference. These comparisons are assessed by using statistical analyses and machine learning techniques, to evaluate the degree of similarity and predictive potential of the recorded signals, enhancing the effectiveness of wearable monitoring systems employed for concussion detection and biomechanical analysis in real-world sports applications.

Within this framework, the study aims to address the following research questions:

- **RQ1:** Which of the helmet-mounted IMUs exhibits signal patterns most similar to those recorded by the reference IMU?
- **RQ2:** Does the direction of impact affect the accuracy of the measurements recorded by the different IMUs?

To support the analysis, the following hypotheses were formulated:

- **H1:** IMUs placed in structurally isolated regions of the helmet are expected to provide signals that better reflect actual head motion.
- **H2:** IMUs located closer to the point of impact are expected to show greater deviations from the ground truth signal.

Chapter 2

Introduction and State of Art

2.1 Definition of Traumatic Brain Injury, Mild Traumatic Brain Injury, and Concussion

Traumatic Brain Injury (TBI) is a complex clinical condition caused by external mechanical forces acting on the head, often as a result of sporting incidents, vehicular collisions, or accidental falls. It is a public health challenge, that affects annually more than 50 million people of all ages [5]. Among the various types of brain injury, TBI covers a broad spectrum of severity, ranging from critical and life-threatening conditions to mild impairments in brain function. One of the most common and widespread forms of TBI is mild traumatic brain injury (**mTBI**), which, despite its name, can result in long-lasting neurocognitive disturbances, frequently underdiagnosed, especially when symptoms appear mild or resolve quickly. Despite the seemingly minor nature of its symptoms, clinical studies have shown associations between mTBI and altered cerebral blood flow, as well as an increased risk of developing chronic neurological conditions [5].

Within the spectrum of mTBI, concussions constitute a well-defined subclass. A **concussion** is typically characterized as a temporary disturbance in brain function caused by rapid acceleration and deceleration of brain tissue, often triggered by rotational or translational forces [6]. Although in everyday language and clinical practice the terms **concussion** and **mTBI** are frequently used interchangeably, this simplification overlooks a fundamental distinction. By medical definition, all concussions are a form of mTBI, but not all mTBIs are concussions [6]. This distinction is critical for an accurate diagnosis, prognosis, and the development of targeted therapeutic strategies. Clarifying this relationship is especially important in contact sports, where athletes are regularly exposed to high-energy collisions and repetitive head impacts.

Figure 2.1 illustrates the continuum of traumatic brain injuries (TBI), showing how head impacts can range in severity from subclinical or asymptomatic events that require no diagnosis to more serious conditions, including moderate and severe TBI.

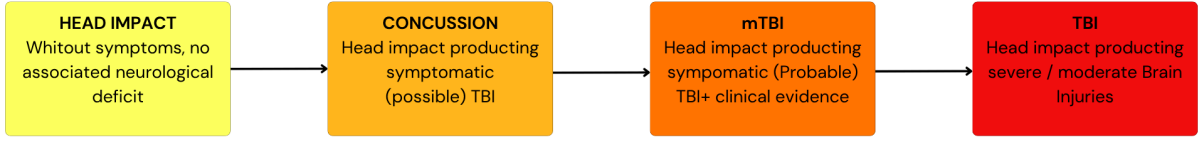


Figure 2.1: Progression of TBI severity.

2.2 Current Methodologies for Head Impact Assessment

2.2.1 Inertial Measurement Units (IMUs)

Inertial Measurement Units (IMUs) are compact devices that integrate multiple inertial sensors, typically triaxial accelerometers and gyroscopes, and in some cases magnetometers, to measure a body's linear acceleration and angular velocity. These devices have become fundamental tools in fields such as biomechanics, clinical monitoring, and movement analysis because of their small size, low energy consumption, and affordability. Their wide adoption has been made possible by advancements in MEMS (Micro-Electro-Mechanical Systems) technology, which enables the fabrication of micro-scale mechanical components integrated with electronics [7].

One of the key advantages of IMUs is their ability to capture motion in real-time through wearable configurations, providing continuous monitoring in naturalistic environments. From a biomechanical perspective, their role is particularly relevant in the study of head injuries. Since the human brain is especially sensitive to rotational kinematics, abrupt angular accelerations or decelerations of the head can lead to shear forces within the brain tissue, potentially causing traumatic brain injuries (TBIs) such as diffuse axonal injury (DAI). Research has shown that six-axis IMUs, combining three-axis accelerometers and three-axis gyroscopes, can significantly improve the prediction and analysis of these injuries [8].

Physical Principles

IMUs operate on the basis of the principle of inertia, which is the tendency of a body to resist changes in its state of motion. The motion of a rigid body is described by the cardinal equations of dynamics.

First cardinal equation (translational motion):

$$\sum \vec{F} = m \cdot \vec{a} \quad (2.1)$$

The sum of external forces acting on a body is equal to the product of its mass and the

acceleration of its center of mass.

Second cardinal equation (rotational motion):

$$\sum \vec{M} = \mathbf{I} \cdot \vec{\alpha} + \vec{\omega} \times (\mathbf{I} \cdot \vec{\omega}) \quad (2.2)$$

where $\mathbf{I}$ is the inertia tensor, $\vec{\alpha}$ is the angular acceleration, and $\vec{\omega}$ is the angular velocity. The sum of external forces acting on a body is equal to the product of its mass and the acceleration of its center of mass.

Accelerometers

Accelerometers are sensors that measure the proper acceleration, i.e., the acceleration experienced by the sensor excluding gravity. Most IMUs use triaxial accelerometers, which can detect motion along three orthogonal axes.

The working principle of an accelerometer is often modeled as a second-order mass-spring-damper system. A proof mass is suspended within the sensor structure through compliant springs and dampers. When the device is accelerated, the inertia of the mass causes it to move relative to the frame. This displacement alters the capacitance between the conductive plates, and the resulting signal is proportional to the applied acceleration [7], as shown in Figure 2.2.

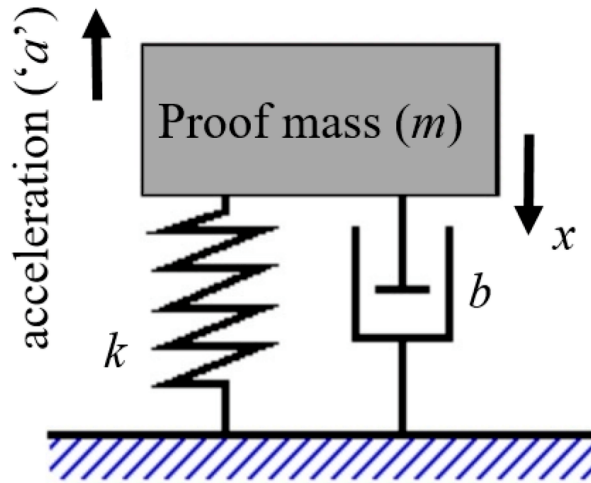


Figure 2.2: Second-order mass-spring-damper model [9].

There are various types of accelerometer, each exploiting different physical phenomena:

- Capacitive (most commonly used in wearable devices)
- Piezoresistive
- Piezoelectric
- Optical
- Tunneling-based

Among these, capacitive accelerometers are favored for wearable and mobile applications due to their high sensitivity, low power consumption, and ease of integration.

Gyroscopes

Gyroscopes measure the angular velocity of the sensor and are essential for capturing rotational motion. The most widespread MEMS gyroscopes are based on the Coriolis effect, an apparent force observed in rotating reference frames.

As shown in Figure 2.3, to understand this effect, a small mass oscillates within the sensor, when the device undergoes rotation, the Coriolis force causes a perpendicular deflection of the mass. This deflection is proportional to the angular velocity and can be measured to infer rotational motion [7], [10]:

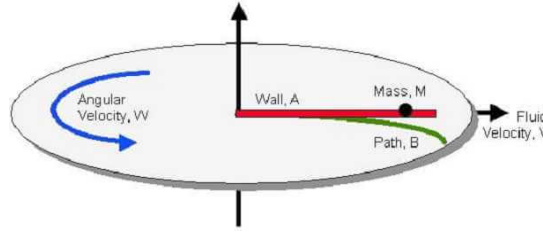


Figure 2.3: Coriolis force acting on an oscillating mass inside a MEMS gyroscope [10].

The gyroscope structure can be conceptualized as a second-order mechanical system, with stiffness elements along two axes: the drive axis, where motion is actively induced, and the sense axis, where deflection due to the Coriolis force is detected.

2.3 In-Lab Methods for Head Impact Evaluation

In the context of head impact data collection in American football and ice hockey, the current state of the art in the literature is characterized by a continuous refinement of methodologies and technologies aimed at achieving accurate and reliable measurements of impact kinematics.

Helmet impact tests play a key role in the design and regulation of protective equipment. When evaluating helmet performance, it is important to recognize that a helmet equipped with inertial sensors does not directly measure the true head kinematics. This limitation arises from the relative motion between the helmet and the skull, which prevents helmet-mounted sensors from serving as a definitive ground truth. For this reason, reference systems are required to obtain accurate head kinematics during impacts. This role is fulfilled by anthropomorphic *headforms*, which are designed to replicate the physical and dynamic properties of the human head.

Modern impact evaluations extend this concept by incorporating a deformable neck into the headform assembly. The purpose is to reproduce the natural coupling between the head and cervical spine. By doing so, it allows the headform to exhibit angular kinematics representative of real-world impacts, in addition to linear accelerations already measurable with a rigid headform configuration.

Linear accelerations of the head are typically measured through a tri-axial accelerometer or an IMU embedded within the headform, positioned in close proximity to the estimated center of mass (CoM) of the head. This location provides a physiologically meaningful reference point, as it approximates the true dynamics of the head's translational motion, enabling the headform to serve as a gold-standard reference against which helmet-mounted sensors can be evaluated.

The Hybrid-III head and neck model (HIII-HN) is among the most widely adopted systems in this context and is extensively used to assess the effectiveness of professional helmets under controlled impact conditions. For this purpose, the Hybrid-III head and neck model (HIII-HN) is widely used in impact tests to assess the effectiveness of professional helmets.

Figure 2.4

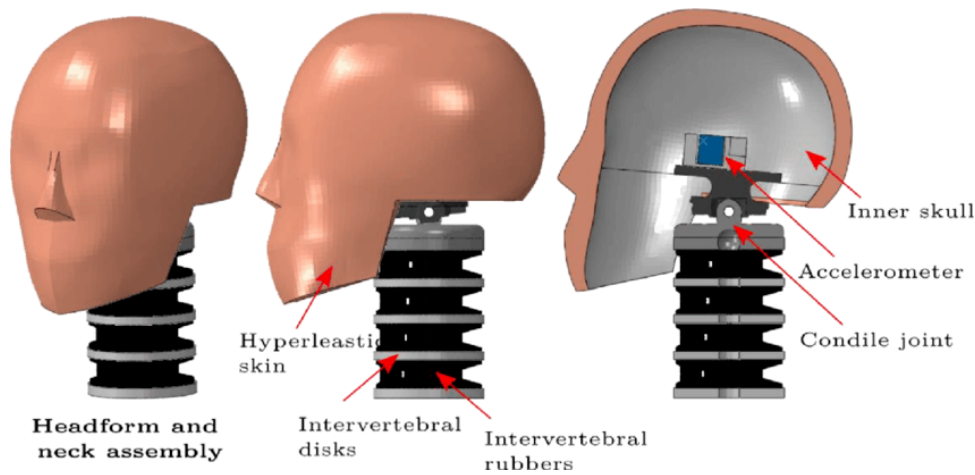


Figure 2.4: Hybrid-III head and neck model [11]

Two main impact test configurations are primarily referenced in the literature:

- **Rigid Pendulum Impacts:** In this setup, the pendulum consists of a rigid arm with a mass at its end, which is released from a fixed height to swing and strike a helmeted Hybrid III (HIII-HN) headform mounted on a test fixture [12].
- **Linear Impacts:** The impactor typically consists of a cylindrical ram equipped with a deformable end cap (often made of elastomeric material). This ram is driven forward using a pneumatic or spring-loaded mechanism along a guided rail, striking the headform at pre-defined impact velocities (5.5, 7.5, and 9.3 m/s) [13].

Figure 2.5 shows these two common impactor configurations used in standardized helmet testing protocols.

2.3.1 Real-Time Acquisition Systems

Numerous wearable sensors have been developed to measure head impacts, these systems can generally be categorized into two main types:

- **Helmet-Based Systems:** These are among the most commonly used tools to monitor head impact biomechanics in high-contact sports such as football and ice



(a) Rigid pendulum impactor setup [13].



(b) Linear impactor setup [13].

Figure 2.5: Common laboratory impact configurations for helmet evaluation: (a) pendulum impactor, (b) linear impactor.

hockey. Two widely cited examples in the literature are the Head Impact Telemetry (HIT) system and the gForce Tracker (GFT), which differ substantially in sensor placement, configuration, and performance.

The *HIT* system, integrates six single-axis linear accelerometers directly into the inner padding of the helmet. This embedded configuration promotes tight coupling with the athlete's head, allowing for better estimation of true head motion during impacts [14]. The system transmits data wirelessly in real time to a computer, making it valuable for live impact monitoring. However, it only measures linear acceleration and cannot capture rotational kinematics. *gForce Tracker (GFT)* includes a triaxial accelerometer and a triaxial gyroscope. It is typically mounted inside or on the helmet shell using adhesives or brackets, rather than integrated into

the padding [15], which makes the GFT a more modular and easy-to-deploy system. However, because it is mounted rather than embedded, relative motion between the helmet and the head can lead to measurement artifacts or reduced accuracy. In summary, the HIT system provides better head coupling but limited kinematic data (linear only), while the GFT offers more comprehensive kinematic measurements but may suffer from decoupling effects due to its mounting configuration. These trade-offs must be considered when selecting a helmet-based system for biomechanical monitoring in sports.

- **Instrumented Mouthguards:** Are advanced wearable devices equipped with inertial sensors, typically accelerometers and gyroscopes, embedded within a mouthpiece worn by the athlete. Located in direct contact with the upper dentition, these devices provide measurements that closely approximate the center of mass of the head, offering accurate and reliable motion data directly related to head kinematics [16]. However, a major limitation of instrumented mouthguards lies in their interaction with the athlete: players may remove them, bite on them, or reposition them during gameplay or practice. These voluntary or incidental movements can introduce artifacts or noise into the recorded data, potentially resulting in false positives or degraded measurement accuracy. Therefore, although instrumented mouthguards offer promising precision in capturing head impact biomechanics, their practical deployment requires careful validation and filtering techniques to mitigate these motion-induced errors.

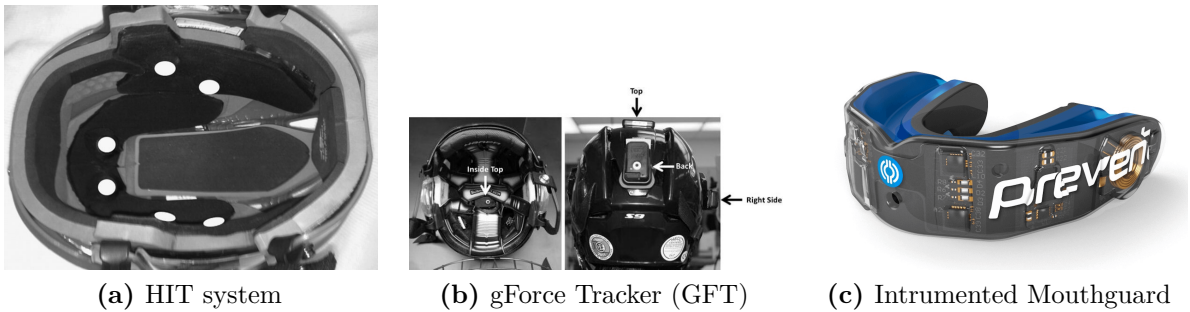


Figure 2.6: Examples of head impact sensor configurations.
[14, 15, 16].

To visually illustrate the main types of wearable sensor systems used for head impact monitoring in sports, including helmet-based systems and instrumented mouthguards, Figure 2.6 presents examples of the Head Impact Telemetry (HIT) system with embedded accelerometers, the gForce Tracker (GFT) with combined inertial sensors and an instrumented mouthguard designed for biomechanical data collection.

Chapter 3

Methods

The primary objective of this study is to identify the optimal placement of an Inertial Measurement Unit (IMU) on an ice hockey helmet, among three possible positions, to ensure reliable detection of head impacts. This choice has significant practical implications: a well-placed sensor is crucial for improving the accuracy of data acquisition systems, particularly with a view to potential real-time use during games to support medical staff in the timely diagnosis of concussions.

The analyses presented in this work were conducted in a controlled laboratory environment using a pendulum impactor designed to generate repeatable impacts on a helmeted Anthropomorphic Test Device (ATD). Three IMUs were mounted along the central axis of the helmet (from the forehead to the nape), in three distinct locations: front, middle-center, and lower rear. An additional IMU was positioned inside the dummy’s head and served as a reference, considered as ground truth for evaluating the accuracy of the signals recorded by the other sensors.

Impacts were applied from various directions, simulating realistic collision scenarios. Although none of the helmet-mounted IMUs were directly struck, their distance from the point of impact varied depending on the direction, allowing for analysis of whether and how proximity influences the fidelity of the recorded signal with respect to the ground truth.

The methods described in the following sections provide a detailed explanation of the entire experimental and analytical process, including the experimental setup, data collection protocol, signal preprocessing, and application of machine learning algorithms to evaluate sensors performance under various impact conditions.

3.1 Experimental Setup

3.1.1 Impact Generation System

To simulate head impacts under controlled conditions, a custom-designed pendulum system was employed. The pendulum impactor consists of a rectangular steel tube with a

cylindrical impactor mounted at its distal end. The impact surface is made of ultra-high molecular weight polyethylene (UHMWPE) and has a mass of approximately 0.6 kg. The total mass of the impactor is 15.5 kg, while the overall mass of the pendulum amounts to 37 kg [12].

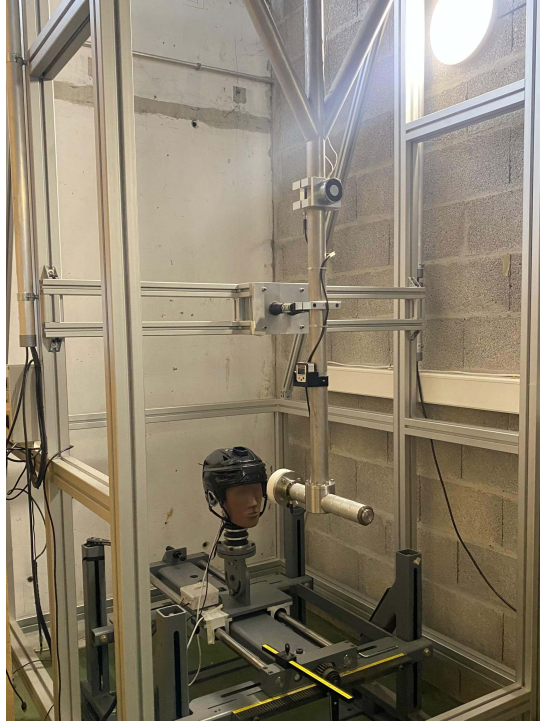


Figure 3.1: Experimental Setup (pendulum impactor, sliding table, and headform assembly).

3.1.2 ATD and Headform Sensor Positioning

The dummy used for impact testing is a Hybrid III (HN-III) model, widely employed in the automotive and sports industries for safety assessments. The head of the dummy consists of three main components: a rigid skull, an external deformable rubber skin, and a mounting system. The outer skin is modeled using a hyperelastic material to replicate the mechanical behavior of human soft tissue. The neck of the dummy was left unconstrained to better simulate natural head kinematics during impact.

An IMU was placed inside the dummy head, in a position that approximates the center of mass of the head. This IMU, referred to as the Headform, was considered the ground truth for all comparative analyses. Data acquisition was carried out using the software SLICEWare v1.08.0868, to which the sensor was connected via cable to ensure signal stability and synchronization.

3.1.3 Sensor Specifications and Helmet IMUs Positioning

This section provides a detailed overview of the technical specifications of the sensors employed in the study, organized according to their type and placement. Understanding these characteristics is essential for evaluating sensor performance and interpreting the data collected during the experiments.

Table 3.1: Technical specifications of the inertial sensors used in the experiments.

Sensor	Type	Model	Manufacturer	Country	Range	Sampling Rate
Headform	Accelerometer	Endevco 7264C-2000	DTS	USA	± 500 g	10 kHz
	Gyroscope	ARS 8000	DTS	USA	± 8000 °/s	10 kHz
Physilogs	Accelerometer	ADXL375	Analog Devices	USA	± 200 g	1024 Hz
	Gyroscope	LSM6DSOX	STMicroelectronics	Switzerland	± 2000 °/s	512 Hz

The helmet used for the experiments is a rigid plastic model manufactured by *Bauer*, a brand widely adopted in the context of ice hockey. Three plastic housings were glued on the helmet to securely and consistently accommodate the three IMU sensors under investigation.

The three sensors, referred to as *Physilog Top*, *Physilog Middle*, and *Physilog Bottom*, were positioned as follows:

- **Physilog Top - BM B49J** : front external part of the helmet;
- **Physilog Middle - BM B4D3**: upper central external part of the helmet;
- **Physilog Bottom - BM 0230**: between the rigid outer shell and the inner foam layer in the lower rear section.

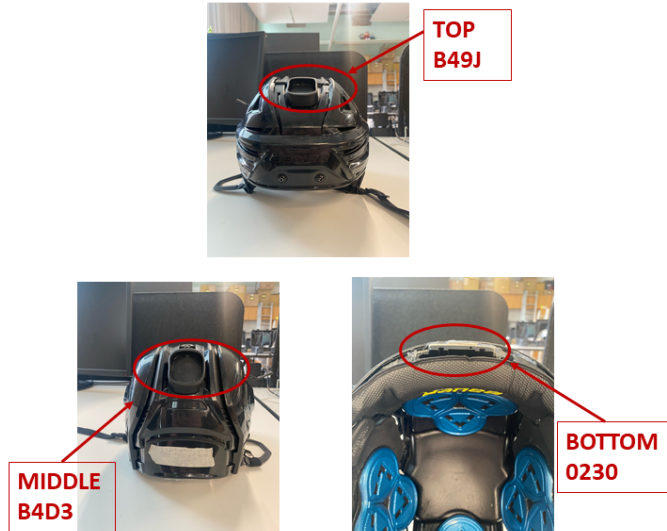


Figure 3.2: Sensor placement on the helmet: Physilog Top, Middle, and Bottom.

The sensors used are Physilog devices, produced by the Swiss company *MindMaze*, which

specializes in wearable technologies for motion analysis and is commonly used in clinical, sports and rehabilitation settings. In this study, these sensors were adapted to capture high-dynamic impacts.

The same firmware was installed on the Physilog devices used in this study. This firmware was specifically developed to support high-frequency data acquisition, optimized for short-duration, high-intensity dynamic events such as head impacts.

3.2 Protocol – Data Collection

The data collection phase was conducted following a rigorous experimental protocol, designed to ensure repeatability and controlled impact conditions. The experimental variables considered include the impact direction, the impact angle (and consequently the transferred impact energy), and the IMU sensor position on the helmet, as previously described in Section 3.2.3.

3.2.1 Impact Directions

Four representative impact directions were selected for the analysis: **Front**, **Front-Oblique** (or Front-Boss) , **Side**, and **Back-Oblique** (or Back-Boss), to replicate typical impact scenarios associated with sports-related head trauma, particularly in ice hockey. Each impact configuration was obtained by rotating the headform around its vertical axis (y-axis). The **Front** direction, in which the headform directly faces the impactor, was defined as the reference orientation (0°). The subsequent directions were generated by incrementally rotating the headform by 45° relative to the reference.

Since impacts can occur on either side of the helmet, both left- and right-side configurations were considered for each nominal direction. The complete naming convention and corresponding rotation angles with respect to the frontal orientation are summarized in Table 3.2.

Table 3.2: Impact directions and corresponding rotational angles relative to the frontal position.

Impact Direction	Angle ($^\circ$)
Front	0°
Left Front Boss	45°
Left Side	90°
Left Back Boss	135°
Right Back Boss	225°
Right Side	270°
Right Front Boss	315°

To provide a visual representation of the impact locations on the helmet, Figure 3.3 illustrates the directions defined in Table 3.2. For the sake of clarity, symmetric directions (e.g., left and right) are grouped together and not distinguished in the illustration.

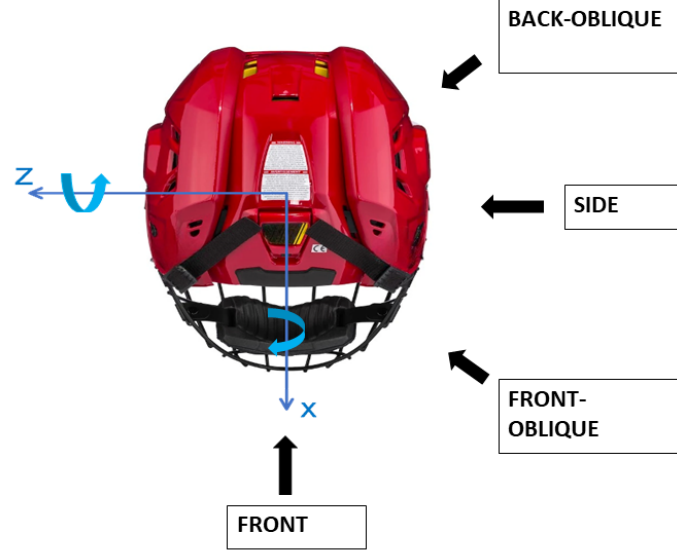


Figure 3.3: Impact Directions.

3.2.2 Impact Angles

For each impact direction, impacts were performed at three distinct release angles of the pendulum, corresponding to increasing levels of kinetic energy transferred to the headform. The selected angles (30° , 50° , 70°) cover a range of impact severities, from mild to high energy events. Table 3.3 reports the estimated kinetic energy associated with each angle.

Table 3.3: Impact angles and corresponding kinetic energy levels.

Impact Angle ($^\circ$)	Kinetic Energy (J)
30°	30 J
50°	79 J
70°	144 J

3.2.3 Experimental Protocol

Each acquisition consisted of a sequence of five consecutive impacts on the same configuration (defined by the combination of impact direction, angle, and IMU sensor location). A 20-second interval was maintained between impacts to allow the system to reset and stabilize. The total duration of each acquisition was approximately 120 seconds.

The protocol was structured as follows:

- **Configurations tested per IMU:** 4 impact directions (Front, Side, Front-Boss, Back-Boss) $\times$ 3 impact angles $\times$ 6 repetitions = **72 configurations**
- **Impacts per configuration:** 5 consecutive impacts per acquisition
- **Total impacts per IMU:** 72 configurations $\times$ 5 impacts = **360 impacts**
- **Total number of recorded impacts:** 4 IMUs (1 Headform + 3 Physilog) $\times$ 360 = **1440**

This experimental design enables a robust comparative analysis across IMU placements while ensuring statistical reliability and full coverage of the defined experimental conditions.

Chapter 4

Data Processing

The data acquisition process involves the collection of raw signals from both the Headform and the helmet-mounted IMUs, which are used to capture the full kinematic profile of head impacts. These sensors provide measurements of angular velocity and linear acceleration along the three spatial axes (x , y , z). However, in their raw form, the signals are not directly interpretable for analytical purposes, therefore, a structured and multi-step analytical workflow is required to extract meaningful information and enable comparisons between helmet-mounted IMUs and the ground-truth Headform reference. The complete processing pipeline — from signal acquisition to statistical validation — is summarized in Figure 4.1.

The analytical pipeline consists of a sequence of well-defined steps designed to isolate and analyze signals collected during impact events. Raw kinematic signals are first subjected to a signal preprocessing phase, aimed at isolating the segments of interest surrounding each impact. Next, a feature selection step is performed, these features are heuristically chosen to best capture aspects relevant to sensor comparability, in a way that enhances their ability to quantify the similarity between each helmet-mounted IMU and the reference Headform. The resulting set of features serves a dual purpose. First, it is used in a statistical analysis framework to objectively evaluate the degree of similarity between the signals recorded by each IMU and those from the Headform. Second, the same features are used as inputs to machine learning algorithms, whose goal is to provide additional similarity, this time in terms of their decouplability. Ultimately, the goal is to determine whether one of the IMUs, based on its position on the helmet and the characteristics of its recorded signals, offers the highest degree of similarity to the Headform. Such a finding would help guide optimal sensor placement strategies in future helmet-based monitoring systems.

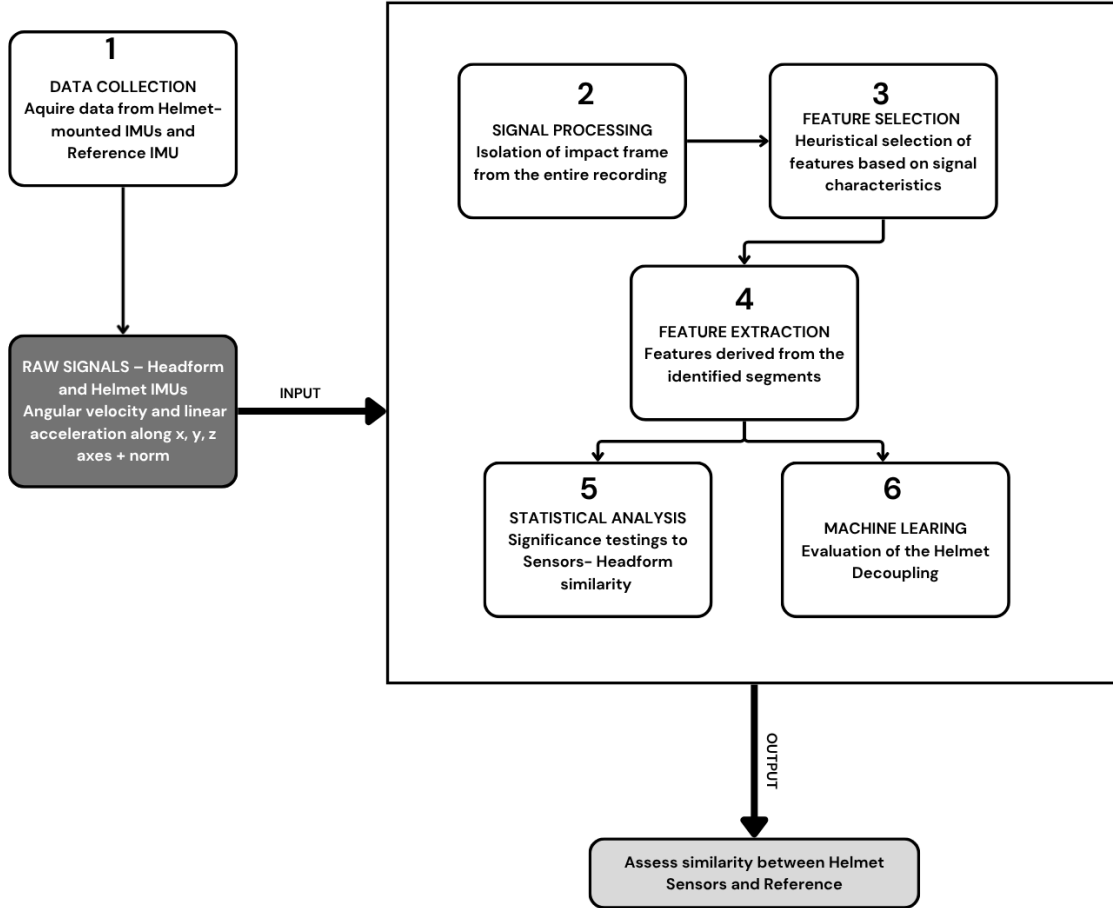


Figure 4.1: Overview of the analytical pipeline.

4.1 Preprocessing

Preprocessing is a crucial step to ensure that the signals recorded by the various sensors during experimental acquisitions can be compared consistently. Each recording lasts approximately 120 seconds and contains five impacts. The goal of this phase is to isolate, for each impact, a 100 ms time window containing the event, allowing detailed analysis of the impact dynamics and direct comparison between sensor outputs.

4.1.1 Functional Calibration

The analysis is performed using data from four devices:

- the **Headform**, which serves as the ground truth;
- three **Physilog IMUs**, mounted externally on the helmet.

Since each sensor is mounted in a different position and orientation, it is necessary to transform all signals into a **common reference frame**.

The helmet's **Functional Frame (FF)** was defined as:

- X_{FF} = forward
- Y_{FF} = vertical (upward)

- Z_{FF} = lateral (to the right)

Helmet Sensors Calibration Procedure: Each IMU mounted on the helmet has an unknown initial reference frame (Technical Frame, TF). A functional calibration was performed by placing the helmet on the Headform and performing:

- 5 seconds rotation around the Y_{FF}
- 5 seconds rotation around the Z_{FF}

The gyroscope signals were segmented for the two rotations and Principal Component Analysis (PCA) was applied to each segment to extract the main axis of rotation. The resulting rotation matrix was used to transform each IMU from its TF to the helmet FF. In the current project, experimental data preprocessing does not recompute the transformation matrix. Instead, the matrix is loaded directly from the calibration file and applied to the raw data recorded during the impact, projecting them into the predefined right-handed reference frame.

Headform Calibration Procedure: For the Headform, no empirical calibration was needed because its reference frame is known in advance. However, its native TF was originally defined as a *left-handed* coordinate system:

- X_{TF} pointing inward,
- Y_{TF} to the left,
- Z_{TF} upward.

Therefore, a transformation was required to align it with the right-handed FF used for the IMUs.

The following transformations were applied:

- a -90° rotation around X_{TF} , followed by a 180° rotation around Z_{TF} , to realign the axes;
- inversion of the Y_{TF} sign to convert the system to right-handed;
- the same rotations were applied to the angular velocity signals, followed by a sign inversion, since the Headform gyroscopes use the left-hand rule.

These operations allow the Headform signals to be projected into the common FF, making them directly comparable with the IMU signals.

4.1.2 Up-Sampling

In the experimental context, the Headform records data at 10000 Hz, while the Physilog sensors acquire data at different frequencies depending on the measurement type: 1024 Hz for linear acceleration and 512 Hz for angular velocity.

This step of preprocessing consists of resampling the Physilog sensor data to the Headform's 10,000 Hz frequency using linear interpolation. This procedure increases the temporal resolution of the data and temporally aligns the signals for subsequent synchronized

analyses.

Linear interpolation was performed using the *interp* function from the Python *numpy* library, which estimates new values by calculating, for each target time point, a weighted average between samples immediately adjacent to the original signal.

Important Note: The operations of up-sampling, segmentation, and synchronization operations were performed on all three axes (X_{FF} , Y_{FF} , Z_{FF}) for both acceleration and angular velocity data, as well as on the vector norm of the signals. This methodological choice was made to preserve all potentially relevant information regarding impacts, avoiding the omission of events that might manifest along nonprimary directions of motion.

4.1.3 Segmentation and Synchronization of Impacts

Impacts Detection

Impact events were identified by analyzing the amplitude peaks in the signals' resultants. For this purpose, a custom function was implemented, inspired by methods available in the Python library `scipy.signal` (such as the `find_peaks()` function), which allows the identification of local maxima representative of impact events.

During the analysis, the first 6 seconds of each recording were excluded because this time window contained artifacts caused by the positioning and stabilization of the sensors on the helmet. Although these artifacts exhibited amplitudes higher than the baseline, they did not correspond to true impacts and were therefore discarded to avoid false positives. The identification of impact events is based on predefined amplitude thresholds, which differ depending on the signal type:

- **Linear acceleration threshold:** A peak is considered an impact if the acceleration signal exceeds **10 g** for the Headform and **15 g** for the helmet-mounted IMUs, and then returns below this threshold. This criterion helps to distinguish true impacts from background noise or transient, non-significant spikes.
- **Angular velocity threshold:** An additional filter is applied to the angular velocity data, with a threshold set at **400 °/s** for all sensors. This step reduces *false positives*, for example signals caused by minor adjustments or movements of the helmet between impacts, which may exceed the acceleration threshold but do not correspond to real impact events.

Only peaks that satisfy both thresholds—linear acceleration and angular velocity—are considered true impact events, ensuring a more reliable detection.

Extraction of Relevant Signal Segments

The extraction of signal segments representative of impact events was conducted through a multi-step strategy designed to maximize temporal precision and ensure coherence among signals acquired from different sensors. The processing workflow is organized as follows:

1. Initial Isolation of a Wide Window Around the Peak

As explained in the previous section Section 4.1.3, the impacts were detected by first identifying the points where the signals exceed and return to a pre-defined threshold. From these points, time offsets were applied: 200 ms before the peak and 300 ms after the peak. The total signal duration of 500 ms was intentionally chosen in order to assure that the whole dynamic of impacts is captured.

2. Determination of the Impact Onset

From the previously extracted wide segments, the custom function *find_start_of_impact()* is applied with the goal of accurately identifying the actual onset of the impact. This function is applied independently to each signal and sensor, allowing for the correction of potential temporal misalignments among devices.

The logic of the function is based on a threshold criterion derived from the local peak:

- It computes the vector norm of the three-dimensional signal (X_{FF} , Y_{FF} , Z_{FF});
- It removes the initial mean to stabilize the baseline;
- From the baseline, the point where the signal exceeds 10% of the absolute peak, is extracted as start of the impact;

This approach enables a robust and adaptive detection that is independent of the absolute signal magnitude, providing a reliable and consistent temporal reference across different sensors.

3. Extraction of the Final 100 ms Window

Once the impact onset has been identified for each signal, the final window is extracted. This window has a total duration of **100 ms**, covering 30ms and 80ms after the onset, which is sufficient to capture the core impulsive phase of the event. It is defined as:

- 0.02 s (20 ms before the onset)
- 0.08 s (80 ms after the onset)

4.2 Feature Selection and Feature Extraction

4.2.1 Translational and Rotational Contributions to Brain Injury

The process of feature selection and extraction was designed to identify signal descriptors that are both biomechanically meaningful and sensitive to the injury mechanisms under investigation. While both linear acceleration and angular velocity signals were initially preprocessed, the feature extraction, as well as the final comparison between helmet and

headform IMUs were applied on the angular velocity. This choice was motivated by both biomechanical evidence and signal characteristics observed in the dataset.

Rotational kinematics are widely recognized as a key contributor to mild traumatic brain injuries (mTBIs), particularly diffuse axonal injury (DAI), which results from shear strains caused by rapid head rotations. Studies have shown that up to 90% of brain tissue strain during impact is due to rotational motion, with only a minor role played by translational components [17]. While linear acceleration has been associated with focal injuries like contusions, angular velocity better captures the dynamics underlying diffuse injuries [18]. Moreover, when evaluating the performance of helmet-mounted IMUs against Headform, it is important to consider the biomechanical differences in the signals they capture. In the case of linear acceleration, discrepancies between the two systems can be primarily attributed to attenuation effects of the helmet, which reduces the measured peak values. This effect is not observed for rotational motion, as angular velocity measurements are less influenced by helmet damping. Therefore, any differences observed in angular velocity signals are more likely to arise from signal noise, including helmet decoupling or vibration, rather than true biomechanical attenuation.

This rationale further supports the choice of focusing on angular velocity as the primary parameter for meaningful comparison between helmet and headform IMUs. In addition to the biomechanical perspective, the nature of the acquired signals also justifies this focus. Due to the short duration and high intensity of impacts, linear acceleration signals are often more susceptible to high-frequency noise, whereas angular velocity provides a more stable and informative representation of the impact dynamics, enabling a more robust feature extraction process.

4.2.2 Feature Selection

After the preprocessing phase — which included signal segmentation and synchronization — features were extracted mainly from the angular velocity signals recorded for each impact. These signals include components along the three spatial axes: X_{FF} (antero-posterior), Y_{FF} (inferior-superior), and Z_{FF} (medio-lateral), as well as the vector norm.

The feature selection process followed a **heuristic approach**: the features were selected to be as meaningful and interpretable as possible from a biomechanical point of view, considering the complexity and nonlinearity of the recorded signals. Additionally, features were selected in order to avoid introducing redundant or uninformative features that could lead to overfitting in the machine learning models (which will be explained later). This is especially important, since the selected features will be used to build the dataset for the predictive algorithms.

The selected features are grouped into three main domains of analysis: time domain, frequency domain, and time-frequency domain. Each domain provides complementary information on the morphology and dynamics of the signal. Figure 4.2 shows an overview

of the selected feature set, organized by domain.

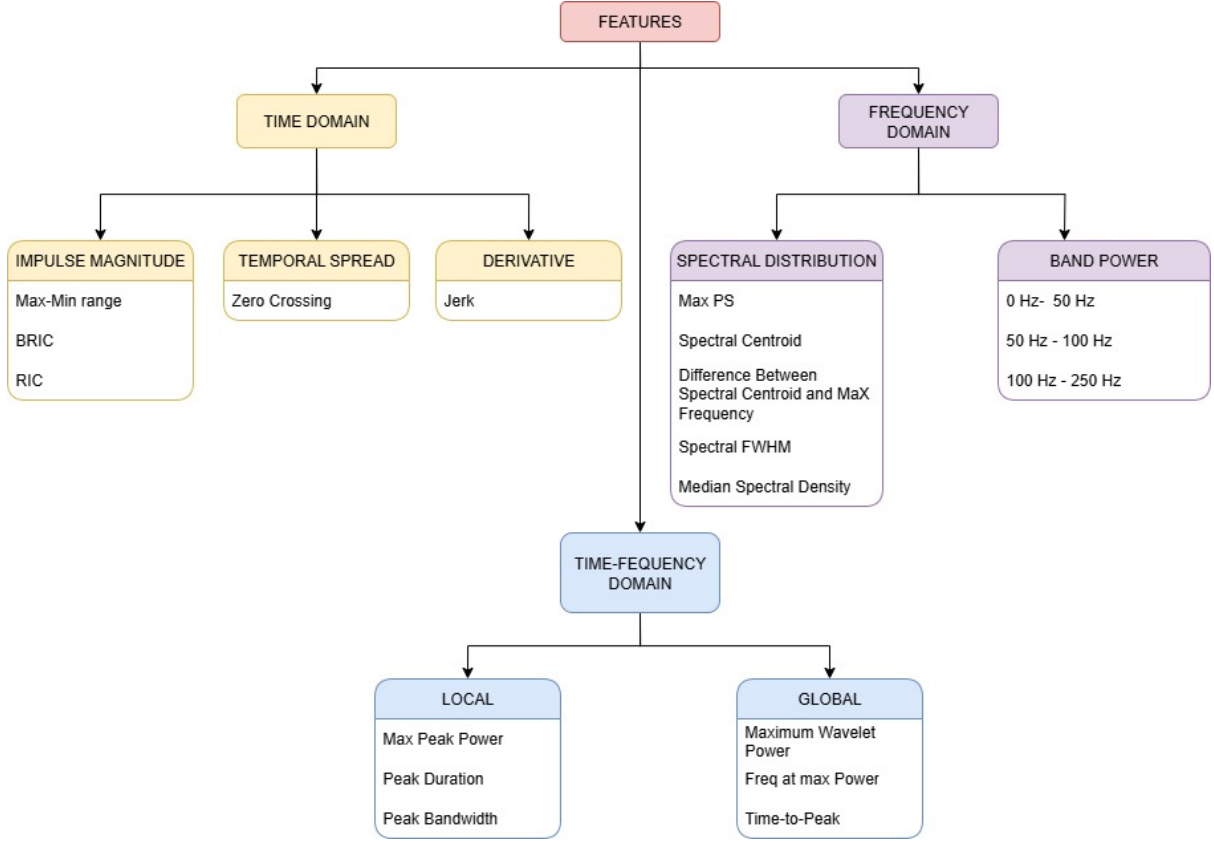


Figure 4.2: Feature selection overview.

Time-Domain Features

In the time domain, features are extracted based on the shape of the signal and its variations over time [16]. This approach allows for the quantification of morphological properties that are visually observable in the signals through objective and repeatable measures. The selected features, in addition to being informed by scientific literature, were chosen for their biomechanical relevance.

1.BrIC – Brain Injury Criterion The Brain Injury Criterion (BrIC) is a metric developed to correlate peak angular velocity of the head with the risk of brain injury [19],[20]. BrIC is calculated using the maximum angular velocities along the three axes, normalized by experimentally derived critical thresholds:

$$\text{BrIC} = \sqrt{\left(\frac{\omega_x}{\omega_{xC}}\right)^2 + \left(\frac{\omega_y}{\omega_{yC}}\right)^2 + \left(\frac{\omega_z}{\omega_{zC}}\right)^2} \quad (4.1)$$

where ω_x , ω_y , and ω_z represent the maximum angular velocities along the three axes, instead ω_{xC} , ω_{yC} , and ω_{zC} are the corresponding critical thresholds, respectively equal to 66.25 deg/s, 44.87 deg/s, 56.45 deg/s. This index is justified by evidence showing that the

rotational components of head kinematics are strongly associated with mTBI and DAI, compared to the translational components.

2. RIC – Rotational Injury Criterion

The Rotational Injury Criterion (RIC) [20] follows a similar logic to the well-known Head Injury Criterion (HIC) [20], but replaces linear acceleration with angular acceleration. RIC is computed as:

$$\text{RIC} = \max_{t_2 - t_1 \leq 36 \text{ ms}} \left[\left(\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} |\alpha(t)| dt \right)^{2.5} \cdot (t_2 - t_1) \right] \quad (4.2)$$

where $\alpha(t)$ is the magnitude of the angular acceleration vector. The 36 ms integration window represents a period long enough to capture the impulsive dynamics typical of an impact event and is consistent with standard biomechanical criteria such as HIC [20].

3. Min–Max Range (Amplitude)

The amplitude represents the difference between the maximum and minimum values of the angular velocity signal within the selected time window [16]. This metric provides a direct estimate of the maximum angular excursion recorded during the impact:

$$\text{Amplitude} = \max(\omega(t)) - \min(\omega(t)) \quad (4.3)$$

This value is useful for assessing the overall intensity of the rotational motion.

4. Zero Crossing Rate (Adaptive Threshold)

The classical zero-crossing metric, which counts the number of times the signal crosses the zero line, was adapted using a dynamic, adaptive threshold. This approach was introduced to address the higher instability and variability present in signals acquired from helmet-mounted sensors compared to the reference Headform, which produces cleaner signals with fewer spurious oscillations.

The adaptive threshold was computed as half of the mean of the absolute peak angular velocities recorded by the three helmet-mounted sensors. The number of crossings with respect to this threshold provides a measure of the signal’s oscillatory behavior, potentially associated with the sensor quality or its coupling effectiveness with the actual head motion.

5. Maximum of Angular Jerk

Jerk is defined as the first derivative of angular acceleration, or equivalently, the third derivative of angular position with respect to time [3]. It is calculated as follows:

$$\text{Jerk}(t) = \frac{d\alpha(t)}{dt} = \frac{d^2\omega(t)}{dt^2} \quad (4.4)$$

In our case, the maximum jerk was computed over the first 35 ms of the signal, with the aim of capturing the initial variability of the motion. High jerk values indicate a rapid change in acceleration, typically associated with impulsive and less smooth dynamics.

Frequency-Domain Features

Frequency domain analysis is a technique used to study time-series signals by transforming them from the time domain (values recorded over time) into the frequency domain, where signals are expressed as a combination of sinusoidal components. This transformation is useful for identifying hidden characteristics in the signal, such as periodic behavior, high frequency noise, or dominant oscillations, that may not be evident in the time domain [21].

In this work, frequency analysis is performed using the Fast Fourier Transform (FFT), specifically through the *rfft* function from the *numpy.fft* module. This function efficiently computes the Discrete Fourier Transform (DFT) for real-valued input signals. The FFT is a fast algorithm for computing the DFT, which decomposes a signal into its frequency components and allows conversion between the time and frequency domains [22].

The Power Spectrum (PS) is computed by taking the squared magnitude of the output of the Fast Fourier Transform (FFT), which decomposes a time-domain signal into its constituent frequency components. This provides a measure of how the signal's energy is distributed across frequencies.

The FFT provides a discrete representation of the Fourier Transform of a signal $x(t)$, sampled at regular intervals. Let $X(f)$ denote the Fourier Transform of the signal:

$$X(f) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad (4.5)$$

For discrete sampled signals, the FFT approximates this integral. The **Power Spectrum (PS)** is then given by:

$$P(f) = |X(f)|^2 \quad (4.6)$$

This expression represents the energy content of the signal at each frequency f , without normalization.

In contrast, the **Power Spectral Density (PSD)** is defined as:

$$\text{PSD}(f) = \frac{P(f)}{\Delta f} \quad (4.7)$$

where Δf is the frequency resolution of the FFT, calculated as:

$$\Delta f = \frac{f_s}{N} \quad (4.8)$$

with f_s being the sampling frequency and N the number of samples in the signal segment [23].

Considering $f_s=10$ kHz and the length of the segment $T=100$ ms, N is equal to:

$$N = f_s \times T = 10,000 \times 0.1 = 1,000 \quad (4.9)$$

The PSD normalizes the energy per unit frequency (e.g., per Hz), which is essential when comparing signals of different durations or resolutions.

In this study, PSD normalization was deemed unnecessary because:

- All signals were acquired at the same sampling frequency (10 kHz).
- All signals were segmented to the same temporal duration (100 ms) during preprocessing.
- Therefore, all segments contain the same number of points and share the same frequency resolution (Δf), making them directly comparable in terms of energy.

As a result, using the Power Spectrum is sufficient to extract consistent and meaningful frequency-based features across all sensor signals. PSD estimation would instead be required in scenarios with variable sampling rates or durations, where normalization is needed across frequency resolution.

Therefore, the features selected in the frequency domain are:

1. Maximum Value of the Power Spectrum (Max PS) This feature represents the maximum value of the power spectrum, i.e., the highest energy level associated with a specific frequency. It is related to the dominant frequency in the signal, which corresponds to the frequency with the greatest concentration of energy. This is useful for identifying the most energetically significant components in the signal.

2. Spectral Centroid The spectral centroid represents the center of the power spectrum. It is calculated as the weighted average of the frequencies, where the weights are the power values associated with each frequency. It provides a synthetic measure of the spectral distribution and helps us to understand whether the energy is concentrated at low, high, or spread-out frequencies.

3. Centroid–Peak Difference This feature quantifies the difference (in Hz) between the frequency corresponding to the maximum peak of the power spectrum and the spectral centroid. The motivation behind this feature lies in the observation that, in some signals, the peak of the power spectrum does not coincide with the centroid. This suggests a broader and less focused energy distribution, which can be due to the presence of high-frequency noise or complex dynamic behaviors.

4. Spectral Full Width at Half Maximum (FWHM) The Spectral FWHM is a feature that measures the bandwidth of the main lobe of a power spectrum, that is how widely the energy is spread around the dominant frequency. It is defined as the width of the frequency band where the power remains above half of the maximum value of the spectrum.

In signal processing, the FWHM is a useful metric to assess spectral sharpness [24]:

- A narrow FWHM indicates that the signal’s energy is concentrated in a narrow band (sharp peak), typical of periodic or pure-tone signals.
- A broad FWHM suggests that the energy is spread across a wider frequency range, which may reflect more complex or noisy signals.

5. Median Spectral Density The median is a descriptive statistic that identifies the

central value of an ordered data set and is robust against outliers and abnormal values. The Median Spectral Density is used to provide a robust estimate of the signal's energy distribution across frequencies. By calculating the median of the spectral densities, we get a measure that represents the typical energy present in the frequency bands.

6. Spectral Band Power To evaluate how the signal's energy is distributed across different frequency bands, the Power Spectrum was divided into three bands:

- 0–50 Hz (Low-frequency band)
- 50–100 Hz (Mid-frequency band)
- 100–250 Hz (High-frequency band)

For each band, the spectral energy was calculated by numerical integration of the area under the spectrum within that interval.

The choice of these bands is based on practical considerations and references in the literature [5]:

- The low and mid-frequency bands are generally associated with the main informative content of the signal, such as slow oscillations or significant dynamics.
- The high-frequency band tends to capture contributions related to noise, mechanical vibrations, or high-frequency disturbances, which are often less relevant from an interpretative point of view.

This feature therefore allows assessing how much of the signal's energy is concentrated in the more informative bands compared to that which might be attributable to noise or artifacts.

Time-Frequency Domain - Wavelet Transform

The Continuous Wavelet Transform (CWT) represents a signal as a linear combination of base wavelet functions, known as mother wavelets, which are **scaled and shifted** to analyze the frequency content of the signal at different time points. Specifically, the mother wavelet $\psi(t)$ is scaled by a factor a and shifted by b according to the formula:

$$WT(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t - b}{a} \right) dt,$$

where a controls the *scale* (related to frequency) and b controls the *translation* in time. Small values of a correspond to high frequencies with fine temporal resolution, while large values of a correspond to low frequencies with better frequency resolution. This adaptive trade-off makes the Wavelet Transform an ideal tool for highlighting local energy peaks, transient oscillations, and brief but significant events that would not emerge through a standard global spectral analysis.

This adaptive trade-off makes the Wavelet Transform an ideal tool for highlighting local energy peaks, transient oscillations, and brief but significant events that would not emerge through a standard global spectral analysis.

In this study, we used the complex *Morlet* wavelet [16], particularly suitable for analyzing signals with oscillatory content, applied over a scale range from 1 to 255. From the resulting power matrix, two types of features were extracted: global and local.

Global features are extracted from the entire time-frequency map of the WT and provide an overview of the energy distribution in the signal [16]:

1. **Maximum Global Power (Max WT Power):** The highest power value in the WT power map.
2. **Frequency at Maximum Power:** The frequency corresponding to the point of maximum energy, useful for identifying the dominant spectral band.
3. **Time Difference Between WT Peak and Time-Domain Peak:** Measures the time shift between the peak detected in the WT map and the peak in the original signal. This parameter reflects the temporal alignment between the WT representation and the actual signal.

Local features: For each of the first three local energy peaks found in the WT map [16] for every axis (X_{FF} , Y_{FF} , Z_{FF} and norm).

In this context, a local peak refers to one of the three most prominent energy concentrations in the WT map, identified using masks. The procedure is as follows:

1. Find the **maximum absolute value** in the WT power map. This represents the highest energy point in the matrix.
2. Create a **mask** around this maximum, including all points in the WT map that are above a certain threshold (e.g., 10% of the peak value). This mask defines the spatial "territory" of the peak in time and frequency.
3. Set all points within the mask to zero in a copy of the WT map, effectively removing the peak from consideration.
4. Repeat the steps above two more times to extract the second and third highest local peaks, generating three separate masks, each isolating a local energy peak.

Three descriptive metrics were computed to characterize the nature of localized energy events:

1. **Maximum Power of the Local Peak** This represents the highest energy value within the local peak. It indicates how intense the energy concentration is in that specific transient event.
2. **Temporal Duration of the Peak** This metric measures how long the energy of the peak remains above a predefined threshold set to 50% of the peak's maximum value. It reflects the persistence of the event over time.
3. **Bandwidth of the Peak** Similar to temporal duration, this is measured along the frequency axis, defining the frequency range where the peak's energy remains above the 50% threshold.

4.2.3 Overview of Extracted Features

For each detected impact, a total of 18 distinct features were computed across three different analytical domains:

- 5 features in the Time Domain
- 8 features in the Frequency Domain
- 6 features in the time-Frequency Domain (via Wavelet Transform)

These features were extracted from the three accelerometric axes (X_{FF} , Y_{FF} , Z_{FF}) and the signal norm, with some exceptions:

Time Domain:

- **Axis-specific features:** 3 features — Amplitude, Zero Crossing with adaptive threshold, and Jerk — were computed separately for each of the four signal components (X_{FF} , Y_{FF} , Z_{FF} and norm), resulting in $3 \times 4 = 12$ features.
- **Composite features:** 2 features — RIC and BrIC — were computed once per impact, as their formulations inherently include contributions from all three axes.

Total time-domain features per sensor: 12 (axis-specific) + 2 (composite) = **14**

Frequency Domain: All 8 features were calculated separately for each of the three signal components and norm, giving: Total frequency-domain features per sensor: $8 \times 4 = 32$ features per sensor

Time-Frequency Domain (Wavelet Transform)

- **Global features:** 3 features — maximum WT power, dominant frequency, and time-to-peak delay — were extracted from the entire time-frequency representation.
- **Peak-based features:** 3 features (peak power, temporal duration, spectral bandwidth) were computed for each of the top 3 energy peaks: $3 \times 3 = 9$ features per signal component.

Total time-frequency features per sensor: $(3 \text{ global} + 9 \text{ peak-based}) \times 4 = 48$ features

Total number of Features extracted: $14 + 32 + 48 = 94$ features, across 4 sensors $94 \times 4 = 376$ features

Before proceeding with the statistical analysis, an exploratory investigation was conducted to evaluate whether the features extracted from each sensor exhibited any degree of correlation, both within the same sensor and across different sensors. The rationale behind this preliminary step was to assess whether the information provided by each sensor could be considered independent, or whether redundant information might be present due to inter-sensor correlation.

For this purpose, the Pearson correlation coefficient was employed [5], as it quantifies the linear relationship between two variables on a scale ranging from -1 to $+1$. In this context, only positive correlations were expected, since identical features compared with

themselves necessarily yield a maximum correlation coefficient of 1.

4.3 Statistical Analysis

As mentioned in Chapter 3, Section 3.1, one of the main goals of this work is to answer the following research question: which of the helmet-mounted sensors provides measurements that are most similar to those recorded by the reference unit (IMU) placed inside the headform.

To address this question, a quantitative comparison was carried out between the signals acquired from the helmet sensors and the ground truth. For this purpose, descriptive features were extracted to summarize the main characteristics of the signals in three different domains: time, frequency, and time-frequency. The aim of the statistical analysis is to determine whether there are statistically significant differences between the sensor measurements and the ground truth, in order to assess which sensor provides more accurate estimates. The method selected for this analysis falls within the domain of statistical inference, which represents the branch of statistics concerned with drawing conclusions about a population based on observations from a representative sample. In essence, statistical inference offers a rigorous framework for extrapolating the findings obtained from a limited dataset (sample) to the broader population from which the sample was drawn [25]. The specific test adopted is the Paired Sample t-test (or dependent samples t-test), implemented in Python using the *ttest_rel* function from the *scipy.stats* library. The Paired Sample t-test is a statistical tool used to estimate whether the mean difference between two paired sets of observations is significantly different from zero. It is particularly appropriate in experimental settings where each observation is acquired under two conditions on the same subject or event, as in this study.

4.3.1 Hypotheses and Errors

Like many statistical methods, the paired t-test relies on two opposite hypotheses [26, 27]:

- **Null Hypothesis (H0):** the average difference between the paired measurements is zero. In other words, there is no significant difference between the features measured by the helmet sensor and those from the Headform. Any differences are due to chance or natural variability.
- **Alternative Hypothesis (H1):** the average difference between the two data sets is not equal to zero, suggesting a real and systematic difference between the sensor measurements.

These hypotheses can be expressed mathematically as:

$$\begin{cases} H_0 : \mu_1 = \mu_2 & \text{(the means are equal)} \\ H_1 : \mu_1 \neq \mu_2 & \text{(the means are different)} \end{cases}$$

where:

- μ_1 is the mean value of a given feature measured by the Headform,
- μ_2 is the mean value of the same feature measured by one of the helmet sensors.

Once the hypotheses are defined, it is important to consider the possible types of statistical errors [27]:

- **Type I Error (α):** occurs when the null hypothesis is rejected even though it is actually true. In this case, one concludes that there is a difference, when in fact the two sets of measurements are statistically equivalent. This is also known as a *false positive*.
- **Type II Error (β):** occurs when the null hypothesis is not rejected even though it is actually false. This means concluding that the two sets of data are similar when a real difference exists. This is also known as a *false negative*.

To decide whether to reject the null hypothesis, the paired t-test provides a statistical parameter known as the **p-value**. This value represents the probability of observing a difference between two measurements, assuming that the null hypothesis is true. The decision rule is defined as follows:

- If $p < \alpha$ (probability of Type I error), the null hypothesis is rejected: the difference is considered statistically significant.
- If $p \geq \alpha$ (probability of Type I error), the null hypothesis is not rejected: there is not enough statistical evidence to conclude that the two measurements differ significantly.

In the context of this study, the objective is to identify those features for which the helmet-mounted sensors provide measurements statistically equivalent to the ground truth (Headform). Therefore, the features of interest are those for which the p-value is **greater than or equal to** the significance level α , indicating no significant difference between the two measurement systems.

Assumptions and Validity of the Test: The paired t-test assumes that the distribution of the differences between paired observations is approximately normal. In this study, 360 impacts were recorded for each sensor, for a total of 1440 impacts.

Thanks to this large sample size, we can rely on the Central Limit Theorem (CLT)[28], which states that as the number of observations increases, the distribution of sample means (or differences) tends to become normal, regardless of the original data distribution. This allows us to legitimately use the t-test.

Calculated Parameters: For each comparison between the Headform and one of the three sensors (Physilog Top, Physilog Middle, Physilog Bottom), the following statistical indicators were calculated using the paired t-test:

- The mean difference for each feature between the two sensors
- The standard deviation of the differences
- The t-value, which is the test statistic of the Student's t-test

- The p-value, which indicates the statistical significance of the observed difference
- The 95% confidence interval for the mean difference
- Cohen’s d value was also computed. This is a measure of effect size, which will be further discussed in the next section.

4.3.2 Sensor Pairs and Test Design

Three sensor pairings were analyzed:

- Headform vs. Physilog Top
- Headform vs. Physilog Middle
- Headform vs. Physilog Bottom

For each feature, a separate Paired t-test was performed across the following configuration:

- 3 angles of impact (30°, 50°, 70°)
- 4 impact locations (Front, Front-Boss, Side, Back-Boss)
- 3 sensor pairs

This results in a total of:

$$3 \times 4 \times 3 = 36 \text{ Paired t-tests}$$

To evaluate statistical significance, the p-value was compared against a corrected threshold derived from the Bonferroni correction method. This correction is used to control the family-wise error rate when multiple hypotheses are tested simultaneously, thus reducing the risk of Type I errors (false positives).

The adjusted significance level α was calculated following the Bonferroni’s correction [29]:

$$\alpha = \frac{\alpha'}{k}$$

where:

- α' is the standard significance level (typically 0.05),
- k is the number of tests performed (in this case, $k = 36$).

Therefore:

$$\alpha = \frac{0.05}{36} = 0.0014 \approx 0.001$$

For each extracted feature:

- If $p < 0.001$: the null hypothesis is rejected, and the difference in feature values between the headform and the helmet-mounted sensor is considered **statistically significant**.
- If $p \geq 0.001$: the null hypothesis is **not rejected**, indicating that there is **not enough statistical evidence** to conclude that the two measurements differ significantly.

In this study, features with $p \geq 0.001$ are of particular interest, as they suggest that the helmet-mounted sensor produces measurements statistically comparable to those of the

Headform.

In addition to statistical significance (p-value), it is crucial to consider the magnitude of the observed differences. For this reason, Cohen’s d was computed for each paired comparison.

Effect size represents the main outcome of a quantitative analysis, as it quantifies how substantial the observed difference is. While the p-value indicates whether a difference exists, the effect size assess *how large* that difference is [30].

Cohen’s d is commonly interpreted using the following thresholds:

- **Small effect:** $d = 0.2$
- **Medium effect:** $d = 0.5$
- **Large effect:** $d = 0.8$
- **Very large effect:** $d \geq 1.3$

Table 4.1: Summary of Paired t-test analysis settings

Test Type	Paired Sample t-test (dependent samples)
Comparison Pairs	Headform vs. Physilog Top Headform vs. Physilog Middle Headform vs. Physilog Bottom
Total Number of Tests	3 impact angles $\times$ 4 channels $\times$ 3 comparisons = 36 tests
Significance Level (α)	0.05 (before correction)
Adjusted Significance (α_{adj})	$\alpha/k = 0.05/36 \approx \mathbf{0.001}$
Null Hypothesis (H_0)	$\mu_1 = \mu_2$ (No difference between means)
Alternative Hypothesis (H_1)	$\mu_1 \neq \mu_2$ (Means are significantly different)
Criteria for Rejection	If $p < 0.001$, reject H_0 (significant difference)
Effect Size Metric	Cohen’s d

4.4 Machine Learning

Following the feature extraction and selection phases, an additional objective of this work was to assess the level of decoupling between the helmet and the Headform during impact events. This aspect was investigated in order to evaluate whether the position of a sensor on the helmet affects its sensitivity to relative motion, and therefore its reliability in capturing true head kinematics.

In ideal conditions, the helmet and the headform would behave as a rigidly coupled system, meaning that they would move as a single solid body during impact. However, due to imperfect contact and possible slippage or rotation between the helmet and the head, relative motion can occur. This phenomenon is known as helmet-head decoupling. The decoupling introduces discrepancies between the actual head motion and the motion recorded by sensors placed on the helmet, potentially compromising the accuracy of biomechanical measurements. For this reason, identifying sensor positions more susceptible to decoupling can help guide the design of improved instrumentation strategies in head

impact monitoring.

To analyze this problem, a supervised Machine Learning (ML) approach was adopted to automatically classify each recorded impact based on the observed level of decoupling between the helmet and the headform.

Using a set of extracted signal features as input, the ML model aims to distinguish whether the helmet behaved coherently (low decoupling) or independently (high decoupling) from the head during impact. Each impact in the dataset was manually labeled as **high** or **low** decoupling based on visual inspection. The classification criterion is based on the relative displacement and rotation of the helmet with respect to its initial position before impact. Specifically:

- **Low Decoupling** is defined when the helmet maintains an alignment close to its pre-impact configuration, with no appreciable displacement or rotation relative to the headform. In these cases, the helmet and headform behave approximately as a rigidly coupled body.
- **High Decoupling** is observed when the helmet shows a clear deviation from its initial position, either by forward slippage (exposing a larger portion of the dummy's forehead) or by rotational misalignment with respect to the headform. This condition indicates a relative motion between the helmet and the head, consistent with a loss of rigid coupling.

An illustrative example of these two scenarios is reported in Figure 4.3, where panel (a) corresponds to a low decoupling condition and panel (b) to a high decoupling condition.

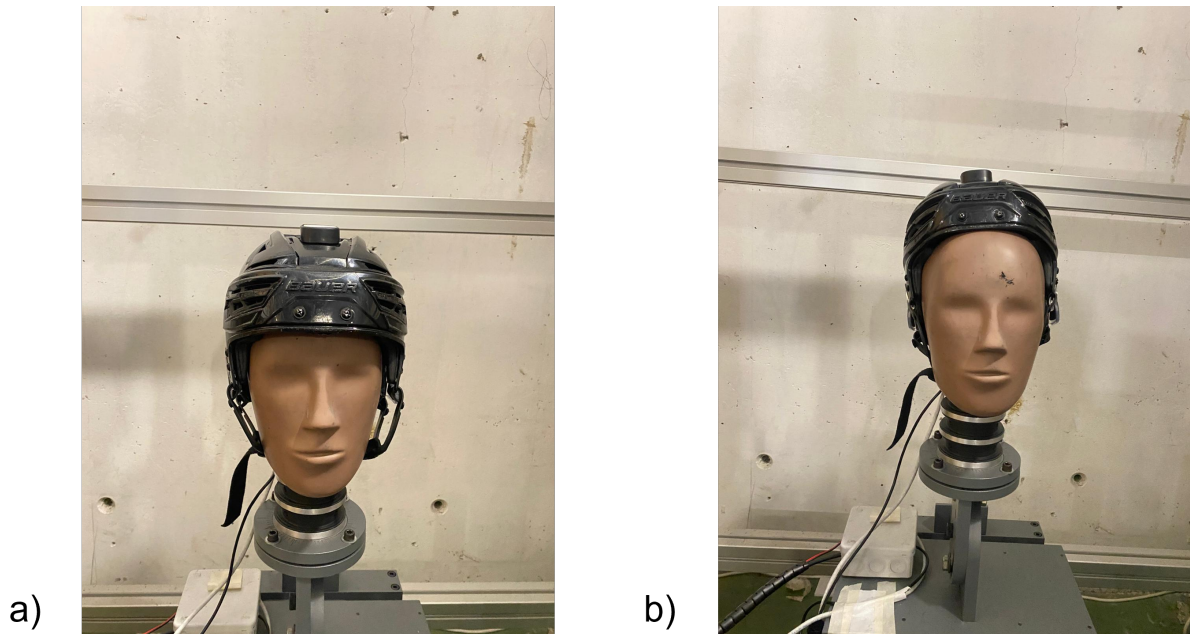


Figure 4.3: Low vs High Decoupling.

In this framework, a sensor that allows the ML model to easily discriminate between high and low decoupling events can be interpreted as more sensitive to the decoupling

phenomenon, and therefore less reliable in reproducing true head kinematics. Conversely, sensors showing a reduced sensitivity to decoupling can be regarded as less affected by this phenomenon and therefore more suitable for accurate motion tracking.

4.4.1 ML Overview

Machine Learning (ML) is a branch of artificial intelligence that enables computational systems to infer patterns and structures directly from data. Unlike traditional programming, where explicit rules are predefined by the developer, ML algorithms iteratively adjust their internal parameters through exposure to training data, thereby progressively improving their predictive accuracy and generalization capability [31]. Depending on the type of problem and the availability of labeled data, ML algorithms are generally categorized into two main groups:

- **Supervised learning:** the model is trained on a labeled dataset, where each sample is associated with a known output (label). The objective is to learn a mapping function that can predict the correct label for new, unseen data. Supervised learning tasks are typically divided into classification problems, where the output variable is categorical, and regression problems, where the output variable is continuous.
- **Unsupervised learning:** the model is trained on unlabeled data, where the structure of the dataset is unknown. The goal is to uncover hidden patterns or groupings within the data, often through clustering algorithms or dimensionality reduction techniques.

Figure 4.4 [31] illustrates the basic subdivision of Machine Learning techniques into supervised and unsupervised approaches.

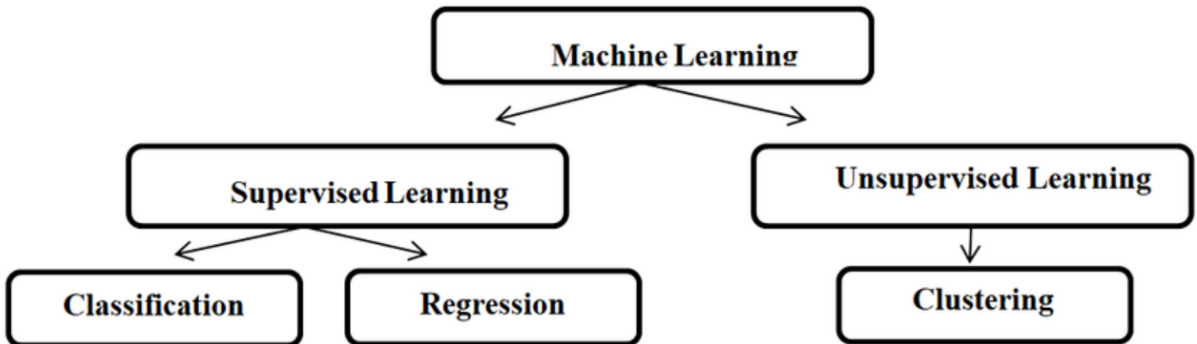


Figure 4.4: Machine Learning approaches.

In this study, a supervised learning framework was adopted, since each impact event was manually labeled as High or Low Decoupling. Two distinct ML techniques were selected and compared [31]:

- **Decision Trees and Random Forests** A Decision Tree is a flowchart-like structure where data is recursively split based on feature values. At each step (node) a condition is checked, each branch is a possible answer, and at the end (leaf) the sample is

assigned to a class [31]. The Random Forest builds many decision trees using random parts of the data and features, and then combines their results. The final decision is made by majority vote, which makes the model more accurate and stable than a single tree.

- **Support Vector Machines (SVMs)** Support Vector Machines are powerful classifiers that work by finding the optimal hyperplane that separates data points belonging to different classes with the maximum margin. Training data are mapped into a high-dimensional feature space, and the SVM algorithm seeks the boundary that maximizes the separation between the two classes. New observations are then classified based on which side of the hyperplane they fall. This approach is particularly effective in binary classification problems, such as the present case.

By employing these two supervised learning algorithms, it was possible to quantitatively assess how well the helmet-mounted sensors capture head motion relative to the ground truth, and to evaluate the impact of sensor position on susceptibility to decoupling.

4.4.2 Datasets

For the classification task, four datasets were prepared, each containing the features extracted from the helmet-mounted IMUs after the feature selection step:

- **All IMUs without Ground Truth:** includes all the selected features computed from the three helmet-mounted sensors (Physilog Top, Physilog Middle, Physilog Bottom), for a total of 360×282 entries.
- **Physilog Top Dataset:** includes only the features computed from the Physilog Top sensor (360×94).
- **Physilog Middle Dataset:** includes only the features computed from the Physilog Middle sensor (360×94).
- **Physilog Bottom Dataset:** includes only the features computed from the Physilog Bottom sensor (360×94).

Each dataset contains 360 impacts, with each impact labeled as *High* or *Low Decoupling* according to the visual inspection criteria described in Section 4.4. Importantly, the datasets are balanced in terms of class distribution, with approximately 47% of samples labeled as *High Decoupling* and 53% as *Low Decoupling* (H: 169, 46.94%; L: 191, 53.06%). This balance ensures that the classification algorithms are not biased toward one of the two classes, a common issue in imbalanced datasets.

4.4.3 Train-Test Splitting

To evaluate the generalization performance of the models, each dataset was split into a **Training Set (80%)** and a **Test Set (20%)**. The split was performed using *stratified sampling*, meaning that the proportion of *High* and *Low* labels was preserved in both sets.

This ensures that both training and test sets remain representative of the overall dataset. Table 4.2 reports the size of the datasets before and after splitting.

Table 4.2: Dataset sizes before and after Train-Test splitting.

Dataset	Original Size	Training Set	Test Set
All Sensors (282 features)	360 samples	288 samples	72 samples
Single Sensor (94 features)	360 samples	288 samples	72 samples

The splitting procedure was implemented in Python using the function `train_test_split` from the `scikit-learn` library, with the parameter `stratify=y` to guarantee class balance. The splits for each dataset were stored in serialized files using the `joblib` library, enabling consistent reuse across model training and testing phases.

4.4.4 Random Forest

The Random Forest (RF) algorithm is a supervised Machine Learning method that can be applied to both classification and regression tasks. In classification, the model predicts a categorical outcome (class labels), whereas in regression it outputs a continuous numerical value [31]. In this work, the RF was employed as a classifier to model the relationship between the extracted features and the target outcome.

Random Forest belongs to the family of ensemble methods [5], as it combines the predictions of multiple base learners (decision trees) in order to improve generalization performance and mitigate overfitting. Each decision tree is built by recursively partitioning the data according to feature values until terminal nodes, representing predictions, are reached. The ensemble nature of the RF relies on the bootstrap aggregating (bagging) technique: each decision tree is trained on a bootstrap sample of the training set, generated by sampling with replacement. Some features may appear multiple times in a sample while others may be excluded, leading each tree to grow differently. The final prediction is obtained by aggregating the outputs of all trees: through majority voting in classification tasks, or by averaging in regression tasks. This procedure reduces the variance of the model compared to a single decision tree, enhances stability, and improves robustness against noise and overfitting.

In this study, the Random Forest classifier was configured with the following fixed hyperparameters [8]:

- `n_estimators` = 100, corresponding to the construction of 100 decision trees;
- `random_state` = 42, ensuring reproducibility of the results;
- `max_depth` = None, allowing each tree to expand until its maximum depth.

The experimental workflow involved splitting the dataset as described in Section 4.4.3. The training set was used both for model development and for applying a 5-fold cross-validation procedure [8]. In k-fold cross-validation, the training set is partitioned into k

equal subsets (folds): at each iteration, the model is trained on $k-1$ folds and validated on the remaining one. The process is repeated k times, and performance metrics are averaged across folds to provide a reliable estimate of the model's generalization ability. This approach eliminates the need for a separate validation set, since each fold acts as validation once.

It is important to note that cross-validation was not used for hyperparameter tuning in this case, as the values of `n_estimators`, `random_state`, and `max_depth` were predetermined. After cross-validation, the final model was retrained on the entire Training Set and subsequently evaluated on the independent Test Set, which had not been used during training or validation.

This methodology ensured a robust evaluation of the classifier's performance [8] while maintaining a strict separation between Training and Test data, thereby preserving the reliability of the generalization assessment.

4.4.5 Support Vector Machines

Support Vector Machines (SVMs) are widely used supervised learning algorithms, particularly effective for binary classification tasks. The fundamental objective is to identify a hyperplane that best separates data points belonging to two different classes[31]. Among all possible separating hyperplanes, SVM aims to construct a separating hyperplane that maximizes the distinction between the two classes of data points [32]. A wider margin generally enhances the model's ability to generalize to unseen data.

However, datasets are not always linearly separable, meaning that no straight line (in two dimensions), plane (in three dimensions), or hyperplane (in higher-dimensional spaces) can perfectly separate the classes without misclassifications. In such cases, SVMs rely on kernel functions [32], which map the original data into a higher-dimensional feature space where linear separation becomes possible. Yet, by applying a radial basis function (RBF) kernel, the data can be projected into a higher-dimensional space where a linear separating hyperplane can be identified.

The SVM model was implemented in Python using the `scikit-learn` library. As a first step, the dataset was divided into Training and Test Sets, as described in Section 4.4.3. The features were then standardized using `StandardScaler`, ensuring that each variable had zero mean and unit variance. This preprocessing step is essential for distance-based algorithms such as SVMs, since it guarantees that all features contribute equally to the definition of the decision boundary.

For classification, a Radial Basis Function (RBF) kernel was employed. The RBF kernel enables the algorithm to handle cases in which the data are not linearly separable in the original feature space, by projecting them into a higher-dimensional space where separation between classes is more feasible [33].

Two main hyperparameters were optimized. The first is the **regularization parameter**

C , which controls the trade-off between the simplicity of the decision boundary and the ability to correctly classify all training samples. Low values of C allow the model to tolerate some misclassifications while producing smoother and more stable boundaries, whereas high values of C enforce stricter separation of training samples, possibly at the cost of overfitting[33]. In this study, several increasing values of C were tested [0.5, 1, 1.8, 2, 5] to identify the optimal compromise.

The second hyperparameter is **gamma**, which defines the width of the RBF kernel and thus the influence of individual data points on the decision boundary. Small gamma values produce smoother and simpler boundaries, while larger values generate highly complex boundaries that may overfit the data. In **scikit-learn**, gamma can be set in three ways:

- "scale": which computes $1/(n_features \cdot \text{Var}(X))$ and is generally robust;
- "auto": which uses $1/n_features$;
- positive float: allows direct control over the model complexity.

The best combination of C and gamma was determined through a Grid Search, an exhaustive exploration of predefined parameter values. Model evaluation during this process was performed using a **Leave-One-Out Cross-Validation (LOO)** applied exclusively to the training set. At each iteration, the model was trained on all samples except one, which was used for validation; this process was repeated until every sample had served once as validation data. The aggregated results provided a robust estimate of model performance.

Finally, after the optimal hyperparameters had been identified, the model was retrained on the entire training set and subsequently evaluated on the test set. This step ensured an unbiased assessment of the generalization ability on unseen data.

In summary, the implementation of supervised Machine Learning methods, namely Random Forests and Support Vector Machines, provided a systematic framework to evaluate the susceptibility of helmet mounted sensors to head-helmet decoupling. These models offered a quantitative means to assess sensor reliability and to support the identification of the most suitable sensor placement for accurate motion tracking.

4.4.6 Evaluation Metrics

To assess model performance, **Accuracy** and **F1 Score** were used as primary evaluation metrics.

Accuracy represents the proportion of correctly classified impacts over the total number of impacts and is mathematically defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where:

- TP = True Positives (high decoupling correctly classified)
- TN = True Negatives (low decoupling correctly classified)

- FP = False Positives (low decoupling misclassified as high)
- FN = False Negatives (high decoupling misclassified as low)

The **F1 Score** provides a balanced measure of a classifier's precision and recall, defined as the harmonic mean of these two quantities:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

where:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}$$

The F1 Score is particularly useful in cases where the dataset is not perfectly balanced, as it accounts for both false positives and false negatives, providing a more comprehensive assessment of classification performance than accuracy alone.

Chapter 5

Results and Discussions

In the Results and Discussions chapter, the main objective is to clearly present the data obtained from the experimental acquisitions, highlighting both the main characteristics of the signals recorded by the sensors and the differences between devices and impact conditions.

5.1 Visual Inspection of Angular Velocity Signals

As described in the previous section, the signals acquired from the Headform and the three helmet-mounted IMUs were carefully preprocessed, including functional calibration, up-sampling, and manual synchronization of individual impacts into 100 ms time windows. This approach allowed obtaining comparable and reliable data for the analysis of impact dynamics, keeping the information from the three Cartesian axes and the vector norm separate, in order to preserve all motion components.

A crucial first step in this process was the **functional calibration**, which ensured that all sensors operated in a common reference frame. Since the helmet-mounted IMUs were attached in arbitrary positions and orientations, their local coordinate systems were initially unknown. To make the signals comparable, each IMU output was transformed into a predefined right-handed frame where the **$\mathbf{X}_{FF}$ point forward, $\mathbf{Y}_{FF}$ upward, and $\mathbf{Z}_{FF}$ to the left.**

In contrast, the Headform reference frame was directly aligned to the Functional Frame (FF) for comparison.

Looking at Figures 5.1, 5.2 and 5.3, a typical impact is characterized by a short and intense impulsive phase, which ideally appears as a sharp rise followed by a maximum peak and a rapid decay (see red signals for the reference). This phase corresponds to the moment when the pendulum makes contact with the helmet and the headform starts to accelerate. The relevant dynamics are fully captured within the first 100 ms after the onset of the impact, which represents the time window of interest in this study. Beyond this interval, helmet signals often show oscillatory behavior related to helmet vibrations

or secondary motion, which are not the focus of the present analysis. When looking at the signals, it should be noticed that the primary contribution to rotation is along the axes aligned with the main rotation axis, as will be explained in detail later in this section. Instead, on the secondary axes, small oscillations or low-amplitude fluctuations can be observed. These do not follow the impulsive profile of the impact and are mainly due to mechanical vibrations of the helmet, indirect force transmission, or measurement noise.

The vector norm, computed as the magnitude of the three-dimensional signal, provides an orientation-independent descriptor of the motion. In the plots, it generally appears as a smoother peak compared to the single Cartesian components. The agreement of the signal norm across the Headform and the helmet-mounted IMUs demonstrates the effectiveness of segmentation and synchronization in producing temporally aligned and comparable data.

For illustration, three representative impact conditions are reported in the following figures, each showing the angular velocity signals recorded by the Headform and the three helmet-mounted IMUs. In all cases, the four subplots represent the three Cartesian axes (X_{FF} , Y_{FF} , Z_{FF}) and the vector norm.

- **Frontal impact, 32° (Figure 5.1):** the dominant rotational contribution is observed along the positive Z_{FF} , consistent with the impact direction. The signal on this axis exhibits the expected impulsive shape, with a sharp rise, a maximum peak, and a fast decay within the 100 ms analysis window. The other axes show only small fluctuations, attributable to noise or secondary helmet motion. The vector norm displays a clear and consistent peak across all sensors, confirming proper synchronization and segmentation.
- **Left-side impact, 50° (Figure 5.2):** in this configuration, the main rotational contribution is along the positive X_{FF} . The angular velocity profile on this axis is characterized by a sharp impulsive peak, while Y_{FF} and Z_{FF} components exhibit only minor oscillations. As in the previous case, the vector norm captures the overall dynamics of the impact, showing a single smooth peak aligned across all devices.
- **Right Front-Boss impact, 70° (Figure 5.3):** the dynamics are more complex, with the primary contribution distributed between the positive X-axis and the negative Z_{FF} . Both components show impulsive peaks of comparable magnitude, reflecting the combined rotation induced by the oblique rearward strike. The Y_{FF} component remains less affected, showing only minor oscillations. The signal norm once again provides a compact description of the overall event, showing a well-defined impulsive peak across all sensors.

These three representative cases demonstrate how the preprocessing pipeline and functional calibration allow consistent comparison between the Headform and helmet-mounted sensors, while also highlighting the expected dependence of the rotational kinematics on the impact direction and angle.

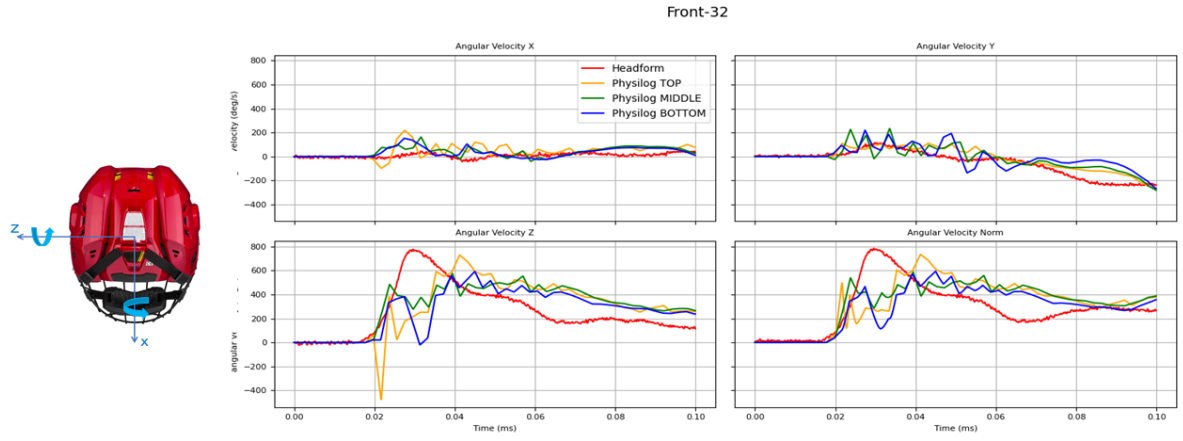


Figure 5.1: Location: Front, Pendulum arm angle: 32.

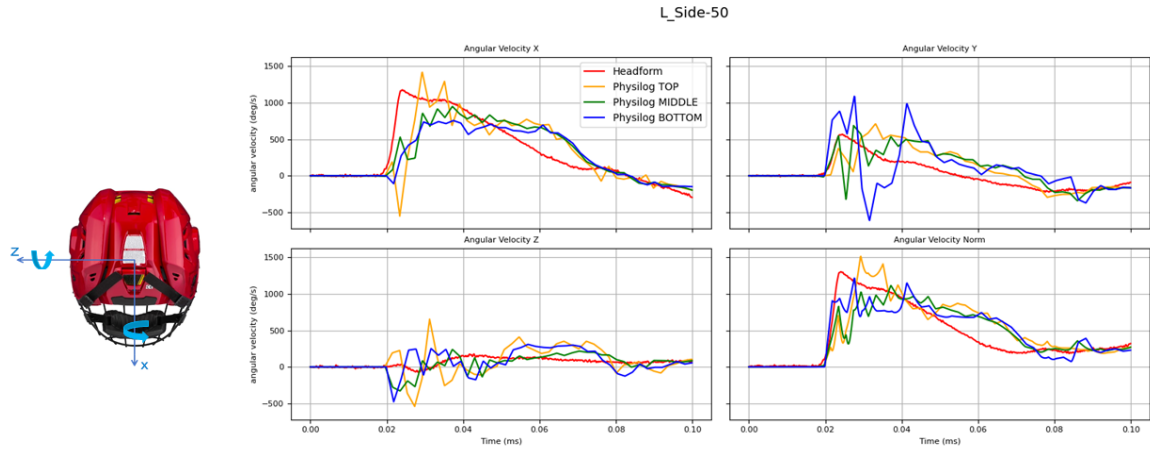


Figure 5.2: Location: Side, Pendulum arm angle: 50.

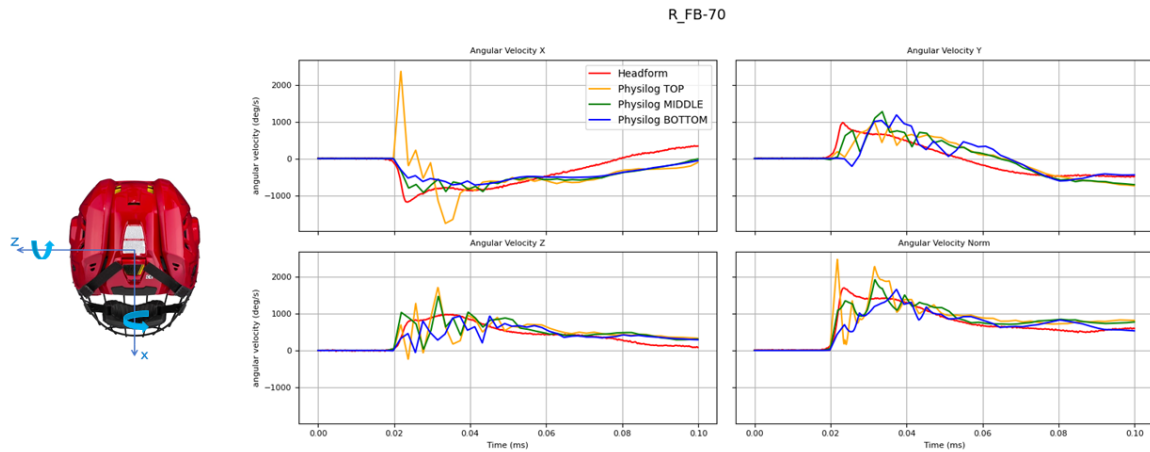


Figure 5.3: Location: Right Front-Boss, Pendulum arm angle: 70.

5.2 Feature Extraction Analysis

It is important to recall that to ensure a fair and reproducible comparison, the same set of features was extracted from all sensors (Headform reference IMU, Physilog Top, Middle, and Bottom) under identical impact conditions. This consistent approach allowed the evaluation of how closely each helmet-mounted IMU reproduces headform dynamics (RQ1) and whether the direction of impact influences measurement reliability (RQ2). Figure 5.4 is an illustrative example: in the time domain the Amplitude feature was

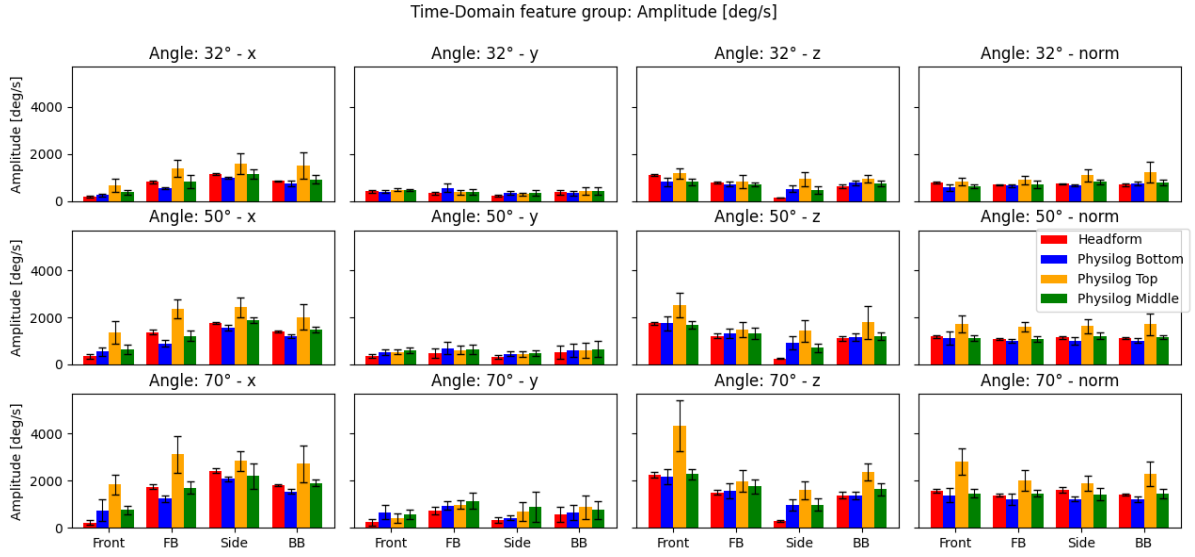


Figure 5.4: Amplitude of angular velocity signals across impact directions and angles, for the Headform reference and helmet-mounted sensors (Top, Middle, Bottom).

considered, defined as the difference between the maximum and minimum values of the angular velocity signal, reflecting the overall excursion of the dynamic response. The plot is organized as a grid of 3×4 subplots, where the three rows correspond to the three tested impact angles, and the four columns represent the four columns represents X_{FF} , Y_{FF} , Z_{FF} and the norm. Within each subplot, the x-axis reports the four impact directions (Front, Front-Oblique, Side, and Back-Oblique), while the y-axis shows the mean value of the Amplitude feature, defined as the difference between the maximum and minimum values of the angular velocity (expressed in degrees per second).

For each impact direction, the bar plots display the results of the four sensors considered: the red bars correspond to the Headform reference IMU, whereas the helmet-mounted sensors are represented in orange (Physilog Top), blue (Physilog Middle), and green (Physilog Bottom). This visual encoding enables direct comparison between the reference and the helmet-mounted sensors.

To correctly interpret this plot, two fundamental considerations must be considered (as mentioned in the previous section). First, each impact direction is associated with a main axis of angular velocity: for frontal impacts the dominant axis is Z_{FF} , for side impacts it

is X_{FF} , while for oblique impacts both X_{FF} and Z_{FF} contribute significantly. Conversely, the Y_{FF} is generally expected to show the smallest amplitudes, since it is not directly excited by the impacts and should mainly reflect noise.

From this perspective, the plot can be interpreted in light of the research questions. With respect to RQ1, which concerns the similarity of helmet-mounted IMU signals with the Headform reference, the relevant comparison is between the mean amplitude values of the sensors and the corresponding values of the Headform. Ideally, the bars of the helmet-mounted sensors should align closely with the red bar of the Headform, indicating minimal deviation from the reference dynamics. Instead, the Physilog Top sensor systematically exhibits higher amplitude values in the X_{FF} , Z_{FF} , and norm components compared to both the Physilog Middle and Bottom and to the Headform. This systematic deviation highlights the Top sensor's greater tendency to amplify angular excursions, which may be interpreted as a higher sensitivity of that location to the shock caused by the impact.

Regarding RQ2, which investigates whether the impact direction influences the degree of similarity between sensors and the Headform, the results do not reveal a consistent pattern. No specific impact direction appears to systematically enhance or reduce similarity across sensors, suggesting that deviations are primarily linked to sensor positioning rather than to the orientation of the impact. In summary, the analysis reveals two key findings:

- The vertical component (Y_{FF}) consistently shows smaller amplitude values across all sensors, confirming its limited contribution to the dynamic response.
- The Physilog Top sensor is systematically less aligned with the Headform reference. A possible explanation is its mounting location: being positioned on the upper shell, farther from the main structural contact points with the headform, it may be more exposed to local deformations and relative motion of the helmet. This is in line with Hypothesis H2, which associates reduced agreement with mounting sites that are mechanically less constrained.

For the Frequency-domain analysis, among the several features extracted, three frequency bands of interest: 0–50 Hz (low-frequency band), 50–100 Hz (mid-frequency band), and 100–250 Hz (high-frequency band) were selected as example for detailed discussion, as it is particularly informative for distinguishing between the meaningful content of the signal and potential contributions from noise or high-frequency vibrations. Figures 5.5, 5.6, 5.7 illustrate the results for each frequency band, organized in a 3×4 grid of subplots. The structure mirrors that used for the previously analyzed Amplitude feature: the rows correspond to the three tested impact angles, while the columns represent the four reference axes (X_{FF} , Y_{FF} , Z_{FF} , and norm). Within each subplot, the x-axis reports the four considered impact directions, whereas the y-axis displays the mean value of the band power (expressed in Hz) for the specific condition.

The colored bars allow distinguishing the different sensors: the reference Headform in red and the three helmet-mounted sensors in blue, green, and orange, respectively.

The analysis of the plots highlights several key findings:

- **Low-frequency band (0–50 Hz):** The majority of the informative content of the signal is concentrated in this band, as expected from the literature. As the pendulum angle increases, and thus the kinetic energy of the impact, the spectral power recorded by the sensors also increases. The dominant rotation axes are clearly visible: for the Front direction, the main contribution lies along the Z_{FF} , while for the Side it is along the X_{FF} . This behavior is consistent with the physical dynamics of the impacts, since the axes directly solicited by the collision are associated with the highest energy content, whereas the Y_{FF} systematically remains the least relevant, as it is not directly involved.
- **Mid and High-frequency bands (50–100 Hz and 100–250 Hz):** In these bands, the spectral content is generally lower, suggesting that the signals do not exhibit significant contributions at higher frequencies and that the level of noise remains limited compared to the Headform. However, one relevant aspect emerges: the Physilog Top sensor (orange bars) systematically shows higher values than the other two helmet-mounted sensors (Physilog Middle and Bottom). This behavior indicates a greater susceptibility of the Top sensor to introduce unwanted contributions in the mid-to-high frequency range, thereby highlighting a higher noise level in its signals.

In conclusion, the results confirm that the most meaningful portion of the angular velocity signal content is concentrated in the low-frequency band (0–50 Hz), consistent with the real dynamics of impacts. The higher-frequency bands, on the other hand, are mainly associated with noise or mechanical disturbances, with the Physilog Top sensor standing out as the least reliable in spectral terms.

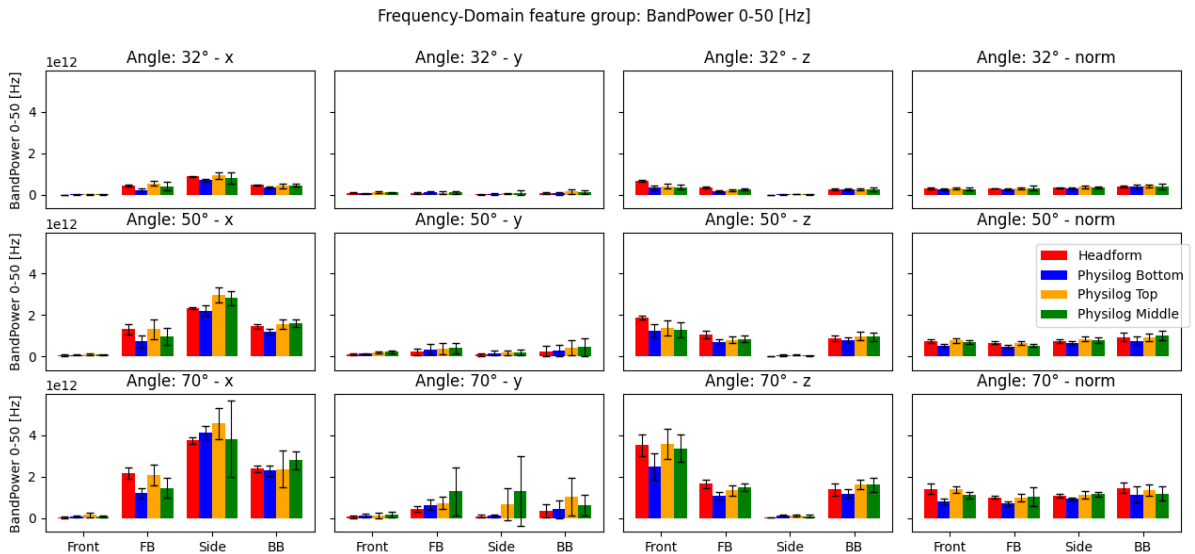


Figure 5.5: Band power in the 0–50 Hz range: distribution of spectral energy across impact directions.

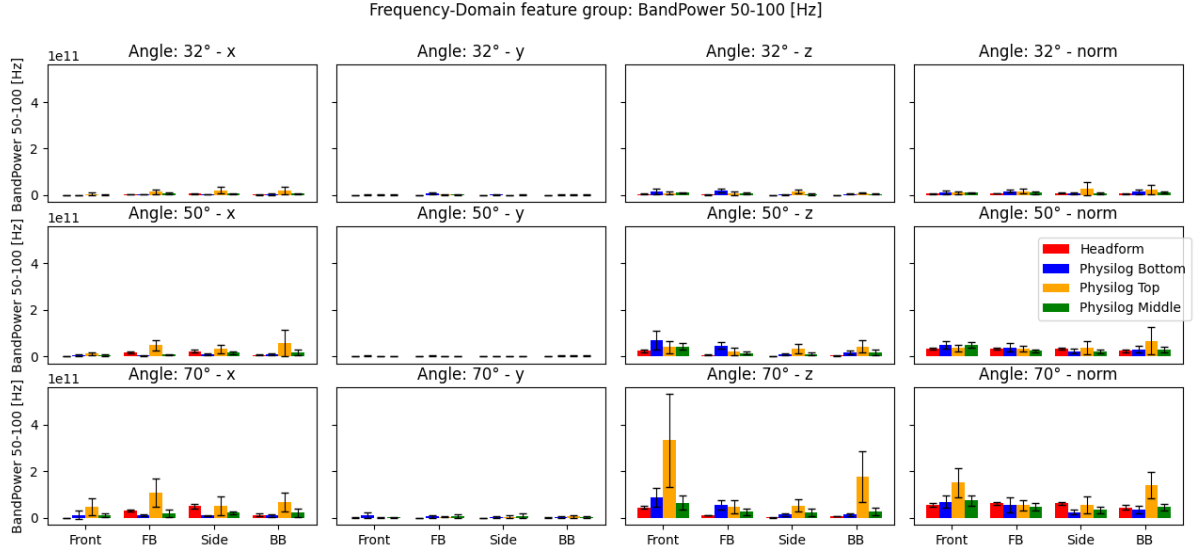


Figure 5.6: Band power in the 50-100 Hz range: distribution of spectral energy across impact directions.

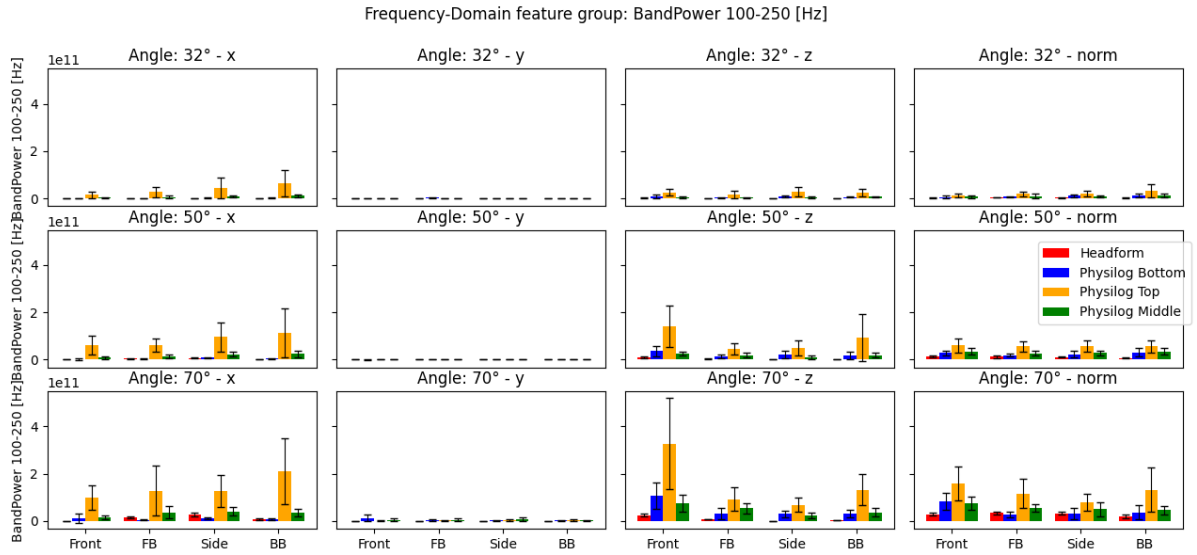


Figure 5.7: Band power in the 100-250 Hz range: distribution of spectral energy across impact directions.

5.3 Quantitative Comparison and Statistical Results

5.3.1 Correlation between extracted features

The correlation matrices were computed for each experimental configuration (impact direction and angle) and visualized in the form of heatmaps.

An example of these visualizations is reported for each of the three feature domains analyzed Figures 5.8, 5.9, 5.10 . In the heatmaps, each block along the main diagonal represents the correlation of a sensor's features with themselves, which consistently results

in a Pearson coefficient equal to 1, confirming the correctness of the implementation. More interestingly, the off-diagonal blocks represent the correlations either between different features of the same sensor, or between features extracted from different sensors. In particular, the correlation between the reference sensor (Headform) and the three Physilog sensors was the focus of interest, as it may indicate the degree to which the helmet sensors capture similar information to the reference measurement.

From the visual inspection of the heatmaps, it is evident that correlations vary considerably depending on both the feature domain and the specific experimental configuration (impact angle and direction). In some cases, inter-sensor correlations were observed, while in others the degree of similarity appeared weaker. Notably, no consistent or repeatable pattern emerged across conditions, suggesting that correlation strength is highly context-dependent.

Given the variability observed in the exploratory heatmap analysis and the qualitative nature of these inspections, a more rigorous statistical approach was deemed necessary to objectively quantify the similarity between the reference Headform sensor and the Physilog sensors, which can be considered statistically meaningful or negligible.

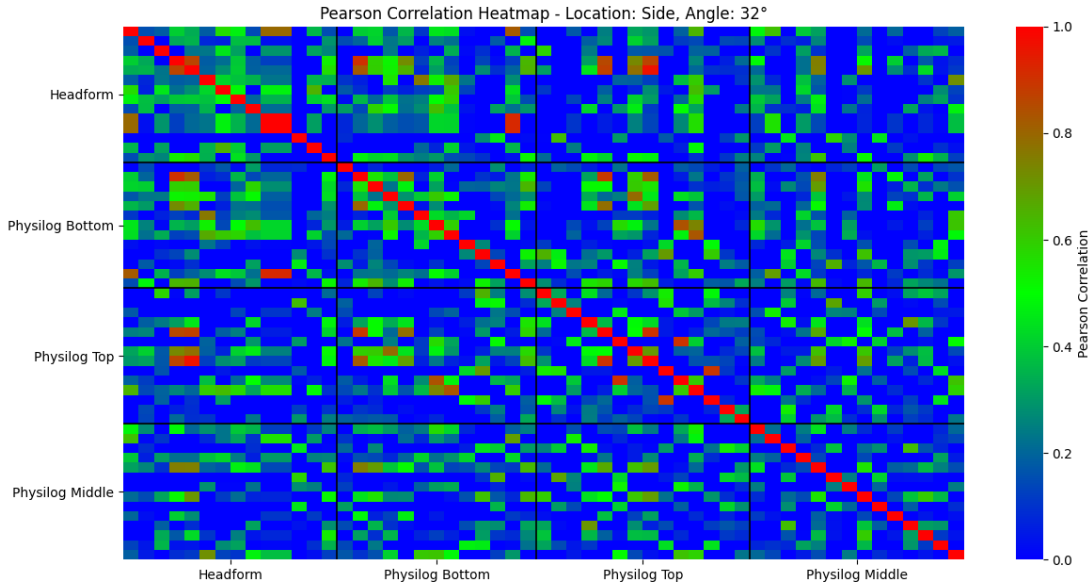


Figure 5.8: Heatmap Time Domain (Side, 32°)

5.3.2 Interpretation of the Statistical Results

Figure 5.11 summarizes, for each frequency-domain feature, the outcome of the paired comparisons between the Headform and the three helmet-mounted sensors (Physilog Top, Middle, and Bottom). On the $\mathbf{X}_{FF}$, the three sensor pairs are reported, while the $\mathbf{Y}_{FF}$ lists the extracted features. Each square in the plot represents the result of a paired comparison for a given feature–sensor pair.

The color encodes the outcome of the statistical test:

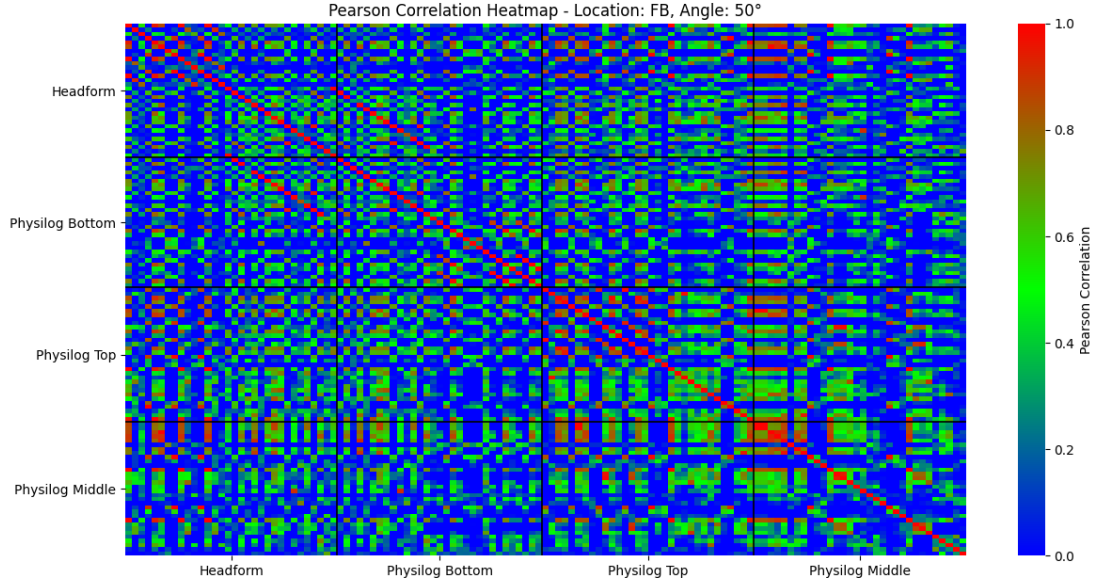


Figure 5.9: Heatmap Frequency Domain (Front Boss, 50°)

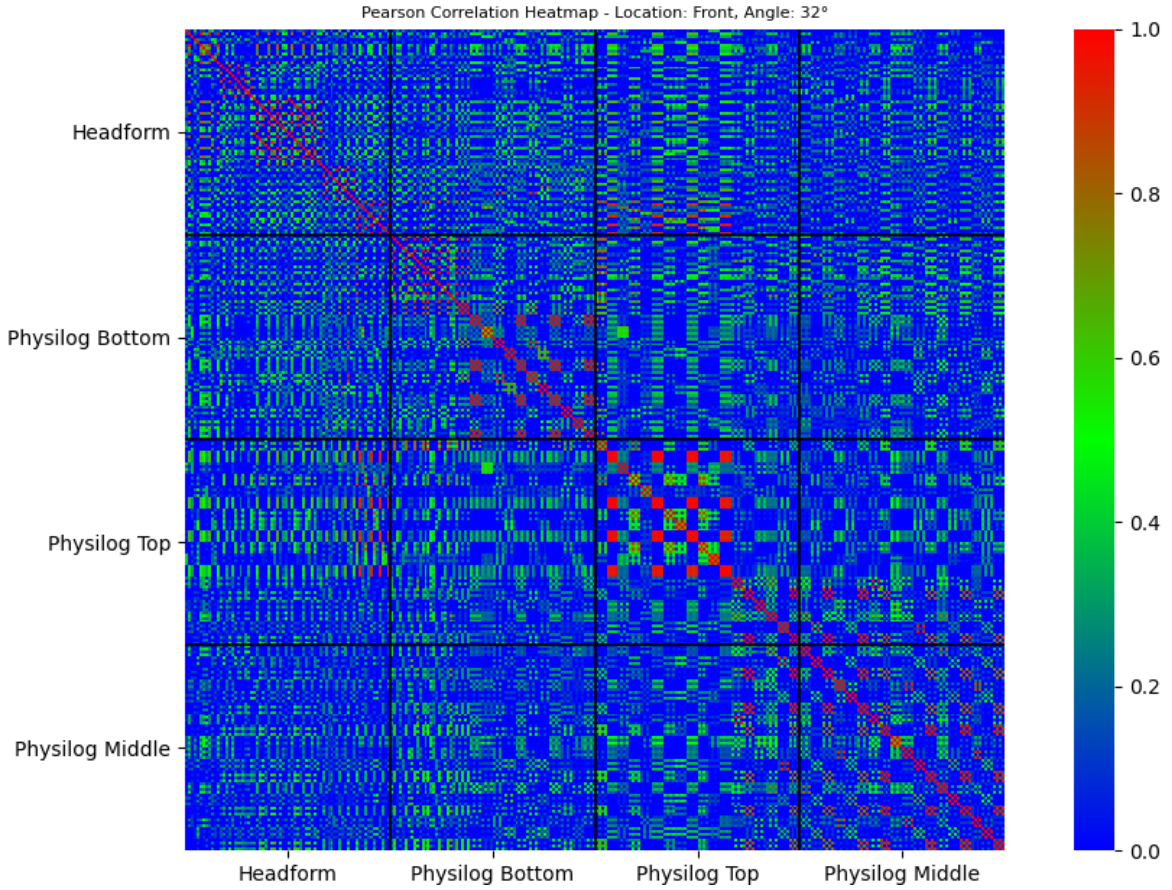


Figure 5.10: Heatmap Time-Frequency Domain (Front, 32°)

- **Green squares:** the helmet sensor measurements are statistically comparable to the Headform, i.e., no systematic deviation was detected.
- **Red squares:** a systematic difference was identified, indicating that the helmet sensor did not fully reproduce the Headform reference for that feature.

In addition to statistical significance, the size or shading of the markers reflects the **effect size (Cohen's d)**, which provides information on the practical relevance of the difference. Small values indicate negligible deviations even if significant, while large values highlight substantial discrepancies that may impact the interpretation of sensor performance.

This visualization therefore enables a rapid identification of:

- which features are reliably captured by helmet sensors across locations;
- where systematic deviations occur;
- whether detected differences are minor or potentially relevant in biomechanical terms.

Taking as an example the **BandPower 0–50 Hz** feature, which represents the impact information content in the main frequency band, the following can be observed:

The impact direction is Side, and the main signal axis is X_{FF} , which captures the predominant motion contribution.

In the row corresponding to BandPower 0–50 Hz on the X_{FF} :

- Headform vs. Physilog Middle: green square, $p \geq 0.001$, Cohen's $d = 0.22$
- Headform vs. Physilog Top: green square, $p \geq 0.001$, Cohen's $d = -0.35$

This indicates that the measurements from the Physilog Middle and Top sensors are statistically comparable to those from the Headform for this feature. The negative sign of Cohen's d merely indicates the direction of the difference (Headform < Physilog or vice versa), but the magnitude is assessed using the absolute value, which in both cases falls within the medium effect range.

Conversely, red squares in higher frequency bands or along other axes highlight statistically significant differences, likely due to noise or minor signal contributions.

In contrast, examining the **BandPower 50–100 Hz** feature in the **norm**, all three helmet-mounted sensors—Top, Middle, and Bottom—show green squares, indicating $p \geq 0.001$, for each sensor pair. This result signifies that, for this particular feature, the measurements from all helmet-mounted sensors are statistically comparable to the Headform, suggesting a high degree of consistency and robustness across sensor positions. Such a finding highlights the reliability of this frequency band in capturing relevant impact information without significant sensor-dependent variation.

In this case (Figure 5.12), the impact direction is **Front Boss** with an incident angle of 50° , and the feature under analysis is the **Median Spectral Density**. The MSD represents the frequency value that divides the power spectrum of the signal into two equal halves, providing an estimate of the dominant spectral content and the central tendency of energy distribution across frequencies. In other words, it reflects the central frequency around which most of the signal energy is concentrated, offering insight into the main oscillatory components of the impact response.

For this impact direction, the contributions of the impact are notable on both the X_{FF} and Z_{FF} axes. The comparison between the Headform and the helmet-mounted sensors reveals the following:

Paired T-Test p-values and Cohen's d | Location: Side | Angle: 32°

	Physilog Bottom	Physilog Middle	Physilog Top
BandPower 0-50Hz Norm	p=0.0000 d=1.37	p=0.1936 d=-0.24	p=0.0003 d=-0.76
BandPower 0-50Hz X	p=0.0000 d=3.64	p=0.2367 d=-0.22	p=0.0040 d=-0.55
BandPower 0-50Hz Y	p=0.0001 d=-0.80	p=0.0023 d=-0.91	p=0.0000 d=-0.88
BandPower 0-50Hz Z	p=0.0000 d=-3.02	p=0.0000 d=-2.99	p=0.0000 d=-3.58
BandPower 100-250Hz Norm	p=0.0000 d=-1.15	p=0.0000 d=-1.57	p=0.0000 d=-1.44
BandPower 100-250Hz X	p=0.0031 d=-0.99	p=0.0000 d=-1.64	p=0.0000 d=-0.93
BandPower 100-250Hz Y	p=0.0000 d=-1.79	p=0.0000 d=-0.88	p=0.0000 d=-1.37
BandPower 100-250Hz Z	p=0.0000 d=-1.51	p=0.0000 d=-1.15	p=0.0000 d=-1.33
BandPower 50-100Hz Norm	p=0.3036 d=0.15	p=0.7516 d=0.05	p=0.0030 d=-0.95
BandPower 50-100Hz X	p=0.0000 d=1.73	p=0.8218 d=-0.04	p=0.0000 d=-1.15
BandPower 50-100Hz Y	p=0.0000 d=1.03	p=0.0002 d=-0.78	p=0.0000 d=-1.98
BandPower 50-100Hz Z	p=0.0000 d=-1.48	p=0.0000 d=-1.30	p=0.0000 d=-2.16
Diff Between Max and Centroid Norm	p=0.0000 d=1.04	p=0.0000 d=0.97	p=0.0000 d=0.95
Diff Between Max and Centroid X	p=0.0000 d=1.63	p=0.0143 d=-0.48	p=0.0000 d=1.05
Diff Between Max and Centroid Y	p=0.0006 d=-0.71	p=0.0026 d=0.60	p=0.0000 d=1.38
Diff Between Max and Centroid Z	p=0.8598 d=-0.08	p=0.0000 d=1.23	p=0.0068 d=-0.53
Max Ps Norm	p=0.0358 d=0.40	p=0.0006 d=-0.71	p=0.0001 d=-0.84
Max Ps X	p=0.0000 d=1.37	p=0.9618 d=0.01	p=0.0026 d=0.60
Max Ps Y	p=0.0635 d=-0.35	p=0.0078 d=-0.52	p=0.0008 d=-0.68
Max Ps Z	p=0.0000 d=-0.98	p=0.0000 d=-1.32	p=0.0000 d=-1.72
MedianSpectralDensity Norm	p=0.0773 d=0.33	p=0.0001 d=-0.84	p=0.0000 d=-1.18
MedianSpectralDensity X	p=0.0019 d=0.62	p=0.1036 d=-0.31	p=0.0000 d=-0.92
MedianSpectralDensity Y	p=0.0000 d=-1.19	p=0.0010 d=-0.67	p=0.0000 d=-1.01
MedianSpectralDensity Z	p=0.0000 d=2.80	p=0.0000 d=3.08	p=0.0000 d=1.68
SpectralCentroid Norm	p=0.0000 d=-0.92	p=0.0004 d=-0.74	p=0.0000 d=-1.00
SpectralCentroid X	p=0.0000 d=1.70	p=0.0147 d=-0.47	p=0.0000 d=-1.05
SpectralCentroid Y	p=0.0001 d=0.85	p=0.1977 d=0.24	p=0.0000 d=-1.09
SpectralCentroid Z	p=0.0757 d=-0.34	p=0.0002 d=0.77	p=0.0000 d=-1.28
SpectralFWHM Norm	p=0.0000 d=1.01	p=0.0000 d=0.99	p=0.0000 d=0.99
SpectralFWHM X	p=0.0003 d=0.75	p=0.0013 d=0.62	p=0.0013 d=0.62
SpectralFWHM Y	p=0.0000 d=-0.99	p=0.0003 d=-0.76	p=0.0003 d=-0.76
SpectralFWHM Z	p=0.0000 d=-1.40	p=0.0000 d=-0.91	p=0.0000 d=-0.91

Figure 5.11: Visual summary of paired t-test results comparing the Headform reference sensor with helmet-mounted Physilog sensors (Top, Middle, Bottom) for Frequency-domain features, Front impact direction, 32°.

- **Physilog Bottom:** green squares for all three axes (X_{FF} , Y_{FF} , Z_{FF}), $p \geq 0.001$, indicating that the Median Spectral Density measured by the bottom sensor is statistically comparable to the Headform across all axes.
- **Physilog Top and Middle:** green squares only for the Y and Z axes, $p \geq 0.001$, suggesting that on these sensors, the Median Spectral Density is comparable to the Headform only along the vertical and lateral directions, but not along the X_{FF} .

The similarity of the Median Spectral Density in this impact direction indicates that the helmet-mounted sensors are capturing the dominant spectral content of the impact signal in a way that is largely consistent with the Headform reference. For the Bottom sensor, this consistency across all three axes suggests a robust replication of the headform dynamics in multiple directions, which could enhance the reliability of derived features such as impact intensity, energy distribution, and directional analysis.

In contrast, the Top and Middle sensors show agreement only along Y_{FF} and Z_{FF} axes, which may reflect local variations in sensor mounting or minor directional biases. These differences highlight that, depending on sensor position, some directional components of the signal may be less accurately captured, which could affect subsequent analyses that rely on the spectral content along the X_{FF} .

Two key time-domain features derived from the angular velocity signals are analyzed: **Amplitude (Min–Max Range)** and **Zero Crossing Rate (Adaptive Threshold)**. The analysis focuses on different impact configurations and axes to highlight how helmet-mounted sensors replicate the Headform measurements in the time domain.

Amplitude (Min–Max Range) The Amplitude is defined as the difference between the

Paired T-Test p-values and Cohen's d | Location: FB | Angle: 50°

	Physilog Bottom	Physilog Middle	Physilog Top
BandPower 0-50Hz Norm	p=0.0000 d=2.58	p=0.0000 d=2.13	p=0.1576 d=0.26
BandPower 0-50Hz X	p=0.0000 d=4.30	p=0.0000 d=1.68	p=0.9722 d=0.01
BandPower 0-50Hz Y	p=0.0000 d=-1.09	p=0.0000 d=-1.43	p=0.0000 d=-1.29
BandPower 0-50Hz Z	p=0.0000 d=2.32	p=0.0000 d=1.18	p=0.0000 d=2.02
BandPower 100-250Hz Norm	p=0.0004 d=0.73	p=0.0000 d=-1.54	p=0.0000 d=1.86
BandPower 100-250Hz X	p=0.0109 d=0.46	p=0.0000 d=-1.39	p=0.0000 d=-1.99
BandPower 100-250Hz Y	p=0.0039 d=-0.54	p=0.0107 d=-0.50	p=0.0004 d=-0.73
BandPower 100-250Hz Z	p=0.0000 d=-1.39	p=0.0000 d=-1.83	p=0.0000 d=-1.71
BandPower 50-100Hz Norm	p=0.1868 d=-0.25	p=0.0000 d=1.17	p=0.8655 d=0.03
BandPower 50-100Hz X	p=0.0000 d=4.12	p=0.0000 d=1.99	p=0.0000 d=-1.36
BandPower 50-100Hz Y	p=0.0000 d=-0.93	p=0.0070 d=-0.53	p=0.0001 d=-0.81
BandPower 50-100Hz Z	p=0.0000 d=-3.01	p=0.0000 d=-1.10	p=0.0000 d=-1.06
Diff Between Max and Centroid Norm	p=0.0094 d=-0.54	p=0.0000 d=-1.68	p=0.0000 d=-2.48
Diff Between Max and Centroid X	p=0.0000 d=0.89	p=0.0000 d=-1.02	p=0.0000 d=-1.90
Diff Between Max and Centroid Y	p=0.0004 d=-0.72	p=0.4198 d=0.15	p=0.5358 d=0.11
Diff Between Max and Centroid Z	p=0.0000 d=-1.99	p=0.0000 d=-1.43	p=0.0000 d=-2.07
Max Ps Norm	p=0.0000 d=2.54	p=0.0000 d=2.16	p=0.0000 d=1.62
Max Ps X	p=0.0000 d=3.12	p=0.0000 d=1.22	p=0.0102 d=0.50
Max Ps Y	p=0.0000 d=-0.93	p=0.0000 d=-1.40	p=0.0000 d=-1.31
Max Ps Z	p=0.0000 d=1.75	p=0.0000 d=1.02	p=0.0000 d=1.53
MedianSpectralDensity Norm	p=0.0531 d=-0.37	p=0.0000 d=-1.39	p=0.0000 d=-1.42
MedianSpectralDensity X	p=0.2681 d=0.20	p=0.0002 d=-0.79	p=0.0003 d=-0.75
MedianSpectralDensity Y	p=0.0892 d=0.31	p=0.1912 d=0.24	p=0.0017 d=-0.63
MedianSpectralDensity Z	p=0.1740 d=-0.25	p=0.0027 d=-0.60	p=0.0115 d=-0.46
SpectralCentroid Norm	p=0.0000 d=-1.75	p=0.0000 d=-2.41	p=0.0000 d=-2.32
SpectralCentroid X	p=0.0000 d=0.93	p=0.0000 d=-1.04	p=0.0000 d=-1.90
SpectralCentroid Y	p=0.0006 d=-0.71	p=0.4323 d=0.15	p=0.4085 d=0.13
SpectralCentroid Z	p=0.0000 d=-1.98	p=0.0000 d=-1.43	p=0.0000 d=-2.06
SpectralFWHM Norm	p=0.0000 d=-2.23	p=0.0000 d=-1.59	p=0.0000 d=-1.59
SpectralFWHM X	p=0.0000 d=0.99	p=0.0067 d=0.53	p=0.0067 d=0.53
SpectralFWHM Y	p=0.0047 d=-0.56	p=0.0713 d=-0.34	p=0.0713 d=-0.34
SpectralFWHM Z	p=0.0602 d=-0.36	p=0.0045 d=-0.56	p=0.0045 d=-0.56

Figure 5.12: Visual summary of paired t-test results comparing the Headform reference sensor with helmet-mounted Physilog sensors (Top, Middle, Bottom) for Frequency-domain features, Front-Boss impact direction, 50°.

maximum and minimum angular velocity within the selected time window:

$$\text{Amplitude} = \max(\omega(t)) - \min(\omega(t)).$$

This metric provides a direct estimate of the maximum angular excursion recorded during the impact, reflecting the intensity of the rotational motion experienced by the head.

For the **Front impact at 50°** Figure 5.13, the principal axis of interest is the Z_{FF} , which captures the dominant component of the rotational movement. Paired statistical analysis between the Headform and the helmet-mounted sensors reveals the following:

- **Physilog Bottom and Middle:** paired t -tests yield $p \geq 0.001$, indicating that the null hypothesis of no systematic difference cannot be rejected. Cohen's d values fall within the small to medium range, confirming that the magnitude of any observed differences is negligible.
- **Physilog Top:** statistically significant differences are observed along the Z_{FF} ($p < 0.001$), suggesting a systematic deviation from the Headform measurement.

These results indicate that the rotational dynamics captured by the Bottom and Middle sensors along the Z_{FF} closely reproduce the Headform reference in terms of angular velocity amplitude. Consequently, these sensors can reliably represent the intensity of the impact, supporting their use in estimating head kinematics during frontal collisions.

Zero Crossing Rate (Adaptive Threshold) The Zero Crossing Rate (ZCR) quantifies the number of times the signal crosses a threshold, which in this case is an adaptive value computed as half of the mean of the absolute peak angular velocities from the helmet sensors. This adaptation compensates for the higher variability and spurious oscillations present in helmet-mounted signals compared to the cleaner Headform measurements. The

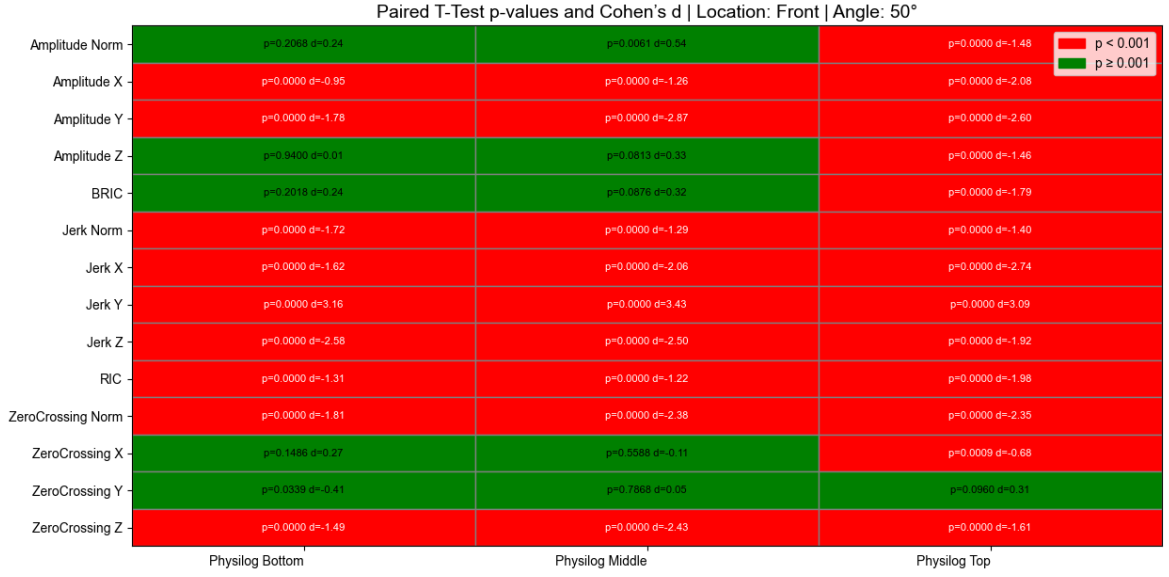


Figure 5.13: Visual summary of paired t-test results comparing the Headform reference sensor with helmet-mounted Physilog sensors (Top, Middle, Bottom) for Time-domain features, Front impact direction, 50°.

ZCR provides insight into the oscillatory behavior of the signal, potentially reflecting sensor quality or coupling effectiveness.

For the **Front impact at 50°** 5.14 analyzed along the **Side direction**, the principal axis is the X_{FF} . The paired t -tests indicate:



Figure 5.14: Visual summary of paired t-test results comparing the Headform reference sensor with helmet-mounted Physilog sensors (Top, Middle, Bottom) for Time-domain features, Side impact direction, 50°.

- All three helmet-mounted sensors (Top, Middle, Bottom) yield $p \geq 0.001$, with small Cohen's d values, indicating that the oscillatory behavior measured by the sensors is statistically indistinguishable from the Headform along this axis.

This finding suggests that, in terms of signal oscillations relative to the adaptive threshold, all helmet-mounted sensors adequately replicate the Headform dynamics along the X_{FF} for this impact configuration. Such consistency implies that the temporal characteristics of the rotational signal, including the frequency of angular velocity fluctuations, can be reliably captured by each sensor in the sideward component of frontal impacts.

The combined analysis of Amplitude and Zero Crossing Rate across different axes and impact directions demonstrates that helmet sensor placement and the choice of feature critically influence measurement fidelity. While the Bottom and Middle sensors reproduce angular velocity amplitudes accurately along the Z_{FF} , all sensors capture the oscillatory behavior along the X_{FF} for the Side component. These results emphasize the importance of axis-specific evaluation when validating wearable sensor performance against a reference Headform.

The Wavelet Time-to-Peak feature represents the moment in the time–frequency domain when the maximum energy of the signal occurs, thus indicating the temporal alignment between the sensor and the reference (Headform).

In the Figure 5.15 for the Front impact at 32° , where the main axis of interest is Z, the paired t-test analysis shows that the Physilog Top and Physilog Middle sensors do not differ significantly from the Headform. This means that, in this configuration, their time-to-peak estimates are statistically equivalent to the reference, highlighting their ability to accurately capture the temporal dynamics of the impact.

Paired T-Test - Location: Front, Angle: 32

wt_max_pow_freq_norm	p=0.0000 d=-1.60	p=0.0000 d=-0.87	p=0.0000 d=-1.06	p < 0.001 p ≥ 0.001
wt_max_pow_freq_x	p=0.0000 d=-0.34	p=0.0000 d=-0.28	p=0.0000 d=-1.5	
wt_max_pow_freq_y	p=0.0000 d=-0.18	p=0.0000 d=-0.16	p=0.0000 d=-1.0	
wt_max_pow_freq_z	p=0.0000 d=-2.45	p=0.0001 d=-0.92	p=0.0000 d=-1.8	
wt_max_pow_norm	p=0.0000 d=-1.45	p=0.0000 d=-2.04	p=0.0000 d=-1.5	
wt_max_pow_x	p=0.0000 d=-1.95	p=0.0000 d=-1.86	p=0.0000 d=-1.05	
wt_max_pow_y	p=0.0000 d=-1.26	p=0.0000 d=-1.03	p=0.0000 d=-0.87	
wt_max_pow_z	p=0.0000 d=-1.12	p=0.0000 d=-2.14	p=0.0000 d=-1.12	
wt_max_time_to_peak_norm	p=0.0000 d=-0.17	p=0.0000 d=-0.28	p=0.0000 d=-0.31	
wt_max_time_to_peak_x	p=0.0000 d=-1.26	p=0.0000 d=-1.38	p=0.0000 d=-1.41	
wt_max_time_to_peak_y	p=0.0000 d=-1.06	p=0.0017 d=-0.63	p=0.0000 d=-0.59	
wt_max_time_to_peak_z	p=0.0000 d=-1.33	p=0.0000 d=-1.12	p=0.0000 d=-1.01	
wt_peak_0_bandwidth_norm	p=0.0000 d=-1.33	p=0.0000 d=-1.12	p=0.0000 d=-1.49	
wt_peak_0_bandwidth_x	p=0.0000 d=-0.24	p=0.0000 d=-0.11	p=0.0000 d=-0.17	
wt_peak_0_bandwidth_y	p=0.0000 d=-1.85	p=0.0000 d=-1.83	p=0.0000 d=-1.06	
wt_peak_0_bandwidth_z	p=0.0000 d=-2.26	p=0.0000 d=-0.38	p=0.0000 d=-1.25	
wt_peak_0_duration_norm	p=0.0000 d=-1.33	p=0.0000 d=-1.12	p=0.0000 d=-2.29	
wt_peak_0_duration_x	p=0.0000 d=-4.60	p=0.0000 d=-0.30	p=0.0000 d=-0.70	
wt_peak_0_duration_y	p=0.0000 d=-1.45	p=0.0000 d=-2.04	p=0.0000 d=-1.52	
wt_peak_0_max_pow_norm	p=0.0000 d=-1.95	p=0.0000 d=-1.86	p=0.0000 d=-1.05	
wt_peak_0_max_pow_x	p=0.0000 d=-1.26	p=0.0000 d=-1.03	p=0.0000 d=-0.87	
wt_peak_0_max_pow_y	p=0.0000 d=-1.12	p=0.0000 d=-2.14	p=0.0000 d=-1.12	
wt_peak_0_max_pow_z	p=0.0000 d=-1.16	p=0.0000 d=-0.94	p=0.0000 d=-1.02	
wt_peak_1_bandwidth_norm	p=0.0000 d=-1.47	p=0.0000 d=-1.29	p=0.0000 d=-1.48	
wt_peak_1_bandwidth_x	p=0.0000 d=-1.48	p=0.0000 d=-1.29	p=0.0000 d=-1.02	
wt_peak_1_bandwidth_y	p=0.0000 d=-1.01	p=0.0000 d=-1.17	p=0.0000 d=-0.89	
wt_peak_1_bandwidth_z	p=0.0000 d=-0.83	p=0.0000 d=-1.17	p=0.0000 d=-1.61	
wt_peak_1_duration_norm	p=0.0000 d=-0.34	p=0.0000 d=-0.55	p=0.0000 d=-0.68	
wt_peak_1_duration_x	p=0.0000 d=-1.47	p=0.0000 d=-2.36	p=0.0000 d=-1.27	
wt_peak_1_duration_y	p=0.0000 d=-1.38	p=0.0000 d=-1.82	p=0.0000 d=-1.51	
wt_peak_1_max_pow_norm	p=0.0000 d=-1.92	p=0.0000 d=-1.82	p=0.0000 d=-1.05	
wt_peak_1_max_pow_x	p=0.0000 d=-1.26	p=0.0000 d=-1.13	p=0.0000 d=-1.02	
wt_peak_1_max_pow_y	p=0.0000 d=-1.03	p=0.0000 d=-2.50	p=0.0000 d=-1.10	
wt_peak_1_max_pow_z	p=0.0000 d=-0.98	p=0.0000 d=-1.08	p=0.0002 d=-0.78	
wt_peak_2_bandwidth_norm	p=0.0000 d=-0.12	p=0.0000 d=-0.20	p=0.0000 d=-1.67	
wt_peak_2_bandwidth_x	p=0.0000 d=-0.12	p=0.0000 d=-0.20	p=0.0000 d=-1.35	
wt_peak_2_bandwidth_y	p=0.0000 d=-1.52	p=0.0000 d=-1.06	p=0.0000 d=-1.16	
wt_peak_2_bandwidth_z	p=0.0000 d=-1.07	p=0.0000 d=-0.94	p=0.0000 d=-1.86	
wt_peak_2_duration_norm	p=0.0000 d=-1.02	p=0.0000 d=-1.41	p=0.0000 d=-1.86	
wt_peak_2_duration_x	p=0.0000 d=-0.83	p=0.0000 d=-0.54	p=0.0000 d=-1.35	
wt_peak_2_duration_y	p=0.0000 d=-0.92	p=0.0000 d=-1.18	p=0.0004 d=-0.73	
wt_peak_2_max_pow_norm	p=0.0000 d=-1.97	p=0.0000 d=-1.81	p=0.0000 d=-1.12	
wt_peak_2_max_pow_x	p=0.0000 d=-1.46	p=0.0000 d=-1.10	p=0.0000 d=-0.95	
wt_peak_2_max_pow_y	p=0.0000 d=-0.65	p=0.0000 d=-1.14	p=0.0000 d=-0.94	
wt_peak_2_max_pow_z	p=0.0000 d=-0.65	p=0.0000 d=-1.14	p=0.0000 d=-0.94	

Headform vs Physilog Bottom

Headform vs Physilog Middle

Headform vs Physilog Top

Figure 5.15: Visual summary of paired t-test results comparing the Headform reference sensor with helmet-mounted Physilog sensors (Top, Middle, Bottom) for Time-Frequency domain features, Front impact direction, 32° .

Overall, the statistical analyses performed across the time, frequency, and time–frequency domains demonstrate that the degree of agreement between helmet-mounted sensors and the Headform varies depending on both the feature considered and the impact configuration.

In several cases, the sensors—particularly in the Middle and Bottom positions—provided measurements statistically equivalent to the reference, faithfully reproducing specific aspects of the impact dynamics. However, systematic differences also emerged in other conditions, highlighting the influence of sensor placement and axis-specific sensitivity. These findings provide the necessary context for introducing the final and most critical plot, which condenses the key outcomes of the statistical investigation and identifies the features with the highest reliability for impact assessment.

5.3.3 Statistical Summary: Sensor Comparison and Impact Direction Influence

The plot in Figure 5.16 represents the final synthesis of the statistical analysis, where the information from the Time, Frequency, and Time-Frequency domains was integrated to provide a comprehensive overview of the performance of the Helmet-mounted sensors.

For each combination of impact direction and angle (x-axis of the plot), the total number of features (y-axis) was computed that simultaneously satisfied two statistical criteria:

- $p \geq 0.001$
- Cohen’s $d < 0.5$ (at most medium effect size).

This procedure identifies features for which the sensor measurements are statistically indistinguishable from those of the Headform (ground truth).

The three colored curves represent the different sensors (Physilog Bottom in blue, Physilog Top in orange, and Physilog Middle in green). The points plotted on the curves indicate, for each impact condition, the number of statistically equivalent features, while their red shading reflects the magnitude of Cohen’s d : from light red (small effect) to dark red (very large effect).

The total number of features that satisfied the criteria for each sensor is reported in the legend (top-right corner):

- Physilog Top: 268 features
- Physilog Bottom: 331 features
- Physilog Middle: 395 features

These aggregate values provide a quantitative index of each sensor’s overall ability to approximate the Headform measurements.

From the analysis of the plot, two main insights emerge:

1. Relative Sensor Performance

- The Physilog Middle shows the highest number of statistically equivalent features, suggesting a slightly more reliable performance in capturing impact dynamics.
- However, no sensor demonstrates consistent or pronounced superiority across all experimental conditions. The observed differences are not sufficient to indicate a universally best-performing sensor.

2. Impact Direction Influence

- The initial hypothesis predicted that sensors closer to the impact point would deviate more from the Headform. This is not fully confirmed by the results.
- In particular:
 - **Physilog Bottom:** expected to perform well in frontal impacts and poorly in back impacts, but in reality it performs satisfactorily in both.
 - **Physilog Top:** expected to perform well in back impacts and poorly in front impacts, but instead shows suboptimal results in both cases.

Overall, the Physilog Middle, positioned in a more isolated area of the helmet, appears to be the relatively most robust and consistent sensor. Nevertheless, the advantage is not decisive: no helmet-mounted IMU consistently reproduces the Headform signals across all impact configurations.

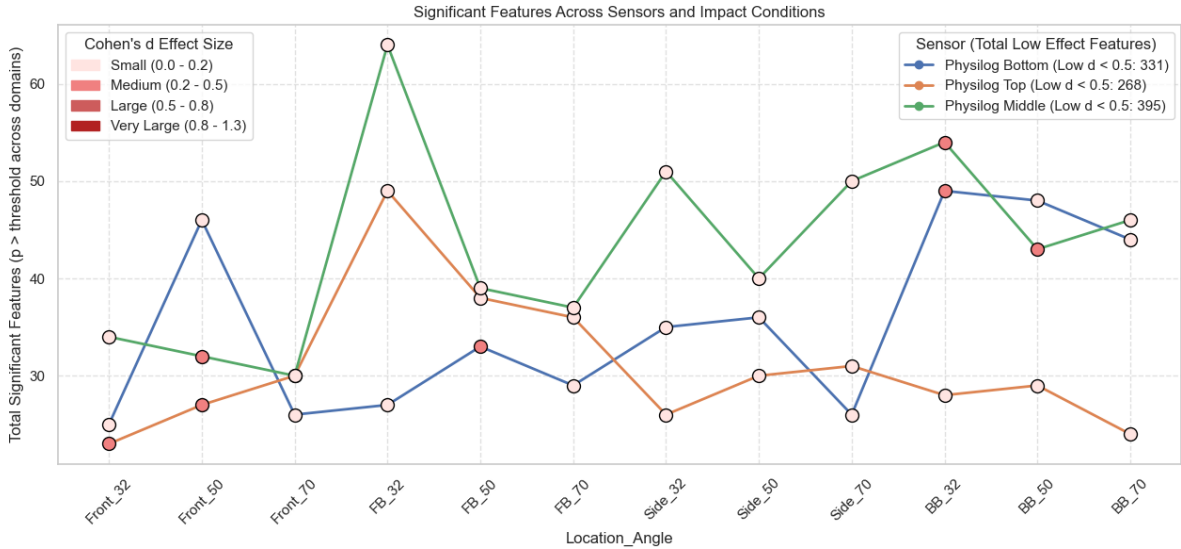


Figure 5.16: Final statistical synthesis: total number of features per impact condition satisfying both criteria ($p \geq 0.001$ and $d < 0.5$). The three curves represent the different helmet-mounted sensors, while the point colorization indicates Cohen’s d effect size.

5.4 Machine Learning Results

The Machine Learning analysis was performed to quantitatively assess the reliability of helmet-mounted sensors in capturing true head kinematics during impact events. Specifically, ML models were employed to classify each impact according to the level of helmet–head decoupling, providing an objective measure of sensor sensitivity to relative motion.

While traditional statistical analyses offered an initial insight into sensor performance—showing that the middle sensor tended to perform slightly better than the others—no single sensor

demonstrated a clear and consistent superiority across all measured features. Therefore, the application of supervised Machine Learning methods, such as **Random Forests (RF)** and **Support Vector Machines (SVMs)**, was particularly useful in complementing the statistical approach. By integrating multiple signal features simultaneously, these models enable a more nuanced evaluation of each sensor’s ability to detect high versus low decoupling events, highlighting subtle differences that might not be apparent from univariate statistics alone.

In this context, model performance can be interpreted as a measure of sensor sensitivity to helmet–head decoupling:

- **Higher model performance** → Indicates that the sensor captures motion more independently from the head, suggesting higher decoupling sensitivity.
- **Lower model performance** → Suggests reduced sensitivity to decoupling, implying that the sensor better reflects the true motion of the head.

Machine Learning models were used to complement the statistical analysis and provide an integrated view of sensor sensitivity to helmet–head decoupling. In this framework, higher classification performance indicates stronger sensitivity to decoupling (i.e., the sensor captures motion that deviates from true head dynamics), whereas lower performance suggests closer alignment with the Headform reference.

The results (Table 5.1) show that:

- The **Physilog Bottom** achieves the highest performance with Random Forests, while the **Physilog Top** is favored by SVMs, both reflecting higher susceptibility to decoupling.
- The **Physilog Middle** consistently exhibits the lowest performance across models, indicating reduced sensitivity to decoupling and a closer approximation to the Headform signal. However, its advantage remains modest and not uniform across all metrics.

Table 5.1: Accuracy and F1 Score for RF and SVM across datasets

Dataset	Model	Test Accuracy	Test F1 Score
No Headform	RF	0.8493	0.8139
	SVM	0.7778	0.7895
Physilog Bottom	RF	0.8116	0.8171
	SVM	0.7917	0.8052
Physilog Top	RF	0.7945	0.8104
	SVM	0.8056	0.8205
Physilog Middle	RF	0.7887	0.8024
	SVM	0.7778	0.7895

In line with the findings from the statistical analysis, these results confirm that no single helmet-mounted IMU can be considered unequivocally superior. While the Physilog Middle demonstrates a relatively more favorable behavior by capturing motion dynamics more faithfully, its improvement over the other sensors is not decisive. This highlights the

intrinsic complexity of accurately replicating Headform dynamics through helmet-mounted sensors, as their performance is inevitably influenced by both sensor location and the mechanical interaction between helmet and head during impact.

5.5 Limitations

When interpreting the results presented in this work, it is essential to discuss the main limitations that characterize the study. These concern both methodological aspects and technological or material constraints, and they help to define more precisely the reliability and scope of the conclusions drawn.

An important limitation concerns the experimental setup and the headform used for testing [34]. While the laboratory environment allows for controlled and repeatable measurements, the headform itself cannot fully capture the diversity of human heads. It has a simplified anatomy and its facial characteristics are basic. Moreover, the material covering the headform can influence how the helmet interacts with the head, for example by altering friction properties, which may differ from real-world conditions. These factors mean that the dynamics of helmet-head interaction observed in the lab may not entirely reflect those experienced by real users. Future studies could address this limitation by testing multiple headform sizes, adding layers such as skull caps or wigs to simulate different levels of helmet fit, and exploring alternative materials to better represent human anatomy.

Another limitation relates to the range of impact conditions tested. In this study, impacts were applied at three discrete kinetic energy levels. Although this allows for comparisons between lower, medium, and higher severity impacts, it does not cover the full spectrum of impacts that can occur in practice. Future work could expand the range of impact energies and orientations to better capture the variability of real-world scenarios.

The helmet itself introduces structural limitations. A first constraint lies in the fact that only one model, produced by Bauer, was used in this study; testing multiple hockey helmets could have provided a broader basis for comparison and potentially different outcomes. Beyond this, the shell of the helmet is not uniform, featuring softer and stiffer regions that differently affect shock propagation. Local deformations and vibrations generated by an impact can influence sensors even when they are positioned far from the contact point. Additionally, while sensors were placed in multiple positions within the helmet, their placement was not strictly standardized relative to anatomical landmarks, introducing further variability. At the same time, this flexibility may be seen as an opportunity for designers, allowing optimization of sensor placement for both protective and monitoring purposes.

Finally, another limitation concerns the labeling process used for training machine learning algorithms. At present, labels are based on the final position of the helmet at the end of the impact. In some cases, however, the helmet undergoes substantial displacement during the impact dynamics—with a high degree of decoupling from the head—before

returning to its initial position. In such situations, the event is still classified as “low,” despite the fact that significant movement occurred in the intermediate phase. This labeling strategy introduces a potential bias in the dataset, as it neglects the temporal evolution of the impact and instead focuses solely on the static final state. Consequently, the dataset provides only a partial representation of impact dynamics, which may affect the performance of predictive algorithms.

In conclusion, these limitations do not diminish the value of the findings obtained, but they help delineate their scope of validity. At the same time, they indicate promising directions for future research: the adoption of sensor technologies specifically designed for high-dynamic impacts, the development of labeling strategies that account for the full temporal evolution of impact events, and a more systematic consideration of material heterogeneity in helmets. Conversely, the observation that there is no strict constraint in sensor placement represents a concrete opportunity for designers, who may balance engineering requirements with ergonomic considerations in the development of intelligent helmets for both protection and monitoring purposes.

Chapter 6

Summary and Conclusions

This study investigated the reliability of helmet-mounted IMUs in capturing true head kinematics during impact events, focusing on the influence of sensor placement and impact direction. Two research questions guided the work: evaluating which IMU provides the most accurate signals compared to the Headform reference, and assessing whether impact direction affects measurement fidelity. Associated hypotheses anticipated that structurally isolated sensors would better reflect head motion (H1), and that sensors closer to the impact would show greater deviations (H2).

This study provides evidence that helmet-mounted IMUs are capable of capturing head kinematics with a level of reliability sufficient for practical use. Several important findings emerged from the analyses.

First, regarding sensor placement, the Middle sensor—positioned in a structurally isolated region—consistently showed the strongest agreement with the Headform. While differences compared to the Top and Bottom sensors were not large enough to indicate a uniquely optimal location, this result demonstrates that accurate measurements can be achieved from multiple positions, offering valuable flexibility in helmet design. Importantly, the observation that no sensor placement systematically outperformed the others leaves room for manufacturers to prioritize ergonomics and player comfort without compromising data fidelity.

Second, the results did not reveal a systematic influence of impact orientation on measurement accuracy. Rather than undermining the study, this finding is informative: it shows that helmet-mounted sensors maintain stable performance across a range of impact directions, underscoring their robustness in realistic scenarios where impacts occur with varying orientations.

Third, machine learning analyses highlighted differences in sensitivity to helmet–head decoupling. While the Top and Bottom sensors were more affected by relative motion, the Middle sensor consistently demonstrated reduced sensitivity and remained closely aligned with the reference. This finding not only supports the importance of sensor placement but also points to the Middle region as a promising location for applications requiring high-fidelity data.

Finally, the study confirmed that shock propagation within the helmet is not localized but reflects the integrated mechanical behavior of the shell, padding, and material properties. Recognizing this complexity opens the door to more accurate sensor data interpretation and improved helmet design. A summary of the main findings for this study, with the corresponding research questions and hypotheses can be found in the Table 6.1.

Table 6.1: Summary of Research Questions (RQ), Hypotheses (H), and Results

Research Question	Hypothesis	Results and Interpretation
RQ1: Which of the helmet-mounted IMUs exhibits signal patterns most similar to those recorded by the internal reference IMU?	H1: IMUs placed in structurally isolated regions of the helmet are expected to better reflect actual head motion.	• Middle sensor, located in a structurally isolated region, showed relatively better agreement with the Headform. • Differences with Top and Bottom sensors were not decisive.
RQ2: Does the direction of impact affect the accuracy of the measurements recorded by the different IMUs?	H2: IMUs closer to the point of impact are expected to show greater deviations from the ground truth.	• No consistent effect of impact direction was observed. • Proximity to the impact point did not systematically worsen sensor performance. • ML confirmed that the Middle sensor is the least sensitive to decoupling, though not decisively superior.

Appendix A

Impact Synchronization Analysis

In this appendix, a dedicated analysis is presented to investigate the synchronization of impact events recorded by the Physilog sensors and the Headform reference system. The purpose of this study was to better understand whether the observed discrepancies between the signals of the helmet-mounted IMUs and the Headform were due to systematic errors in the acquisition procedure or to the physical dynamics of the helmet relative to the headform during an impact.

A.1 Motivation and Objectives

During the main data processing pipeline, impact onset was defined as the instant at which the signal exceeded 10% of the baseline value around the identified peak. While this criterion allowed a consistent segmentation of impacts, it did not guarantee perfect synchronization across all sensors. Minor misalignments were observed, particularly when comparing the Headform signals to those acquired by the Physilog sensors.

To address this limitation, a manual synchronization procedure was carried out. The objective of this analysis was twofold:

- To determine whether the differences observed between the Headform and Physilog signals originate from systematic acquisition errors.
- To evaluate whether these differences are instead the result of the relative motion of the helmet with respect to the Headform during the initial phase of impact.

A.2 Methodology

A.2.1 Synchronization among Physilog Sensors

Synchronization among the three Physilog sensors (Top, Middle, and Bottom) was performed using two manual shaking events:

- **Initial Shake:** executed before the first trigger event.

- **Final Shake:** executed after the second trigger event.

During both events, the three sensors (not mounted on the helmet) were shaken simultaneously (Figure A.1). This procedure generated a common oscillatory pattern clearly visible across all sensors, which was used as a reference signal for temporal alignment.

The first prominent peak of the oscillatory signal was identified for each sensor. The signals from the Top and Bottom sensors were then aligned with respect to the Middle sensor, chosen as the reference. The procedure followed two cases:

- If a sensor's first peak occurred earlier than that of the Middle sensor, the signal was padded with zeros at the beginning.
- If a sensor's first peak occurred later, the initial part of the signal—containing only noise and no relevant dynamics—was trimmed.

This ensured that all three Physilog signals shared a common temporal reference based on the onset of the shaking event.

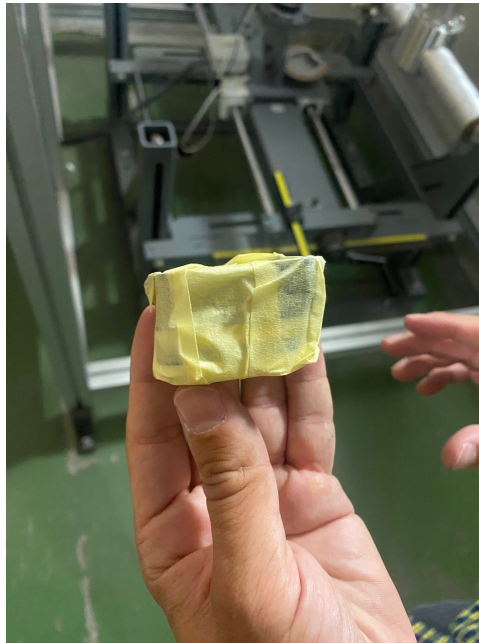


Figure A.1: Positioning of the Physilog sensors during shaking for synchronization.

A.3 Synchronization between Physilog Middle and Headform

The synchronization between the Physilog Middle sensor and the Headform sensor required a more complex procedure, due to their different sampling frequencies (512 Hz for Physilog, 10,000 Hz for Headform). Controlled frontal impacts, hereafter referred to as *trigger events*, were used as reference points. The procedure was structured as follows:

1. After the initial shaking event, the Middle sensor was temporarily mounted on the dummy's face to avoid attenuation effects caused by the helmet (Figure A.2).

2. A first frontal impact at 50° (*Initial Trigger*) was delivered to the dummy's head. This produced a sharp and distinct peak in both the Headform and Middle sensor signals, which served as the primary synchronization reference.
3. The Middle sensor was then mounted on the helmet, together with the Top and Bottom sensors.
4. Eight frontal impacts were recorded, generated by a pendulum impactor inclined at 50° .
5. Following this impact sequence, all sensors were removed, and the Middle sensor was repositioned on the dummy's face.
6. A second frontal impact at 50° (*Final Trigger*) was applied, producing another distinct synchronization reference point.
7. A final shaking event was executed, providing an additional validation reference for the three Physilog sensors.

This structured approach ensured that both trigger events (initial and final) were identical and could be reliably used to synchronize the signals from the two acquisition systems, despite the differences in sampling rates.

A.4 Data Processing and Alignment

A.4.1 Trigger Detection

Trigger events were identified by detecting prominent peaks in the norm of the angular velocity signals acquired by the Middle and Headform sensors. To facilitate peak detection in the Headform signal, a 200 Hz low-pass filter was applied, reducing high-frequency noise while preserving the impact dynamics.

Since the Headform recordings had shorter duration compared to the Physilog signals, they were left-padded with zeros to match the time base of the Middle sensor.

A.4.2 Drift Correction and Interpolation

The first trigger peak showed good alignment across both signals (Figure A.3 a)), while a slight temporal drift was observed at the second trigger (Figure A.3 b)). This drift was attributed to desynchronization between the acquisition systems.

To correct for this misalignment, the Middle sensor signal was linearly interpolated onto the time base of the Headform signal. The interpolation was restricted to the segment bounded by:

- the first trigger peak, and



Figure A.2: Positioning of the Physilog Middle sensor on the dummy's face during trigger events.

- the first local minimum following the second trigger peak.

The selection of this minimum was motivated by its stability and reproducibility, ensuring improved correspondence in both the temporal dynamics and the overall shape of the signals.

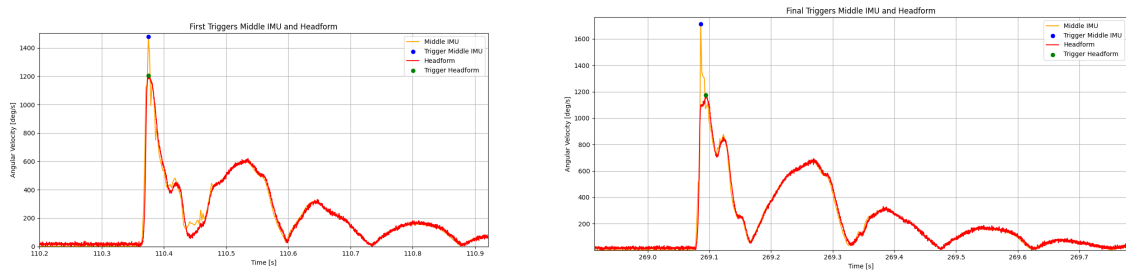


Figure A.3: a) Alignment at the first trigger event; b) Drift observed at the second trigger event.

A.4.3 Impact Segmentation

Each trigger event (initial and final) and each impact detected during the acquisition was segmented by extracting a time window of ± 0.5 s around the event, ensuring that the complete dynamics of the impact were captured.

Impact peaks were then identified using the following parameters:

- Minimum peak height: 1000
- Minimum distance between peaks: 0.5 s

This processing pipeline ensured that all events were consistently and robustly identified for subsequent analysis.

A.5 Results and Discussions

From the synchronized data, as Figure A.4 shows, the following observations were made:

1. In the first milliseconds of impact, the helmet-mounted IMUs exhibited strong variability and deviated from the Headform signal, particularly during the explosive onset of the impact.
2. After this initial phase, the Physilog signals converged to a behavior more similar to each other and to the Headform, although consistently delayed in time with respect to the Headform.
3. A systematic order of activation was observed across all impacts: the Top sensor detected motion first, followed by the Middle sensor, and finally the Bottom sensor. This sequence was reproducible across trials and may be explained by the kinematic response of the helmet when struck, where the upper part begins to move before the lower parts.

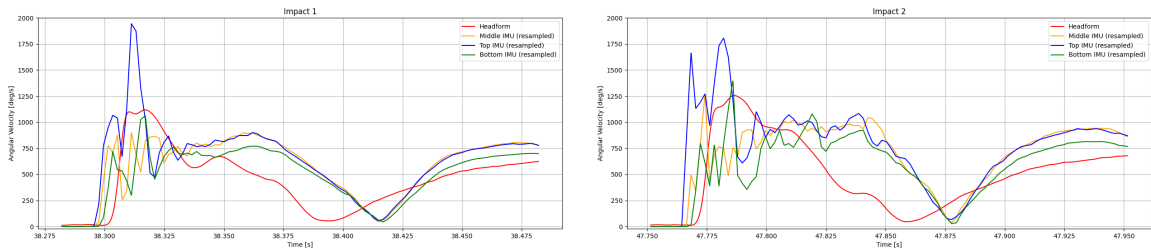


Figure A.4: a) First impact event; b) Second impact event.

The manual synchronization study suggests that the differences between Headform and Physilog signals are not simply artifacts of acquisition or systematic synchronization errors. Rather, they likely reflect the relative motion of the helmet with respect to the Headform during impact. In particular:

- The divergence observed in the early milliseconds may correspond to the helmet's independent motion immediately after the impact.
- The delayed convergence of the Physilog signals towards the Headform response indicates that the helmet subsequently couples more tightly with the dummy's head, reproducing its dynamics with a temporal shift.

A.6 Conclusion

This additional synchronization analysis provides important insights into the physical interaction between helmet and headform during impacts. By going beyond the automated 10% threshold method and introducing a manual alignment procedure, it was possible to demonstrate that the discrepancies observed are consistent with the relative movement of the helmet and not with systematic errors in data acquisition. Furthermore, the repeatable activation order of the sensors (Top–Middle–Bottom) suggests a characteristic sequence in the helmet's kinematic response to frontal impacts, which could inform future studies on helmet dynamics and sensor placement.

Appendix B

Principal Component Analysis for Impact Direction Verification

In this appendix, is reported the investigation conducted to assess whether Principal Component Analysis (PCA) could be effectively employed to identify the primary direction of impacts in our dataset, using data collected from multiple sensors. While PCA was ultimately not the main approach for impact classification due to potential error propagation, this analysis provides useful insights on sensor performance and impact direction estimation.

B.1 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique used to reduce the dimensionality of multivariate data while retaining the directions of maximum variance. In the context of impact analysis, PCA can be applied to accelerometer and gyroscope signals to determine the principal direction of the measured impact:

- **Accelerometers:** Applying PCA to the 3-axis accelerometer signals allows an initial estimate of the main direction of impact. However, accelerometer data are subject to noise due to external shocks, helmet movement, and vibrations, which can reduce the accuracy of this estimation.
- **Gyroscopes:** Gyroscope data provide reliable information on rotational movements. PCA applied to gyroscope signals can identify the principal axis of rotation associated with the impact, offering a more robust reference for verifying the impact direction.

B.2 Methodology

B.2.1 Preprocessing and Impact Segmentation

The starting point of the analysis was the raw gyroscope signal acquired along the three axes (x, y, z), together with the corresponding norm. Consistently with the procedure adopted in the main analysis pipeline (see Chapter 3), the signal was segmented to isolate individual impact events. Segmentation was carried out by identifying the instant of impact onset, defined as the time at which the angular velocity exceeded a threshold relative to the baseline level. This onset time was used to synchronize the signals across different sensors and to extract a time window centered on the impact. For each segmented window, the three components of the angular velocity vector and the corresponding norm were retained for further analysis.

B.2.2 Principal Component Analysis for Impact Direction

To characterize the dominant direction of each impact, Principal Component Analysis (PCA) was applied to the segmented three-dimensional gyroscope data. The objective of the PCA was to reduce the dimensionality of the signal while preserving the direction of maximum variance, which corresponds to the primary orientation of motion during the impact. Specifically, the first principal component was extracted and interpreted as the *impact vector*, providing a compact representation of the main direction of angular velocity induced by the collision. This vector was then expressed in terms of its x , y , and z components, along with its norm.

B.2.3 Impact Vector and Reference Directions

The extracted impact vector was compared against a set of predefined reference directions, which were defined in Chapter 3, Section 4.1.3. These reference vectors correspond to the main impact orientations (e.g., Front, Side Left, Back), such as $[-1, 0, 0]$ for the Front direction. The comparison was performed by calculating the angle between the impact vector and each reference vector.

B.2.4 Angle Calculation

The angle θ between the measured impact vector $\mathbf{v}_1$ and a reference vector $\mathbf{v}_2$ was calculated using the normalized dot product:

$$\theta = \cos^{-1} \left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} \right). \quad (\text{B.1})$$

The resulting angle, expressed in degrees, quantifies the deviation of the measured impact direction from the corresponding reference direction.

B.2.5 Classification by Angular Ranges

The calculated angle was then compared to predefined angular ranges associated with each reference direction. Table B.1 reports the classification ranges used to assign each impact vector to a specific direction. The direction label was determined by the interval in which the computed angle fell.

Table B.1: Angular ranges for impact direction classification.

Direction	Angular Range (°)
Front	0–22.5, 337.5–360
Front Oblique Left	22.5–67.5
Side Left	67.5–112.5
Back Oblique Left	112.5–157.5
Back	157.5–202.5
Back Oblique Right	202.5–247.5
Side Right	247.5–292.5
Front Oblique Right	292.5–337.5

This classification framework allowed each segmented impact to be systematically mapped to one of the main directions, enabling consistent comparison across sensors and impact events.

B.3 Results and Limitations

Figure B.1 presents the classification matrices related to the classification of impact directions, obtained by applying PCA analysis to the signals acquired from the different sensors. The horizontal axis represents the *Predicted Labels*, i.e., the directions estimated by the algorithm from the input data, while the vertical axis indicates the *True Labels*, corresponding to the actual impact directions.

The matrix associated with the Headform, shown in red, exhibits a perfectly correct classification, with all values aligned along the main diagonal. The matrices in blue refer instead to the three Physilog sensors, where some discrepancies between true and predicted directions can be observed.

The main diagonal of each matrix corresponds to correctly classified impacts; high values along this diagonal are indicative of strong sensor accuracy. Deviations from the diagonal represent misclassifications, the severity of which can be categorized as follows:

- **Minor (adjacent) errors:** occur when the predicted direction differs only slightly from the true one, while still remaining consistent with the overall impact orientation. Examples include a true impact in the Left Front Oblique direction classified as Front, or a true Right Front Oblique impact classified as Right Side. Although such cases are technically misclassifications, they do not significantly compromise the

reliability of the estimation.

- **Severe (distant) errors:** occur when the predicted direction falls into a completely different angular sector compared to the actual one. A typical case would be a *Front* impact being classified as *Side* or *Back*. These errors are particularly critical, as they reflect a complete loss of spatial coherence in the classification.

When comparing the different sensors, it becomes evident that the Physilog Bottom demonstrates the best performance, showing the fewest misclassifications and the highest concentration of correct predictions along the diagonal. The other Physilog sensors exhibit more frequent errors, though in most cases these are minor misclassifications limited to adjacent directions.

In conclusion, this analysis confirms that sensor placement plays a crucial role in the quality of classification. In particular, the Bottom sensor proves to be the most reliable in estimating the direction of impacts, demonstrating strong robustness and consistency with respect to the ground truth. While PCA provided an initial estimate of impact direction,

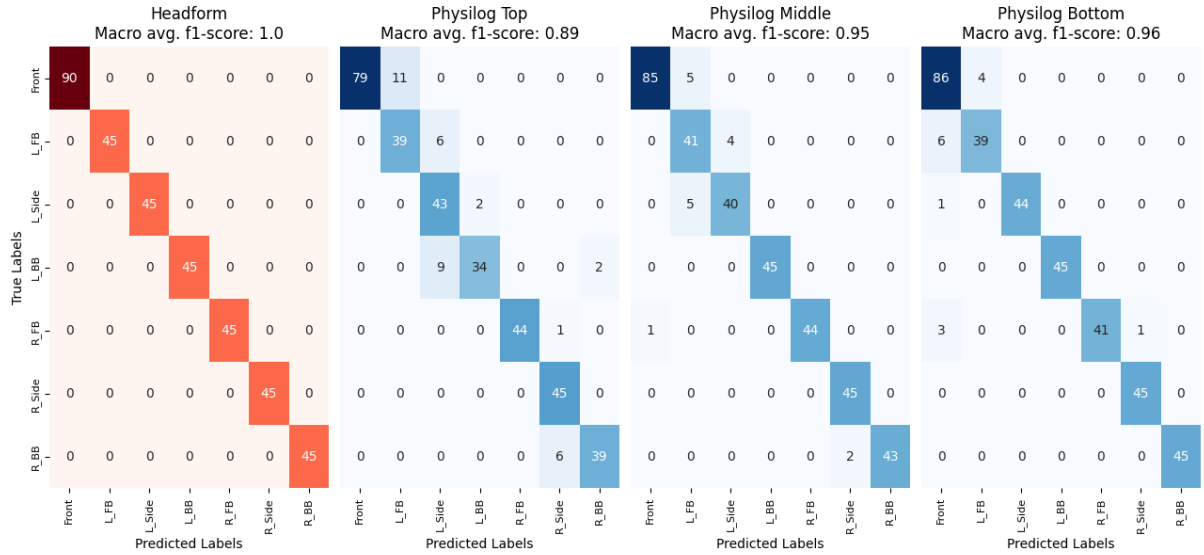


Figure B.1: Principal Component Analysis.

several limitations were observed:

- **Error Propagation:** Misclassification of even a few impacts could propagate through the dataset, leading to significant errors in overall direction estimation.
- **Accelerometer Noise:** Due to external disturbances, accelerometer-based PCA often yielded inconsistent directions.
- **Gyroscope Reliability:** PCA applied to gyroscope data offered more stable estimates of the principal rotation axis but still required careful alignment with accelerometer vectors to define the true impact direction.

Ultimately, relying solely on PCA for impact direction classification was not ideal. Instead, a more robust approach based on direct sensor measurements and angular thresholds was used for final classification.

The following table summarizes the macro-average F1-scores obtained by different sensors:

Table B.2: Macro-average F1-scores for impact classification using different sensors.

Sensor	Macro-average F1-score
CapSense	0.96
Headform	1.0
Physilog Top	0.89
Physilog Middle	0.95
Physilog Bottom	highest accuracy

These results confirm that sensor placement significantly influences classification accuracy. Specifically, the bottom-mounted Physilog sensor provides the most reliable data for determining impact direction.

B.4 Conclusion

PCA was investigated as a potential tool to simplify impact direction estimation. Although it provided valuable insights into the principal axes of acceleration and rotation, practical limitations—such as noise in accelerometer data and error propagation—prevented PCA from becoming the primary classification method. Nevertheless, the study highlighted the superior performance of the Physilog Bottom sensor achieved slightly better performance compared to the other sensors, thereby reinforcing the importance of sensor placement for accurate impact detection.

Bibliography

- [1] Stuart MJ Anderson GR Melugin HP. «Epidemiology of Injuries in Ice Hockey». In: *Sports Health* (Nov. 2019), pp. 514–519 (cit. on p. 1).
- [2] UK Research and Innovation (UKRI). *Concussion in Sport Report*. Sept. 2024. URL: <https://www.ukri.org/publications/concussion-in-sport/> (cit. on p. 1).
- [3] Shin S Broglio SP Sosnoff JJ. «Head impacts during high school football: a biomechanical assessment». In: *Journal of Athletic Train* (2009), pp. 342–9 (cit. on pp. 1, 23).
- [4] Mattson J Camarillo DB Shull PB. «An instrumented mouthguard for measuring linear and angular head impact kinematics in American football». In: *Ann Biomed Eng* (2013). DOI: 10.1007/s10439-013-0801-y. (cit. on p. 1).
- [5] Yuzhe Liu Xianghao Zhan Yiheng Li. «Machine-learning-based head impact subtyping based on the spectral densities of the measurable head kinematics». In: *Journal of Sport and Health Science* (2023), pp. 619–629. URL: <https://www.sciencedirect.com/science/article/pii/S2095254623000236> (cit. on pp. 3, 26, 28, 36).
- [6] Giuseppe Musumeci, Silvia Ravalli, Angela Maria Amorini, and Giuseppe Lazzarino. «Concussion in Sports». In: *Journal of Functional Morphology and Kinesiology* (2019), pp. 37–. URL: <https://www.mdpi.com/2411-5142/4/2/37> (cit. on p. 3).
- [7] Advanced Navigation. *Inertial Measurement Unit (IMU): An Introduction*. 2023. URL: <https://www.advancednavigation.com/tech-articles/inertial-measurement-unit-imu-an-introduction/#h-the-technology-behind-imu-sensors> (cit. on pp. 4–6).
- [8] A. Strandberg. «Head impact detection with sensor fusion and machine learning». MA thesis. Luleå University of Technology, 2022 (cit. on pp. 4, 36, 37).
- [9] Zakriya Mohammed, Ibrahim (Abe) M. Elfadel, and Mahmoud Rasras. «Monolithic Multi Degree of Freedom (MDoF) Capacitive MEMS Accelerometers». In: *Micro-machines* (2018). URL: <https://www.mdpi.com/2072-666X/9/11/602> (cit. on p. 5).
- [10] IdeeGreen.it. *Effetto Coriolis: definizione e spiegazione*. 2018. URL: <https://www.ideegreen.it/effetto-coriolis-107756.html> (cit. on p. 6).

- [11] Marcos Rodríguez Millán, T. Ito, and J. Olmedo Loya. «Development of numerical model for ballistic resistance evaluation of combat helmet and experimental validation». In: *Materials and Design* (2016). DOI: 10.1016/j.matdes.2016.08.015 (cit. on p. 7).
- [12] J.S. Giudice, G. Park, K. Kong, A. Bailey, R. Kent, and M.B. Panzer. «Development of Open-Source Dummy and Impactor Models for the Assessment of American Football Helmet Finite Element Models». In: *Annals of Biomedical Engineering* (2019). DOI: 10.1007/s10439-018-02155-3 (cit. on pp. 7, 11).
- [13] Steven B. Rowson and Stefan M. Duma. «A Review of On-Field Investigations into the Biomechanics of Concussion in Football and Translation to Head Injury Mitigation Strategies». In: *Annals of Biomedical Engineering* (2020). DOI: 10.1007/s10439-020-02684-w. URL: <https://doi.org/10.1007/s10439-020-02684-w> (cit. on pp. 7, 8).
- [14] Bolte JH 4th Allison MA Kang YS. «Validation of a helmet-based system to measure head impact biomechanics in ice hockey». In: *Medicine, Science in Sports and Exercise* (2014). DOI: 10.1249/MSS.0b013e3182a32d0d (cit. on pp. 8, 9).
- [15] Maltese MR Allison MA Kang YS. «Measurement of Hybrid III Head Impact Kinematics Using an Accelerometer and Gyroscope System in Ice Hockey Helmets». In: *Annals of Biomedical Engineering* (2015). DOI: 10.1007/s10439-014-1197-z (cit. on p. 9).
- [16] Loza J Wu LC Kuo C. «Detection of American Football Head Impacts Using Biomechanical Features and Support Vector Machine Classification». In: *Scientific Reports* (2017). DOI: 10.1038/s41598-017-17864-3 (cit. on pp. 9, 22, 23, 27).
- [17] rett D. Stemper, A.S. Shah, and Pintar. «Head Rotational Acceleration Characteristics Influence Behavioral and Diffusion Tensor Imaging Outcomes Following Concussion». In: *Annals of Biomedical Engineering* (2015). URL: <https://doi.org/10.1007/s10439-014-1171-9> (cit. on p. 21).
- [18] B. Hoshizaki, A. Post, M. Kendall, and C. Kartn. «The Relationship between Head Impact Characteristics and Brain Trauma.» In: *Journal of Neurology and Neurophysiology* (2013). DOI: 10.4172/2155-9562.1000181 (cit. on p. 21).
- [19] Takhounts EG, Craig MJ, Moorhouse K, and McFadden J. «Development of brain injury criteria (BrIC)». In: *Stapp Car Crash J* (2013). DOI: 10.4271/2013-22-0010 (cit. on p. 22).
- [20] Xianghao Zhan, Yiheng Li, and Yuzhe Liu. «The relationship between brain injury criteria and brain strain across different types of head impacts can be different». In: *Journal of Royal Society* (2001). URL: <https://doi.org/10.1098/rsif.2021.0260> (cit. on pp. 22, 23).

- [21] Milvus AI Quick Reference. *What is the role of frequency domain analysis in time series?* 2025. URL: <https://milvus.io/ai-quick-reference/what-is-the-role-of-frequency-domain-analysis-in-time-series> (cit. on p. 24).
- [22] NumPy Developers. *NumPy: Fast Fourier Transform (numpy.fft)*. Available at: <https://numpy.org/doc/stable/reference/routines.fft.html>. 2024 (cit. on p. 24).
- [23] Richard N. Youngworth, Benjamin B. Gallagher, and Brian L. Stamper. «An overview of power spectral density (PSD) calculations». In: *SPIE* (2005). DOI: 10.1117/12.618478 (cit. on p. 24).
- [24] Edutinker. *Full Width at Half Maximum (FWHM)*. URL: <https://edutinker.com/glossary/full-width-at-half-maximum-fwhm/> (cit. on p. 25).
- [25] Luisa Zanolla and Maria Graziani. «Glossary for the reader of a scientific paper. Part II: Inferential statistics». In: *Biochimica Clinica* (2014) (cit. on p. 29).
- [26] Kent State University. *SPSS Tutorials: Paired Samples t Test*. URL: <https://libguides.library.kent.edu/spss/pairedsamplesttest> (cit. on p. 29).
- [27] Park JH Kim TK. «More about the basic assumptions of t-test: normality and sample size». In: *Korean J Anesthesiol* (2019). DOI: 10.4097/kja.d.18.00292 (cit. on pp. 29, 30).
- [28] Kim JH. Kwak SG. «Central limit theorem: the cornerstone of modern statistics.» In: *Korean J Anesthesiol* (2017). DOI: 10.4097/kjae.2017.70.2.144 (cit. on p. 30).
- [29] Schulz P. Curtin F. «Multiple correlations and Bonferroni’s correction.» In: *Biol Psychiatry* (1998). DOI: 10.1016/s0006-3223(98)00043-2 (cit. on p. 31).
- [30] Feinn R. Sullivan GM. «Using Effect Size-or Why the P Value Is Not Enough.» In: *J Grad Med Educ.* (2012). DOI: 10.4300/JGME-D-12-00156.1. (cit. on p. 32).
- [31] Esraa Najjar and Aqeel Breesam. «Supervised Machine Learning a Brief Survey of Approaches». In: *Al-Iraqia Journal of Scientific Engineering Research* (2023). DOI: 10.58564/IJSER.2.4.2023.121 (cit. on pp. 34–37).
- [32] MathWorks. *What Is a Support Vector Machine (SVM)?* URL: <https://ch.mathworks.com/discovery/support-vector-machine.html> (cit. on p. 37).
- [33] Scikit-learn. *Support Vector Machines*. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html> (cit. on pp. 37, 38).
- [34] Ionescu A. Sciacca D. «Helmet–Head Decoupling in Ice Hockey Impacts: An In-lab Exploratory Study Using Autoregressive Modeling of Acceleration Data Measured from a Helmet-Mounted Inertial Measurement Unit (IMU).» In: *Ann Biomed Eng* (2025). DOI: <https://doi.org/10.1007/s10439-025-03848-2> (cit. on p. 57).