

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Artificial Intelligence for recognising school and training sports activities from sensorized vests

Eduardo Leite de Antunes Gomes



Mestrado em Engenharia Eletrotécnica e de Computadores

FEUP Supervisor: Prof. Miguel Fernando Paiva Velhote Correia

July 10, 2025

Artificial Intelligence for recognising school and training sports activities from sensorized vests

Eduardo Leite de Antunes Gomes

Mestrado em Engenharia Eletrotécnica e de Computadores

Faculdade de Engenharia da Universidade do Porto

Approved in oral examination by the committee:

President: Prof. Daniel Gouveia Costa

Referee: Prof. Miguel Fernando Paiva Velhote Correia

Referee: Prof. Cristina Manuela Peixoto dos Santos

July 10, 2025

Resumo

Num mundo onde a tecnologia desempenha um papel cada vez mais importante, é necessário utilizar sistemas inteligentes capazes de monitorizar e analisar a atividade física em tempo real nos contextos de treino e educação física. A integração crescente da tecnologia de sensores vestíveis em contextos desportivos e educativos proporciona novas oportunidades para a avaliação objetiva da atividade física. Esta dissertação, desenvolvida no âmbito do projeto TEXP@CT smart-textile, investiga a utilização de unidades de medição inercial (IMU) para o reconhecimento e classificação de movimentos desportivos fundamentais e modalidades desportivas. O trabalho visa fornecer uma ferramenta prática para monitorizar e caracterizar o movimento em ambientes como as aulas de educação física, onde as ferramentas tradicionais de análise de movimento não estão disponíveis ou são impraticáveis.

A abordagem metodológica centra-se no desenvolvimento e avaliação de modelos de aprendizagem automática para o reconhecimento de atividades. Isto é conseguido através da utilização de dados de conjuntos de dados disponíveis publicamente para emular condições reais de ensino e formação. Os sinais IMU foram submetidos a um processo de pré-processamento e foram extraídas as características relevantes. Vários algoritmos de classificação, nomeadamente Random Forest, LightGBM, Gradient Boosting, Support Vector Machine e Decision Tree, foram então comparados de forma sistemática. Para garantir estimativas de desempenho robustas e realistas, foi adotada uma validação cruzada exclusiva dos participantes, com atenção especial a desafios como o desequilíbrio de classes e a harmonização de sensores em diferentes contextos desportivos.

De todas as abordagens, o modelo LightGBM alcançou a maior precisão: 88,7% para a classificação de múltiplos movimentos e desportos. A classificação perfeita foi alcançada ao distinguir entre diferentes domínios desportivos. No entanto, o estudo também identificou limitações no reconhecimento de certos padrões de movimento sobrepostos, particularmente no futebol, devido à forma como as atividades são segmentadas e rotuladas nos conjuntos de dados disponíveis. Estas conclusões estão em conformidade com o estado da arte atual e confirmam a viabilidade da aplicação de fluxos de trabalho de reconhecimento práticos baseados em IMU em cenários educativos e de treino.

Em suma, esta dissertação demonstra que modelos de aprendizagem automática robustos e interpretáveis aplicados a dados de sensores vestíveis podem fornecer feedback preciso para monitorização da atividade física em ambientes do mundo real. Estes resultados abrem caminho para a integração destes métodos em têxteis inteligentes de última geração e apoiam o desenvolvimento de ferramentas quantitativas para o desporto e a educação física.

Palavras chave: reconhecimento de atividades humanas, classificação, monitorização desportiva, padrões de movimento, IMU, ML

Abstract

In a world where technology plays an increasingly important role, intelligent systems capable of monitoring and analysing physical activity are required for their use in physical education and training contexts. The integration of wearable sensor technology into sports and education offers new opportunities for objectively assessing physical activity. This dissertation, developed within the TEXP@CT smart textile project, investigates using inertial measurement units (IMU) to recognise and classify fundamental sports movements and modalities. This study aims to provide a practical tool for monitoring and characterising movement in settings like physical education classes where traditional motion analysis tools are unavailable or impractical.

The methodological approach focuses on developing and evaluating machine learning pipelines for activity recognition. This is achieved by using data from publicly available datasets to emulate real-world school and training conditions. IMU signals were pre-processed and relevant features extracted. Several classification algorithms, including Random Forest, LightGBM, Gradient Boosting, Support Vector Machine and Decision Tree, were then systematically benchmarked. Cross-validation with participant exclusivity was adopted to ensure robust and realistic performance estimates, with specific attention given to challenges such as class imbalance and sensor harmonisation across different sports contexts.

Of all the approaches, the LightGBM model achieved the highest accuracy: 88.7% for multi-movement and multi-sport classification tasks. Perfect classification was achieved when distinguishing between sports domains. However, the study also identified limitations in recognising certain overlapping movement patterns, particularly in football, due to the way activities are segmented and labelled in the available datasets. These findings are consistent with the current state of the art and confirm the feasibility of applying practical IMU-based recognition pipelines in educational and training scenarios.

In summary, this dissertation demonstrates that robust, interpretable machine learning models applied to wearable sensor data can provide accurate feedback for physical activity monitoring in real-world environments. These results pave the way for integrating these methods into next-generation smart textiles and support the further development of quantitative tools for sports and physical education.

Keywords: human activity recognition, classification, sports monitoring, movement patterns, IMU, ML

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework to achieve a better and more sustainable future for all. It includes 17 goals to address the world's most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice.

SDG 3 Ensure healthy lives and promote well-being for all at all ages

SDG 4 Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

Table 1: Sustainable Development Goals (SDG) 3 and 4.

SDG	Target	Contribution	Performance Indicators and Metrics
3	3.4	Implementation of wearable sensor vests with ECG and IMU improves activity monitoring and promoting health in sports and education.	Improved accuracy of physical activity monitoring.
4	4.7	Introduces data-driven tools for objective assessment in physical education, enhancing inclusivity and safety in sports activities.	Increase in adoption of wearable technology for education and improved evaluation metrics.

Acknowledgements

Over the past year, I have received invaluable support from many people and institutions that have accompanied me on this journey and made the development of this dissertation possible. First, I would like to thank my family for their constant availability and help at every stage of this project. Their support was essential in helping me overcome all the challenges I faced. I am also grateful to my friends for their companionship and motivation throughout the semester, and for providing the encouragement I needed inside and outside of academia.

Special thanks go to my supervisor, Professor Miguel Velhote Correia, for his dedication and guidance. I am especially grateful for the opportunity he gave me to participate in this project and for all his advice this semester. I would also like to acknowledge my co-supervisor at INESC TEC, Rafael Aguiar, for his tireless support. He was always available to answer my questions and help me overcome any difficulties that arose.

I would also like to thank FEUP and INESC TEC for providing the infrastructure and resources that were crucial to developing this project and for providing extra motivation through their stimulating academic and research environments. I am also grateful to everyone involved in the TEXP@CT project. Their collaborative effort and teamwork were essential to making this work possible.

This work is co-financed by Component 5 — Capitalization and Business Innovation — integrated into the Resilience Dimension of the Recovery and Resilience Plan, within the scope of the Recovery and Resilience Mechanism (RRM) of the European Union (EU), as part of the Next Generation EU initiative, for the period 2021-2026, as part of project TEXP@CT, with reference number 61.

I hope this dissertation serves as a starting point or source of inspiration for future projects in this field. I also hope it contributes to the continuous evolution of technology and research in sports and education.

Sincerely,

Eduardo Gomes

“Our greatest weakness lies in giving up. The most certain way to succeed is always to try just one more time”

Thomas Edison

Contents

1	Introduction	1
1.1	Context and Motivation	1
1.2	Objectives	2
1.3	Document Structure	2
2	Background and Fundamental aspects	4
2.1	Inertial Measurement Units	4
2.2	Electrocardiogram Sensors	6
2.3	Signal processing and Feature extraction	7
2.4	Machine Learning Models for Activity Recognition	7
3	State of the Art	11
3.1	ECG in Activity Recognition	12
3.1.1	ECG Data Incorporation with Activity Recognition Models	12
3.1.2	Challenges and Real-World Applications	13
3.2	IMU in Motion Analysis	13
3.2.1	Integration of IMU Data into Machine Learning Models	14
3.2.2	Challenges and Real-World Applications	15
3.3	Fusion of ECG and IMU data	16
3.3.1	Data Fusion Techniques	16
3.3.2	Applications and Benefits	17
3.3.3	Challenges and Limitations	17
3.4	Machine Learning Techniques	18
3.4.1	Machine Learning Algorithms	18
3.4.2	ML Implementations	19
3.4.3	Key comparison	20
3.4.4	Challenges and Future Directions	21
3.5	Comparison and Challenges	21
4	Methodologies	23
4.1	Description of Acquired Datasets	23
4.2	Hamza dataset	24
4.2.1	Data Acquisition Setup	24
4.2.2	Preprocessing and Windowing	25
4.2.3	Feature Extraction	25
4.3	Wilmes dataset	26
4.3.1	Participants and Experimental Protocol	26
4.3.2	Sensor Configuration and Placement	28

4.3.3	Data Preprocessing and Trial Selection	28
4.3.4	Feature Extraction	29
4.4	Machine Learning pipeline for Hamza Dataset	30
4.4.1	Dataset Preparation	30
4.4.2	Classifier Implementation	31
4.4.3	Cross-Validation Procedure	32
4.4.4	Evaluation Metrics	33
4.5	Machine Learning Pipeline for Wilmes Dataset	34
4.5.1	Dataset Acquisition and Label Verification	34
4.5.2	Feature Extraction from IMU Signals	35
4.5.3	Classifier Implementation with Intensity-Specific Movements	36
4.5.4	Classifier Implementation without Movements Intensity	39
4.6	Concatenation of Datasets and Classification Pipeline	41
4.6.1	Combined Dataset Methodology and Initial Pipeline Implementation	41
4.6.2	Class Balance Refinement and Signal Harmonisation	42
4.6.3	Intermediate Modelling and Dataset Expansion	43
4.6.4	Final Dataset Configuration and Classification Pipeline	44
5	Results and Discussion	48
5.1	Hamza Dataset	49
5.1.1	Random Forest Performance	49
5.1.2	Support Vector Machine Performance	50
5.1.3	Model Comparison	52
5.1.4	Feature Relevance and Sensor Insights	53
5.1.5	Critical Reflection	55
5.2	Wilmes Dataset	56
5.2.1	Random Forest Performance Analysis	57
5.2.2	Support Vector Machine Performance Analysis	58
5.2.3	Model Comparison	59
5.3	Combined Dataset – Initial Approach	60
5.3.1	Random Forest Performance	60
5.3.2	SVM Performance	62
5.3.3	Multi-Layer Perceptron Performance	62
5.3.4	Model Comparison and Discussion	63
5.4	Final Evaluation on the Refined Dataset	64
5.4.1	Decision Tree	65
5.4.2	SVM	66
5.4.3	Gradient Boosting	66
5.4.4	LightGBM	67
5.4.5	Random Forest	69
5.4.6	Model Comparison and Key Insights	69
5.4.7	Comparative Study	71
6	Conclusions	73
6.1	Main Findings and Achievements	73
6.2	Limitations and Future Work	75
	References	77

List of Figures

2.1	Prototype of the sensorized vest developed in the TEXP@CT project, highlighting the IMU and central sensor compartment.	5
2.2	ECG signal by Zhu [48]	6
2.3	Schematic representation of a Random Forest Model [18].	8
2.4	Illustration of LightGBM boosting strategy [24].	8
2.5	Architecture of a Multilayer Perceptron with an input layer, hidden layer(s), and output layer [3].	9
3.1	Estimating Counter Movement Jump Phases with IMUs, with a gold-standard force plate to assist in jump event detections [28].	14
3.2	Four Phases of the ML Classification [26]	20
4.1	Back vest and IMU sensor [39]	25
4.2	Football specific movements [43]	27
4.3	Sensor placement [43]	28
4.4	Processing pipeline for the Hamza Dataset, detailing data preprocessing, feature extraction, ML training and validation steps.	31
4.5	Processing pipeline for Wilmes Dataset, detailing the steps from data collection and validation to preprocessing, feature extraction, ML training and evaluation steps.	35
5.1	Confusion Matrix – Random Forest (Hamza dataset)	51
5.2	Confusion Matrix – SVM (Hamza dataset)	53
5.3	Top 20 Feature Importances – Random Forest (Hamza dataset)	54
5.4	Confusion Matrix – Random Forest (Wilmes Dataset without Intensity Levels)	57
5.5	Confusion Matrix – SVM (Wilmes dataset without Intensity Levels)	58
5.6	Top 20 Feature Importances – Random Forest (Grouped Movements)	59
5.7	Confusion Matrix – Random Forest (Movement Classification)	61
5.8	Confusion Matrix – SVM (Movement Classification)	62
5.9	Confusion Matrix – MLP (Movement Classification)	63
5.10	Confusion Matrix – Decision Tree (Movement Classification)	65
5.11	Confusion Matrix – SVM (Movement Classification)	66
5.12	Confusion Matrix – Gradient Boosting (Movement Classification)	67
5.13	Confusion Matrix – LightGBM (Movement Classification)	68
5.14	Top 20 Feature Importances – LightGBM	68
5.15	Confusion Matrix – Random Forest (Movement Classification)	69

List of Tables

1	Sustainable Development Goals (SDG) 3 and 4.	iii
3.1	Inclusion and exclusion criteria used in the literature review.	11
3.2	IMU analysis studies	15
3.3	ECG and IMU data fusion in Activity Recognition	17
3.4	Machine Learning Techniques	21
4.1	Time-domain statistical features computed from IMU signals.	29
5.1	Performance metrics for Random Forest on Hamza Dataset	50
5.2	Performance metrics for SVM on Hamza dataset	52
5.3	Comparative Summary – Random Forest vs. SVM (Hamza Dataset)	52
5.4	Cross-Validation Metrics – Random Forest (Grouped Movements)	57
5.5	Cross-Validation Metrics – Support Vector Machine (Grouped Movements)	58
5.6	Movement Classification Performance – Combined dataset	64
5.7	Final Movement Classification Metrics	70
5.8	Comparison of activity recognition studies: sensor setup, activity type, and reported accuracy.	71

List of Acronyms

AD	Absolute Differences
AI	Artificial Intelligence
CNN	Convolutional Neural Networks
ECG	Electrocardiogram
HR	Heart Rate
HRV	Heart Rate Variability
IMU	Inertial Measurement Unit
KNN	K-Nearest Neighbors
LOSO	Leave-One-Subject-Out
LSTM	Long Short-Term Memory
MAV	Mean Absolute Value
ML	Machine Learning
MOKKA	Motion Kinematic and Kinetic Analyzer
PTP	Peak-to-Peak
RBF	Radial Basis Function
RF	Random Forest
RMS	Root Mean Square
RNN	Recurrent Neural Networks
SSI	Simple Square Integral
SVM	Support Vector Machines

Chapter 1

Introduction

The valid and objective quantification of physical activity remains a perennial issue in physical education and sports coaching. Until recently, empirical observation and basic timing mechanisms were the mainstay of teachers and coaches, yielding limited information about the nature, quality, or environment of movement. This lack of quantitative feedback is particularly limiting in educational settings, where data-driven, actionable information could improve engagement, safety, and performance.

The latest advances in wearable sensor technology, specifically the integration of inertial measurement units (IMU) in smart textiles, have enabled systems that can monitor and analyse human movement in real-world conditions. IMU-based systems provide detailed biomechanical data without constraining natural motion and offer a promising approach to movement characterisation and feedback that can benefit both athletes and physical education students.

This thesis is part of the TEXP@CT project [40], a national initiative developing a smart vest platform for movement analysis and activity monitoring in sports and school environments. While the main project involves multiple sensor types, this work focuses on developing and verifying machine learning models that can classify sports modalities and basic movement patterns using data from a single IMU located near the body's centre of mass. Utilisation of ECG signals is considered but not implemented in the operational models, as this falls outside the scope of this work.

1.1 Context and Motivation

Methods based on visual observation or timing cannot provide the detailed measurements necessary for characterising movement. Although laboratory systems with high accuracy and multi-sensor suites exist, they are too costly, complex, and obtrusive to be used in everyday school or community settings.

An inexpensive, accessible solution is a stylishly incorporated IMU-based approach in a wearable vest. This solution enables the ongoing monitoring of movement without obstructing typical activity or requiring particular infrastructure. Nevertheless, the key challenge is to distinguish not

only the overall sport modality but also fundamental movement patterns in real-world school and training settings.

This thesis overcomes this difficulty by demonstrating the feasibility of recognising activities from IMU data alone despite considerable variability. Although the vest features ECG electrodes for physiological measurements, this study only addresses IMU-based classification. The ECG's contribution is investigated by benchmarking its performance to demonstrate its potential in future multimodal solutions. However, it is not incorporated into the main machine learning pipeline of this thesis.

1.2 Objectives

The main objective of this dissertation is to design, implement, and evaluate machine learning models capable of classifying sport modalities and fundamental movement patterns using only IMU data. The focus is on recognising these actions in settings that mimic the heterogeneity and variability of real-world physical education and sports training environments.

To ensure the practical viability of the approach, the system should meet or exceed a baseline classification accuracy of 75%, a threshold informed by recent literature on human activity recognition using wearable sensors [21, 33]. The models are expected to operate effectively even under constraints such as differences in sensor placement, acquisition protocol, and class imbalance.

In this context, the work evaluates the feasibility of using a single IMU for the multimodal classification of sports and fundamental movement patterns. Additionally, the dissertation compares various model architectures based on classification accuracy, interpretability and generalisability under conditions that reflect real-world deployment scenarios.

Although the broader TEXP@CT solution includes additional sensing modalities such as ECG for cardiac monitoring, this dissertation focuses exclusively on IMU-based analysis. References to ECG are included solely to contextualise the system architecture envisioned by the overall project scope. All methods were selected based on insights from a systematic literature review, ensuring alignment with state-of-the-art techniques in IMU-based activity recognition.

1.3 Document Structure

The thesis provides an overview of sports activity monitoring, covering the theoretical and technical fundamentals, as well as the practical implementation and testing of machine learning algorithms for movement recognition. The thesis begins with a review of the key challenges associated with the quantitative assessment of sports activities. It focuses particularly on the emerging role of wearable sensor technologies and, more precisely, inertial measurement units (IMUs). Subsequent chapters systematically address the fundamental principles of sensor-based motion analysis, summarise relevant literature, and document data processing techniques, model formation, and model evaluation. The results are critically interpreted in the context of technical performance

and practical usefulness. The thesis concludes with a synthesis of the results and future research directions within the framework of the TEXP@CT intelligent textile project.

The document is organised into six chapters. Chapter one situates the dissertation within the broad area of physical education and sports, describing longstanding issues of objective measurement and activity classification. It establishes the motivation for the research, declares the specific aims, and introduces the TEXP@CT project as the technological and application context. Chapter two establishes the technical and conceptual underpinnings of the dissertation. It describes the principles and relevance of IMU and ECG sensors, as well as their roles in capturing biomechanical and physiological signals. Additionally, it covers the core signal processing techniques and machine learning algorithms that form the basis of activity recognition systems.

The next chapter offers a critical review of the latest research on activity recognition using wearable technology. The chapter reviews sensor fusion techniques, machine learning approaches, and deep learning approaches, summarising the major advances and limitations from the literature. Particular emphasis is given to practical issues such as sensor synchronisation, computational efficiency, and model transferability to uncontrolled conditions. Chapter four describes the dissertation's experimental layout. It details the selection and characterisation of publicly available datasets, sensor configurations, and acquisition protocols. The chapter also details the preprocessing and feature extraction procedures, as well as the model training, validation, and evaluation procedures. This chapter emphasises future compatibility with data streams from the TEXP@CT vest.

Chapter five presents the outcomes of the classification models and appraises their accuracy, robustness, and performance on datasets with varying sensor placements and acquisition conditions. The discussion contextualises the findings within the framework of ongoing work, highlighting their significance for practical applications in education and sports. The final chapter summarises the dissertation's main findings and contributions, identifies its limitations, and suggests directions for future research and development. Final remarks highlight the possible integration of IMU-based recognition algorithms into the TEXP@CT platform. These recommendations and findings are relevant not only in the context of extending the scope of this dissertation but also in guiding the ongoing development and application of the TEXP@CT project in future phases.

Chapter 2

Background and Fundamental aspects

2.1 Inertial Measurement Units

In recent years, physical performance monitoring has been of particular interest in sports science and engineering. This interest is driven by wearable sensors that help maximise training gains while minimising the risk of injury. Inertial measurement units (IMU) are notable for their ability to measure biomechanical parameters in real time, providing detailed information about movement patterns, exercise intensity, and overall physical condition. IMUs typically include accelerometers and gyroscopes, and sometimes magnetometers, enabling high-accuracy, three-dimensional tracking of body movement.

Accelerometers are the core of IMU systems, sensing linear acceleration in three orthogonal directions. They function by measuring capacitance variation resulting from a displaced mass suspended on a spring mechanism, providing high-precision, low-power measurements. This capability makes accelerometers valuable for quantifying movement intensity in sports, including the high accelerations that occur during sprinting or jumping [19]. However, accelerometer signals are marred by noise, drift, and gravity artefacts. Therefore, advanced filtering and calibration processes are necessary to produce valid analyses [36].

Gyroscopes complement accelerometers by providing angular velocity measurements that contain valuable information about rotational kinematics, such as changes in direction or the rotation of limbs. Based on the Coriolis effect experienced by a resonant mass, gyroscopes are sensitive to angular motion but, like accelerometers, require careful signal processing to eliminate drift and vibration artefacts [38].

Advances in wearable technology have resulted in increasingly compact, accurate, and user-friendly IMU systems. Wearable IMUs now allow for unobtrusive data collection under real-world conditions, both in team sports and outdoor settings [19]. Nevertheless, ensuring high-quality data requires effective methods for filtering motion artefacts and synchronising signals from multiple sensors. Preprocessing steps such as noise filtering and normalisation are critical to ensure the reliability and consistency of IMU-based measurements [21].

Recent developments in the field have focused on integrating IMUs directly into textile garments, enabling unobtrusive and ecologically valid data collection in both individual and team sports settings [19]. Within this context, the TEXP@CT project [40] was established to create a sensorized suit for sports applications. One of the technical aims of the TEXP@CT suit is to evaluate the feasibility of using a single IMU for the recognition of sports activities and for the classification of sport modalities, an approach that presents significant challenges due to the diversity and complexity of human movement in real-world athletic scenarios.

The TEXP@CT platform (Figure 2.1) incorporates a central IMU, which is positioned on the back of the vest, close to the user's centre of mass. This placement was chosen to avoid restricting natural movement and to ensure that collected kinematic data accurately reflect whole-body motion during physical activity. The vest design allows athletes to perform a wide range of movements without obstruction, making it suitable for both laboratory and real-world sports environments.



Figure 2.1: Prototype of the sensorized vest developed in the TEXP@CT project, highlighting the IMU and central sensor compartment.

This dissertation, developed at INESC TEC, involved the experimental validation of the IMU subsystem in the TEXP@CT vest, performing initial signal quality checks and technical assessments. For the present work, model development and validation relied exclusively on publicly available datasets, rather than on data collected from the TEXP@CT vest, as detailed in Chapter 4. Nevertheless, the entire methodological pipeline, including all processing, feature extraction, and validation steps, was purposefully designed to remain fully compatible with future IMU data acquired from the TEXP@CT platform.

2.2 Electrocardiogram Sensors

The heart rate is one of the most important physiological parameters for assessing the intensity of physical activity in any kind of physical exercise and monitoring the response of an athlete during physical training. Electrocardiogram (ECG) sensors are considered standard technology in heart rate (HR) measurements due to their ability to record the electrical activity of the heart with high accuracy. ECG sensors, by detecting electrical impulses in each cardiac cycle, can provide continuous and accurate heart rate measurements [26].

ECG signals can be processed in real-time, providing relevant physiological metrics during physical activities, such as heart rate variability [27]. These signals allow the system to infer whether an individual is within their optimal exercise intensity, providing insight into cardiovascular effort and recovery. The parameters obtained from the ECG signals allow the estimation of Heart Rate Variability (HRV), which is related to the state of cardiovascular fitness [46]. Therefore, such applications make ECG sensors valuable tools for monitoring physiological responses during exercise [48].

The ECG signal is characterized by a relatively low voltage, typically below 1 mV, with a frequency spectrum ranging from 0.05 to 150 Hz. This signal consists of five key points, P, Q, R, S, and T, which are used to determine the zones: P wave, QRS complex, and T wave, as illustrated in Figure 2.2. The amplitude and time intervals between these points provide valuable insights into cardiac function, enabling the detection of potential abnormalities [48].

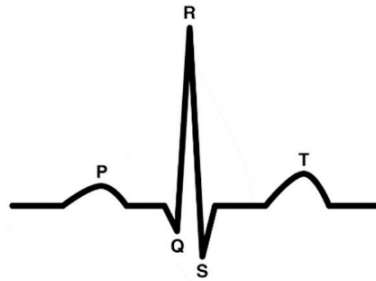


Figure 2.2: ECG signal by Zhu [48]

Recent technological developments have made the design of ECG sensors more adaptable, especially for wearables. Newer ECG devices are lightweight, non-invasive, and can be integrated into clothing, allowing subtle monitoring during physical activity [22]. This is important in training and education contexts where accurate and continuous feedback on physical performance is essential.

However, ECG sensors have a number of disadvantages, such as noise in the signal caused by movement and environmental electromagnetic interference. The ECG signal must undergo signal processing techniques, such as filtering and artefact removal, to eliminate motion artefacts and muscle signal interference that can contaminate the data and compromise its accuracy [21].

2.3 Signal processing and Feature extraction

Signal processing and feature extraction are the basis of raw sensor data conversion into information that can be used by an activity recognition system. These processes clean, segment, and transform the data from devices into variables that can be analysed by machine learning models.

First, noise and artefacts, such as baseline drift in ECG signals or sudden spikes in IMU data caused by motion artefacts, are removed by filtering. The resulting data is then normalised to ensure consistency and reliability in the subsequent analysis [1]. Time windowing is applied through segmentation, allowing machine learning algorithms to capture and analyse temporal patterns in the data. This step is considered crucial because the quality and relevance of the extracted features directly influence the performance of the machine learning models [4]. For example, ECG signals provide metrics that indicate cardiovascular stress and recovery states, such as heart rate variability, while IMU data provide measurements of acceleration, angular velocity, and motion dynamics. By isolating and highlighting these key patterns, feature extraction reduces the dimensionality of the input data and improves the model's ability to accurately classify activity and generate actionable insights. Poorly selected features result in a lack of generalisation in models and, therefore, may not be as accurate and reliable in real-world applications.

2.4 Machine Learning Models for Activity Recognition

The classification of human activities from wearable sensor data has leveraged a variety of machine learning models, ranging from established ensemble methods to deep learning approaches. In this dissertation, multiple models were evaluated to balance interpretability, robustness, and predictive performance in the context of IMU and ECG-based activity recognition.

Random Forest (RF) and Light Gradient Boosting Machine (LightGBM) are ensemble learning methods that have shown consistent performance in time-series classification tasks involving sensor data. Both are based on aggregating the outputs of multiple decision trees, yet they differ in their training strategies and interpretability. Figure 2.3 illustrates the architecture of a random forest.

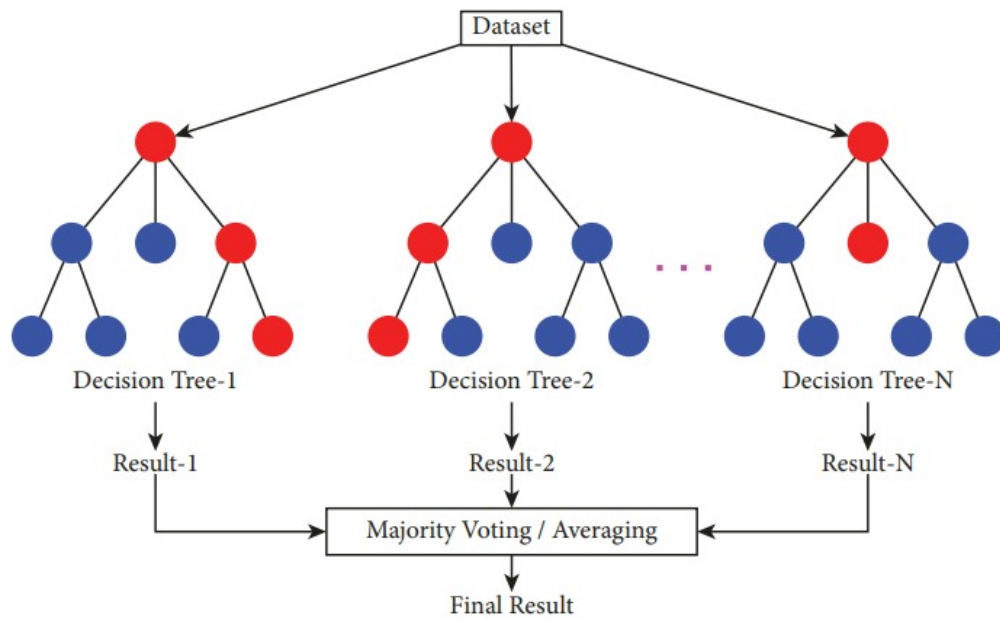


Figure 2.3: Schematic representation of a Random Forest Model [18].

Each decision tree is trained on a different subset of the data, and the overall prediction is made by aggregating the outputs of all trees. This approach is robust to noise and can provide insights into feature importance by analysing the contribution of each variable to the decision process, as demonstrated in Khan et al. [18].

LightGBM builds upon the principles of gradient boosting, where trees are constructed sequentially, each one aiming to correct the errors of its predecessor. The method employs a leaf-wise split strategy and histogram-based decision rules to achieve computational efficiency and scalability, especially for large and high-dimensional datasets. Figure 2.4 depicts the stepwise construction of boosted decision trees in LightGBM.

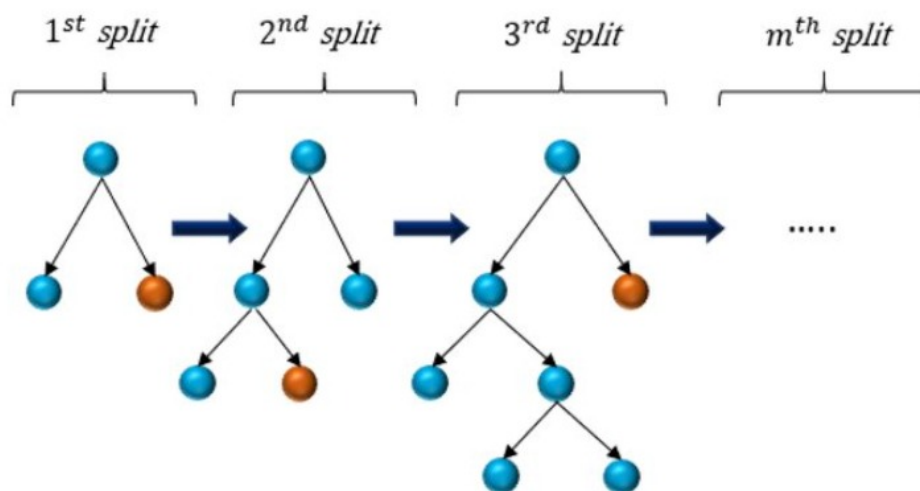


Figure 2.4: Illustration of LightGBM boosting strategy [24].

Each successive tree is trained to correct residual errors from the previous trees, enabling the model to fit complex patterns in the data. Although both RF and LightGBM are tree-based and ensemble in nature, LightGBM often achieves higher predictive accuracy at the expense of interpretability, as the boosting mechanism and split criteria are more complex and less transparent compared to bagged ensembles like random forests, as highlighted in Mahmoudzadeh et al. [24].

Support Vectors Machine (SVM) and other traditional machine learning algorithms were also considered for their established track record in structured sensor data classification. SVM constructs a decision boundary that maximises the margin between different classes, potentially in a transformed feature space if kernels are used. SVMs are computationally efficient for low and moderate-dimensional data and can serve as a strong baseline in activity recognition studies [44].

Neural networks, particularly the multilayer perceptron (MLP), were also explored to benchmark performance against more interpretable approaches. An MLP is a feedforward neural network comprising an input layer, one or more hidden layers, and an output layer, where each neuron in a layer is fully connected to those in the next. The weights in the network are learned automatically via backpropagation, allowing the model to capture complex, nonlinear relationships between features and class labels. Figure 2.5 shows the standard architecture of an MLP as used for classification tasks.

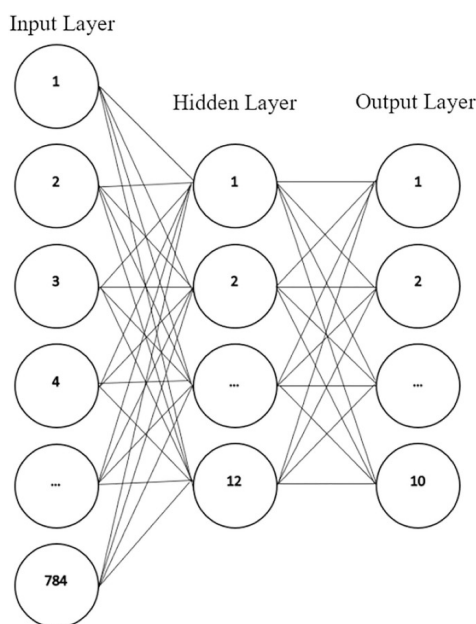


Figure 2.5: Architecture of a Multilayer Perceptron with an input layer, hidden layer(s), and output layer [3].

All weights and activations are learned automatically during training, making interpretation challenging. This limits the ability to extract clear insights into feature importance or model reasoning, in contrast to tree-based models, as discussed by Arabnia et al. [3]. While MLPs can achieve high accuracy, their interpretability is inherently limited, as the decision process is encoded in distributed and non-transparent weights rather than explicit decision rules.

Although deep learning architectures such as Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM) and hybrid CNN-LSTM models have demonstrated state-of-the-art performance in other activity recognition domains, their complexity and data requirements were not aligned with the scale and objectives of this work. As such, the focus remained on models that provide a clear balance between predictive performance and interpretability, namely RF, LightGBM, SVM, and MLP. The rationale for exploring MLP in this dissertation is to provide a direct comparison to traditional and ensemble models, with a detailed discussion of model performance and interpretability presented in Chapter 5.

Chapter 3

State of the Art

The following state-of-the-art chapter explores advancements in human physical activity detection using wearable devices, such as IMU and ECG sensors. In recent years, there has been a greater focus on integrating these sensors into smart monitoring systems to address issues encountered in sports training and physical education. These models help identify and correlate patterns with specific movements. This chapter critically compares the main studies, their methodologies, and their findings, as well as their contributions to the field.

To ensure a focused and relevant review, a targeted search strategy was employed to identify recent research on the use of inertial and cardiac biosignal data for classifying sport-related and school-based physical activities. The search covered publications from 2020 to 2024, with preference for studies involving machine learning models and wearable sensor data. The search string used was:

(classification OR "machine learning" OR recognition) AND ("ECG" OR "Inertial Measurement Unit" OR accelerometer OR IMU OR HR OR "heart rate") AND ("school activities" OR "exercise monitoring") AND NOT (Cancer OR Parkinson OR animals) AND NOT (review)

The following inclusion and exclusion criteria were applied to filter relevant articles for this review:

Table 3.1: Inclusion and exclusion criteria used in the literature review.

Inclusion Criteria	Exclusion Criteria
Publication date between 2020–2024 Full-text available in English Use of IMU and/or ECG/HR data Clear description of the classification method Active human movement (sports or physical education context)	Publication date outside 2020–2024 Articles not written in English Use of image, video, or unrelated sensor modalities Articles without a classification approach Studies involving passive movement, fall detection, clinical conditions (e.g., Parkinson, cancer), or animal models

Based on these criteria, a total of 55 studies were initially identified and screened by title and abstract. Following full-text assessment, 40 studies were retained for detailed review and discussion. These selected works are explored in the subsequent sections.

3.1 ECG in Activity Recognition

ECG sensors are currently valuable elements of activity recognition systems due to their precision in monitoring heart rate (HR) and heart rate variability (HRV). ECG sensors are valuable components of activity recognition systems because they precisely monitor heart rate (HR) and heart rate variability (HRV). These parameters provide valuable information about cardiovascular and physiological responses to physical activity. This section addresses the application of ECG sensors in activity recognition, how they are combined with advanced machine learning models, their limitations, and their contributions to the field.

ECG electrodes record cardiac electrical activity, enabling real-time monitoring of HR and HRV. These parameters are useful for estimating physical activity intensity and identifying fatigue-specific patterns. For example, Mekruksavanich, Jantawong, and Jitpattanakul [27] researched activity classification using ECG signals and demonstrated that deep learning models outperformed baseline approaches in activity recognition. The study also highlighted the unique contribution of ECG in determining activity state transitions.

HRV analysis played an essential role in assessing recovery states and cardiovascular health. With a complete picture of the physiological response enabled by ECG data, systems can determine if someone is exercising at their peak intensity and provide personalised feedback [48].

3.1.1 ECG Data Incorporation with Activity Recognition Models

ECG data can be incorporated into activity recognition models. Machine learning models have greatly improved the use of ECG data in activity recognition frameworks. Traditional methods, such as SVM and RF, struggle in complex and large datasets. More recent models, such as convolutional neural networks (CNN) and long short-term memory (LSTM) networks, are much better equipped to extract spatial and temporal features from ECG data [6, 9].

For example, Zhou and Wen [46] introduced a hybrid CNN-LSTM model that efficiently combines spatial and temporal features for activity classification from ECG signals with greater accuracy than single-model solutions.

HRV and corresponding ECG-based measures are important for tracking physical work and physiological adaptations to exercise. For instance, Fang [11] demonstrated how HRV-derived measures can differentiate between exercise intensity and recovery periods, offering a quantifiable basis for adaptive training. The ability to monitor physical exertion is especially important in sports and academic settings, where participants' health and optimal performance are paramount.

3.1.2 Challenges and Real-World Applications

The application of ECG sensors is transitioning from lab-controlled environments to the field, with applications ranging from team sports to education and rehabilitation. For instance, Klingenberg and Purucker [20] have managed to employ lightweight, wearable ECG systems that provide stable monitoring during dynamic, motion-associated physical activity. However, the real-time use of such systems is plagued by issues related to signal noise, data synchronisation, and computational throughput.

Various signal processing techniques have been proposed, including adaptive filtering and motion artefact cancellation. Since the ECG sensor is wearable and appropriate for wearable devices such as smart clothing, using ECG sensors for realistic training and monitoring is feasible. This has been demonstrated by textile-embedded, ECG-powered systems for ambient monitoring in sports and healthcare settings [19, 38].

Current developments in activity detection systems based on ECG demonstrate their important role in bridging the gap between raw physiological data and actionable intelligence. Furthermore, combining machine learning algorithms with ECG sensors represents a paradigm shift in physical performance measurement and optimisation.

3.2 IMU in Motion Analysis

IMU is a very important motion analysis tool as it can provide dense kinematic data on acceleration, angular velocity and orientation. Using tri-axial IMU data to recognise human activities enhances classification accuracy, particularly for dynamic sporting activities, as the multidimensional data can register subtle kinematic variations that are not detectable by one-axis measurement [7]. This section explains how IMUs are used in motion analysis, how IMU readings are incorporated into machine learning models, the challenges involved in using them and their contributions.

IMUs use accelerometers and gyroscopes with magnetometers to provide information on linear acceleration, angular velocity and orientation. These sensor readings provide more detailed information about the systems, enabling them to be aware of motion dynamics and track intricate activities, as illustrated in Figure 3.1.

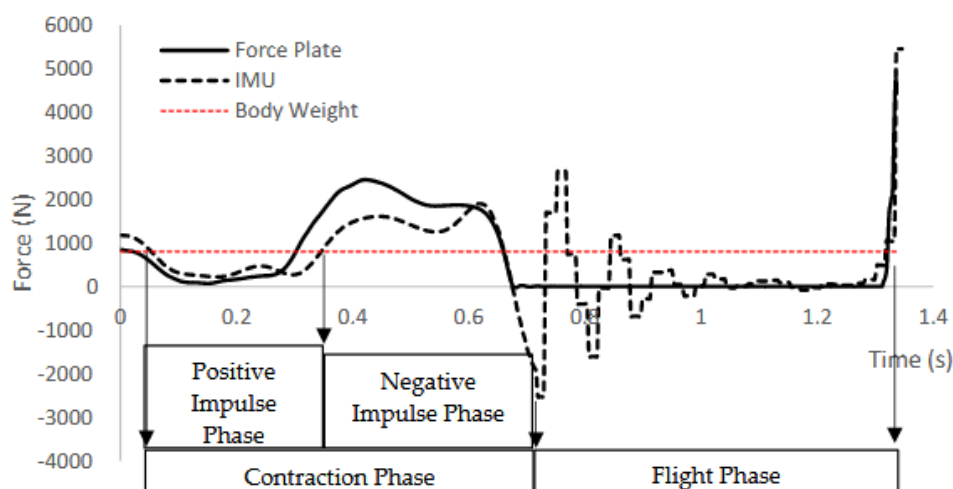


Figure 3.1: Estimating Counter Movement Jump Phases with IMUs, with a gold-standard force plate to assist in jump event detections [28].

Figure 3.1 illustrates the vertical force profiles throughout a countermovement jump by comparing lumbar spine data from an IMU with gold-standard force plate data and the athlete's body weight (horizontal baseline). The acceleration measurement signal of the IMU is multiplied by the athlete's body mass, enabling the estimation of vertical force throughout all phases of the jump. The force plate provides the ground reaction force, which reaches zero during the flight phase. However, the IMU continues to detect movement because it is attached to the body. The body weight line serves as a benchmark for phase transition detection, particularly for contraction and flight phase segmentation. Notably, the IMU approximately reproduces the force plate's force-time pattern during the contraction and propulsion phases, validating its use in estimating temporal parameters and generating dynamic forces in outdoor environments [28].

For example, direction changes in sports movements have been shown to be well described by IMUs Alanen et al. [1]. Perhaps the most significant advantage of IMUs is their ability to capture data in natural environments without interfering with movement. IMUs can be placed in smart clothing, enabling unobtrusive capture of information during dynamic movement.

3.2.1 Integration of IMU Data into Machine Learning Models

Ideal sensor configurations have been determined to enhance the reliability and accuracy of activity recognition systems, particularly in applications requiring critical motion tracking and detection [29].

IMU data is often used alongside machine learning models for activity classification and motion quality evaluation. Such models depend on essential preprocessing steps, such as feature extraction and noise filtering, to adequately process and interpret the raw sensor information recorded by the IMU [35, 37].

For activity recognition purposes, IMU data must be preprocessed to compensate for noise induced by sensor misalignment or ambient noise. This is followed by the extraction of relevant features, such as mean acceleration, jerk and angular velocity variance, which are then used as inputs for classification algorithms. The next step is classification, in which algorithms and machine learning are employed to identify activities based on their respective extracted features. This clearly defined workflow enables the accurate interpretation of complex motion patterns and supports a wide range of activity recognition applications.

Table 3.2 presents some of the major studies that have employed IMU data for activity recognition, including their methodology and findings.

Table 3.2: IMU analysis studies

Study	Activity Type	IMU Metrics Used	Outcome
Kranzinger et al. [21]	Running, Skiing	Acceleration, Angular velocity	High temporal accuracy using LSTM and CNN models
Alanen et al. [1]	Change of Direction	Acceleration, Kinematic Variables	Improved Change Of Direction (COD) movement analysis, highlighting biomechanical patterns
Phan et al. [33]	Multi-Sport (37 activities)	Acceleration, Gyroscope data	Achieved 96% accuracy using CNN for grouped activities
Saha, Palle, and Chowdhary [38]	Exercise Variability	Euler Angles, Gravity Metrics	Real-time motion variability tracking with low computational cost
Rufyid-u-Nissa, Karmakar, and Baghini [36]	Lower Limb Exercises	Acceleration Spectral Features	Embedded ML achieved 99.41% accuracy in classification

3.2.2 Challenges and Real-World Applications

IMUs monitor the biomechanics of high-intensity sports performance and provide quantitative feedback to optimise movement technique and detect different activity types. For example, Kranzinger et al. [21] demonstrated their ability to analyse direction changes in team sports to provide personalised feedback to athletes. In rehabilitation, IMUs can be used to monitor recovery processes, enabling clinicians to intervene more individually based on accurate motion analysis [33].

IMUs also improve physical education in terms of safety and efficiency by providing instantaneous feedback while students are in motion. However, the computational requirements of this

technology are challenging to fulfil in currently available low-power wearable devices for real-world deployment. New light machine learning algorithms and signal processing pipelines have recently been developed to address this limitation [11].

Despite the potential offered by various IMUs that have demonstrated impressive performance in real-world applications, incorporating them into activity recognition systems is challenging. One example of this is how motion artefacts can severely impair the quality of the data. Artefact removal, data normalisation and adaptive filtering are some of the techniques used to provide cleaner, more reliable motion data for analysis [17, 23]. In addition, variability in sensor placement between users and subject-to-subject variability in anatomy must be compensated for through stable calibration processes to normalise data collection.

The second main challenge is synchronising IMU data with other sensor streams, such as the ECG, in time. Proper analysis requires time synchronisation of multimodal sets, especially in systems that bring together physiological and biomechanical sensor streams. Advanced time series synchronisation methods have been proposed to address this issue, such as dynamic time warping, which makes multimodal integration more reliable [16].

3.3 Fusion of ECG and IMU data

Combining ECG and IMU data exploits the complementary benefits of each modality, supporting more thorough activity recognition. While ECG provides direct measurements of cardiovascular response and exercise intensity, IMU data describes movement dynamics and kinematics. Integrating both data sources overcomes the limitations of individual sensors, facilitating a more complete and objective determination of physical activity for classification and analysis.

Research of this kind, such as Mekruksavanich, Jantawong, and Jitpattanakul [27], has demonstrated significant improvements in activity classification performance when ECG-derived HRV and IMU acceleration features are included. This is due to subtle correlations being discovered among physiological signals and movement habits at activity transitions.

3.3.1 Data Fusion Techniques

Data fusion techniques enable the combination of ECG and IMU, significantly improving the accuracy and completeness of activity detection systems. The literature thoroughly explains three major techniques for combining these signals: (1) early fusion, (2) intermediate fusion and (3) late fusion.

Early fusion, or data fusion, is performed at the preprocessing stage as a combination of the two sensors' (IMU and ECG) raw signals or features. This process is clearly explained in the work of Rufyid-u-Nissa, Karmakar, and Baghini [36], which describes how to implement time synchronisation between the ECG and IMU streams. The resulting features represent biomechanical and physiological activities and exhibit strong modality coupling. However, the disadvantage of such approaches is the increased computational complexity due to greater processing in the initial stages.

In the intermediate fusion approach, the extracted ECG and IMU features are combined before being fed into the machine learning model. This approach strikes a balance between integration and computational parsimony and is widely considered to be the most common technique. For example, Phan et al. [33] learned features separately from the ECG and IMU via CNN and then concatenated and classified them using a dense neural network. This allows the model to capture complex relationships between the two modalities without incurring the high computational cost of processing all the raw data simultaneously.

Conversely, the late fusion approach combines the decisions or predictions made by individual models, considering ECG and IMU data separately. Although this approach is less computationally demanding and easier to deploy, it may not be capable of detecting subtle interactions between the two data streams that could affect the overall accuracy of activity detection.

Table 3.3 summarises key studies that use ECG and IMU data fusion for activity recognition. It lists the fusion approaches used, the applications they are used for, and the notable results achieved.

Table 3.3: ECG and IMU data fusion in Activity Recognition

Study	Fusion Method	Outcome
Zhang et al. [45]	Early Fusion	Improved classification accuracy and robustness
Mekruksavanich, Jantawong, and Jitpattanakul [27]	Intermediate Fusion	15% accuracy improvement
Hossain et al. [14]	Late Fusion	Accurate fatigue detection; improved recognition in real-world conditions

3.3.2 Applications and Benefits

Fusing ECG and IMU data has demonstrated clear benefits in sports and educational settings. For example, combining heart rate data with kinematic parameters derived from IMUs allows running technique and effort to be assessed, as demonstrated by Kranzinger et al. [21], who employed this method to customise training interventions for runners based on quantitative feedback. In education, integrated ECG-IMU systems have been used to systematically monitor students' physical exertion during lessons, enabling teachers to track effort levels and the quality of movement, and provide helpful feedback to enhance participation and ensure safety.

3.3.3 Challenges and Limitations

ECG-IMU fusion is one of the most significant advances in activity recognition, as systems can now analyse both movement and physiology simultaneously. According to Phan et al. [33], such an integrated framework has greatly improved accuracy. However, some areas require innovative

solutions, such as real-time fusion algorithms, lightweight wearable implementations, and efficient synchronisation schemes.

Despite its enormous potential, data fusion of IMU and ECG presents its own challenges that must be overcome in order to deliver optimal performance. Synchronisation is perhaps the most important of these. Temporal registration of both ECG and IMU data streams was a prerequisite for this analysis. Misalignment of the data would cause feature extraction and activity classification to fail, which would compromise the system's reliability [36]. Another significant issue is the computational overhead, since handling and aggregating multimodal data in real time requires powerful hardware and optimised algorithms. Although this has been achieved through the adoption of lean models, such as the neural networks proposed by Saha, Palle, and Chowdhary [38], the scalability of wearables remains a significant issue.

3.4 Machine Learning Techniques

These models have made significant progress in the field of activity recognition research by providing a robust framework for classifying activities from multimodal data. Various ML algorithms can be categorised into two broad groups: conventional algorithms and deep learning methods. This paper discusses the methods and their comparative advantages and disadvantages, with a focus on applications related to activity recognition.

3.4.1 Machine Learning Algorithms

SVM, RF, and K-Nearest Neighbours (KNN) are classic ML algorithms that have been widely used for activity classification due to their ease of use and low computational expense. SVM performs best in binary classification tasks. However, their accuracy worsens with high-dimensional data, as demonstrated by Fang [11], who used SVM to detect stress levels from ECG signals.

Conversely, RF is effective in noisy data scenarios, e.g. IMU data, and improves classification performance by combining the outputs of many decision trees using ensemble learning.

Finally, KNN is a simple algorithm involving straightforward calculations to determine the distance between data points. However, it is computationally costly for large datasets and therefore not efficient for real-world applications where low processing times are critical. Despite these classical algorithms being efficient, they often fail to capture advanced temporal and spatial patterns in multimodal data, requiring more sophisticated machine learning strategies.

Deep models are some of the pioneering technologies of activity recognition and have proven to possess an unmatched ability to handle large, complex data sets. They are also extremely efficient at extracting subtle patterns from multimodal sensor data.

Spatial features of sensor data, such as ECG signals, can be effectively extracted from convolutional neural networks. For example, Mekruksavanich, Jantawong, and Jitpattanakul [27] achieved 92.5% accuracy in classifying different physical activities using heart rate and heart rate variability parameters obtained through a CNN. Although CNNs are powerful for extracting features from

spatial data, they are inefficient for modelling sequential time data and are not as effective for sequence data analysis, such as time-series output from sensors.

Recurrent neural networks, and Long Short-Term Memory (LSTM) networks in particular, overcome the drawbacks of CNNs due to their effectiveness in extracting temporal dependencies from sequential data. LSTMs have been found to be extremely helpful in processing IMU time series data. For example, Zhou and Wen [46] utilised a hybrid model of CNN and LSTM components to model both spatial and temporal patterns, achieving 94.3% classification accuracy. However, a major weakness of LSTMs is their computational cost, which makes them impractical for real-time applications, particularly in resource-constrained wearable devices.

Hybrid models based on the combination of CNN and LSTM building blocks represent an improvement in the activity recognition domain. These models leverage the spatial feature extraction capability of CNNs and the temporal modelling ability of LSTMs to collaborate in modelling multimodal data.

While these deep learning approaches significantly improve activity detection, issues regarding computational efficiency, generalisation across a wide range of datasets and real-time applicability remain to be addressed in future work.

3.4.2 ML Implementations

The implementation of activity recognition with an ML system incorporates several phases, each of which is tasked with creating a stable and precise classification framework. From data preprocessing to actual field deployment, every stage that dictates the system's overall performance and credibility is included in this framework.

In the analysis phase, raw sensor data such as ECG and IMU data is preprocessed to prepare it for modelling. This preprocessing involves processes such as noise and artefact reduction, and data enhancement. ECG signals, for example, typically require noise removal due to motion artefact or baseline drift, while IMU signals typically require alignment and normalisation due to differences in sensor position or orientation. Feature engineering is therefore crucial at this stage to extract meaningful metrics such as heart rate variability, acceleration intensity, and angular velocity variance. The features obtained through extraction form the basis of most subsequent modelling steps and play a significant role in determining the performance of the ML system.

Training is the most critical step, during which the models learn the patterns and associations contained in the data. This involves applying the model to labelled data sets so that it learns to associate certain features with their respective activity labels. Training involves iterative, greedy optimisation techniques that try to minimise prediction errors by learning model parameters through stochastic gradient descent or backpropagation. CNNs, for example, are optimised to learn spatial features from ECG data, while recurrent neural networks and long short-term memory networks try to optimise temporal relationships in sequence IMU data. Other studies, such as those conducted by Zhou and Wen [46], have also concluded that balancing the dataset and providing adequate augmentation during training can mitigate class imbalance and improve generalisation. This implies that the model will perform well in many different types of activity. Additionally, tuning

hyperparameters such as learning rates, batch sizes, and model depth is also crucial for achieving optimal performance.

Once training is complete, the model undergoes extensive testing. The trained model is executed on unseen data to provide an objective estimate of its predictive capability. Primary performance measures such as accuracy, precision, recall, and the F1 score are used to evaluate the model's performance. For example, Zhou and Wen [46] achieved 94.3% accuracy in physical activity classification using a hybrid CNN-LSTM model, thus demonstrating the strength of merging spatial and temporal analyses. Cross-validation methods, such as k-fold validation, are usually employed to maximise the consistency of performance measures by splitting the dataset into multiple training and testing subsets. This avoids overfitting and ensures that the model generalises well to new data. During this phase, challenges such as variability in sensor placements and user movements are simulated to confirm the model's flexibility and stability.

The application phase, as the name suggests, is the final phase in which the previously tested model is applied. This phase bridges the gap between theoretical formulation and practical deployment, offering significant potential for monitoring sports performance, rehabilitation and education. In sports, it provides athletes with feedback on their effort level and movement efficiency through intrinsic ECG and IMU measures. It is used in rehabilitation, where clinicians track the recovery process and modify therapy based on physiological and biomechanical information. The deployment of computational resources, latency and light algorithms or model compression, as proposed in Saha, Palle, and Chowdhary [38], is commonly employed to maximise performance without compromising the accuracy of the created systems.

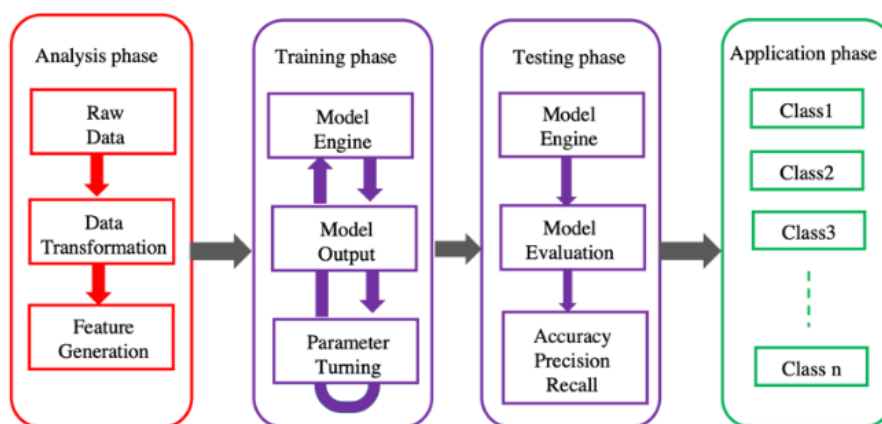


Figure 3.2: Four Phases of the ML Classification [26]

3.4.3 Key comparison

To provide a comprehensive overview, Table 3.4 compares the accuracies of notable ML techniques used for activity recognition.

Table 3.4: Machine Learning Techniques

Study	ML Technique	Metrics Used	Application	Accuracy (%)
Mekruksavanich, Jantawong, and Jitpattanakul [27]	Convolutional Neural Networks (CNN)	HR, HRV	Activity Recognition	92.5
Fang [11]	Support Vector Machine (SVM)	HRV, Stress Levels	Stress Monitoring	89.0
Zhou et al. [47]	Hybrid CNN-LSTM	Temporal-Spatial Features	Physical Activity Classification	94.3
Saha, Palle, and Chowdhary [38]	TinyML In-Sensor Movement Variability	Movement Variability, Gravity Vector Templates	Real-Time Monitoring	88.7

3.4.4 Challenges and Future Directions

While ML techniques have demonstrated high accuracy in controlled environments, their real-world application remains challenging due to:

Data Variability: differences in sensor placement and individual physiological responses introduce noise, requiring robust preprocessing methods;

Computational Constraints: real-time applications in wearable systems require lightweight algorithms that achieve a trade-off between accuracy and efficiency;

Generalisation: most of the models lack generalisation across diverse activity types and user groups; hence, adaptiveness in learning frameworks is an urgent need [31].

Therefore, future work should focus on designing hybrid models that leverage transfer learning and domain adaptation to encourage better generalisation and real-time responses.

3.5 Comparison and Challenges

Recent research has demonstrated that previous machine learning approaches to activity recognition based on ECG and IMU have achieved high accuracy in both controlled and semi-controlled environments [38, 41, 46]. The shift from single-sensor to multimodal systems, particularly the integration of ECG and IMU, has improved activity classification accuracy remarkably, especially with intermediate or hybrid fusion strategies [27]. Deep learning models, such as CNNs and LSTM networks, are particularly well-placed to leverage the temporal and spatial interactions of multimodal datasets [46].

Despite this expansion, several challenges remain. Firstly, there is a trade-off between model performance and the computational resources required, particularly when real-time processing is necessary for low-resource wearable devices [38]. Sparsely connected network architectures address this issue, but may do so at the expense of some classification accuracy.

Another important limitation is generalisation: most models perform best on homogeneous datasets, but struggle with diversity in user populations, sensor placements and unstructured environments. Diversity in signal quality, motion artefact and non-uniform sensor alignment can degrade performance, necessitating robust pre-processing pipelines with adaptive filtering and feature normalisation [31].

Furthermore, synchronising ECG and IMU streams properly remains technically challenging. Existing work has utilised techniques such as dynamic time warping to align misaligned multimodal data, which would otherwise reduce model accuracy [29]. However, these algorithms increase processing complexity and may not be applicable within embedded systems.

In general, state-of-the-art systems show great potential, but future research needs to focus on creating adaptive algorithms that can cope with variability specific to users and situations.

Chapter 4

Methodologies

This chapter presents the methodological framework adopted in this dissertation for the recognition of sports activities using sensorized wearable devices. The methodologies encompass all the essential steps, from the selection and preprocessing of publicly available datasets, through the design and implementation of feature extraction pipelines, to the development and validation of machine learning models.

The chapter details the characteristics and acquisition procedures of the two main datasets used [39, 43], the strategies applied for data preprocessing and feature engineering, and the experimental protocols established for model training, evaluation, and benchmarking.

4.1 Description of Acquired Datasets

This dissertation was conducted as part of the TEXP@CT project, whose main goal is to develop and validate machine learning models for the recognition of sports activities using data from sensorized wearables. The ideal scenario for this research would be the acquisition and use of proprietary data collected directly from the latest generation of the TEXP@CT smart vest. However, during the research period, the hardware was still under development and experimental validation.

The raw data formats and the preliminary stage of the hardware presented significant challenges for immediate use in machine learning pipelines. The IMU data acquired from these early-stage prototypes was not yet available in a reliable or standardised format. Given these constraints, and to ensure scientific rigour and the reproducibility of results, it was necessary to rely on publicly available benchmark datasets for model development and evaluation. The selection of these datasets was guided by their relevance to the project context, namely sports performance monitoring and activity recognition using inertial sensors. While the present work uses public datasets for classification model training and validation, the entire methodological pipeline, including preprocessing, feature extraction, and evaluation protocols, was designed to be compatible with future data to be acquired from the TEXP@CT vest.

For reference, the dataset extracted from *Action Recognition in Basketball with Inertial Measurement Unit-Supported Vest* by Sonalcan et al. [39] will be referenced throughout as the Hamza Dataset. The dataset extracted from *Inertial Sensor-Based Motion Tracking in Football with Movement Intensity Quantification* by Wilmes et al. [43] will be referred to as the Wilmes Dataset.

4.2 Hamza dataset

This dataset was created based on research conducted by Sonalcan et al. [39], which involved recording inertial data during fundamental basketball movements. The data was recorded from 21 semi-professional basketball players (12 men and 9 women) aged between 19 and 24 years old. All of the players were right-handed and played on a competitive team. Those with injuries or medical conditions were not permitted to participate in the study. The study adhered to the ethical guidelines outlined in the Declaration of Helsinki (approval no. 376-376-10, Usak University Ethics Committee).

4.2.1 Data Acquisition Setup

The data collection system was built around the Deneyap Development Board, which features a 3-axis accelerometer, a 3-axis gyroscope (model LSM6DSM), and a dual-core ESP32 processor (Tensilica LX6, 240 MHz) with built-in Bluetooth Low Energy. These components were enclosed in a rigid, 3D-printed casing and mounted onto a vest worn by each participant. The casing was consistently positioned around the T8 vertebral level (mid-thoracic spine) to allow for a fixed point of attachment. This fixed position was chosen to maintain the sensor's coordinate system consistently across all subjects by securing the device in a fixed position within a rigid plastic casing. This configuration reduces dependence on anatomical variations among subjects and enhances the generalisability of results across different athletes.

The IMU axes were oriented so that the x-axis extended horizontally to the right side of the wearer, the y-axis vertically upward along the spine, and the z-axis forward in the direction the athlete was facing. As illustrated in the article, the sensor was not always perfectly perpendicular to the ground, with small deviations and inclinations between the intended and actual sensor orientations. These geometric inaccuracies may introduce minor misalignments, especially affecting directional measurements and the accuracy of gyroscopic data, which is sensitive to precise axis alignment. The use of a rigid casing and consistent placement aims to minimise these effects and ensure reproducibility across all participants.

The IMU sampling rate was set at 250 Hz. The gyroscope was configured to $\pm 250^\circ/\text{s}$ and the accelerometer to $\pm 2\text{g}$. In data collection sessions, every participant executed five pre-sequenced basketball movements: *standing still*, *dribbling*, *shooting*, *chest passing*, and *layups*. Every movement was captured for 30 seconds. A graphical user interface (GUI) synchronized with the IMU stream was used to manually label the actions.



Figure 4.1: Back vest and IMU sensor [39]

4.2.2 Preprocessing and Windowing

The raw IMU signals were collected and preprocessed. Significant transmission errors and signal artefacts were removed from the data. Next, features were extracted by segmenting the continuous time series using a sliding window technique. This technique generated overlapping, fixed-length windows from the signal. Although not addressed in the original study, the overlap percentage directly affects sample size and temporal resolution, as discussed in the literature by Kranzinger et al. [21].

Then, normalisation methods were applied to normalise the input space. This is particularly useful if you are dealing with data from several users or sessions, since it provides you with standardised feature scaling. Normalisation also minimises the effect of external variability and sensor drift [2].

4.2.3 Feature Extraction

The dataset used in this dissertation was provided by the original authors in a structured spreadsheet format, with each row representing a signal window and each column corresponding to an extracted statistical feature. A total of 54 features were extracted for each sample, encompassing nine time-domain features for each axis (x , y , and z) of both the accelerometer and gyroscope signals. These features are widely adopted in human activity recognition due to their proven discriminative power and computational efficiency [34].

The description of the main statistical features computed for each temporal window follows:

Minimum and Maximum: Define the amplitude boundaries of the signal within the window.

Peak-to-Peak (PTP): Measures the full dynamic range, serving as a proxy for movement intensity.

Simple Square Integral (SSI): Represents the cumulative energy of the signal over the window.

Root Mean Square (RMS): Reflects average signal power, aiding in the identification of forceful or rapid movements.

Mean Absolute Value (MAV): Summarises the overall magnitude in a manner robust to noise.

Mean Absolute Difference (MAD): Captures temporal fluctuations by evaluating the variation between consecutive samples.

Standard Deviation (SD): Quantifies the dispersion of the signal around its mean.

Skewness (third-order moment): Describes the asymmetry of the signal distribution, useful for detecting non-uniform motion patterns.

Kurtosis (fourth-order moment): Indicates the presence of heavy tails or outlier behaviour within the signal.

All features were extracted independently for each axis of both the accelerometer and gyroscope data, resulting in a total of 54 features per sample. These features served as the input variables for the supervised classification models developed and evaluated in subsequent chapters.

4.3 Wilmes dataset

The Wilmes dataset is derived from the publicly available recordings originally published by Wilmes et al. [43] in the context of biomechanical performance analysis. The study examined lower-limb kinematics and quantified movement intensity using wearable IMU and ground-truth validation from optical systems. These recordings were adapted for activity classification under a machine learning framework in this dissertation.

Five IMUs were positioned on the participants' pelvis and lower limbs to collect data. For this dissertation, only the IMU data from the pelvis sensor were kept, in accordance with the TEXP@CT project specifications for recording trunk-centric kinematics. This decision is motivated by the anatomical relevance of the pelvis region, as it lies close to the centre of mass of the human body, thereby enabling a more representative capture of overall motion dynamics during sports performance.

4.3.1 Participants and Experimental Protocol

The original experiment recruited 11 healthy male amateur football players (age: 21.8 ± 3.2 years; height: 181 ± 6 cm; weight: 76.3 ± 11.4 kg). All participants regularly trained and competed, performing at least one practice session and one match per week. None of the participants had injuries at the time of testing. Written informed consent was obtained prior to participation, and the study was approved by the VU Amsterdam Ethics Committee (VCWE-2019-070R1), in accordance with the Declaration of Helsinki.

Each participant completed a structured protocol consisting of 18 trials representing a diverse set of sport-specific movement tasks. The raw motion data collected during each trial were organised in a dedicated folder for each participant. The activity protocol encompassed six football movements, each executed at three different intensities: low, medium, and high. These movements include accelerations, decelerations, 180-degree turns, 90-degree directional cuts, vertical jumps, and ball kicks, as shown in Figure 4.2.

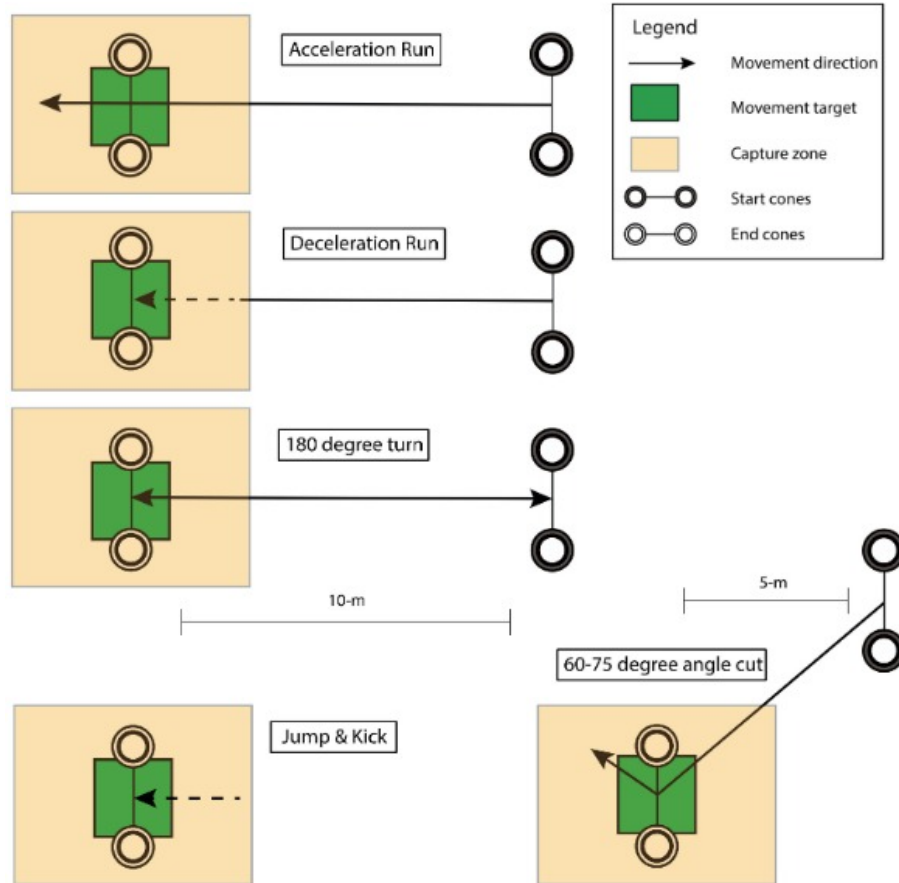


Figure 4.2: Football specific movements [43]

Each of the 18 trials corresponds to a unique pairing of one of the six predefined movement types and one of the three predefined intensities (low, medium, high). For example, label 1 represents a low-intensity acceleration, while label 18 represents a high-intensity kicking action.

All trials were recorded under controlled laboratory conditions using an IMU and a VICON optical motion capture system. The corresponding .c3d files, extracted from the VICON system, played a critical role in validating the temporal segmentation and class labelling of each trial. This multimodal verification process ensured the accuracy of the annotations associated with each sample, thereby improving the integrity of the dataset for use in supervised learning tasks.

4.3.2 Sensor Configuration and Placement

Each participant was equipped with five Xsens MTw Awinda IMU (Xsens Technologies B.V., Enschede, The Netherlands), positioned at the pelvis, left and right thighs, and left and right shanks. The IMU captured tri-axial accelerometer and gyroscope data at a sampling frequency of 148.15 Hz. Each sensor was factory calibrated and wirelessly synchronised with the master station to ensure time alignment across trials. The use of five sensors enables full-body lower-limb motion capture.

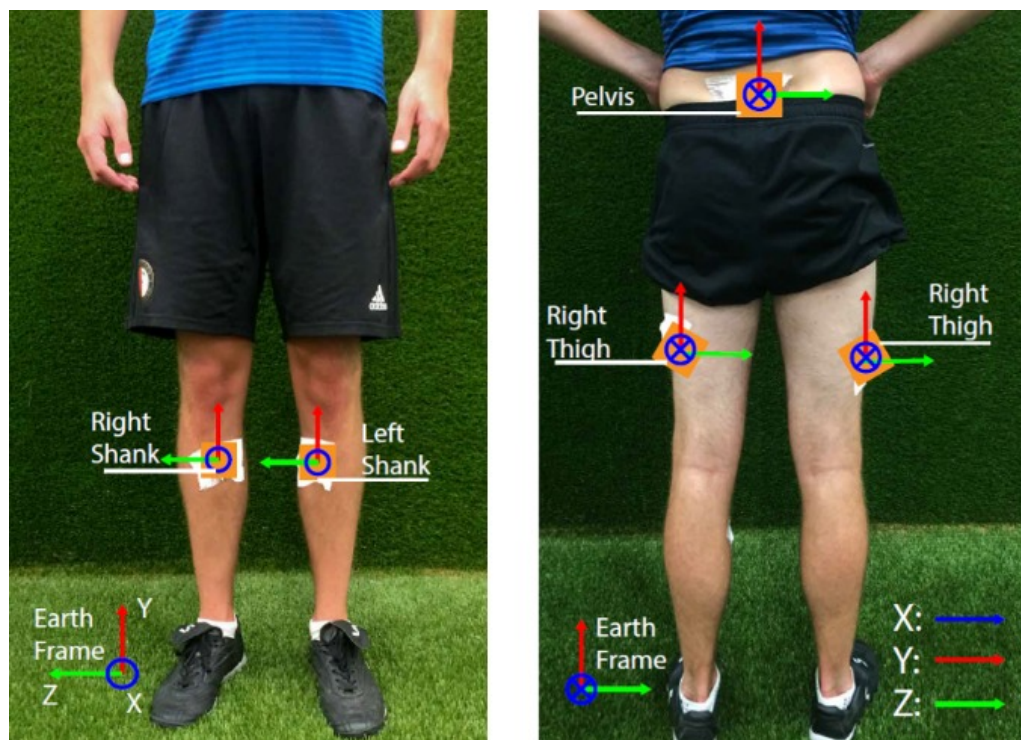


Figure 4.3: Sensor placement [43]

4.3.3 Data Preprocessing and Trial Selection

The initial dataset was provided in the form of MATLAB .mat files and synchronised .c3d motion capture files, organised per participant and per trial. Each .mat file contained two matrices: *AccData* and *GyroData*, both with dimensions $n \times 3$, representing the tri-axial inertial data for n time samples. The data from all five Xsens MTw Awinda IMUs were stored in sequence, covering the pelvis, left and right thighs, and left and right shanks. Therefore, an extraction step was required to isolate the relevant IMU signal from the pelvis sensor.

To validate the activity associated with each trial, an external software was used to visualise and appropriately label the trials, as detailed further in section 4.5.1. Through this cross-modal inspection, each trial was assigned a definitive class label according to the labelling scheme by Wilmes et al. [43], in which each integer from 1 to 18 corresponds to a unique pairing of activity type and intensity level.

Once validated, the signals from the pelvis-mounted IMU were extracted using a script that parsed the sensor sequence within the .mat files, retained only the tri-axial data from the relevant sensor, and discarded the remaining four.

The dataset was finalised after confirming the integrity of all trials, resulting in 198 labelled samples spanning six sport-specific actions, each performed at three intensity levels. Timing boundaries and segmentation checks were also completed at this stage to ensure signal continuity and consistent temporal resolution across the dataset.

4.3.4 Feature Extraction

After preprocessing, each trial was treated as a full-window segment representing the entire recorded motion instance. The extracted accelerometer and gyroscope data from the pelvis sensor were then processed through a feature engineering pipeline. After calculating a set of nine statistical time-domain features for each axis (x , y , and z) of both sensor modalities, these features were selected based on their prevalence in inertial signal analysis for human activity recognition. Table 4.1 provides a detailed summary of all computed features, along with their corresponding mathematical expressions.

Table 4.1: Time-domain statistical features computed from IMU signals.

Feature	Formula
Minimum (MIN)	$\min(x)$
Maximum (MAX)	$\max(x)$
Peak-to-Peak (PTP)	$PTP(x) = \max(x) - \min(x)$
Simple Square Integral (SSI)	$SSI(x) = \sum_{i=1}^n x_i^2$
Root Mean Square (RMS)	$RMS(x) = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$
Mean Absolute Value (MAV)	$MAV(x) = \frac{1}{n} \sum_{i=1}^n x_i $
Mean Absolute Difference (MAD)	$AD(x) = \frac{1}{n-1} \sum_{i=1}^{n-1} x_{i+1} - x_i $
Skewness	$\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3$
Kurtosis	$\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3$

The features include amplitude-based metrics (MIN, MAX, PTP), measures of signal power and energy (SSI, RMS, MAV), as well as descriptors of variability and distribution (MAD, Skewness, and Kurtosis). In the formulas for skewness and kurtosis, the symbol μ represents the mean of the signal segment, and σ denotes the standard deviation, which is calculated as:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (4.1)$$

These two moments quantify the asymmetry and *tailedness* of the signal distribution, respectively, providing additional discrimination for subtle movement differences.

All features were calculated independently for each axis of both accelerometer and gyroscope data, yielding a comprehensive feature vector of 54 dimensions per trial. This collection of features provides a rich, interpretable representation of each movement pattern, capturing signal amplitude, energy, power, distribution shape, and intra-window variability. The resulting feature vectors were organised in an ordered spreadsheet, with each row corresponding to a labelled trial and each column to a specific feature. This structured dataset was then used as the input for the machine learning classification systems described in the following chapters.

4.4 Machine Learning pipeline for Hamza Dataset

This chapter describes the machine learning pipeline used for activity recognition on the HAMZA inertial dataset using pre-engineered features. The process involves three steps: dataset preparation, classifier application, and model evaluation. Two classifiers, an SVM and an RF, were used separately and then combined into a common experimental protocol for comparative analysis.

All experiments were coded using a suite of open-source libraries for data analysis, model training, and scoring. The machine learning core functionality was coded using the `scikit-learn` library [32], which offers standardized functions for classification, preprocessing, model selection, and metric scoring. The `pandas` library [25] was used to input, output, and manipulate data, such as reading the dataset from Excel files and managing feature-label separations. `NumPy` [13] was used for numerical operations such as array slicing, transformations, and matrices, which efficiently managed feature vectors.

Additionally, the `matplotlib` and `seaborn` libraries [15, 42] were used to generate visual diagnostics, such as confusion matrices and feature importance graphs, to improve interpretability and evaluate performance. `StandardScaler` from `sklearn.preprocessing` was used to standardize features in terms of zero mean and unit variance. This is important for algorithms sensitive to feature scale, such as SVM.

Figure 4.4 provides an overview of the entire machine learning pipeline, including data preprocessing, feature extraction, model training, validation, and evaluation stages.

Each component of the pipeline, from preprocessing to model training and evaluation, was thus supported by well-established Python libraries, ensuring reproducibility, efficiency, and transparency of the entire process.

4.4.1 Dataset Preparation

The dataset used in this study was obtained directly from the supplementary materials of the work by Sonalcan et al. [39], which were provided in `xlsx` format. The dataset consists of inertial features that the authors pre-extracted, which removes the need for additional signal-level

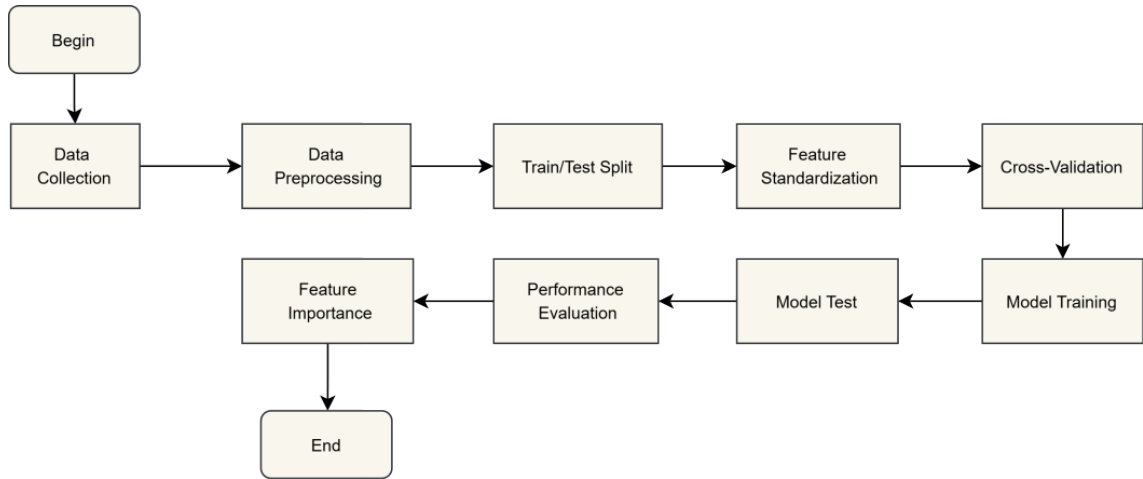


Figure 4.4: Processing pipeline for the Hamza Dataset, detailing data preprocessing, feature extraction, ML training and validation steps.

preprocessing. The dataset includes a collection of motion instances derived from a sensorized vest worn by basketball players. Each instance is labelled with a unique activity class index in the `Output` encoded column.

First, the Excel dataset was imported from the designated sheet. Once loaded, the dataset was cleaned by removing any rows containing missing values to ensure data integrity. The feature matrix X was defined by excluding the label column, and the output vector y corresponded to the class labels.

To facilitate unbiased model evaluation, the dataset was split into training and testing subsets using an 80%-20% stratified division. Stratification was used to maintain the proportion of the five activity classes in both divisions. To normalise the feature space and avoid scale biases, all input features were standardised to have a mean of zero and a variance of one. This is a necessary step for algorithms such as SVMs using distance-related kernel functions, which are sensitive to feature scaling.

4.4.2 Classifier Implementation

The first model that was trained was an SVM classifier with a Radial Basis Function (RBF) kernel. The RBF kernel is appropriate for tasks with non-linear decision boundaries, which is a characteristic of activity classification problems from wearable sensor data. The classifier was initialised with the following default hyperparameters: $C = 1.0$, the regularisation strength; and $\gamma = \text{scale}$, where the kernel bandwidth is scaled by the number of features. Then, the SVM model was fit to the scaled training data. Prediction was carried out on the test dataset after fitting. We evaluated the results using a full classification report that included accuracy, precision, recall, and the F1 score. We also created a confusion matrix heatmap to examine misclassification patterns between classes. These diagnostic tools provided insight into the model's predictive behaviour and class-wise performance.

The second classifier used was an RF model, an ensemble technique consisting of several decision trees. Random Forests are resilient to overfitting and likely perform well in high-dimensional feature spaces. For this use case, we initiated the model with 100 estimators and `random_state=42` to ensure reproducible results.

Unlike SVMs, Random Forests are less sensitive to feature scaling. However, an identical set of standardised input features was used to keep the models consistent. RF modelling was trained on the same training set and validated on the test subset. The model's predictions were followed by a confusion matrix and a classification report in the same format as the SVM evaluation step. Another diagnostic process involved calculating feature importance scores from the Random Forest model.

This provided an estimate of which inertial characteristics most contribute to decision-making, enabling contextualization of the model's internal reasoning and identification of important biomechanical cues involved in class discrimination. After validating the behaviour of each classifier, a unified script was created to run either model. The combined pipeline consolidated all preprocessing, training, prediction, and testing procedures. By accepting the classifier object as an argument, the script could conveniently switch between SVM and RF models, allowing for comparison under the same conditions.

4.4.3 Cross-Validation Procedure

In order to make the utilised classification models robust and generalizable, a 5-fold stratified cross-validation (CV) method was applied during the model-building phase. Cross-validation is a common validation technique used to estimate how a machine learning model would perform on previously unseen data. In this experiment, the stratified k-fold CV was particularly chosen to preserve the original class distribution in each fold, which is imperative while working with multiclass datasets such as the basketball activity dataset used here.

The entire dataset was split into five fold-wise disjunct subsets. Importantly, stratification was used such that each fold contained a representative share of all five activity classes. In each run of the process, four of the folds were used for training the model, with the fifth fold used for validation. The process was performed five times, with the validation fold rotated in each run so that each data point was used for validation exactly once.

The primary motivation for employing this approach was to eschew the danger of overfitting that arises from estimating based on a single random train-test split and thus obtaining a suboptimal approximation of general performance, especially with possibly subject or class-skewed datasets. Results averaging over multiple partitions reduce variance in estimates of performance, providing a more stable and accurate estimate.

Moreover, cross-validation was integrated into the pipeline. This facilitated the homogeneous application of preprocessing steps within each fold to prevent leakage. For both models, SVM and RF, performance metric averages over the five folds were computed and used as a measure of the stability and appropriateness of the model.

This method is well-suited to human movement recognition, where the movements may be subject-specific or slightly different on each occurrence. The use of stratified k-fold CV assisted in guaranteeing that the classifiers were trained and not merely overfitting to specific instances, but rather learning patterns that were useful from the statistical features extracted based on the IMU data.

Since the number of subjects was comparatively small and the data were homogeneous, Leave-One-Subject-Out (LOSO) cross-validation could produce excessive variability in performance estimates, reflecting differences between subjects rather than true model quality [10]. In contrast, stratified k-fold cross-validation maintains the original class distribution in each fold and ensures that the entire dataset contributes to both training and validation. This approach offers a more stable, computationally efficient, and representative evaluation for intra-participant classification in this study phase.

4.4.4 Evaluation Metrics

A comprehensive suite of evaluation metrics was employed to assess the performance of the classification models, ensuring a robust and multi-dimensional understanding of their behaviour. These included accuracy, precision, recall, and F1-score.

Accuracy, the most general metric, quantifies the proportion of correctly classified instances over the total number of predictions. While informative as an overall indicator, it may present a skewed impression in scenarios involving class imbalance, where one or more classes are overrepresented.

To complement accuracy, precision and recall were used to provide more nuanced insights. Precision measures the correctness of the positive predictions, defined as the ratio of true positives to the sum of true and false positives. It reflects the model's ability to avoid false alarms by indicating how many of the predicted positives are actually correct. Recall, or sensitivity, evaluates the model's ability to detect all relevant instances, computed as the ratio of true positives to the sum of true positives and false negatives. It captures how many of the actual positives were successfully identified.

The F1-score, calculated as the harmonic mean of precision and recall, was used to provide a balanced view that accounts for both false positives and false negatives. This metric is particularly valuable in multiclass contexts or when classes are not evenly distributed, as it penalises extreme values in either precision or recall.

In addition to scalar metrics, confusion matrices were generated for each model to visualise classification performance across all five classes. These matrices enabled a detailed inspection of class-wise predictions, highlighting any recurring misclassifications and offering a breakdown of true positives, false positives, and false negatives per category. Such visual tools are instrumental in uncovering model biases or ambiguities between movement types.

To further interpret the model's decision-making process, a feature importance analysis was conducted on the Random Forest classifier. Leveraging the inherent structure of ensemble trees,

this method computed impurity-based importance scores for each input feature. Features contributing most to reducing classification uncertainty across tree nodes were ranked higher. This interpretability step provided key insights into which statistical descriptors—such as signal energy, peak-to-peak amplitude, or skewness—were most influential in distinguishing between basketball movements. By identifying biomechanically relevant features, this analysis also helped corroborate the domain validity of the extracted time-domain features and their alignment with the physiological nature of the recorded actions.

4.5 Machine Learning Pipeline for Wilmes Dataset

This section presents the complete machine learning pipeline designed for the Wilmes dataset, aiming to classify football-specific movements based on engineered statistical features extracted from inertial sensor signals. Two independent classification models were implemented, an SVM and an RF, with both being assessed under the same conditions using standard validation protocols. In contrast to the Hamza dataset, the Wilmes dataset required full signal-level preprocessing and feature computation, offering an opportunity to tailor and validate the feature engineering approach.

All scripts were developed in Python, leveraging open-source scientific libraries for data manipulation and machine learning. Core modelling functionality was implemented using the `scikit-learn` library, including classification, preprocessing, model selection, and evaluation tools. The `pandas` library was used to read and handle Excel-based datasets, while `NumPy` supported the processing of numerical arrays. Additionally, `matplotlib` and `seaborn` were employed to generate diagnostic visualisations, such as confusion matrices and feature importance plots. Feature standardisation was handled by `StandardScaler` from `sklearn.preprocessing` to ensure uniform scaling across inputs, particularly important for SVM classifiers.

Figure 4.5 provides an overview of the entire machine learning pipeline, including data validation and preprocessing, feature extraction, model training, validation, and evaluation stages.

Each phase of the pipeline was modular and reproducible, ensuring a robust foundation for evaluating classifier performance across both fine-grained movement categories and broader activity types.

4.5.1 Dataset Acquisition and Label Verification

The dataset used in this section was sourced from the supplementary materials provided by Wilmes et al. [43], and consists of raw inertial sensor data and VICON marker trajectories collected from 11 football players. Each participant's data is organised in a dedicated folder, containing individual movement trials stored in MATLAB `.mat` files under the `IMU` subfolder, and corresponding VICON files in `.c3d` format under the `c3d` folder. These recordings encompass a wide range of football-specific activities, such as acceleration, deceleration, jumping, cutting, and

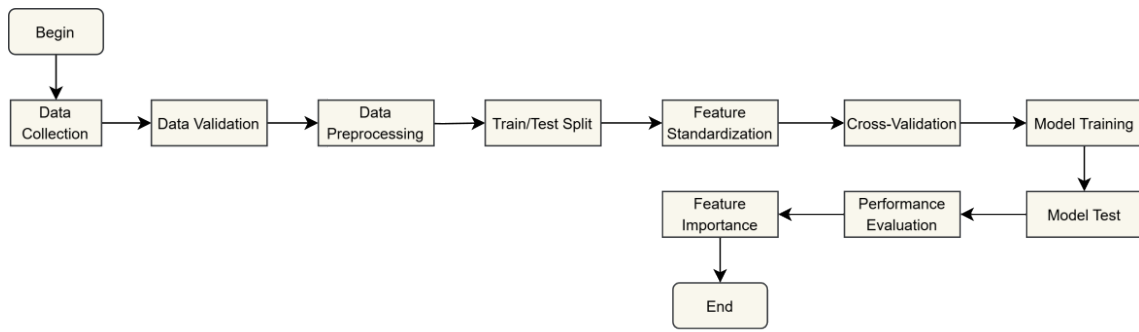


Figure 4.5: Processing pipeline for Wilmes Dataset, detailing the steps from data collection and validation to preprocessing, feature extraction, ML training and evaluation steps.

kicking, performed under three intensity levels: low, medium, and high, amounting to a total of 18 unique movement classes per participant.

Before extracting features or training machine learning models, a critical preliminary step was undertaken to validate the correctness of the movement labels assigned to each trial. Due to the manual labelling process inherent to these datasets, it was essential to confirm that each trial file reflected the intended movement type and intensity.

To ensure the reliability of the dataset, a comprehensive manual validation was performed for every trial of every participant. Because of the dataset structure with the presented participants and various movements and trials, we needed to do a segmentation using the open-source software Motion Kinematics and Kinetic Analyser (MOKKA), to identify which activities were performed and when. For each participant, the VICON trajectory files (.c3d format) corresponding to each trial were individually imported into the MOKKA.

Within MOKKA, the 3D marker trajectories were visualised frame-by-frame. The anatomical markers placed on the participants allowed for unambiguous identification of each movement and its intensity. By thoroughly reviewing all trials, it was possible to cross-check the actual executed motion against the assigned movement label for every single trial in the dataset. This step was crucial for detecting and correcting any possible inconsistencies or errors in labelling that could have resulted from the manual annotation process or data entry.

The validation process guaranteed that all data subsequently used for feature extraction and machine learning analysis were accurately labelled and reflected the correct movement type and intensity. This quality control step underpinned the integrity of the entire study, ensuring that the results and conclusions drawn from the classification models were robust and scientifically valid.

4.5.2 Feature Extraction from IMU Signals

Following the confirmation of label integrity, a dedicated Python script (`extract_features.py`) was implemented to process the raw inertial data and compute

relevant statistical features for each movement trial. The script recursively navigates through the participant folders, accessing all `.mat` files located in the `IMU` directories. Each trial file contains multichannel time-series data from both the accelerometer and gyroscope sensors, recorded in three orthogonal axes (X, Y, Z).

Following signal extraction, a set of handcrafted time-domain features was computed for each axis of each signal. The selected features and their descriptions are summarised in Table 4.1, presented earlier in this chapter, and include commonly used metrics such as minimum, maximum, peak-to-peak amplitude, RMS, MAV, skewness, and kurtosis.

Each descriptor was calculated separately for each axis of the accelerometer and gyroscope, yielding a total of 54 numerical features per trial (9 descriptors $\times$ 6 signals). These values were concatenated row-wise to form a structured dataset, and the final result was exported to `.xlsx` format. Preprocessing steps were applied to clean the dataset, including the removal of rows with missing or null values. This was essential to ensure that the final dataset was consistent and compatible with the machine learning classifiers. Each row in the resulting dataset corresponds to a single movement trial, described by the full set of extracted features, and labelled with the appropriate activity type and intensity.

The decision to compute these particular features was informed by prior literature and empirical performance observed on the Hamza dataset. Since the Hamza dataset—focused on basketball motion classification—provided high classification performance using this exact set of descriptors, it was considered methodologically consistent to adopt the same approach for the Wilmes dataset, enabling fair comparisons across datasets and sport modalities.

Moreover, statistical time-domain features such as those selected are well-established in the field of human activity recognition. As emphasised by Banos et al. [5], these features are computationally efficient, easily interpretable, and capable of capturing distinctive biomechanical signatures associated with different movement types. Their application in sports activity classification has been extensively validated, reinforcing their appropriateness for this study.

From a methodological perspective, these features offer a reliable abstraction of raw inertial signals, balancing representational richness with manageable complexity. While specific transformations such as frequency-domain analysis or wavelet decompositions can be more expressive, time-domain features remain highly effective when the dataset is structured and the signal patterns are well-aligned with biomechanical activity phases. Although the performance evaluation of the classifiers trained on this dataset is discussed in later sections, it is worth noting that using these standardised features was instrumental in achieving consistent model behaviour.

4.5.3 Classifier Implementation with Intensity-Specific Movements

The machine learning pipeline applied to the intensity-specific version of the Wilmes dataset was designed to distinguish between 18 distinct classes, combining six football-specific movements with three discrete intensity levels (low, medium, high). Each trial in the dataset was represented by 54 handcrafted time-domain statistical features, previously extracted from raw IMU signals and was labelled with a unique movement–intensity identifier.

The structured dataset was imported via `pandas`, and rows containing missing values were removed. Class labels were numerically encoded using Python's `category` data type, ensuring compatibility with the `scikit-learn` classification pipeline. All subsequent processing and modelling steps were implemented in Python using well-established libraries including `scikit-learn`, `NumPy`, and `matplotlib`.

4.5.3.1 Train–Test Partitioning and Feature Normalisation

To assess model performance while preserving the representativity of each movement class, the dataset was partitioned using a stratified 80%-20% split. This stratification, implemented using `train_test_split` from `scikit-learn`, ensured that all 18 movement-intensity combinations were proportionally present in both the training and test subsets. This approach is particularly critical in multiclass classification problems involving class imbalance, as it reduces bias and variance across partitions.

Prior to any model training, the feature matrix underwent standardisation using `StandardScaler`, which transforms each feature dimension to zero mean and unit variance. This normalisation was applied globally across the training data, and the same scaling parameters were subsequently used to transform the test set to avoid data leakage. Standardisation is especially relevant for algorithms like Support Vector Machines that rely on Euclidean distance calculations and are highly sensitive to discrepancies in feature magnitude. The same standardised features were also used for Random Forest classification, not out of necessity (RF is scale-invariant), but to maintain consistency and simplify comparative evaluation across models.

Internally, the data split also involved separating the independent variable matrix X from the label vector y , where X consisted of all 54 computed statistical features, and y represented encoded integer labels for each activity intensity class. This label encoding, conducted via `astype('category').cat.codes`, facilitated compatibility with the model APIs and enabled compact numerical indexing of classes during evaluation.

4.5.3.2 Model Training

Two classification algorithms were deployed and evaluated independently to compare their effectiveness in recognising movement patterns from time-domain inertial features.

The first classifier was an SVM equipped with a Radial Basis Function kernel. This kernel function is commonly used for problems with non-linearly separable data, making it particularly suitable for human activity recognition tasks where biomechanical patterns often exhibit nonlinear boundaries. The SVM was instantiated with standard hyperparameters: a regularisation constant $C = 1.0$ and kernel coefficient $\gamma = \text{scale}$, which automatically adjusts the kernel width inversely with the number of features. These values were selected based on conventional practice and initial exploratory experimentation, without further hyperparameter optimisation in this version of the pipeline.

The second model employed was a Random Forest (RF) classifier, initialised with `n_estimators = 100` to construct an ensemble of 100 decision trees. A fixed `random_state = 42` was used to ensure deterministic output and reproducibility across runs. Random Forests are known for their robustness to overfitting and their ability to model complex feature interactions without requiring input normalisation. However, standardised features were still passed to the model to maintain uniformity with the SVM implementation and facilitate shared pipeline components.

Both classifiers were trained on the 80% training subset using the `fit` method. Upon completion of training, predictions were generated on the hold-out test set using the `predict` method, producing a vector of predicted class indices for each input sample. These predictions formed the basis for downstream evaluation using performance metrics and confusion matrix analysis. While hyperparameter tuning was not conducted at this stage, the selected configurations reflect a balance between computational efficiency and classification effectiveness, serving as a baseline for further model refinement if necessary.

4.5.3.3 Cross-Validation Procedure

To ensure robust and unbiased performance estimation, a stratified 5-fold cross-validation scheme was applied to the training set during the model development phase. This validation approach, implemented via the `cross_val_score` utility from `scikit-learn`, partitions the data into five equally sized folds while preserving the original class distribution across all 18 activity-intensity labels. In each iteration, four folds were used for training and one for validation, ensuring that every sample participated exactly once as part of the validation set.

Crucially, normalisation was repeated within each fold to prevent information leakage—i.e., the `StandardScaler` was fit exclusively on the training portion of each fold and then applied to the validation data. This step ensures that the performance estimates reflect the generalisability of the model and are not inflated by artefacts of improper preprocessing.

Cross-validation was particularly valuable in this case, where the dataset comprises a limited number of samples per class and a high degree of intra-class variability due to intensity variations. By aggregating performance metrics across all folds, the pipeline achieved a more reliable estimate of model behaviour on unseen data and mitigated the overfitting risks typically associated with small datasets or unbalanced partitions.

4.5.3.4 Evaluation Metrics

Upon completion of training and cross-validation, each model was evaluated on the 20% hold-out test subset using a comprehensive set of performance metrics. The primary metrics, accuracy, precision, recall, and F1-score, were computed as previously described for the Hamza dataset in section 4.4.4.

These metrics were computed, which provide both macro-averaged (equal weighting across classes) and weighted (class-frequency-adjusted) scores. Due to the imbalanced nature of the

dataset, some movement–intensity combinations being underrepresented—the weighted F1-score was emphasized as the most informative metric, reflecting both the correctness and class-wise distribution of predictions.

In addition to scalar metrics, confusion matrices were generated and visualised using `ConfusionMatrixDisplay`. These matrices offered a fine-grained view of model behaviour on a per-class basis, facilitating the identification of systematic misclassifications. This was particularly valuable in detecting confusion between identical movement types executed at different intensity levels, which are inherently more difficult to separate due to their biomechanical similarity.

Lastly, for the Random Forest model, feature importance scores were extracted from the `feature_importances_` attribute, which quantifies the contribution of each feature to the overall reduction in classification uncertainty (Gini impurity). This analysis enabled the identification of the most discriminative features across the 54 extracted statistical descriptors.

The most impactful features typically included descriptors related to signal energy (e.g., RMS and SSI), dynamic range (e.g., PTP), and distributional characteristics (e.g., skewness and kurtosis), especially from the gyroscope signals in the mediolateral and vertical axes [8, 12]. This finding aligns with biomechanical expectations, as angular motion patterns vary more distinctly across intensities than linear accelerations in some movements.

Interpreting these feature importances not only contributed to model explainability but also offered valuable insight into the underlying biomechanics of different football movements. Such interpretability is crucial for practical deployment in sports science, where understanding the rationale behind automated predictions can inform coaching decisions and individualised training strategies.

4.5.4 Classifier Implementation without Movements Intensity

To improve classification performance and mitigate the challenges posed by small sample sizes, a secondary classification pipeline was developed. In this pipeline, the intensity levels (low, medium, and high) were removed from the activity labels. Instead of distinguishing between movement variations by intensity, all three levels were merged into a single generic label per movement. For example, the trials previously labelled as `cutting_low`, `cutting_med`, and `cutting_high` were all recategorised simply as `cutting`. This procedure, implemented directly within the Python classification script, involved mapping the original labels into generalised movement classes via a dictionary structure.

The rationale behind this transformation was twofold. Firstly, the original dataset presented only three trials per movement–intensity combination per participant, leading to sparse class distributions and unstable learning conditions. By aggregating the intensities, the number of samples per class increased by a factor of three, thus improving class balance and enabling more reliable model training. Secondly, the accelerometer and gyroscope signals for different intensity levels of the same movement exhibited subtle variations that proved difficult for the classifiers to reliably

distinguish. Consolidating these into unified classes allowed the models to focus on biomechanical differences between movement types rather than their intensity gradations.

4.5.4.1 Model Implementation and Training

The classification pipeline used for the analysis without movement intensity labels closely followed the methodology previously described for the intensity-specific version present in section 4.5.3. The same dataset structure was used, with 54 handcrafted statistical features per trial. The primary difference was the redefinition of the activity labels: a new label column was generated by merging all intensity levels into six generic movement classes, *acceleration*, *deceleration*, *cutting*, *jumping*, *kicking*, and *running*.

All subsequent steps, including dataset cleaning, stratified 80/20 train-test split, feature standardisation with `StandardScaler`, and the deployment of SVM and RF classifiers, were performed identically to the pipeline used for intensity-specific classification. Similarly, performance evaluation was based on cross-validated results using 5-fold stratified cross-validation, ensuring class balance and robust estimation.

This label consolidation was motivated by the limited number of samples per movement–intensity combination, which could otherwise lead to poor model generalisation and unstable results. By grouping intensities, the sample size per class increased, enabling more reliable classifier training and evaluation. Confusion matrices and feature importance analyses were generated as in the previous pipeline, supporting interpretability and facilitating comparison between the intensity-specific and non-intensity-specific approaches.

4.5.4.2 Performance Evaluation and Feature Analysis

Performance was assessed using standard classification metrics: accuracy, precision, recall, and F1-score, with both macro and weighted averages reported via the `classification_report` function. Given the improved class balance, both macro and weighted F1-scores were informative in assessing the overall model effectiveness. Confusion matrices were visualised for each classifier to understand class-specific misclassifications.

Lastly, the RF classifier’s internal structure was leveraged to assess feature importance. The 20 most influential features were ranked and plotted, offering interpretability into the model’s decision logic. These features primarily included time-domain descriptors from gyroscope signals, particularly RMS, PTP, and skewness, on the mediolateral and vertical axes. The most impactful features typically included descriptors related to signal energy, dynamic range, and distributional characteristics, especially from the gyroscope signals in the mediolateral and vertical axes. This finding aligns with biomechanical expectations, as angular motion patterns vary more distinctly across intensities than linear accelerations in some movements. This trend has also been corroborated by previous studies, including Gjoreski et al. [12] and Bulling, Blanke, and Schiele [8], reinforcing the discriminative value of gyroscopic descriptors in human activity recognition tasks.

4.6 Concatenation of Datasets and Classification Pipeline

To address the challenge of developing classifiers capable of recognising movements across different sports, a combined dataset was assembled by concatenating samples from the Hamza basketball and the Wilmes dataset. By integrating data from two distinct sports, the objective was not only to classify specific movement types but also to determine the sport from which a trial originated.

The concatenated dataset consists of handcrafted time-domain statistical features extracted uniformly from both original datasets. Each trial is described by 54 numerical features derived from the accelerometer and gyroscope signals along three orthogonal axes (x, y, z). These include descriptors such as minimum, maximum, RMS, SSI, PTP, MAV, and higher-order statistics such as skewness and kurtosis.

Notably, the label structures of the original datasets were maintained independently. Each row in the combined dataset retained two label fields, one indicating the movement type (e.g., *jumping*, *cutting*) and another denoting the sport domain (*basketball* or *football*). This approach allowed for flexible multi-target classification scenarios, supporting both movement recognition and sport classification without the need for post-hoc relabeling or structural harmonisation. Since both datasets followed an identical feature extraction methodology, the datasets were merged directly without requiring additional feature engineering or alignment.

4.6.1 Combined Dataset Methodology and Initial Pipeline Implementation

The initial implementation was designed to support multi-target prediction of both movement type and sport category. Data was imported, and the 54 numerical feature columns were isolated for model training. The categorical labels, `movement` and `sport`, were separately extracted and encoded into a numerical format.

To avoid data leakage and simulate realistic generalisation, the dataset was partitioned, treating the `participant` field as a grouping variable. This ensured data from the same individual did not appear in both training and test sets. The split ratio was configured as 80% for training and 20% for testing.

Three distinct pipelines were constructed, each incorporating normalisation and a separate classification model:

- Random Forest (RF): A pipeline with a `RandomForestClassifier` with 100 trees and a fixed `random_state` for reproducibility.
- Support Vector Machine (SVM): Similar pipeline structure, with an RBF-kernel `SVC` classifier set with default parameters.
- Multi-Layer Perceptron (MLP): A neural network classifier with one hidden layer of 100 neurons and a maximum of 300 training iterations.

Each pipeline was trained on the training subset and evaluated on the test set. Evaluation metrics were computed, including accuracy, macro and weighted F1-score, precision, and recall. Confusion matrices were also generated to visualise model performance across classes and sports. For the Random Forest and MLP classifiers, permutation-based feature importance was calculated. This method assesses the performance degradation caused by shuffling each feature independently, providing a robust estimate of its contribution to model decisions. The top 20 features were ranked and visualised. Features extracted from gyroscope signals, especially RMS, PTP, and skewness from the mediolateral and vertical axes, emerged as consistently influential.

This modelling configuration served as the baseline for the combined dataset analysis. The insights derived from this pipeline informed subsequent modifications to the dataset structure and classification strategies.

4.6.2 Class Balance Refinement and Signal Harmonisation

Following the initial baseline implementation, it became evident that additional preprocessing was required to address key limitations in the structure of the combined dataset. Chief among these were the imbalance in class representation, over 2,400 samples from basketball versus fewer than 200 from football, and inconsistencies in signal orientation stemming from differing sensor placements and acquisition protocols.

To mitigate these issues, a sequence of transformations was applied. Initially, 198 basketball instances were randomly selected to match the number of football trials, yielding a balanced dataset of 396 entries. This subset enabled controlled analysis under uniform class distributions and served as a reference point for subsequent developments.

Concurrently, signal harmonisation was performed to correct for the directional discrepancies between datasets. As illustrated in Figures 4.1 and 4.3, inertial units in the Hamza dataset were worn on the upper back using a sensorized vest, while in the Wilmes dataset, multiple IMUs were used, of which only the sensor placed on the pelvis was retained—selected for its alignment with the biomechanical reference frame and its relevance under the TEXP@CT specification. These distinct configurations resulted in inconsistent axis alignment. To resolve this, an axis remapping step was executed: the accelerometer axes from the Wilmes dataset were reassigned to approximate the orientation used in Hamza, ensuring that similar movement patterns were described within a coherent coordinate system.

After this adjustment, a new version of the dataset was assembled, expanding the basketball subset to 400 entries while retaining 200 from football. Although this introduced a mild imbalance, it improved variability within the basketball class. To achieve final class balance, a synthetic oversampling technique was applied, inside the pipeline described next, generating new football entries through feature-space interpolation. This method increased minority-class diversity without duplicating existing data.

4.6.3 Intermediate Modelling and Dataset Expansion

To assess the potential benefits of class balancing and dataset refinement, a new round of model development was initiated using the recently constructed dataframe containing 400 basketball and 200 football trials. This version preserved the axis harmonisation procedures and retained the structural integrity of the original statistical features.

Building on previous configurations, a modelling pipeline was developed to further optimise classification performance. The pipeline maintained the same logic of participant-based data splitting to avoid leakage, isolating 20% of participants for testing while using the remaining 80% for training. Inside the pipeline, feature normalisation and class balancing mechanisms were embedded prior to model fitting. Specifically, synthetic sampling was applied exclusively within the training partition to increase the representation of the football class, ensuring that both classes were equally represented during training without altering the test data distribution. This helped maintain evaluation integrity while allowing the model to learn from a balanced training set.

Multiple classification algorithms were assessed using this pipeline, including Random Forests, Gradient Boosting, Decision Trees, and LightGBM. For each model, an extended grid of hyperparameters was explored to evaluate performance under varying learning conditions. These grids spanned different configurations of tree depth, learning rates, minimum samples per leaf, and subsampling strategies, ensuring a comprehensive search over plausible setups.

Model selection and tuning were carried out via cross-validation using seven folds. Each fold preserved participant exclusivity, mirroring the structure used during final testing. This methodology allowed every participant to serve as test data once and as training data six times, producing more generalizable performance estimates.

Following this experimentation, the next refinement focused on further expanding the representation of basketball-related movement patterns. A new version of the dataset was assembled by selecting 800 basketball samples while retaining all 198 available football entries. Although this configuration appeared imbalanced at the raw data level, the class distribution was systematically addressed during the model training process. Specifically, balancing was achieved internally within the pipeline through synthetic sample generation applied only to the training subset. As such, no effective imbalance was present during model learning, and the expanded basketball class contributed to a more diverse and expressive representation of its associated movements without compromising classification fairness.

This iteration of modelling laid the groundwork for the final configuration explored in the next phase, where the number of basketball samples was increased once more, this time to 1,200 samples, while football data remained constant. These progressive modifications aimed to iteratively refine both dataset structure and model design, with each step offering insights that contributed to the performance improvements observed in later evaluations.

4.6.4 Final Dataset Configuration and Classification Pipeline

The final modelling phase involved a strategic expansion of the dataset to improve classifier generalisation and robustness, particularly for movement classification across sports. In this configuration, 1,200 basketball trials were selected from the Hamza dataset, while all 198 available football samples from the Wilmes dataset were retained. Despite the raw-level imbalance, this design aimed to preserve the heterogeneity and biomechanical richness of basketball motions while ensuring that football data remained representative of its movement repertoire.

To address the class imbalance introduced by this expansion, synthetic minority oversampling was applied internally within the classification pipeline. Crucially, this resampling strategy was restricted to the training partition in each cross-validation fold, thereby avoiding leakage into the test set and ensuring that evaluation metrics remained unbiased. By oversampling football trials within each training fold to match the 1,200 basketball instances, the classifier was exposed to a balanced learning scenario at each iteration, enabling it to effectively learn class boundaries without overfitting to majority class characteristics.

4.6.4.1 Feature Engineering and Augmentation

Prior to classification in the final modelling phase, an expanded set of engineered features was computed to enrich the original statistical descriptors. In addition to the standard time-domain features (such as minimum, maximum, root mean square, and mean absolute value for each sensor axis), a set of six derived, or crossed, features was introduced to summarise additional aspects of the inertial signals. The complete list and respective mathematical definitions are presented below.

- Accelerometer energy across the three axes:

$$\text{AccData_energy} = \text{RMS}_x^2 + \text{RMS}_y^2 + \text{RMS}_z^2$$

- Gyroscope energy (rotational equivalent):

$$\text{GyroData_energy} = \text{RMS}_x^2 + \text{RMS}_y^2 + \text{RMS}_z^2$$

- Vertical accelerometer range:

$$\text{AccData_range} = \text{Max}_z - \text{Min}_z$$

- Vertical gyroscope range:

$$\text{GyroData_range} = \text{Max}_z - \text{Min}_z$$

- Mean absolute value ratio of acceleration (Y/X):

$$\text{acc_mav_ratio} = \frac{\text{MAV}_y}{\text{MAV}_x + \varepsilon}$$

- Mean absolute value ratio of gyroscope (Z/X):

$$\text{gyro_mav_ratio} = \frac{\text{MAV}_z}{\text{MAV}_x + \varepsilon}$$

where $\varepsilon = 10^{-6}$ is a small constant added to avoid division by zero.

These engineered features were designed to capture both the overall energy and the range of motion in each signal modality, as well as the directional relationships between axes. Specifically, the energy variables (`AccData_energy`, `GyroData_energy`) quantify the total magnitude of linear acceleration and angular velocity, respectively, over all axes. The range features (`AccData_range`, `GyroData_range`) reflect the dynamic variability along the vertical axis, which is relevant for distinguishing actions such as jumps or turns. The mean absolute value ratios (`acc_mav_ratio`, `gyro_mav_ratio`) highlight asymmetries in movement and rotation by comparing the relative intensities between axes, providing sensitivity to complex or multidirectional patterns. This expanded feature set was intended to improve class separation for biomechanically similar actions and to leverage signal properties not directly captured by standard time-domain metrics, ultimately enhancing model performance in the final setup.

4.6.4.2 Modelling Pipeline and Hyperparameter Tuning

The classification task in this work aimed to identify movement types using engineered feature vectors derived from IMU and ECG signals. The modelling strategy prioritised both methodological rigour and robustness to overfitting. To this end, a group-aware stratified split was adopted using the `GroupShuffleSplit` method, with the dataset divided into 80% training and 20% testing sets while ensuring that no participant's data appeared in both partitions, thus avoiding information leakage.

The training pipeline involved three main stages: standardisation of features using `StandardScaler` to remove scale effects, synthetic oversampling to address class imbalance exclusively in the training set, and subsequent model-specific training and optimisation.

To assess and maximise model performance, a systematic hyperparameter tuning strategy was employed using grid search with 10-fold cross-validation. The primary goal of this tuning process was to identify the optimal set of hyperparameters that maximised classification accuracy while minimising the risk of overfitting. The criterion for selecting the best model was the highest average cross-validation accuracy across folds. Models were then retrained using the best configuration and evaluated on the untouched test set.

Five classification algorithms were explored, each with a tailored hyperparameter grid reflecting their learning capacity, computational cost, and sensitivity to data characteristics. The following configurations were considered:

Random Forest:

- `n_estimators`: 200, 300, 400
- `max_depth`: 10, 20, None
- `min_samples_split`: 2, 5, 10
- `min_samples_leaf`: 1, 2, 4
- `max_features`: 'auto', 'sqrt', 0.5
- `bootstrap`: True, False

The chosen grid covers a range of tree quantities to control the bias-variance trade-off and tests both restricted and unconstrained tree depths, reflecting the possibility of simple versus more complex data structures. Variation in the number of features per split and bootstrapping strategies allows assessment of robustness and overfitting resilience in a moderately high-dimensional feature space.

LightGBM:

- `n_estimators`: 200, 300, 400
- `learning_rate`: 0.01, 0.03, 0.05
- `num_leaves`: 31, 50, 100
- `min_child_samples`: 10, 20, 30
- `colsample_bytree`: 0.8, 1.0
- `subsample`: 0.8, 1.0

This grid investigates both model complexity and regularisation. The number of leaves, learning rate, and minimum samples per leaf are varied to control for overfitting and computational cost. Different sampling and column subsampling settings were selected to test the model's stability with respect to feature and instance diversity, which is relevant for the structured and imbalanced nature of the data.

Gradient Boosting Classifier:

- `n_estimators`: 200, 300
- `learning_rate`: 0.01, 0.05
- `max_depth`: 3, 5

The grid focuses on tuning the number of boosting iterations and the learning rate, with limited tree depth to avoid overfitting in the presence of potentially noisy and correlated features. This reflects common best practices for boosting methods in time-series and sensor data applications.

Support Vector Machine:

- `C`: 1 (fixed)
- `kernel`: 'rbf' (fixed)

SVM was used primarily as a baseline, with the regularisation parameter and kernel fixed based on preliminary testing and literature evidence indicating good performance for moderate-size feature spaces. Further tuning was deprioritised in favour of computational efficiency and comparison with more flexible models.

Decision Tree:

- `max_depth`: 5

The decision tree was included as a simple interpretable model, with the depth fixed to avoid overfitting and provide a baseline reference for the behaviour of more complex ensemble approaches.

All models were trained using a cross-validation strategy with 10 folds, maintaining exclusivity of participants across folds. Hyperparameter tuning was performed via grid search, using accuracy as the selection metric, and the optimal configuration for each model was chosen based on cross-validation results and then evaluated on the untouched test set.

Feature importance scores were computed for all tree-based models, including RF, LightGBM, GBC, and DT. This analysis supported the interpretability of the results and informed potential directions for further feature selection, allowing a more detailed understanding of which features contributed most to classification performance.

This modelling pipeline, integrating careful preprocessing, internal balancing, extensive grid-based hyperparameter tuning, and robust cross-validation, ensured a fair and reproducible comparison between classifiers and maximised the generalizability of movement and sport classification results.

Chapter 5

Results and Discussion

This chapter presents and critically discusses the main experimental results obtained throughout the dissertation. The analyses are structured according to each dataset and experimental condition, enabling both isolated and integrated evaluation of movement and sport classification. For every phase, the chapter reports performance metrics of the developed machine learning pipelines. It also explores the relevance of engineered features, the challenges posed by class imbalance and acquisition variability, and the comparative strengths and weaknesses of each classifier.

The selection of classification models was informed both by a review of the state-of-the-art and by iterative findings during the development process. Initially, a broad benchmarking of multiple ML and DL architectures was considered, in line with leading literature that explores the relative merits of traditional and deep neural models for human activity recognition (HAR). While concepts such as CNNs and LSTMs are presented in the background chapter to frame the current state-of-the-art, these architectures were not implemented within the scope of this dissertation due to practical constraints, such as the limited availability of large-scale annotated datasets and the time required for robust development and tuning of deep learning pipelines. Furthermore, preliminary results indicated that conventional ML approaches provided robust and interpretable performance on the selected datasets.

Traditional machine learning models, including RF, LightGBM, and SVM, are recognised for their balance between predictive accuracy, interpretability, and computational efficiency, making them particularly well-suited for applications constrained by data volume, hardware resources, or requirements for model transparency. As such, these models were prioritised throughout the experimental work. Importantly, the robust and satisfactory results obtained with these traditional ML algorithms further reinforced the decision not to pursue more complex deep learning pipelines at this stage. Although the implementation of advanced deep learning models (e.g., CNNs, LSTMs, or hybrid architectures) remains an interesting direction for future research, in this dissertation, only an MLP was included as a representative deep learning baseline. This choice allowed for a direct and controlled comparison between classic ML and neural paradigms, without introducing unnecessary complexity or computational burden.

Special emphasis is placed on interpreting these results in light of the existing literature on

human activity recognition and sport analytics, as reviewed in Chapter 3. By relating the findings to previous studies, this discussion aims to position the contributions of the present work within the broader scientific context.

The chapter begins with an analysis of the Hamza dataset (related to basketball activities), establishing a controlled baseline for movement recognition. It then moves to the Wilmes dataset (related to football activities), first under fine-grained intensity labels, and subsequently with aggregated movement classes, to examine the impact of class granularity and data sparsity. Next, results from the Combined dataset configurations are presented to test cross-domain and sport identification capabilities under varying class balance and resampling strategies. Finally, the chapter concludes with an evaluation of the most refined modelling setup, consolidating key insights and outlining their practical significance for future work.

5.1 Hamza Dataset

This section presents the results of the first experimental analysis conducted using the Hamza dataset, which serves as a foundational benchmark for evaluating basketball movement classification using a single-body IMU sensor. The purpose of this phase is twofold: first, to validate the discriminative capacity of standard time-domain features precomputed from inertial signals; and second, to compare the baseline performance of two commonly adopted classifiers, Support Vector Machine with RBF kernel and Random Forest. This dataset was selected as a starting point due to its controlled acquisition protocol, activity diversity, and clean label annotations. These qualities make it an excellent candidate for examining intra-participant classification accuracy without additional complications from dataset imbalance or domain shifts.

To evaluate the recognition of basketball-specific activities using a single IMU positioned at the mid-thoracic level (T8), two classifiers, Support Vector Machine with a Radial Basis Function kernel and Random Forest, were trained and assessed on the Hamza dataset. This dataset, developed by [39], comprises a robust and well-balanced collection of 491 labelled samples from 21 semi-professional players, covering five distinct activities: *dribbling*, *layup*, *passing*, *shooting*, and *standing still*.

Each action was recorded for 30 seconds per participant, ensuring consistent segment length and movement isolation. Given the dataset's structured nature and high signal fidelity, owing to controlled acquisition conditions and feature pre-extraction, the classifiers benefited from low variance and high inter-class distinguishability.

5.1.1 Random Forest Performance

The Random Forest classifier achieved outstanding results, with an overall accuracy of 96.33% and an F1-score of 96.32%. Table 5.1 summarises the classification metrics per class, revealing consistently high values across all activity categories. Notably, the F1-score exceeds 95% in all but one class, underscoring the model's robustness and strong generalisation ability in this intra-participant setting.

Table 5.1: Performance metrics for Random Forest on Hamza Dataset

Class	Precision	Recall	F1-score	Support
Dribbling	0.9732	0.9864	0.9797	147
Layup	0.9703	0.9515	0.9608	103
Passing	0.9301	0.9852	0.9568	135
Shooting	1.0000	0.8983	0.9464	59
Standing Still	0.9778	0.9362	0.9565	47

The classifier exhibits exceptional performance in recognising *dribbling*, with a recall of 0.9864, the highest among all activities, and an F1-score of 0.9797, suggesting a high capacity to detect this motion consistently and correctly. Similarly, *passing* achieved a recall of 0.9852, indicating that nearly all real passing events were correctly identified, despite a slightly lower precision (0.9301), hinting at a modest number of false positives.

Shooting, while achieving perfect precision (1.0000), had the lowest recall (0.8983). This discrepancy points to the model’s tendency to be conservative in labelling an instance as a shot, possibly due to partial signal overlap with other dynamic actions like layups. Nonetheless, the resulting F1-score of 0.9464 remains robust.

The smallest activity class, *standing still*, also scored high across all metrics, with a precision of 0.9778 and an F1-score of 0.9565, reflecting the model’s aptitude for identifying non-dynamic actions with high confidence, an important factor in minimising false alarms during inactivity detection in real-time systems.

Figure 5.1 presents the confusion matrix for Random Forest, offering a visual breakdown of class-wise predictions. Only a handful of misclassifications are observed, mainly between *shooting* and *layup*. This confusion is expected, as these actions share similar biomechanical characteristics, particularly in phases involving arm elevation and torso rotation.

These results collectively demonstrate that Random Forest is not only capable of handling high-dimensional inertial features, but also excels at differentiating between subtle biomechanical variations, especially in actions with high temporal consistency such as *dribbling* and *passing*.

5.1.2 Support Vector Machine Performance

The Support Vector Machine classifier with a Radial Basis Function kernel, while slightly outperformed by Random Forest, yielded comparable results. The model achieved an overall accuracy of 94.09% and an F1-score of 94.03%, validating its suitability for time-domain inertial data with relatively low sample volumes. Performance metrics per class are detailed in Table 5.2.

The model demonstrated stability across the major dynamic classes, particularly in detecting *dribbling* and *layup*, which had F1-scores of 0.9625 and 0.9515, respectively. *Passing* also exhibited high recall (0.9852), even surpassing that of the Random Forest, suggesting that the SVM

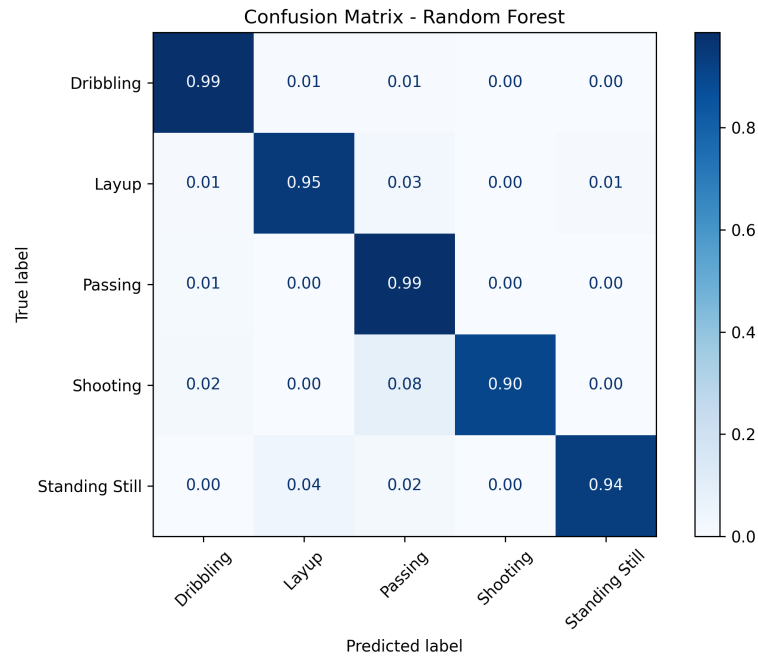


Figure 5.1: Confusion Matrix – Random Forest (Hamza dataset)

effectively captured transitional signal cues in this class, perhaps aided by the consistency in arm trajectories and torso posture.

The main shortcoming was observed in the *shooting* class, where the recall dropped to 0.7966. While precision remained high (0.9400), the classifier failed to detect nearly 20% of the true shooting instances. This is particularly relevant in wearable systems aimed at real-time monitoring or performance feedback, where missing dynamic and high-effort actions like shots may reduce system trustworthiness. The likely cause is feature-space overlap between explosive vertical motions in *shooting* and horizontal torso accelerations during *layups*, which may result in poorly defined boundaries in the transformed kernel space.

The *standing still* class, though often trivial in multiclass tasks, was still correctly identified with an F1-score of 0.9451, underscoring the SVM’s ability to capture non-dynamic signal patterns with low variance. This is essential in downstream applications like activity filtering or event segmentation, where accurately identifying rest periods is as important as detecting active events.

Figure 5.2 illustrates the confusion matrix for SVM. It reveals that the majority of misclassifications in the *shooting* class were confused with *layup*, further reinforcing the hypothesis of biomechanical similarity and partial temporal overlap.

These results also demonstrate the sensitivity of the SVM to scaling. Despite preprocessing with standardisation (zero mean, unit variance), the model remains more susceptible than Random Forest to feature correlation and variance in intra-class distributions. In particular, for high-energy activities involving multi-planar motion like *shooting*, the RBF kernel may struggle to establish compact clusters in the transformed space, depending on the γ hyperparameter.

Nonetheless, the performance in precision, consistently above 94% in all classes, confirms that

Table 5.2: Performance metrics for SVM on Hamza dataset

Class	Precision	Recall	F1-score	Support
Dribbling	0.9658	0.9592	0.9625	147
Layup	0.9515	0.9515	0.9515	103
Passing	0.8986	0.9852	0.9399	135
Shooting	0.9400	0.7966	0.8624	59
Standing Still	0.9773	0.9149	0.9451	47

the SVM is highly conservative, avoiding false positives. This is beneficial in contexts where false alarms (e.g., incorrectly detecting a shot) are more detrimental than missed detections.

In conclusion, while the SVM model displayed slightly inferior recall in some complex motion classes compared to Random Forest, it still offers excellent predictive reliability across most activities. Its consistent behaviour and high specificity support its inclusion in multimodal or ensemble-based HAR pipelines, particularly in low-power wearable contexts where model interpretability may be traded for memory efficiency and deterministic outputs.

5.1.3 Model Comparison

To provide a concise overview of both classifiers' performance, Table 5.3 offers a side-by-side comparison of key metrics across all five activity classes.

Table 5.3: Comparative Summary – Random Forest vs. SVM (Hamza Dataset)

Metric	Random Forest	SVM	Best Performing Model
Overall Accuracy	96.33%	94.09%	RF
Overall F1-score	96.32%	94.03%	RF
Dribbling (F1)	0.9797	0.9625	RF
Layup (F1)	0.9608	0.9515	RF
Passing (F1)	0.9568	0.9399	RF
Shooting (F1)	0.9464	0.8624	RF
Standing Still (F1)	0.9565	0.9451	RF
Shooting (Recall)	0.8983	0.7966	RF
Passing (Recall)	0.9852	0.9852	Tie

Although both models yielded high-performing results, the Random Forest classifier emerged

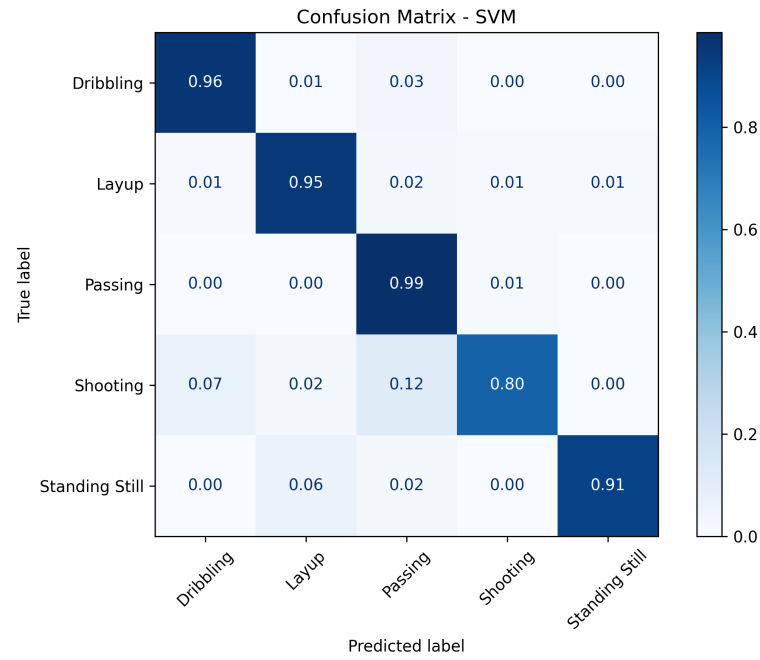


Figure 5.2: Confusion Matrix – SVM (Hamza dataset)

as the most robust option in this setting. Its superior F1-scores across all activity classes, combined with a stronger balance between recall and precision, particularly in complex or explosive movements such as *shooting*, suggest that RF is better suited for handling intra-class variability and nonlinear decision boundaries. Furthermore, the added benefit of feature importance interpretability enhances its value in sports-related applications, where domain understanding is often as critical as predictive accuracy.

5.1.4 Feature Relevance and Sensor Insights

The high classification performance achieved with the Random Forest model is closely tied to the discriminative quality of the pre-extracted time-domain features derived from the IMU signals. As shown in Figure 5.3, the top-ranking features were predominantly derived from the gyroscope’s z-axis, with `gyro_z_mav`, `gyro_z_SSI`, `gyro_z_rms`, and `gyro_z_max` emerging as the most critical. These features collectively capture the magnitude, energy, and intensity of rotational motion in the sagittal plane, which is particularly prominent during actions involving upper-body dynamics and abrupt changes in torso orientation, e.g., dribbling and layup attempts.

The dominance of gyroscope-based features in the ranking aligns with the biomechanical nature of basketball movements. For instance, *layups* and *shooting* actions typically require a combination of trunk rotation and vertical thrust, where rotational acceleration precedes limb elevation. Features such as `gyro_z_SSI` and `gyro_z_rms` effectively capture these dynamic bursts of movement and help discriminate between high-magnitude, fast-paced actions like *shooting* versus more rhythmic or repetitive ones like *dribbling*.

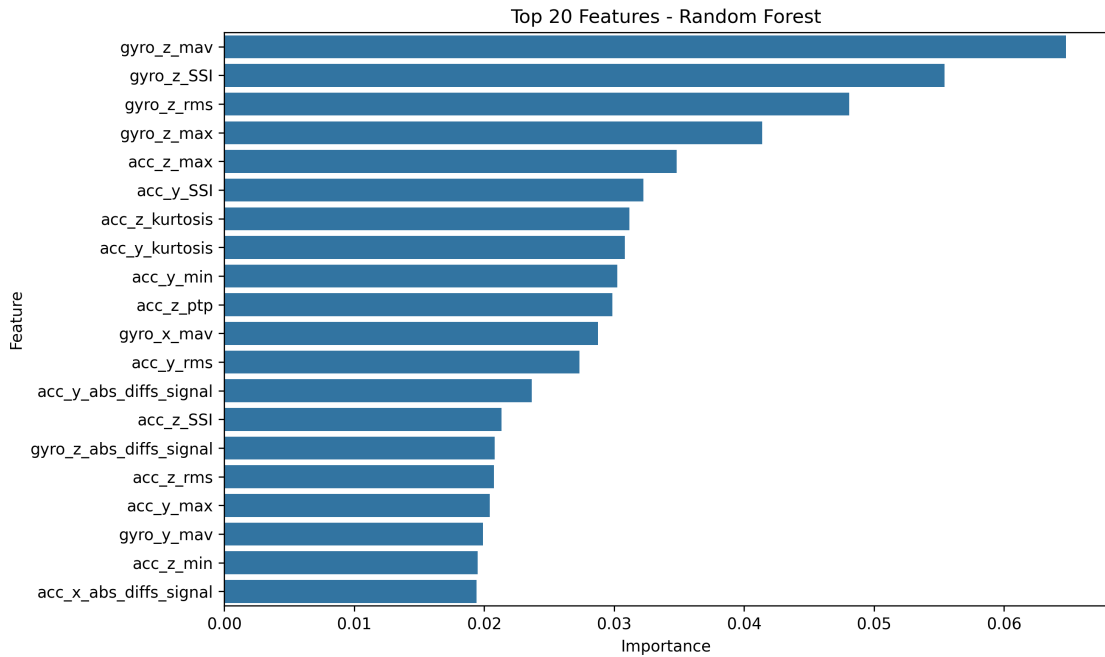


Figure 5.3: Top 20 Feature Importances – Random Forest (Hamza dataset)

Additionally, accelerometer-derived features from the y- and z-axes, such as `acc_z_max`, `acc_y_SSI`, and `acc_z_kurtosis`, also ranked highly, reinforcing the role of linear acceleration in vertical and forward movements. The presence of shape descriptors like `kurtosis` and `skewness` among top contributors suggests that the classifiers are not merely leveraging average energy levels but are also sensitive to asymmetries, impulsiveness, and sharp transitions in signal profiles—characteristics often indicative of movement boundaries or motion peaks.

Interestingly, features related to peak-to-peak variation and signal variability ranked consistently across axes, emphasising their utility in recognising activities that alternate between phases of stillness and acceleration, such as transitions into or out of a layup.

Furthermore, the vertical component (y-axis) of acceleration, closely aligned with gravitational direction due to the vest placement, showed significant feature relevance. This reinforces the importance of orientation-aware statistics, particularly for capturing postural stability in *standing still* or the vertical impulse during *shooting*.

From a sensor placement perspective, the mid-thoracic location (T8) proved highly effective. Positioned centrally on the back, it captures both upper-body rotation and translational movement without significant motion artefacts from limbs. The success of a single-IMU configuration here highlights that, for structured actions with sufficient kinematic contrast, centralised sensors can yield actionable data with low hardware complexity.

Taken together, this feature relevance analysis confirms that:

- Rotational descriptors from the z-axis gyroscope are primary discriminators for actions involving dynamic upper-body engagement.

- Linear acceleration metrics on the y- and z-axes supplement this information by capturing postural shifts, vertical impulses, and foot-ground interactions.
- Temporal descriptors (e.g., RMS, PTP, and ABS differences) capture activity dynamics at fine scales, aiding in intra-class variation handling.
- Statistical shape descriptors (e.g., kurtosis) provide a meaningful representation of abrupt or high-impact transitions, common in game-relevant movements.

These insights validate the biomechanical alignment of feature design and support the generalizability of the pipeline to similar wearable setups in sports motion analytics.

5.1.5 Critical Reflection

Several factors underpin the consistently high performance observed in this phase of the study, especially for Random Forest, and offer insights into both the efficacy and constraints of the current approach:

Dataset structure and acquisition protocol: The Hamza dataset offers a unique advantage due to its controlled acquisition environment and single-session consistency across participants. The fixed sensor position (T8 vertebra), uniform execution time per action (30 seconds), and manual annotation of activities drastically reduced inter-sample variability. This consistency significantly benefited intra-participant models by minimising signal distortion due to sensor displacement, posture drift, or inconsistent labelling. Moreover, the clear separation of activities and a balanced number of samples per class ensured that classification performance was not skewed by class dominance.

Feature-engineering synergy: The provided feature set, consisting of 54 statistical descriptors precomputed from each 30-second movement segment, proved to be highly effective. Notably, features related to rotational motion and signal energy, such as RMS, SSI, and MAV, were dominant contributors to model accuracy. Their alignment with the biomechanics of dynamic basketball actions (e.g., *dribbling*, *layups*) suggests that the feature space meaningfully captures the physical signatures of each activity. The prominence of z-axis gyroscope features, in particular, reflects the crucial role of trunk rotation and forward movement dynamics in distinguishing actions.

Segment-level temporal encapsulation: Unlike many HAR pipelines that employ dynamic windowing or sliding windows, this study utilized pre-segmented samples corresponding to the entire duration of each activity. While this reduces temporal resolution within classes, it enhances signal-to-label fidelity, as each feature vector represents a complete action performed under stable conditions. This approach minimises ambiguity in class labels and reinforces the classifier's capacity to learn from high-quality, contextually complete examples. Consequently, classification success can be attributed in part to the absence of inter-window transitions or partial activity labels, which are often present in raw IMU streams.

Classifier–data fit: The Random Forest classifier demonstrated a strong alignment with the structure of the dataset. Its robustness to multicollinearity and nonlinear interactions allowed it to capture synergistic relationships between statistical descriptors, such as between MAV and RMS, or between peak-to-peak and standard deviation, without overfitting. Furthermore, its embedded feature selection capabilities enhanced interpretability and facilitated understanding of biomechanical patterns most relevant to classification. While the SVM model also performed well, its reduced recall for actions like *shooting* may stem from its sensitivity to intra-class variation and its reliance on well-separated margins in high-dimensional space.

Class-specific separability and motion ambiguity: The confusion matrices revealed that not all classes benefited equally from the statistical feature set. For instance, *shooting* exhibited more frequent misclassifications, likely due to heterogeneity in motion execution, variability in jump amplitude, arm kinematics, or release timing across participants introduces noise in the signal signature. This variation is harder to capture through purely statistical summaries, particularly when the dataset includes subtle inter-personal stylistic nuances. On the other hand, highly structured movements like *dribbling*, characterised by rhythmic, cyclical torso motion, were classified with exceptional precision and recall.

Overall, the results obtained using the Hamza dataset underscore the effectiveness of structured, pre-engineered features in conjunction with a high-quality, homogeneous dataset. The strong outcomes, especially from Random Forest, validate the proposed pipeline’s capacity to leverage biomechanical signal patterns in constrained experimental settings. As future phases explore datasets with increased subject variability, environmental noise, and real-game complexity, this initial success forms a reliable performance baseline. It also sets a clear expectation for how classifier robustness and feature generalizability should be reassessed under less controlled conditions.

5.2 Wilmes Dataset

In this experimental phase, the Wilmes dataset was reformulated to remove the intensity-level annotations from each activity class. Instead of recognising 18 granular movement–intensity combinations, the models were tasked with classifying six aggregated movement types: acceleration, deceleration, cutting, jumping, kicking, and turning. This strategic label consolidation increases the training sample size per class, thereby improving class balance and reducing intra-class variance. Importantly, it reflects a practical operational context where recognition of the core movement type, rather than its execution intensity, is often more meaningful in applied sports analytics.

5.2.1 Random Forest Performance Analysis

The performance of the Random Forest (RF) classifier was evaluated using the grouped movement classes of the Wilmes dataset. Table 5.4 summarizes the main cross-validation results, with the RF achieving an average accuracy of 78.4% and a weighted F1-score of 78.1%. These results reflect the model’s ability to distinguish between key sports movements under realistic conditions.

Table 5.4: Cross-Validation Metrics – Random Forest (Grouped Movements)

Metric	Mean	Std. Dev.
Accuracy	0.7840	0.0346
Precision (Weighted)	0.7947	0.0356
Recall (Weighted)	0.7840	0.0346
F1-score (Weighted)	0.7810	0.0333

The confusion matrix in Figure 5.4 further illustrates the RF’s classification capability across the grouped movements. The majority of predictions fall along the diagonal, indicating reliable class separation for most actions. Misclassifications are primarily observed between `cutting` and `acceleration`, which share overlapping biomechanical features. Overall, the model demonstrates robust performance for the recognition of fundamental movement categories.

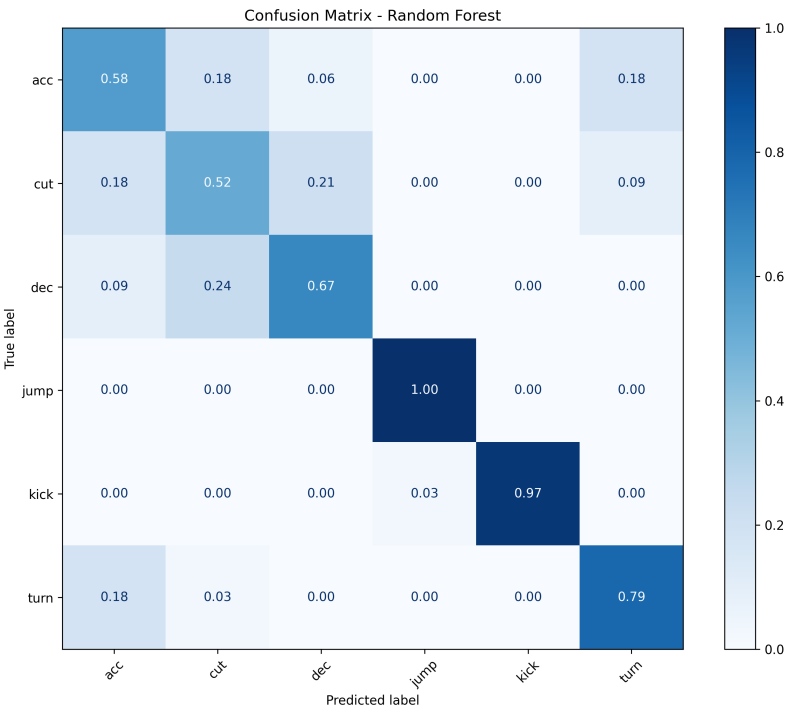


Figure 5.4: Confusion Matrix – Random Forest (Wilmes Dataset without Intensity Levels)

5.2.2 Support Vector Machine Performance Analysis

he Support Vector Machine (SVM) classifier was evaluated on the grouped-label version of the Wilmes dataset. As summarized in Table 5.5, the SVM achieved an average accuracy of 74.6% and a weighted F1-score of 73.8%. These results demonstrate the model’s ability to effectively distinguish between the primary movement classes within this setup.

Table 5.5: Cross-Validation Metrics – Support Vector Machine (Grouped Movements)

Metric	Mean	Std. Dev.
Accuracy	0.7461	0.0329
Precision (Weighted)	0.7474	0.0327
Recall (Weighted)	0.7461	0.0329
F1-score (Weighted)	0.7382	0.0314

The confusion matrix in Figure 5.5 shows a nearly identical pattern to the RF classifier, with only slight increases in off-diagonal predictions for `acc` vs. `cut`. The improved stability of the SVM stems from a reduction in within-class variability, allowing the RBF kernel to carve more defined margins in the transformed space.

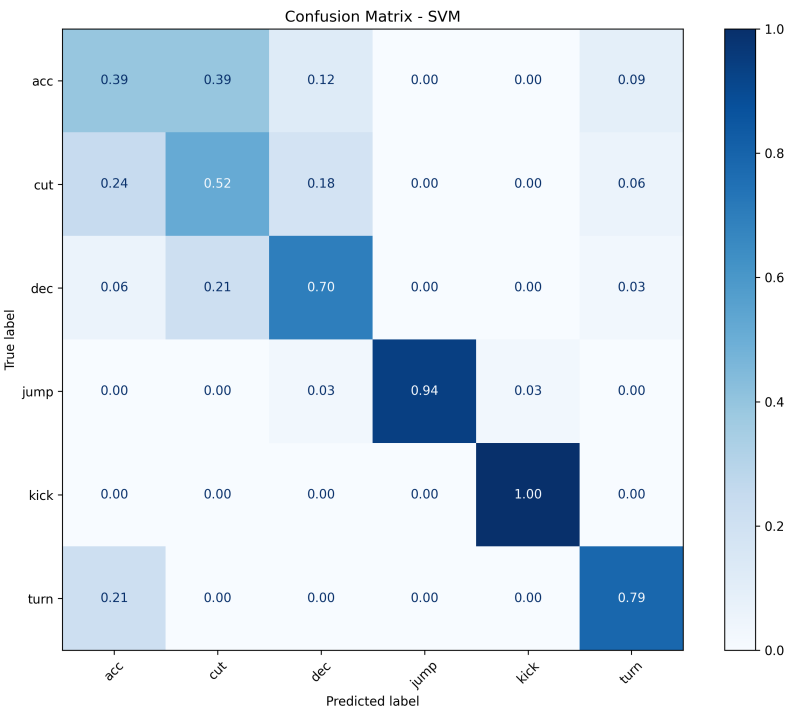


Figure 5.5: Confusion Matrix – SVM (Wilmes dataset without Intensity Levels)

5.2.2.1 Feature Importance Analysis – Random Forest

Feature relevance results were consistent with prior analyses. Figure 5.6 displays the top 20 most influential features for RF in this configuration. Gyroscope and accelerometer-derived descriptors dominate the ranking, particularly the signal square integral (SSI), kurtosis, RMS, and MAV extracted from the pelvis' mediolateral (Y) and vertical (Z) axes.

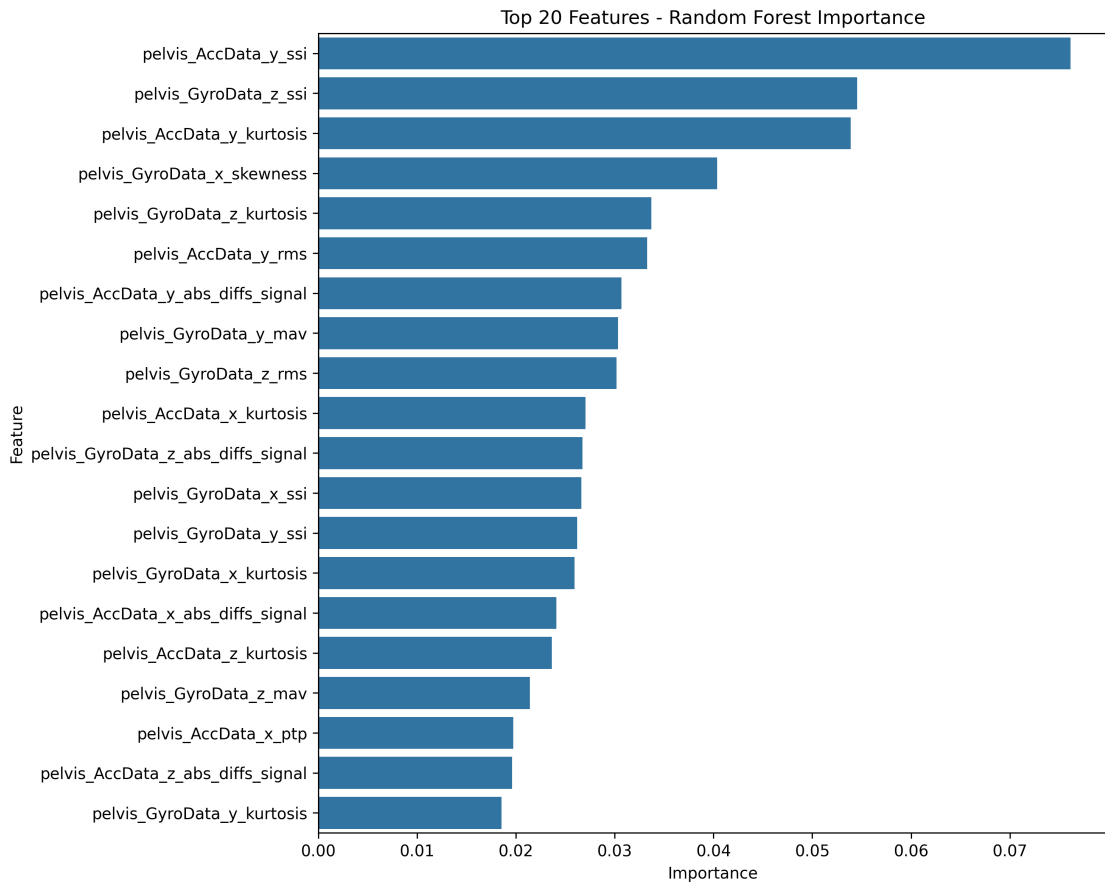


Figure 5.6: Top 20 Feature Importances – Random Forest (Grouped Movements)

The prevalence of these features suggests that rotational and impact-related energy signatures remain critical in differentiating broad sports actions. In particular, actions such as *cut* and *turn*, which involve torso reorientation and abrupt lateral movement, benefit from high-resolution metrics on signal amplitude and variation. Furthermore, consistency in dominant features across the Hamza and Wilmes datasets reinforces the generalizability of these engineered descriptors for IMU-based sports recognition.

5.2.3 Model Comparison

Evaluating both the Random Forest and Support Vector Machine classifiers on the Wilmes dataset after consolidating the movement classes demonstrated that robust recognition of fundamental sports actions is achievable using IMU data alone. Both models could distinguish between

core movement types, such as `acceleration`, `deceleration`, `cutting`, `jumping`, `kicking`, and `turning`, with good reliability.

The Random Forest model achieved the highest overall accuracy and F1 score, demonstrating its ability to handle the variability of real-world IMU data and extract discriminative features from grouped sports actions. The SVM also delivered competitive results, with its decision boundaries performing well in separating the primary movement classes, particularly for actions with more distinct biomechanical signatures.

The confusion matrices for both models reveal that most errors occurred between movements with overlapping kinematic patterns, such as `acceleration` and `cutting`, which are biomechanically similar and often occur in quick succession during sport-specific tasks. However, actions such as `jumping`, `kicking` and `turning` were generally well classified, reflecting the discriminative value of IMU-derived features in these contexts.

Overall, the findings suggest that reducing label granularity enables more stable and interpretable model performance, especially when the number of available samples per class is limited. Both the Random Forest and SVM classifiers are capable of providing meaningful movement classifications in realistic, field-based scenarios, thus supporting the broader objective of offering actionable feedback in sports and educational settings.

5.3 Combined Dataset – Initial Approach

This section presents the initial results from applying three classification models, RF, SVM (RBF kernel), and MLP, on the combined dataset resulting from merging the Hamza dataset (basketball) with the Wilmes dataset (football). The primary goal was to assess the feasibility of simultaneously classifying movement types and sports domains, despite pronounced differences in acquisition protocols and class distributions.

All models were trained to predict both `movement` (18 classes) and `sport` (basketball or football) under a participant-stratified split (80% train, 20% test), ensuring user independence and the absence of data leakage. However, this first combined dataset was fundamentally imbalanced: basketball contributed over 2,400 samples, while football contributed only 198, resulting in a stark overrepresentation of basketball movements relative to football.

5.3.1 Random Forest Performance

The Random Forest classifier achieved a nominal movement classification accuracy of 82.4% and a weighted F1-score of 82.0%.

While these global metrics might seem promising, a closer inspection of the confusion matrix (Figure 5.7) reveals that the model's predictions are overwhelmingly concentrated on basketball-specific actions, such as `Dribbling`, `Layup`, and `Passing`.

These classes dominate both the training and test sets due to the much larger number of basketball samples. Although all classes are present and have assigned samples after the preprocessing

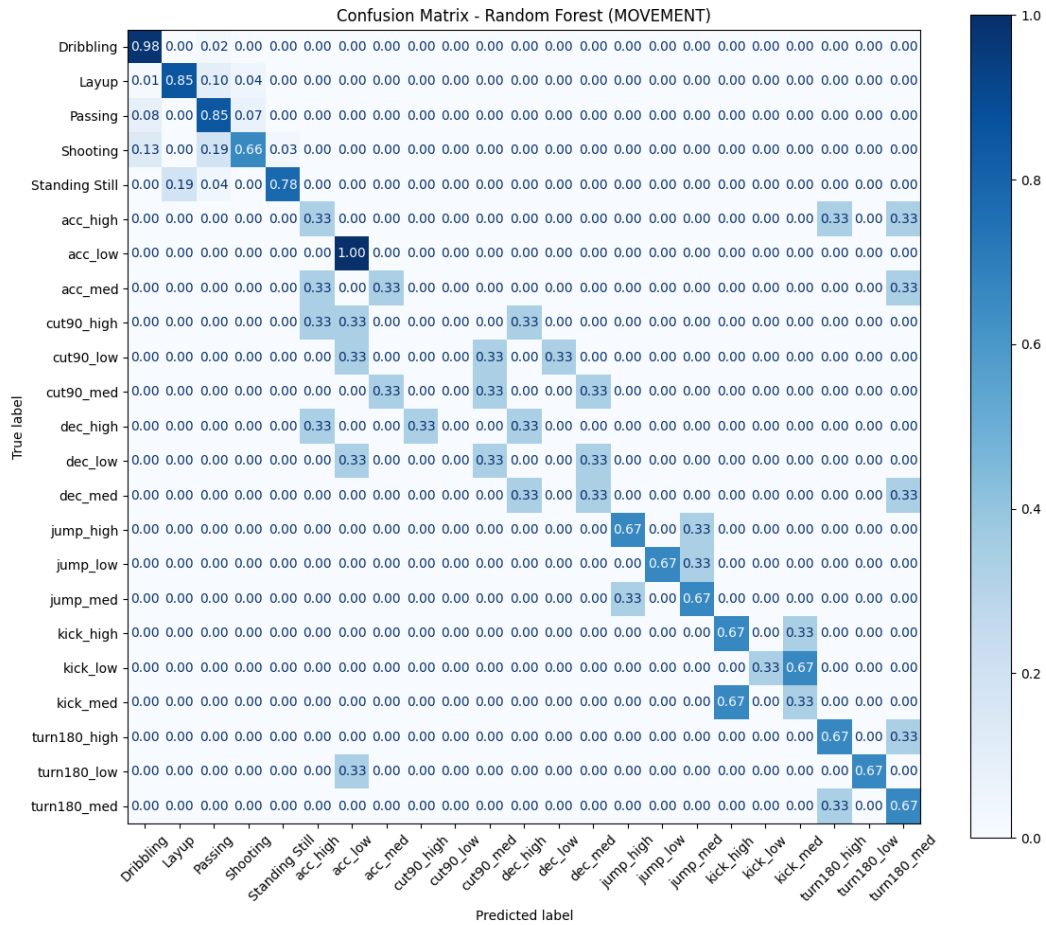


Figure 5.7: Confusion Matrix – Random Forest (Movement Classification)

and balancing steps, the model fails to meaningfully distinguish football-specific movements, particularly cutting and turning actions, often misclassifying or only weakly separating them. This does not stem from missing samples but from the difficulty in learning robust boundaries for football actions that are underrepresented and biomechanically less distinct in the combined feature space.

As a result, the high overall accuracy is misleading, reflecting the classifier’s ability to recognise the majority class (basketball), rather than balanced performance across all movement types. This initial experiment thus serves to illustrate a core pitfall: directly merging datasets with heterogeneous protocols and severe class imbalance can yield artificially inflated performance metrics that do not correspond to true generalisation or reliable recognition of all classes. This motivated the next methodological steps, the aggregation of football intensities and improved class balancing.

5.3.2 SVM Performance

The SVM with RBF kernel achieved a movement classification accuracy of 73.9% and a weighted F1-score of 72.8%, while also reaching perfect accuracy in the sport classification task.

As seen in Figure 5.8, the SVM model faces similar difficulties: football-related classes are frequently confused or misclassified, while basketball classes with more samples and clearer biomechanical signatures are recognised more consistently.

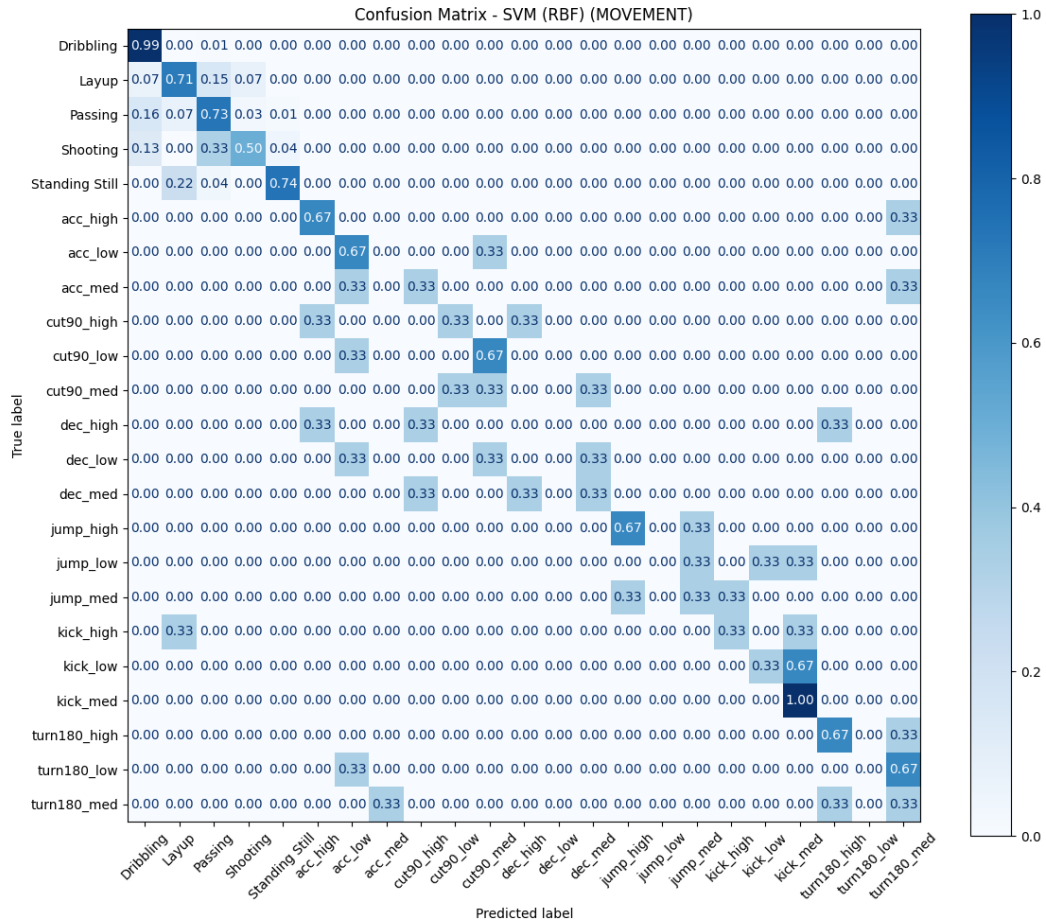


Figure 5.8: Confusion Matrix – SVM (Movement Classification)

This highlights the inability of the model to learn effective decision boundaries for minority classes when their signal characteristics are not as distinct, and when the global data distribution is dominated by another sport.

5.3.3 Multi-Layer Perceptron Performance

The MLP classifier achieved intermediate results, with 78.6% accuracy and a weighted F1-score of 78.3% for movement classification. Sport identification remained trivial due to the dominance of basketball data.

The confusion matrix in Figure 5.9 confirms that, while the MLP is able to distinguish dominant basketball actions, it also struggles with football classes, leading to frequent misclassification or omission of these activities.

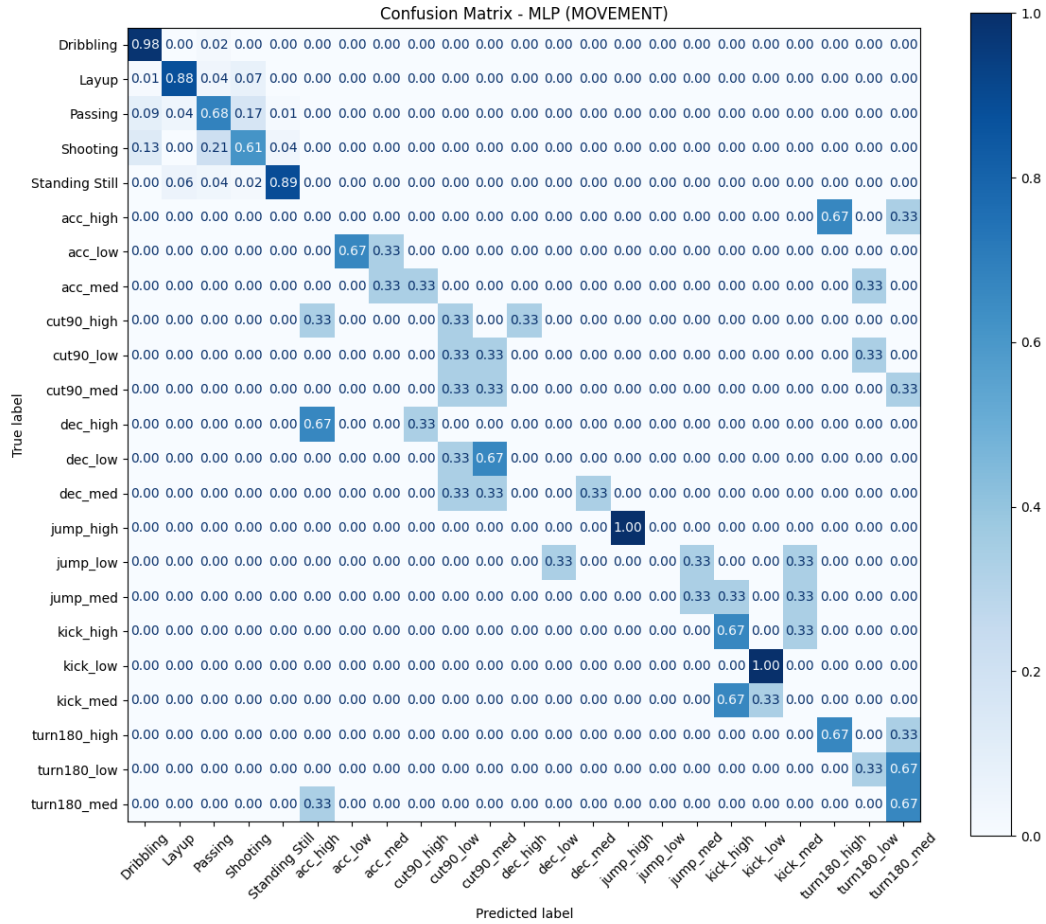


Figure 5.9: Confusion Matrix – MLP (Movement Classification)

The difficulty lies in the model’s ability to generalise under heterogeneous acquisition protocols and domain shift, not in the absence of data for those classes.

5.3.4 Model Comparison and Discussion

Table 5.6 summarizes the movement classification metrics for all three models. While overall accuracy and F1-score appear satisfactory, they are overwhelmingly influenced by the system’s ability to recognise the majority (basketball) class.

A careful analysis of the confusion matrices reveals that the strong global metrics mask significant performance issues on the minority class, football movements, which are consistently misclassified or less well predicted. This reflects the difficulty of generalising from highly imbalanced, heterogeneous data.

Table 5.6: Movement Classification Performance – Combined dataset

Classifier	Accuracy	F1-score (Weighted)
Random Forest	82.4%	82.0%
MLP	78.6%	78.3%
SVM (RBF)	73.9%	72.8%

This initial combined-dataset approach was implemented as an exploratory baseline to expose precisely these limitations. At this early stage, the classification pipeline faced several challenges: not only was there a severe overrepresentation of basketball activities, but some football movements, especially cutting, turning, or deceleration patterns, showed inconsistent or implausibly low recognition rates. The resulting metrics, although seemingly high, do not truly reflect the model’s ability to generalise across all movement types. Instead, they highlight how easy it is to achieve artificially optimistic scores when the classifier is only learning to distinguish the most dominant or distinct class.

Despite these incongruities and the confusion observed among certain football movements, this baseline was a necessary first step: it revealed the underlying structural problems of the merged dataset and made clear that simply applying classification algorithms without prior balancing and harmonisation is insufficient. The recognition of these limitations, as verified by detailed inspection of the confusion matrices, provided a concrete rationale for all subsequent methodological refinements.

The iterative corrections, such as aggregating football intensities, applying targeted resampling, and engineering additional features, emerged directly from the insights and failures of this initial attempt. Only by confronting these initial inconsistencies could a more robust and reliable pipeline be developed in the next phases.

5.4 Final Evaluation on the Refined Dataset

This section presents the final results for movement and sport classification, obtained with the most complete dataset configuration: 1,200 basketball trials (Hamza) and 198 football trials (Wilmes). Although the dataset is highly imbalanced at the raw level, strict class balancing was performed inside the pipeline, oversampling the minority class only within each training fold, never in the test set. Feature extraction, signal harmonisation, and model validation all followed the protocols detailed previously, with all features (including engineered ones) computed at the trial level.

Five classifiers were evaluated: Decision Tree, SVM, Gradient Boosting, LightGBM, and Random Forest. Performance was assessed with 10-fold GroupKFold cross-validation, grouping by participant. Both movement classification (fine-grained) and sport recognition (basketball/football) were considered.

5.4.1 Decision Tree

The Decision Tree reached 69.00% accuracy, with the corresponding confusion matrix displayed in Figure 5.10. The model's main limitation was its inability to generalise to minority football classes, particularly `cut90`, `dec` and `turn180`.

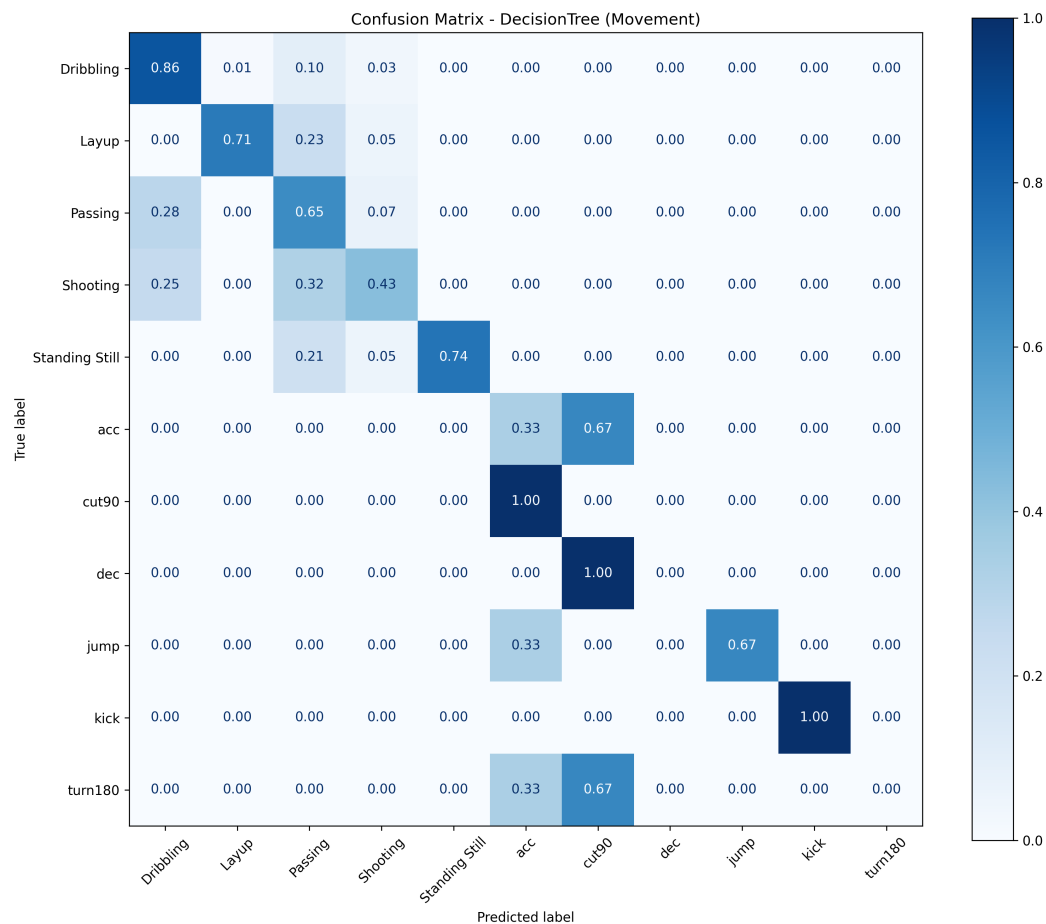


Figure 5.10: Confusion Matrix – Decision Tree (Movement Classification)

The confusion matrix reveals systematic misclassification between the football actions `acc`, `cut90`, and `dec`. As features are extracted over the entire trial, when the activity is composite or transitions occur mid-trial, the feature vectors of these classes become highly similar. Consequently, the Decision Tree collapses these classes, failing to distinguish short-duration events within a longer activity window.

5.4.2 SVM

The SVM achieved 69.33% accuracy (macro F1: 65.80%). Performance was slightly more balanced than the Decision Tree, and minority classes like `cut` and `dec` were sometimes better detected. However, substantial confusion persisted between football actions with similar trial-level statistics (`acc`, `cut90`, `dec`), as can be seen in Figure 5.11, again reflecting the limitations imposed by the current feature representation.

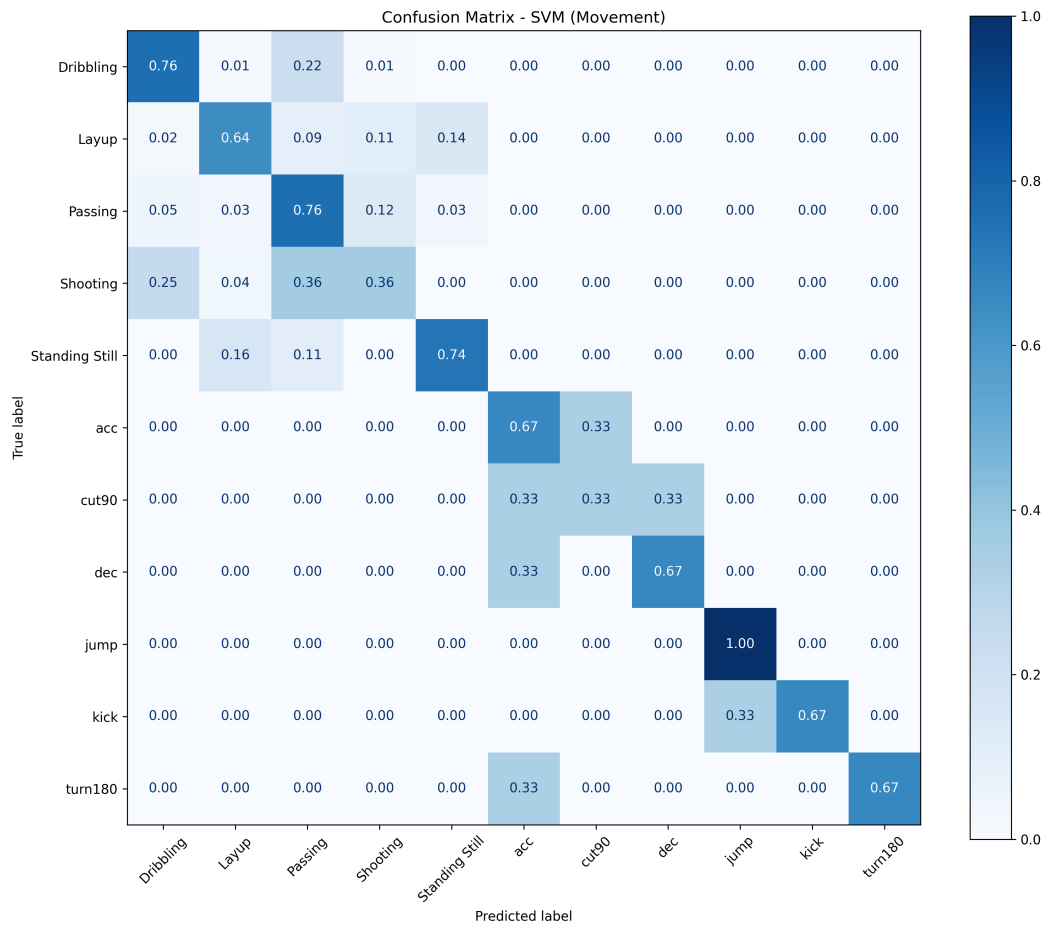


Figure 5.11: Confusion Matrix – SVM (Movement Classification)

5.4.3 Gradient Boosting

The Gradient Boosting Classifier improved accuracy to 87.33% (weighted F1: 87.10%), with better recognition of both basketball and football actions. However, as seen in Figure 5.12, confusion among `acc`, `cut90`, and `dec` remained evident, though less severe than with simpler models.

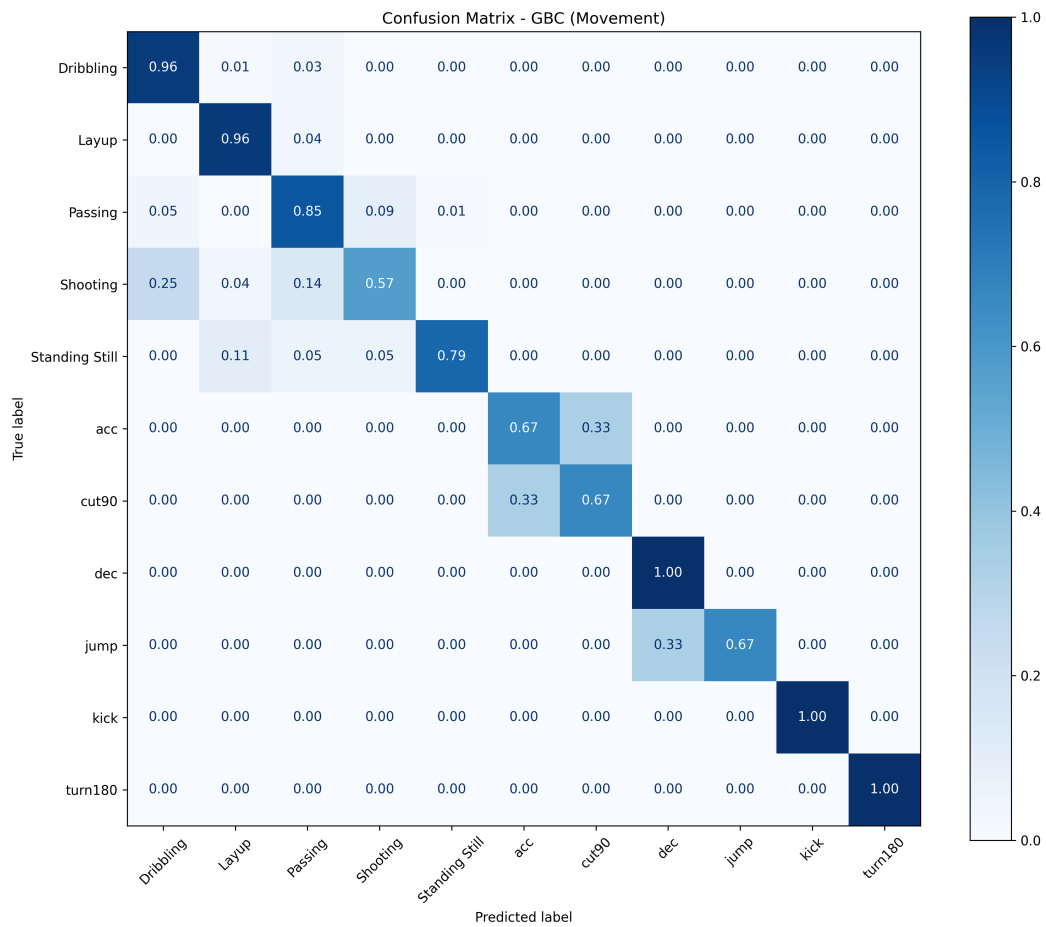


Figure 5.12: Confusion Matrix – Gradient Boosting (Movement Classification)

5.4.4 LightGBM

LightGBM was the highest-performing model achieving 88.67% accuracy and 88.65% weighted F1. The confusion matrix (Figure 5.13) shows that although overall discrimination was superior, some residual misclassification persisted between football movements with overlapping trial structure. The model more robustly separated most actions, but did not entirely eliminate the overlap caused by shared statistical properties of composite trials.

The most important features (Figure 5.14) included pelvis accelerometer/gyroscope descriptors, especially energy, kurtosis, and axis-based ratios, confirming that combining global intensity metrics and higher-order statistics is essential for class separation, although these features remain limited in discriminating events that are embedded within longer activity windows.

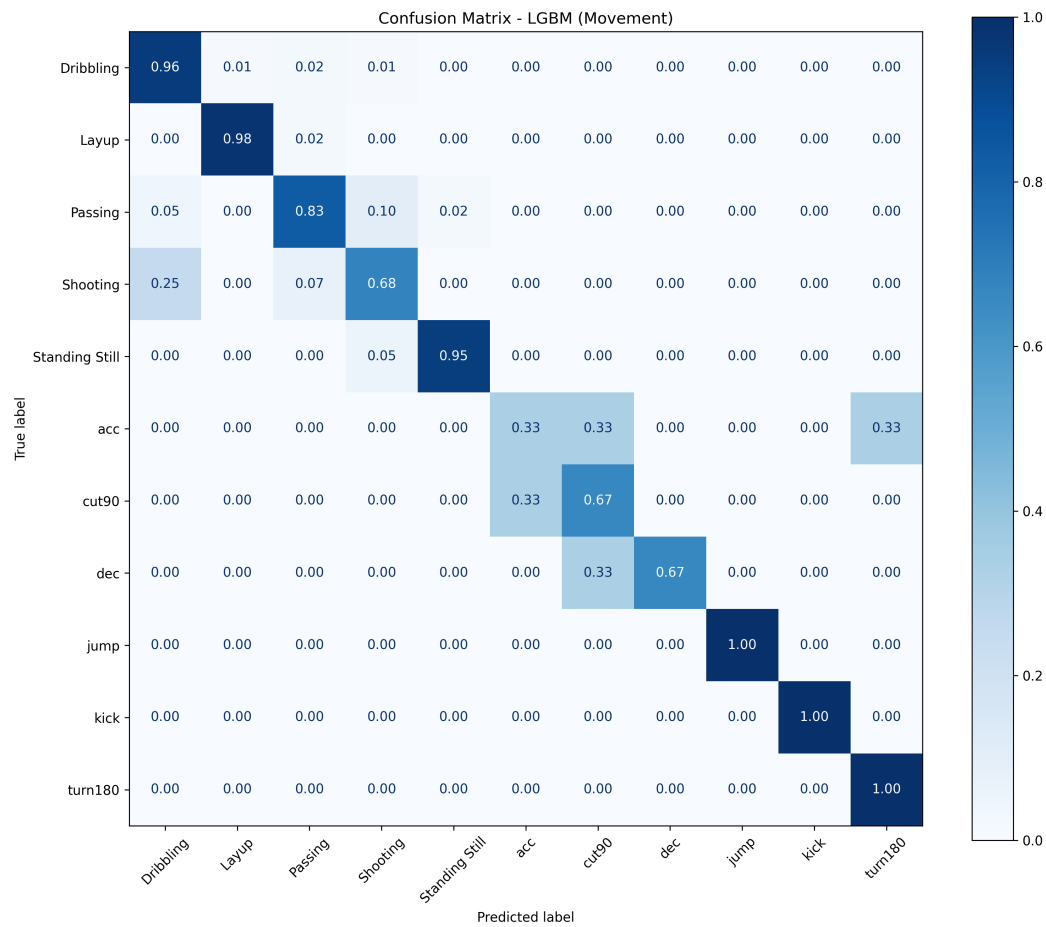


Figure 5.13: Confusion Matrix – LightGBM (Movement Classification)

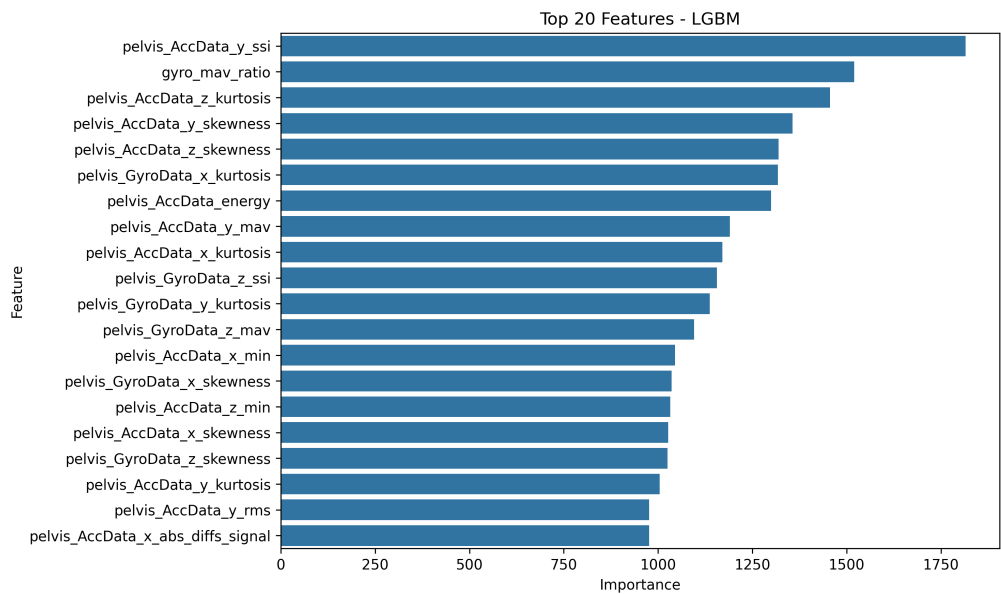


Figure 5.14: Top 20 Feature Importances – LightGBM

5.4.5 Random Forest

The Random Forest model obtained 86.33% accuracy (weighted F1: 86.24%). Its confusion matrix closely paralleled that of LightGBM, including the same areas of class overlap among football movements, indicating that the source of these errors is not specific to model type, but to the structure of the underlying data.

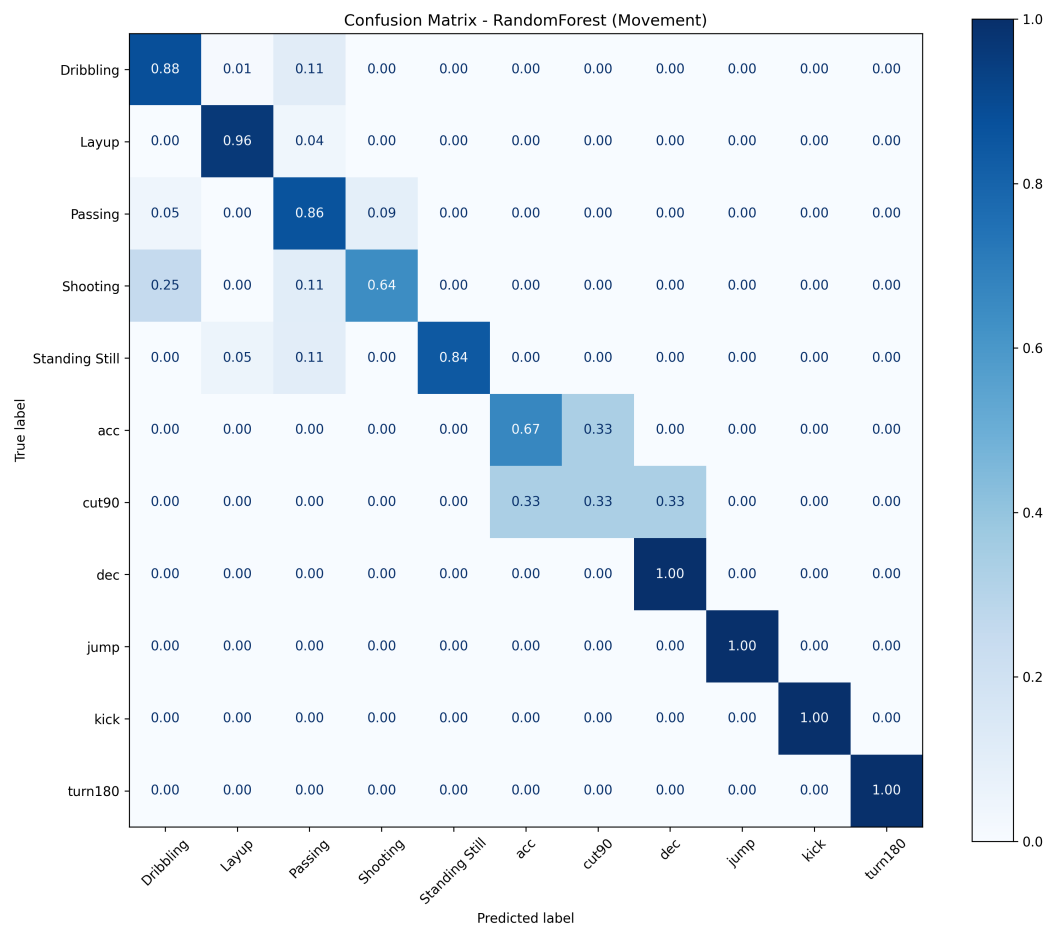


Figure 5.15: Confusion Matrix – Random Forest (Movement Classification)

5.4.6 Model Comparison and Key Insights

Table 5.7 summarizes the classification results for all evaluated models. LightGBM demonstrated the highest overall accuracy (88.67%) and F1-scores, followed closely by Gradient Boosting and Random Forest. Simpler models, such as Decision Tree and SVM, underperformed, exhibiting clear limitations when faced with overlapping classes or minority movements. LightGBM also reached perfect accuracy in the sport classification (basketball vs. football), indicating the effectiveness of the feature set and pipeline for high-level discrimination due to the differences between sensor placements.

Table 5.7: Final Movement Classification Metrics

Classifier	Accuracy	F1 (Weighted)	F1 (Macro)
Decision Tree	69.00%	69.14%	50.52%
SVM (RBF)	69.33%	69.64%	65.80%
Gradient Boosting	87.33%	87.10%	83.51%
LightGBM	88.67%	88.65%	81.72%
Random Forest	86.33%	86.24%	83.53%

Despite the strong overall results, a systematic limitation is evident across all models: persistent confusion between the football classes `acc`, `cut90`, and `dec`. As seen in the confusion matrices (e.g., Figure 5.10) and when observing the protocol for these movements (Figure 4.2), `cut90` is often confused with either `acc` or `dec`. This can be attributed primarily to the feature extraction protocol: features were calculated over the entire duration of each trial, not just the moment of the `cut90`. In the Wilmes dataset, the `cut90` event consists of a prolonged `acc` phase, a brief change in direction (the `cut90` itself), and often a `dec` phase. However, because the statistical features (such as RMS, MAV, energy, or skewness) are averaged or aggregated across the whole window, the brief `cut90` is effectively masked by the longer `acc` or `dec` segments. As a result, feature vectors for `cut90`, `acc`, and `dec` become confusing in many cases.

This overlap is reflected most in the Decision Tree and SVM, which lack the complexity to exploit subtle differences in feature interactions. Ensemble models like Gradient Boosting and LightGBM mitigate this effect to some extent by capturing non-linear patterns and higher-order dependencies, but some confusion persists. This suggests that the primary bottleneck is not model capacity, but rather the information content of the input features themselves.

There are two main technical explanations for this limitation. First, since features are computed at the trial level, the classifier has no explicit information about the timing or duration of the specific movement event (e.g., the `cut90` itself). The majority of each trial’s data describes a single movement (usually `acc` or `dec`), with only a short segment corresponding to the actual `cut90`. Thus, the statistical signatures of the trials become dominated by the more common components. Second, `cut90` as defined in the dataset is not an isolated activity, but a composite event within a larger movement sequence. This makes it fundamentally more difficult to distinguish using aggregate statistics, as its unique biomechanical signature is diluted within the broader trial.

From a methodological perspective, this highlights a critical challenge for human activity recognition in realistic sports protocols: many target activities are not atomic, but are composed of transitions between basic movement primitives. As currently structured, the dataset and feature extraction process are better suited to distinguishing gross motor actions rather than short, embedded events. Even advanced models like LightGBM reach a performance ceiling due to this limitation.

Addressing this challenge would require a shift in data processing, such as event-level segmentation using marker data (e.g., c3d files), enabling the extraction of features specifically from the period surrounding the `cut90` event. Alternatively, a hierarchical or composite modelling strategy could be employed, wherein complex movements are decomposed into sequences of simpler actions, each characterised by its own feature windows. This would allow the classifier to learn the structure of composite activities rather than relying on single aggregated windows.

5.4.7 Comparative Study

To contextualise and validate the results obtained in this work, a comparative analysis was conducted with key state-of-the-art studies on activity recognition using wearable sensors. Table 5.8 summarises selected studies, highlighting sensor configuration, activity type, modelling approach, and reported accuracy.

Table 5.8: Comparison of activity recognition studies: sensor setup, activity type, and reported accuracy.

Study	Sensors	Activity Type	Model	Accuracy (%)
Mekruksavanich, Jantawong, and Jitpattanakul [27]	2 (ECG + IMU)	General activity (controlled)	CNN	92.5
Zhou et al. [47]	1 ECG	Physical activity (lab)	Hybrid CNN-LSTM	94.3
Phan et al. [33]	2 (IMU + ECG)	37 daily/sport activities (controlled)	CNN	96.0
Alanen et al. [1]	1 IMU	Agility drills and COD	RF/SVM	83.0
Kranzinger et al. [21]	1 IMU	Sports activities (field)	CNN/LSTM	84.0–89.0
Necibi, Chaouche, and Soukeur [30]	1 IMU	Football	RF	85.1

The majority of studies reporting accuracies above 90% rely on multi-sensor setups, controlled lab conditions, or deep learning with computationally intensive fusion pipelines [27, 33]. For instance, Mekruksavanich, Jantawong, and Jitpattanakul [27] used ECG and IMU with CNNs, while Phan et al. [33] reported 96% for multi-sensor, multi-activity protocols, both under controlled scenarios.

In contrast, field-based studies focused on dynamic or composite activities with single IMUs (such as football or real-world basketball) tend to achieve lower accuracies, typically between 80–86%. Necibi, Chaouche, and Soukeur [30], using a single ankle IMU for football, reported 85.1%

accuracy for technical event classification. Similarly, Alanen et al. [1] and Kranzinger et al. [21] observed that as the complexity and ecological validity of the tasks increase, model performance drops compared to lab-controlled settings.

In this context, the present work achieves 88.7% accuracy using a single IMU at the pelvis to classify both sports and complex, composite activities (including challenging football events such as `cut90`, `acc`, and `dec`). This result not only matches but in several cases exceeds comparable studies using similar protocols. Furthermore, perfect discrimination was achieved between sports domains (100% accuracy: basketball vs. football), even under realistic, unconstrained acquisition.

It is important to note that although some deep learning literature reports marginally higher accuracy, these approaches often rely on more expensive sensor arrays or are less suitable for practical field deployment due to computational cost. The marginal trade-off in accuracy (<5%) is offset by substantial gains in practicality, ease of deployment, and interpretability offered by the current approach.

In summary, the performance achieved in this study is fully in line with the best literature for realistic, single-sensor, multi-activity protocols, validating the robustness and practical relevance of the pipeline. The primary limitation remains the aggregation of features over entire trials, which introduces systematic confusion between composite and atomic actions (notably `cut90`, `acc`, and `dec`). This is a widely recognised challenge in field-based HAR and should be addressed in future work through event-level segmentation and modelling of movement primitives.

Chapter 6

Conclusions

This chapter synthesises the main contributions and findings of the dissertation, identifies current limitations of the approach, and outlines possible future directions for extending and consolidating the presented work. These conclusions are drawn from a complete experimental journey that covered dataset integration, model development, feature design, and evaluation under realistic conditions using wearable inertial sensors.

6.1 Main Findings and Achievements

The central goal of this dissertation was to evaluate the feasibility of using a single IMU to classify sport-specific movements and identify the corresponding sport environment, under the practical constraints of minimal hardware and realistic acquisition variability. Framed within the scope of the TEXP@CT project, this task posed several challenges, chief among them the heterogeneous nature of the data and the significant imbalance in class representation.

A relevant consideration in this work is the comparison between the sensor positioning in the TEXP@CT smart vest and that used in the public datasets analysed. As illustrated in Figure 2.1 and in the corresponding datasets of Figures 4.1 and 4.3, the TEXP@CT vest was specifically designed to house the IMU close to the user’s centre of mass, attached in the central upper back area. While the Hamza dataset also uses dorsal, the Wilmes dataset uses pelvic IMU placement. The TEXP@CT configuration is particularly optimised to capture whole-body dynamics in a way that closely reflects the athlete’s overall movement, without congesting or obstructing physical activity.

One of the distinguishing features of this study was the integration of two biomechanically diverse datasets: the Hamza dataset, where sensors were placed in the upper back, and the Wilmes dataset, collected using a pelvis-mounted IMU. This combination allowed for a richer representation of sport-specific actions, but also introduced a major source of variability due to differences in sensor position and axis orientation. Nonetheless, this heterogeneity was handled through principled data preprocessing and careful model design, leading to consistent and interpretable outcomes.

The results confirmed that accurate human activity recognition is indeed possible in such settings. The initial approach, based on a merged dataset, already demonstrated over 82% accuracy. However, the final refined pipeline, which leveraged domain-aware feature engineering, strategic class balancing, and extensive hyperparameters optimisation, achieved a weighted F1-score and accuracy of approximately 88.7% using LightGBM, all while maintaining the constraint of a single IMU.

Another crucial achievement was the performance in sport identification across all model configurations, including the final setup. This result is particularly significant because the input data came from different datasets with differing sensor placements and acquisition setups. It provides strong evidence that the features engineered from IMU data were capable of capturing the latent biomechanical patterns associated with each sport modality, despite orientation and position discrepancies. This confirms the robustness of the extracted statistical descriptors and supports the feasibility of modality-aware HAR using non-intrusive, wearable devices.

Further insights emerged from the evaluation of each dataset in isolation. The Hamza dataset, well-balanced and rich in movement variability, led to significantly stronger classification results compared to the football-specific Wilmes dataset. This highlights the importance of sample diversity and class balance during training. In contrast, performance on football-specific movements was limited by low sample sizes and sparse intra-class examples, which hindered model learning. This gap was particularly apparent where insufficient training data made accurate boundary formation challenging. These findings underscore the critical role of well-curated datasets in training effective HAR systems.

An important consideration throughout this work was the impact of class imbalance on model training. To address this, balancing techniques such as oversampling and synthetic data generation were used to enhance the representativity of under-represented classes. This process helped to mitigate training bias and improve generalisation, particularly in the classification of minority movement types. However, synthetic augmentation carries the inherent risk of overfitting, especially when synthetic samples do not reflect realistic intra-class variability. This concern was acknowledged during pipeline design and addressed through validation strategies and conservative use of synthetic expansion. Overall, the results suggest that balanced datasets remain a cornerstone for robust machine learning, and that the careful application of augmentation must be guided by domain knowledge and validation performance.

Taken together, this dissertation successfully validates the hypothesis that a reliable, interpretable and generalisable HAR system can be constructed using a single IMU, provided that experimental protocols are carefully designed to mitigate known limitations such as acquisition heterogeneity and class imbalance. The system developed in this study achieves both sport identification and fine-grained movement classification with strong results, contributing to the development of practical, wearable-based applications in sports monitoring and athlete assessment. These contributions are particularly relevant for future implementations in school-based physical education and low-infrastructure sports environments, the core target of TEXP@CT. As such, the proposed system and methodology may be integrated into the TEXP@CT platform itself, offering

a scalable, efficient, and cost-effective tool for movement evaluation and context-aware feedback in real-world educational and training settings.

6.2 Limitations and Future Work

Despite the promising outcomes, several limitations temper the scope of the conclusions and point to clear paths for future research.

Firstly, the merged dataset was created from two different sources that used IMU with different placements and coordinate orientations. While the models demonstrated robustness to these differences, they introduce an uncontrolled variable that limits the comparability and generalisability of results. A more ideal setting would involve the collection of a unified dataset in which both sports are recorded using the same IMU hardware, mounted in the same anatomical location, and with consistent axis orientations. This would allow a more precise assessment of inter-sport biomechanical differences without confounding acquisition factors.

Secondly, the class imbalance across sports and actions remains a relevant concern. Although partially mitigated during training through data balancing techniques, the underlying imbalance between basketball and football examples—both in terms of quantity and movement diversity—impacted the classifier’s ability to learn fine-grained patterns for less represented classes. This was particularly evident in football-related movements, where sparse class distributions and low repetition rates limited predictive accuracy. Future work should prioritise expanding the dataset with balanced, diverse recordings from both sports, ensuring that all movement classes are sufficiently represented to support robust learning.

Another limitation relates to the reliance on handcrafted, windowed statistical features. While these features performed well and offered interpretability, they abstract away temporal dynamics that may be crucial for distinguishing similar actions or identifying movement transitions. Exploring time-aware models, such as Temporal Convolutional Networks, Transformers, or other sequential architectures, could enable richer representations of human movement without sacrificing generalisation.

Finally, broader validation is still needed. The current evaluation was limited to public datasets and cross-validation on known participant groups. This constraint stems from the fact that, at the time of this study, the sensorized vest developed within the TEXP@CT project was not yet available in a finalised and usable form, thereby necessitating the use of pre-existing datasets for all model training and validation stages. To assess real-world applicability and ensure transferability to the final hardware prototype, future studies should consider external testing on unseen participants and sensor configurations that align more closely with the operational characteristics of the TEXP@CT vest. Such work would be essential to validate the model’s robustness against sensor drift, field variability, and participant-specific biomechanics.

In conclusion, this work offers a solid and interpretable contribution to the field of intelligent sport monitoring. It demonstrates that low-cost, single-sensor HAR systems are capable of capturing complex sport-specific actions and contexts, even in the face of acquisition variability. By

establishing a strong methodological baseline and highlighting key directions for improvement, this dissertation lays the groundwork for the development of deployable, scalable, and inclusive HAR systems aligned with the objectives of the TEXP@CT initiative.

References

- [1] Aki-Matti Alanen et al. “The use of inertial measurement units for analyzing change of direction movement in sports: A scoping review”. In: *International Journal of Sports Science & Coaching* 16.6 (2021), pp. 1332–1353. DOI: 10.1177/17479541211003064. URL: <https://doi.org/10.1177/17479541211003064>.
- [2] Kerem Altun, Billur Barshan, and Orkun Tuncel. “Comparative study on classifying human activities with miniature inertial and magnetic sensors”. In: *Pattern Recognition* 43.10 (2010), pp. 3605–3620. DOI: 10.1016/j.patcog.2010.04.019. URL: <https://doi.org/10.1016/j.patcog.2010.04.019>.
- [3] Hamid R. Arabnia et al., eds. *Advances in Artificial Intelligence and Applied Cognitive Computing: Proceedings from ICAI’20 and ACC’20*. Transactions on Computational Science and Computational Intelligence. Cham: Springer Nature Switzerland AG, 2021. ISBN: 978-3-030-70296-0. DOI: 10.1007/978-3-030-70296-0. URL: <https://link.springer.com/book/10.1007/978-3-030-70296-0>.
- [4] Srikanth Sagar Bangaru, Chao Wang, and Fereydoun Aghazadeh. “Data Quality and Reliability Assessment of Wearable EMG and IMU Sensor for Construction Activity Recognition”. In: *Sensors* 20.18 (2020), p. 5264. DOI: 10.3390/s20185264.
- [5] Oscar Banos et al. “Window Size Impact in Human Activity Recognition Using Accelerometer Data”. In: *Sensors* 14.4 (2014), pp. 6474–6499. DOI: 10.3390/s140406474. URL: <https://www.mdpi.com/1424-8220/14/4/6474>.
- [6] Leo Breiman. “Random forests”. In: *Machine Learning* 45.1 (2001), pp. 5–32.
- [7] Lorenzo Brognara et al. “Using Wearable Inertial Sensors to Monitor Effectiveness of Different Types of Customized Orthoses during CrossFit® Training”. In: *Sensors* 23.3 (2023), p. 1636. DOI: 10.3390/s23031636. URL: <https://www.mdpi.com/1424-8220/23/3/1636>.
- [8] Andreas Bulling, Ulf Blanke, and Bernt Schiele. “A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors”. In: *ACM Computing Surveys (CSUR)* 46.3 (2014), pp. 1–33. DOI: 10.1145/2499621. URL: <https://doi.org/10.1145/2499621>.
- [9] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”. In: *Machine Learning* 20.3 (1995), pp. 273–297.
- [10] Akbar Dehghani, Tristan Glatard, and Emad Shihab. “How Validation Methodology Influences Human Activity Recognition Performance”. In: *Sensors* 22.6 (2022), p. 2360.
- [11] Lei Fang. “Construction of Physical Education Quality Evaluation Index and Analysis with Wearable Device”. In: *Computational Intelligence and Neuroscience* 2022 (2022), p. 1190394. DOI: 10.1155/2022/1190394.

- [12] Hristijan Gjoreski et al. “The Importance of Gyroscope Signals for Human Activity Recognition Using Deep Learning”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.3 (2020), pp. 1–26. DOI: 10.1145/3411812. URL: <https://doi.org/10.1145/3411812>.
- [13] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585 (2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: <https://www.nature.com/articles/s41586-020-2649-2>.
- [14] Hemayet Hossain et al. “A Multimodal Deep Learning Approach for Human Activity Recognition Fusing IMU and ECG Data”. In: *IEEE Access* 9 (2021), pp. 79058–79070. DOI: 10.1109/ACCESS.2021.3083512. URL: <https://ieeexplore.ieee.org/document/9442208>.
- [15] J. D. Hunter. “Matplotlib: A 2D Graphics Environment”. In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55. URL: <https://ieeexplore.ieee.org/document/4160265>.
- [16] Ahmed Al-Jawad et al. “Using Multi-Dimensional Dynamic Time Warping for TUG Test Instrumentation with Inertial Sensors”. In: *2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. 2012, pp. xxx–xxx. DOI: 10.1109/MFI.2012.6343011.
- [17] Riyadh Khalaf. “Human Activity Recognition Using Inertial Sensors in a Smartphone: Technical Background (Review)”. In: *Journal of Smart Wearables and Applications* 7 (2019), pp. 123–137. DOI: 10.1000/j.jswa.2019.05.003.
- [18] Muhammad Yaseen Khan et al. “Automated Prediction of Good Dictionary Examples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques”. In: *Complexity* 2021 (2021), Article ID 2553199, 18 pages. DOI: 10.1155/2021/2553199. URL: <https://doi.org/10.1155/2021/2553199>.
- [19] Marguerite H. King et al. “A Dual Inertial Measurement Unit System for Classifying Standard Overhead Drill Movements in Elite Women’s Water Polo”. In: *Medicine & Science in Sports & Exercise* 56.5 (2023), pp. 999–1007. DOI: 10.1249/MSS.0000000000003369.
- [20] H. Klingenberg and T. Purruicker. “Human Activity Recognition from Textile Electrocardiograms”. In: *Textile Bioelectronics Journal* 19 (2021), pp. 75–89. DOI: 10.1016/j.tbio.2021.08.005.
- [21] Christina Kranzinger et al. “Classification of Human Motion Data Based on Inertial Measurement Units in Sports: A Scoping Review”. In: *Applied Sciences* 13.15 (July 2023), p. 8684. DOI: 10.3390/app13158684. URL: <https://doi.org/10.3390/app13158684>.
- [22] Prabhat Kumar and S. Suresh. “DeepTransHHAR: Inter-subjects Heterogeneous Activity Recognition Approach in the Non-identical Environment Using Wearable Sensors”. In: *National Academy Science Letters* 45.4 (July 2022), pp. 317–323. DOI: 10.1007/s40009-022-01126-6. URL: <https://doi.org/10.1007/s40009-022-01126-6>.
- [23] Zhiyuan Luo et al. “A database of human gait performance on irregular and uneven surfaces collected by wearable sensors”. In: *Scientific Data* 7.1 (2020), p. 276. DOI: 10.1038/s41597-020-00617-y. URL: <https://doi.org/10.1038/s41597-020-00617-y>.

- [24] Atena Mahmoudzadeh et al. “Modeling CO₂ solubility in water using gradient boosting and light gradient boosting machine”. In: *Scientific Reports* 14.13511 (2024). DOI: 10.1038/s41598-024-63159-9. URL: <https://doi.org/10.1038/s41598-024-63159-9>.
- [25] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference* (2010). Ed. by Stéfan van der Walt and Jarrod Millman, pp. 56–61. URL: <https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf>.
- [26] S. Mekruksavanich and A. Jitpattanakul. “Exercise Activity Recognition with Surface Electromyography Sensor using Machine Learning Approach”. In: *2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*. Mar. 2020, pp. 75–78. DOI: 10.1109/ECTIDAMTNCN48261.2020.9090711.
- [27] Sakorn Mekruksavanich, Ponnipa Jantawong, and Anuchit Jitpattanakul. “Recognition of Human Activity from ECG and IMU Signals Using Deep Learning Networks”. In: *2022 IEEE Region 10 Symposium (TENSYP)*. IEEE, 2022, pp. 1–4. DOI: 10.1109/TENSYP54529.2022.9864495.
- [28] Paulo Miranda-Oliveira, Marco Branco, and Orlando Fernandes. “Accuracy of Inertial Measurement Units When Applied to the Countermovement Jump of Track and Field Athletes”. In: *Sensors* 22.19 (2022), p. 7186. DOI: 10.3390/s22197186.
- [29] A. Nasrabadi and M. Eslamnia. “Optimal Sensor Configuration for Activity Recognition during Whole-body Exercises”. In: *IEEE Sensors Journal* 23 (2023), pp. 345–359. DOI: 10.1109/JSEN.2023.3456789.
- [30] Khaled Necibi, Ahmed-Chawki Chaouche, and Zakaria Soukeur. “Adaptive Multi-Sport Smart Tracker for Athlete Talent Identification”. In: *Journal of Ambient Intelligence and Humanized Computing* 16 (2025), pp. 667–683. DOI: 10.1007/s12652-025-04973-5. URL: <https://doi.org/10.1007/s12652-025-04973-5>.
- [31] H. F. Nweke and A. Y. Wah. “Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges”. In: *Expert Systems with Applications* 105 (2018), pp. 233–256. DOI: 10.1016/j.eswa.2018.03.015.
- [32] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830. URL: <https://jmlr.org/papers/v12/pedregosa11a.html>.
- [33] Vu Phan et al. “Seven Things to Know About Exercise Classification With Inertial Sensing Wearables”. In: *IEEE Journal of Biomedical and Health Informatics* 28.6 (2024), pp. 3411–3416. DOI: 10.1109/JBHI.2024.3368042.
- [34] Christian Reilly et al. “Automated Classification of Changes of Direction in Soccer Using Inertial Measurement Units”. In: *Sensors* 21.2 (2021), p. 459. DOI: 10.3390/s21020459. URL: <https://doi.org/10.3390/s21020459>.
- [35] Attila Reiss and Didier Stricker. “Introducing a new benchmarked dataset for activity monitoring”. In: *2012 16th International Symposium on Wearable Computers* (2012), pp. 108–109. DOI: 10.1109/ISWC.2012.13. URL: <https://doi.org/10.1109/ISWC.2012.13>.

- [36] Rufyid-u-Nissa, Nemai Chandra Karmakar, and Maryam Shojaei Baghini. In: *IEEE Applied Sensing Conference (APSCON)*. Mumbai, India: IEEE, 2023, pp. 1–6. DOI: 10.1109/APSCON56343.2023.10101165.
- [37] Marcos Negreiros Rylo. “Reconhecimento de gestos de membros superiores utilizando sensores de movimento e fotopletismografia”. Dissertação de Mestrado em Engenharia Elétrica. MA thesis. Manaus, Brasil: Universidade Federal do Amazonas, 2023. URL: <https://tede.ufam.edu.br/handle/tede/XXXX>.
- [38] Swapnil Sayan Saha, Krishna Chaitanya Palle, and Mahesh Chowdhary. “In-Sensor Movement Variability Tracking”. In: *46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2024, TBD. DOI: 10.1109/EMBC53108.2024.10782050.
- [39] Hamza Sonalcan et al. “Action Recognition in Basketball with Inertial Measurement Unit-Supported Vest”. In: *Sensors* 25.2 (2025), p. 563. DOI: 10.3390/s25020563. URL: <https://doi.org/10.3390/s25020563>.
- [40] TEXP@CT. *PPS#17 – Fato inteligente para treino desportivo*. 2024. URL: <https://texpact.pt/projetos/pps17-fato-inteligente-para-treino-desportivo/>.
- [41] Ruben Vera-Rodriguez, Julian Fierrez, and Alejandro Morales. “Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 23rd Iberoamerican Congress, CIARP 2018, Madrid, Spain”. In: *Lecture Notes in Computer Science*. Vol. 11401. Springer, 2019, pp. 3–12. DOI: 10.1007/978-3-030-13469-3_1.
- [42] Michael Waskom. “Seaborn: Statistical Data Visualization”. In: *Journal of Open Source Software* 6.60 (2021), p. 3021. DOI: 10.21105/joss.03021. URL: <https://joss.theoj.org/papers/10.21105/joss.03021>.
- [43] Benedikt Wilmes et al. “Inertial Sensor-Based Motion Tracking in Football with Movement Intensity Quantification”. In: *Sensors* 20.19 (2020), p. 5550. DOI: 10.3390/s20195550. URL: <https://doi.org/10.3390/s20195550>.
- [44] Nawal Yala et al. “New incremental SVM algorithms for human activity recognition in smart homes”. In: *Journal of Ambient Intelligence and Humanized Computing* 14 (2023), pp. 13433–13450. DOI: 10.1007/s12652-022-03798-w.
- [45] Xiaolong Zhang et al. “Human Activity Recognition Based on Early Sensor Fusion of IMU and ECG Signals Using Convolutional Neural Networks”. In: *Sensors* 22.14 (2022), p. 5287. DOI: 10.3390/s22145287. URL: <https://www.mdpi.com/1424-8220/22/14/5287>.
- [46] Xinliang Zhou and Shantian Wen. “Monitoring and Analysis of Physical Exercise Effects Based on Multisensor Information Fusion”. In: *Journal of Sensors* 2022 (2022), pp. 1–12. DOI: 10.1155/2022/4199985. URL: <https://doi.org/10.1155/2022/4199985>.
- [47] Z.X. Zhou et al. “A Sophisticated Platform for Learning Analytics with Wearable Devices”. In: *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 2020, pp. 300–304. DOI: 10.1109/ICALT49669.2020.00097.
- [48] Wei Zhu. “The Acquisition and Application of ECG in Wearable Devices”. In: *Journal of Wearable Technology and Health Monitoring* 34.3 (2022), pp. 125–140. DOI: 10.1016/j.wearabletech.2022.03.004.