Enhancing Athlete Monitoring:

A Machine Learning Approach to Predicting Ratings of Perceived Exertion and Oxygen Uptake in Team Sports

Dermot Sheridan

MSc

Supervised by Prof Mark Roantree and Prof Niall Moyna

A thesis presented for the degree of Doctor of Philosophy

School of Computing

Dublin City University

July 2025

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Name: Dermot Sheridan

Student Number: 14212510

Date: July 15, 2025

Acknowledgements

First, I would like to express my deepest gratitude to my supervisors, Mark and Niall, for their unwavering support, invaluable guidance, and encouragement throughout this PhD journey. Their expertise, insights, and patience have been instrumental in helping me navigate the challenges of research and achieve this milestone.

I am incredibly grateful to my colleagues, research peers, students, and staff members at the School of Computing, Dublin City University, for creating a supportive and collaborative environment. Your shared knowledge, camaraderie and encouragement have made this journey both productive and enjoyable. Special thanks to Valerio, Aidan, Michael and Robin for their advice, thought-provoking discussions, and shared enthusiasm for our field. I also extend my sincere appreciation to the Centre for Research and Training in Artificial Intelligence (CRT-AI) for their financial and logistical support. I want to thank the athletes, coaches, and sports science professionals who participated in this research, their contributions were essential to the success of this work.

Lastly, to my family, especially my mum, Margaret, my partner, Alex, my sister, Julieanne, and my partner's dad, Con, thank you for your unconditional love, sacrifices, and belief in me. Your support has been my anchor and source of strength. I am deeply grateful for the encouragement you have all given me throughout this journey. To my friends Martin and Mar, thank you for reviewing my text and for your support throughout this process. To everyone who has been part of this journey, directly or indirectly, thank you. This thesis is a testament to your support and belief in me.

Enhancing Athlete Monitoring:

A Machine Learning Approach to Predicting Ratings of Perceived Exertion and Oxygen Uptake in Team Sports

Dermot Sheridan

Abstract

Technological advances have increased data monitoring in sport, offering potential competitive advantages. Wearable technologies, in particular, generate large volumes of data with untapped potential for athlete monitoring. However, practitioners face challenges in managing and interpreting these data. While commonly used descriptively, there is growing interest in leveraging historical data to inform decisions on training loads and performance. This thesis uses training load data from Global Navigation Satellite Systems (GNSS) and Inertial Measurement Unit (IMU) wearable sensors to develop predictive models for Rating of Perceived Exertion (RPE) and Oxygen Uptake (VO_2) in team sports. Feature engineering combined with deep learning was employed to enhance prediction accuracy. Results show that integrating domain knowledge with engineered features improves RPE prediction over traditional metrics. A pilot study also demonstrates that GNSS and IMU data can predict breath-by-breath VO_2, with linear models performing comparably to deep learning approaches. The findings offer practical tools for coaches and sports scientists, supporting more effective load management and non-invasive, evidence-based performance monitoring in elite team sports.

Contents

List of Abbreviations

R^2 Coefficient of Determination. 109, 111, 113

A-V Acceleration-Velocity. 27

ACWR Acute Chronic Workload Ratio. 52, 53, 68, 72, 81

AI Artificial Intelligence. 2–4, 13, 17, 18

Bagging Bootstrap Aggregating. 44, 71, 73, 74, 81, 83, 88

BR Breathing Rate. xii, xv, 61, 62, 141

CK Creatine Kinase. 23

CMJ Countermovement Movement Jump. 23, 24

CNN Convolutional Neural Network. 49, 50, 133, 137, 140

DEXA Dual-energy X-ray Absorptiometry. 52

DOMS Delayed-Onset Muscle Soreness. 23

DT Decision Tree. 17, 21, 43–45, 71, 73–75, 108, 111, 127

ELI External Load Indices. xvi, 5, 20, 21, 66, 67, 69, 70, 81–89, 93

FPW Future Perceived Wellness. 25

GB Gradient Boosting. 25, 28, 45, 71, 75

GBM Gradient Boosting Machine. 21, 108, 111, 113, 127, 146, 147

GNSS Global Navigation Satellite Systems. iii, xii, xiii, xvi, 2, 4–6, 8, 10, 19, 20, 22, 25–29, 31, 33, 34, 38, 39, 44, 50–57, 59, 62, 65, 67–69, 89, 93–99, 102–104, 106, 121, 122, 124, 126–128, 146

GXT Graded Exercise Test. 60

HR Heart Rate. xii, xv, 16, 40, 61, 62, 138, 140–142

IC Individual Characteristics. 21, 94, 105, 106, 125, 126

ILI Internal Load Indices. 21

IMU Inertial Measurement Unit. iii, xii, xvi, 2, 4–6, 8, 10, 26, 27, 29, 33–35, 38, 39, 44, 55–57, 61–63, 65, 129–131, 138, 143, 147

LASSO Least Absolute Shrinkage and Selection Operator. 20

LightGBM Light Gradient Boosting Machine. xii, xiii, 44, 45, 71, 73, 75, 80–83, 86, 88, 108, 111–113, 115

LOSO Leave-One-Subject-Out. 132, 140

LSTM Long-Short-Term Memory. xiv, xv, 34, 49, 129, 133–140, 142, 147

MAD Mean Absolute Deviation. 130, 131, 133, 137

MAE Mean Absolute Error. xv, 28, 72, 75, 109, 111, 113, 125, 132, 133, 136–139, 141

MAS Maximal Aerobic Speed. 16, 52, 56, 68

MEMS Microelectromechanical Systems. 18, 32, 40

ML Machine Learning. xvii, 3, 4, 8–10, 13, 17, 19–22, 25, 26, 28, 29, 33–36, 66, 67, 70, 71, 75, 83–86, 88, 94, 104–106, 111, 113, 121, 129, 130, 132, 134, 136, 141–147

MLP Multi-Layer Perceptron. 48, 133, 140

MLR Multiple Linear Regression. 132, 133, 136–141, 147

MSE Mean Squared Error. 72, 75, 83, 84, 109, 111, 113

NN Neural Networks. 17, 20, 22, 46

PPO Peak Power Output. 24

PPW Presession Perceived Wellness. 25

PRS Perceived Recovery Status. 42, 56, 125, 126, 148

RF Random Forest. 21, 44, 71, 73, 108, 111–113, 127, 146, 147

RFE Recursive Feature Elimination. 106, 127

RFECV Recursive Feature Elimination with Cross-Validation. 21

RMSE Root Mean Square Error. 28, 72, 75, 83, 109, 111, 113, 125, 132, 133, 139

RPE Rating of Perceived Exertion. iii–v, xiii, xiv, xvi, 3–6, 8, 10, 13, 18–22, 27, 28, 35, 36, 41–43, 45, 50, 51, 53–56, 59, 66–68, 70–77, 81–90, 93–95, 105, 106, 109–112, 121, 122, 124, 126–128, 145–148

SmO_2 Muscle Oxygen Saturation. 40

SMP Sequential Movement Patterns. 28

sRPE Session Rating of Perceived Exertion. 12, 21, 42

SVM Support Vector Machine. 17

TCN Temporal Convolutional Networks. 34

TQR Total Quality Recovery. 26

TSfresh Time Series Feature Extraction. vi, xiii, xvi, 27, 94, 95, 99–101, 106, 107, 113, 121, 125–127, 146

TT Time Trial. 52

VIF Variance Inflation Factor. 68, 86

VO_2 Oxygen Uptake. iii, iv, vi, xiv, xv, 3, 4, 6–8, 13, 15, 18, 32, 34–36, 40, 48, 60–63, 128–138, 140–143, 145, 147, 149

VO_2 max Maximal Oxygen Uptake. 15–17, 40, 59, 61, 62

XGBoost eXtreme Gradient Boosting. xii, 25, 45, 71, 73–75, 77, 80, 83, 86, 108, 111, 112, 115, 127, 147

List of Figures

List of Tables

Chapter 1

General Introduction

This dissertation is at the intersection of sports science and machine learning, examining how data-driven techniques can improve the accuracy, efficiency and depth of athlete monitoring. It investigates whether machine learning can accurately model training load data, identify the most critical metrics for performance evaluation and leverage patterns in historical records alongside additional data streams known to influence athletic outcomes. Beyond replicating current practices, the research assesses whether machine learning can refine and enhance these methodologies, ultimately guiding more personalised training prescriptions. In addition, it explores the untapped potential of raw sensor data to produce more granular high-frequency insights, with the goal of predicting the physiological responses of athletes with greater precision than ever before. By integrating principles from data science and sports science, this dissertation transcends traditional disciplinary boundaries, aspiring to establish a more dynamic, data-informed era of decision making in sports training and performance optimisation.

1.1 Introduction

Athlete monitoring has become a cornerstone of high-performance sports, allowing coaches and practitioners to optimise training and reduce the risk of injury. At its core, athlete monitoring involves systematically collecting and analysing data related

to an athlete's training and performance, encompassing both quantitative and qualitative metrics such as training loads, recovery, health, well-being and behaviours (A. J. Coutts et al., 2018; Impellizzeri et al., 2019; Timmerman et al., 2024). These insights inform decision-making processes that improve the effectiveness of training programmes while minimising risks such as overtraining, burnout and injury (A. Coutts, 2019).

The emergence of wearable technology, such as the Global Navigation Satellite Systems (GNSS) and Inertial Measurement Unit (IMU) has revolutionised athlete monitoring by providing detailed real-time data on training loads. Wearable devices enable practitioners to capture a range of movement-related data (Seshadri et al., 2019b). However, estimations of energy expenditure and biomechanics are subject to variability depending on the sensor type and context of use (ODriscoll et al., 2018)(Rana & Mittal, 2021).

Advances in wearable sensors have also facilitated the modelling of dose-response relationships, reducing the need for external tests and supporting real-time fitness and fatigue assessments during training (Helwig et al., 2023). However, the vast amounts of high-frequency multidimensional data generated by these devices are often underused as a result of the absence of advanced analytical frameworks.

Recent technological advances, including cloud computing, wireless connectivity and Artificial Intelligence (AI) have the ability to transform the monitoring of athletes. These innovations have paved the way for sports analytics as a discipline that combines sports science and data science (Passfield & Hopker, 2017; S. West et al., 2024). This interdisciplinary field aims to develop computational tools that address the unique challenges of team sports analysis (Blei & Smyth, 2017; Stein et al., 2017), allowing more objective evaluations of training and performance. Although traditional statistical methods provide a foundation for data analysis (Thornton et al., 2019), many organisations remain data-informed rather than fully data-driven, relying heavily on descriptive analytics that risk confirmation bias. A proposed framework for elite sports analytics categorises analytics into

four stages: descriptive, diagnostic, predictive and prescriptive (Houtmeyers et al., 2021). Moving beyond retrospective analysis, predictive and prescriptive analytics enable forward-looking strategies that optimise athlete monitoring.

AI and Machine Learning (ML) have shown promise as transformative tools in sports analytics, particularly in identifying complex patterns in tactical and performance data (Chmait & Westerbeek, 2021; Claudino et al., 2019). While their use in athlete monitoring is growing, especially through wearable data, their role in physiological monitoring is still evolving and requires further validation (Doherty et al., 2024).. Research has demonstrated their potential to predict injury risk, performance outcomes, and perceived exertion with greater accuracy than traditional approaches (Bartlett et al., 2017; Rossi et al., 2019). ML models excel in integrating and analysing diverse data sources such as external load metrics, internal physiological markers and contextual factors offering a comprehensive view of athlete performance (Geurkink et al., 2019). This capability is particularly valuable for monitoring individual athletes, as ML models can account for the unique interplay of these variables. By streamlining monitoring systems and generating actionable insights, ML enables sports scientists and coaches to personalise training load prescriptions, improve recovery strategies, and enhance overall athlete well-being.

Despite these innovations, significant challenges remain in fully utilising wearable sensor data. Translating high-dimensional data into actionable insights requires sophisticated feature engineering and predictive modelling approaches. This research addresses these gaps by integrating wearable sensor data and machine learning techniques to develop predictive models for key perceptual metrics (Rating of Perceived Exertion (RPE)) and physiological metrics (Oxygen Uptake (VO_2)). Specifically, it aims to determine the key variables that influence perceptual responses, assess the impact of personal and contextual factors on individual training responses, and investigate whether raw sensor data can enhance predictive performance compared to aggregated metrics. Furthermore, this research explores the feasibility of predicting physiological responses, such as VO_2, using traditional and deep learning models

with raw sensor data. Focusing on elite Gaelic football, this work bridges the gap between descriptive monitoring and predictive analytics, providing evidence-based tools to improve training load management and support more effective decision making in team sports.

1.2 Problem Statement

Athlete monitoring is a critical component of high-performance sports, aimed at optimising training processes and reducing the risks of injury, overtraining and burnout. Despite the widespread adoption of wearable technologies, such as GNSS and IMU, and the increasing volume of high-frequency, multidimensional data they generate, many current monitoring systems remain limited to descriptive analytics. These systems primarily offer retrospective information on training loads and athlete performance, failing to provide forward-looking predictive capabilities.

Existing approaches are constrained by several key challenges:

- **Underutilisation of Data:** Wearable sensor data, while detailed and extensive, is often underutilised due to the absence of advanced analytical models capable of extracting actionable insights from raw and complex datasets.
- **Reliance on Descriptive and Aggregate Metrics:** Current methods often rely on aggregated data and subjective assessments, such as the RPE, without adequately accounting for individual variations or contextual factors influencing training responses.
- **Lack of Integration of Advanced Analytics:** The integration of AI and ML into athlete monitoring has been limited. Although these tools have shown potential in other domains, their application in predicting physiological (e.g. VO_2) and perceptual (e.g. RPE) responses remains underexplored.
- **Complexity of Training Load Responses:** The dynamic and individualised nature of the response of athletes to training, particularly in team

sports, requires predictive models that can handle the interaction of external and internal load metrics with contextual and personal variables.

Without addressing these challenges, athlete monitoring systems risk being reactive rather than proactive, limiting their ability to support evidence-based decision making and optimise performance in high-stakes environments.

1.2.1 Hypothesis

Wearable sensor-derived biometric data can accurately predict perceptual responses (RPE) and physiological responses (VO_2) using machine learning models in elite Gaelic football players.

These predictions are made using data from wearable sensors (GNSS and IMU), enriched with contextual information (e.g., training session characteristics) and individual factors (e.g., body composition, positional role).

The integration of granular time-series features and machine learning-based optimisation techniques is expected to support data-driven decision-making in training load planning and athlete monitoring.

1.2.2 Research Questions (RQ)

- **RQ1: Baseline for RPE Prediction.** How do absolute and relative External Load Indices (ELI) compare in their ability to predict RPE using various machine learning algorithms?

- **RQ2: Impact of Additional Variables.** To what extent does incorporating personal characteristics, perceived wellness scores, and contextual factors improve the predictive accuracy of RPE models beyond using ELI alone?

- **RQ3: Time-Series Feature Sets for RPE Prediction.** How do time-series features extracted from raw GNSS data enhance the predictive performance of RPE models compared to traditional models that use only aggregated

metrics? In addition, what insights can these features provide into athlete responses to training?

- **RQ4: Comparing Traditional and Deep Learning Models for Estimating Oxygen Uptake.** To what extent can deep learning models accurately predict VO_2) during simulated team sports activities using detailed GNSS and IMU sensor data, and how does their performance compare to traditional regression models

1.3 Contributions of the Dissertation

This research aims to significant contributions to the field of sports science and athlete monitoring by developing advanced machine learning models and methodologies that enhance the management of training load in team sports. The key contributions are as follows.

1. **Development of Predictive Models for RPE Estimation:**
    - Designed and validated machine learning models to predict RPE using GNSS data.
    - Investigated the relative importance of training volume versus intensity as key predictors of RPE, providing new insights into external load metrics.
    - Demonstrated the value of incorporating additional variables such as individual characteristics, perceived wellness, and contextual factors to improve model performance.
2. **Time-Series Feature Sets for RPE Prediction:**
    - Explored the use of time-series feature extraction techniques to process raw GNSS data, focusing on improving the predictive accuracy of RPE models.

- Validated the effectiveness of time-series feature engineering as a scalable and computationally efficient alternative to deep learning for athlete monitoring

3. **Estimation of Oxygen Uptake Using Wearable Sensor Data:**

   - Developed machine learning models to estimate VO_2 during high-intensity activities using wearable sensor data.
   - Evaluated the effectiveness of different modelling approaches and data representations for physiological monitoring.

1.4 Thesis Structure

This thesis is organised into seven chapters, each contributing to the overarching goal of enhancing athlete monitoring and training load management using advanced machine learning models and wearable sensor data.

Chapter 1: Introduction Introduces the background, motivation, and objectives of the investigation. It presents the problem statement, hypothesis, research questions and key contributions, guiding the reader through the structure of the research.

Chapter 2: Literature Review Reviews key concepts, frameworks and methodologies relevant to athlete monitoring, training load management, and the application of machine learning in sports science. Identifies gaps in the existing literature and establishes the foundation for the research conducted in this dissertation.

Chapter 3: Data and Methods Details the research design, including the common equipment and instruments used, as well as the data collection and preprocessing methods. This chapter also describes the machine learning models employed and provides an overview of the methodological framework underpinning the experiments. The chapter serves as a foundational reference for understanding the data and models used in the thesis.

Chapter 4: Predictive Analysis of RPE Addresses RQ1 and RQ2 by developing baseline predictive models for RPE using aggregated GNSS data. This chapter investigates the roles of external load indicators, individual characteristics and contextual variables in improving the accuracy of the model. It explores whether processed data from device manufacturers sufficiently captures the information needed for accurate RPE predictions, identifying potential limitations and areas for improvement.

Chapter 5: Time-Series Feature Sets for RPE Prediction Addresses RQ3 by investigating the predictive potential of raw GNSS data. This chapter explores the use of time-series features extracted from raw data to estimate RPE. By comparing these time-series features with aggregated metrics, the chapter evaluates the added value of capturing detailed temporal and spatial dynamics to improve predictive accuracy.

Chapter 6: Estimation of Oxygen Uptake ($\mathbf{VO_2}$) Focuses on answering RQ4 by shifting the investigation to an objective physiological response (VO_2) rather than subjective RPE. This chapter explores the use of raw GNSS and IMU data to estimate VO_2, comparing traditional regression models with deep learning techniques. It evaluates the ability of these approaches to capture physiological responses and provides insights into the differences between simple and complex ML methods.

Chapter 7: Conclusion and Future Work Summarises the key findings, emphasises the contributions of the research and discusses implications for athlete monitoring and training load management. It proposes directions for future research, including the exploration of hybrid models and the application of machine learning to new domains in sports science.

This structure provides a cohesive journey through the research, with each chapter building on the last to answer the research questions and demonstrate advances in athlete monitoring, performance optimisation and injury risk reduction in elite team sports.

Chapter 2

Literature Review

In this chapter, the key concepts, methodologies, and research developments that underlie this thesis are introduced. Establishing this background is essential for understanding the subsequent chapters and the research questions they address. The chapter is divided into two main sections: the first section (Section 2.1) examines the evolution of continuous data monitoring in sports, the challenges associated with data collection, interpretation and the relationship between external and internal training load. The second section (Section 2.3) explores the applications of Machine Learning (ML) in athlete monitoring, highlighting both the potential and the current limitations of these emerging approaches.

2.1 Athlete Monitoring

Athlete monitoring can assist in managing training load and informing performance strategies (Ryan et al., 2019). A robust theoretical framework for athlete monitoring involves the quantification of training load and the evaluation of individual responses to the training stimulus (Impellizzeri et al., 2005).

Systematic data collection supports this framework, taking advantage of various technologies and tools designed to monitor and describe the training process in detail (Halson, 2014; Timmerman et al., 2024). Technological advances have expanded the range and precision of tools available for monitoring both internal and external

loads (Bourdon et al., 2017). These innovations enable practitioners to collect data that inform decisions about the planning and delivery of training loads, optimising individual readiness for training and competition (Gabbett et al., 2017). Understanding the dynamic relationship between training load and athlete responses at the individual level is essential to maximise the utility of these data (Impellizzeri et al., 2019).

As the volume and complexity of the athletes' monitoring data continues to grow due to advances in continuous monitoring technologies (S. West et al., 2024), traditional descriptive statistics are often insufficient to identify meaningful patterns (Weaving et al., 2017). Advanced analytical techniques, including data science and ML have emerged as powerful tools to address this challenge.

2.1.1 Framework for Athlete Monitoring: Internal and External Load Measures

Modern athlete monitoring systems rely on a dual framework that integrates *external* and *internal load* measures, providing a comprehensive understanding of training demands and responses. This approach combines physiological, psychological, and performance data, enabling practitioners to assess the effectiveness of training and adapt strategies accordingly (McGuigan et al., 2020; Thornton et al., 2019).

External training load refers to the quantifiable physical work performed by an athlete, often measured through metrics such as distance, speed, and acceleration using tools such as the Global Navigation Satellite Systems (GNSS) and Inertial Measurement Unit (IMU) (Miguel et al., 2021). These metrics provide objective data on the physical demands of training sessions and competitions, forming the "dose" aspect of the training process.

In contrast, the **internal training load** represents physiological and psychological responses to external loads, reflecting the "stress" experienced by the athlete. Internal load is commonly assessed through heart rate monitoring, blood lactate analysis, oxygen consumption measurements and subjective ratings such as the Rat-

ing of Perceived Exertion (RPE) scale (Bourdon et al., 2017). Although external load provides objective measures of physical activity, internal load metrics offer valuable insight into individual variability in training responses (Impellizzeri et al., 2019).

The interplay between external and internal loads is critical for effective load management in training. External load data alone cannot fully capture an athlete's readiness or capacity to perform as it does not account for individual factors such as training status, health or environmental conditions. These contextual elements significantly influence internal load and, consequently, training outcomes (Impellizzeri et al., 2019). Studies have highlighted notable discrepancies between the training intensities prescribed by the coach and the perceived effort levels of the athletes, highlighting the importance of monitoring both external and internal loads to bridge this gap (Inoue et al., 2022).

Due to these factors, accurately estimating the internal load before exercise remains a challenge. For example, two athletes who perform the same external workload may experience vastly different internal responses due to variations in fitness levels, recovery status or environmental stressors (Foster et al., 2017). This variability underscores the need for individualised monitoring approaches to optimise training outcomes.

By combining external and internal load measures within a cohesive framework, practitioners can gain deeper insights into the training process. This integration supports informed decision making allowing adjustments to improve performance, reduce fatigue and mitigate risk of injury (Weaving et al., 2014).

2.1.2 Internal-External Load Relationship

Understanding the relationship between external and internal load is essential to explain why players perceive training sessions or matches as easy, moderate or hard (Wiig et al., 2020). This relationship also serves as a valuable tool for evaluating improvements in player efficiency. For example, an athlete who performs higher ex-

ternal loads compared to a previously lower internal load indicates improved fitness or adaptation (Delaney et al., 2018). Research in team sports consistently reports moderate to significant correlations between external and internal load measures. Metrics such as PlayerLoad and total distance, which reflect external load volume, are strongly associated with internal measures such as Session Rating of Perceived Exertion (sRPE) and variables derived from heart rate (Helwig et al., 2023; McLaren et al., 2018). Similar findings have been observed in elite team sports, including soccer (Gaudino et al., 2015), rugby league (Lovell et al., 2013), and Gaelic football (S. Malone et al., 2020). External load indicators such as total distance covered, high-speed running distances, accelerations, and PlayerLoad a proprietary metric derived from triaxial accelerometer data that quantifies the cumulative rate of change in acceleration across three planes are consistently linked to sRPE (Gómez-Carmona et al., 2019), highlighting their utility in monitoring athlete performance (Casamichana et al., 2013; Gaudino et al., 2015; S. Malone et al., 2020).

Despite these correlations, the complexity of the training load dynamics presents significant challenges for traditional statistical approaches. Methods such as correlation analysis, general linear models and multivariate adjusted within-subject models provide valuable information, but mainly describe associations and lack predictive capabilities (McCall et al., 2017). These approaches also struggle to model the non-linear, interactive and multidimensional relationships inherent in training load data (Weaving et al., 2017). In addition, they often fail to account for critical factors such as the varied training modalities, the characteristics of each athlete and the diverse physiological responses to exercise, limiting their ability to provide personalised information (Helwig et al., 2023; Wiig et al., 2020). To address these limitations, it is increasingly recognised that external-internal load relationships should be analysed at the individual player level. This approach accounts for factors such as training status, personal characteristics and environmental influences, allowing a more nuanced understanding of load responses and their implications for training (Impellizzeri et al., 2019). By tailoring training prescriptions to these individualised

insights, practitioners can enhance training effectiveness and reduce the risk of overtraining or injury.

Given the inherent complexity of these relationships, there is a growing need for advanced analytical techniques capable of modelling non-linear interactions and integrating diverse datasets. Future research should prioritise the development of accurate predictive models that reliably forecast individual outcomes, such as performance and injury risk (McCall et al., 2017). Recent advances in Artificial Intelligence (AI) and ML underscore their potential for analysing training load data in team sports. These approaches provide a robust framework for understanding the intricate interplay between external and internal loads, enabling more precise and individualised load monitoring. Using AI and ML, practitioners can unlock new opportunities to optimise training strategies and improve athlete performance (Claudino et al., 2019).

2.2 Understanding RPE and Oxygen Uptake (VO_2) as Physiological Constructs

Athlete monitoring systems commonly rely on both subjective and objective measures to assess training load and physiological response. In the context of this thesis, RPE and VO_2 serve as the key dependent variables. A clear understanding of these constructs is essential to appropriately interpret the machine learning approaches explored later in the study.

There are two prevailing definitions of perceived exertion (PE), reflecting differing views on its physiological and psychological basis (Halperin & Emanuel, 2020). Borg originally defined PE as the feeling of how heavy and strenuous a physical task is (p. 8), emphasising sensations from the working muscles and cardiorespiratory system during exercise (Borg, 1998). In contrast, Marcora later conceptualised PE as the conscious sensation of how hard, heavy, and strenuous a physical task is, highlighting its cognitive and affective components, particularly linked to breathing

and effort perception (Marcora, 2010, 2019). Although both perspectives converge in their practical applicationoften using Borgs RPE scalethey stem from different theoretical underpinnings. Marcora, for instance, advocates tailoring RPE queries to the activity type to improve measurement accuracy (Halperin & Emanuel, 2020). Additionally, Abbiss et al. distinguish between *effort* (the mental or physical energy required for a task) and *exertion* (the sensation of heaviness and strain), cautioning against the interchangeable use of these terms as it may obscure clarity in research and application (Abbiss et al., 2015).

The concept of perceived exertion as a measurable construct was pioneered by **Gunnar Borg** in the 1970s, driven by the need to quantify an individual's subjective experience of effort during physical activity. Borgs early work focused on creating scales that could reliably capture the intensity of exercise-induced sensations across individuals, even when objective physiological markers such as heart rate or oxygen uptake were not available (Lopes et al., 2022).

Initially, Borg introduced a **7-point category scale**, which he later expanded to a **21-point scale** to provide more nuanced feedback during physical testing. Observing that perceived exertion values often paralleled physiological data, Borg developed the **15-point RPE scale**, ranging from **6 to 20**, with the design intent that the RPE score multiplied by 10 approximates heart rate (e.g., RPE 17 $\approx$ 170 bpm). This relationship made it easier to interpret subjective effort alongside objective physiological measures such as heart rate and VO_2 uptake (Borg, 1998).

To further refine the measurement of exercise intensity and sensory experience, Borg later introduced the **Category-Ratio 10 (CR10) scale**. Unlike the fixed-range 620 scale, the CR10 scale allows for the inclusion of **decimal points** and values exceeding 10, enabling more flexible and **granular reporting** of perceived effort. This was particularly useful in contexts where exertion exceeded traditional limits or required finer discrimination between intensity levels. The CR10 scale was designed to maintain strong correlations with physiological markers, such as heart rate and oxygen consumption, while improving the **sensitivity** of perceived exertion

reporting (Lopes et al., 2022).

Despite the widespread use of RPE scales in sport and exercise science, several methodological concerns remain. These challenges stem from inconsistencies in the definition of perceived effort and the strategies used to measure it (Halperin & Emanuel, 2020). Discrepancies in how perceived exertion is conceptualisedwhether as effort, strain, or a combination of sensory cuescan impact the validity and reliability of RPE data. Moreover, the effectiveness of RPE scales heavily depends on the clarity of instructions provided to athletes. Without precise guidance, individuals may interpret the numerical values differently, leading to measurement error and reduced comparability across studies or applied settings.

Oxygen uptake (VO_2), defined as the volume of oxygen consumed per minute, is a fundamental parameter in exercise physiology. Reflects the rate at which the body uses oxygen during physical activity and serves as a direct indicator of aerobic metabolism. VO_2 is widely considered a key marker of cardiorespiratory fitness, playing a central role in both clinical evaluation and sports performance diagnostics (Guazzi et al., 2017).

Direct measurement of VO_2 is typically performed via open-circuit spirometry, which calculates the difference between inspired and expired oxygen concentrations. Alternatively, it can be estimated using the Fick principle, which considers cardiac output and the arteriovenous oxygen difference (Ferretti, 2015). These laboratory-based methods represent the gold standard but are often impractical in applied sport settings due to their cost, complexity, and requirement for controlled testing environments.

In sport science, VO_2 valuesparticularly Maximal Oxygen Uptake (VO_2 max)are frequently used to guide exercise prescription and monitor training load. VO_2 max represents the maximal rate of oxygen consumption and is positively associated with endurance performance. Elite athletes typically demonstrate significantly higher VO_2 max values, reflecting superior aerobic capacity and enhanced recovery potential between high-intensity efforts (Srivastava et al., 2024). However, its relationship

with performance outcomes is context-dependent. For example, while VO_2 max moderately correlates with repeated sprint performance, locomotor variables such as sprint speed are often stronger predictors (Buchheit, 2012).

Due to the limitations of direct measurement, field-based tests have been widely adopted to estimate VO_2 max. The Yo-Yo Intermittent Endurance Test (YYIET) and the Bronco 1200 m shuttle run are two commonly used protocols. The YYIET has shown strong correlations with VO_2 max in rugby players (Sulaiman et al., 2012), although its validity is reduced in youth and female populations (Castagna et al., 2006; Martínez-Lagunas & Hartmann, 2014). Similarly, the Bronco test showed moderate correlation with maximal aerobic speed (Maximal Aerobic Speed (MAS)) in rugby players ($r = 0.73$) but tended to underestimate VO_2 max (Bennett et al., 2024). Research has further shown poor concordance between YYIR1 test estimates and directly measured VO_2 max in elite soccer players (Cabrera Hernández et al., 2018). These findings highlight the need for sport-specific validation when applying field tests in high-performance contexts.

Recent advances in wearable technology have enabled the estimation of VO_2 max through integrated sensors including HR, GPS, and accelerometry. Consumer-grade devices such as the Garmin *fnix* series, Apple Watch, and Polar Ignite use proprietary algorithms to predict VO_2 max during exercise. Algorithm type plays a crucial role in estimation accuracy: exercise-based algorithms generally yield lower systematic and random error compared to resting-based methods (Molina-Garcia et al., 2022).

Accuracy varies by device and population. For instance, the Garmin *fnix* 6 demonstrated higher accuracy in athletic populations, with a mean absolute percentage error (MAPE) of 6.85%, whereas the Apple Watch Series 7 showed a MAPE of 15.79% in general populations (Carrier et al., 2023; Caserman et al., 2024). The context of use also matters: the Garmin Forerunner 245 showed improved accuracy after repeated assessments, reducing its MAPE from 5.58% after one run to 1.06% after two runs (Düking et al., 2024).

Despite their accessibility, wearable-derived VO_2 max estimates are influenced by device-specific algorithms, user characteristics, and activity patterns. Studies comparing different wearables have demonstrated varying degrees of validity, with some devices overestimating or underestimating VO_2 max depending on the use case (Freeberg et al., 2019; Hernandez et al., 2023). These findings underscore the importance of validating wearables for specific populations and cautioning against overreliance on single-session estimates for training decisions.

2.3 Machine Learning

This section provides an overview of key ML principles, focussing on their application to team sports data. Traditional approaches to sports analytics rely on predefined rules and structured data to solve specific problems. In contrast, ML, a branch of AI, adopts a data-driven approach to uncover patterns and relationships in high-dimensional data. ML offers flexible methods for making predictions and improving decision-making for athletes (Blei & Smyth, 2017). During the past two decades, AI and ML have significantly transformed sports analytics, improving insights into both performance metrics and injury risk (Chmait & Westerbeek, 2021). ML encompasses a variety of algorithms that enable systems to learn from data and improve their performance over time without requiring explicit programming (Bodemer, n.d.). These algorithms are generally classified into *supervised learning*, which uses input-output data pairs to build predictive models (Singh et al., 2016) and *unsupervised learning*, which relies solely on input data to identify patterns or clusters and inform data-driven decisions (Bishop, 2006).

In team sports, a systematic review identified eleven distinct AI techniques applied in twelve sports, with the most widely used methods being Neural Networks (NN), Decision Tree (DT) classifiers, and Support Vector Machine (SVM) (Claudino et al., 2019). Approximately two-thirds of these studies focused on improving sporting performance, while the rest addressed injury risk assessment. This highlights the growing interest in the use of AI and ML to optimise player performance and

mitigate injury risks in team sports. AI tools have the potential to revolutionise the prescription of training load by providing coaches and practitioners with data-driven insights to optimise training plans and manage athlete workloads. By analysing historical and real-time data, these tools can tailor individual training loads, considering variables such as personal characteristics and environmental factors. This level of personalisation not only enhances training results but also reduces the risk of overtraining, contributing to improved athletic performance (Bodemer, n.d.).

2.4 Understanding Athlete Responses to Training Load

Understanding how athletes respond to training is a fundamental aspect of sports science, essential to optimise performance and reduce injury risk. This involves evaluating internal parameters, which measure the physiological response of the body to exercise, in four key areas as outlined by Helwig et al. (2023). Their framework emphasises internal load during exercise, evaluated through metrics such as heart rate indices, RPE, VO_2, and lactate levels, to gauge the intensity of exercise. Long-term adaptations such as improvements in VO_2 max, evaluate how consistent training enhances aerobic capacity. Evaluating current health and fitness status, shaped by factors such as age, gender and genetics establishes a baseline for tailoring training interventions. Crucially, the framework integrates internal parameters with external load data collected through wearable devices, such as Microelectromechanical Systems (MEMS), enabling a deeper understanding of how physiological responses, such as heart rate or RPE, correlate with objective movement and effort data. This integration supports the growing interest in continuous real-time monitoring of internal load during training and match play, facilitating the transition from laboratory-based assessments to practical field applications. This review synthesises empirical studies and methodological frameworks from the last decade, focusing on machine learning applied to internal-external load relationships and

subjective athlete responses.

In the following sections, the ability of ML is explored to enhance our understanding of the relationship between external wearable sensor data and athlete physiological responses. In Section 2.4.1, the role of ML in predicting athlete responses is examined, with a focus on planning training loads through RPE. This is followed by an exploration of recovery status indicators after exercise in Section 2.4.2. Next, in Section 2.4.3, the impact of feature engineering on wearable sensor data to improve the accuracy and reliability of these predictions is reviewed. The potential of ML to estimate fitness status is outlined in Section 2.4.4, specifically through the prediction of oxygen uptake from wearable sensor data, as detailed in Section 2.4.5. Together, these sections provide a comprehensive approach to athlete monitoring and performance optimisation.

2.4.1 Machine Learning Applications in RPE Prediction

RPE has emerged as a widely used proxy of internal load in sport science due to its simplicity and cost-effectiveness. However, predicting RPE from external load data remains challenging due to inter-individual variation and the limitations of linear modelling. Machine learning offers an opportunity to model non-linear, multivariate relationships between athlete workload and perceived exertion.

Accurately prescribing individualised training loads remains a significant challenge in team sports. Athletes often perceive training sessions as more or less intense than the practitioners intended, which can lead to maladaptation or suboptimal performance if not effectively managed (Brink et al., 2014). Despite the recognised differences in athletes' responses to the same external training loads (Impellizzeri et al., 2019), physical training is typically planned around external load variables obtained from GNSS in team sports (Ravé et al., 2020). Individual variations in physiological and psychological responses can result in vastly different outcomes among players (T. Gallo et al., 2015; S. Malone et al., 2020). This variability underscores the need for more sophisticated tools to address inter-individual differences. Predicting

how GNSS data and additional factors influence the RPE is essential for examining athletes' responses. This approach can help in accurately prescribing individualised internal loads by manipulating GNSS variables to drive and promote adaptation.

ML techniques have demonstrated superior performance in predicting athletes' responses to training load compared to traditional methods, enabling more personalised load monitoring (Bartlett et al., 2017; Jaspers et al., 2018). One of the first applications of ML in team sports was conducted by Bartlett et al. (2017), who used NN to predict RPE, highlighting the need for an individual approach to training load monitoring. Their experiments highlighted the ability of GNSS variables such as session distance and speed to predict RPE. However, the exclusion of key GNSS metrics like accelerations and decelerations, along with the absence of individual factors such as fitness or psychological status, limited its application to intermittent team sports. While NN improved predictive accuracy, their black-box nature hindered interpretability for practical use.

Further research sought to address these limitations by analysing a broader range of GNSS variables and incorporating individual characteristics to enhance personalisation. Employing more interpretable ML techniques, such as feature importance analysis, could make findings more actionable while maintaining accuracy. The study by Bastiaansen et al. (2020) provided a foundational framework, but further investigation was needed to maximise its practical utility.

Based on this, Jaspers et al. (2018) incorporated a larger set of external load indicators (67 External Load Indices (ELI)) and compared the predictive performance of the ANN and the Least Absolute Shrinkage and Selection Operator (LASSO) models. By including additional variables, such as accelerations and decelerations, they addressed some limitations of earlier studies. LASSO outperformed the neural network, achieving a 29.8% reduction in mean absolute error over a naive baseline. This demonstrated that simpler, interpretable models can outperform complex ones when variable selection and multicollinearity are well-managed. However, the model still lacked individual characteristics such as fitness or recovery status, limiting its

ability to personalise training prescriptions.

Advancing the field, Geurkink et al. (2019) modelled the sRPE using Gradient Boosting Machine (GBM), integrating variables beyond ELI. This research was among the first ones to incorporate Internal Load Indices (ILI), individual characteristics, and supplementary variables into predictive modelling. ELI accounted for 61.5% of the model's predictive power, with the remainder from supplementary variables. However, player-specific markers contributed only 17%, raising concerns about overfitting and generalisability. ILI and Individual Characteristics (IC) were minimally predictive (1.0% and 4.5%, respectively). The limited range of extreme sRPE values further restricted the models applicability.

Rossi et al. (2019) made significant progress by incorporating contextual factors, workload metrics, and individual characteristics to predict RPE and sRPE. It was the first study to compare the predictive performance of RPE and sRPE, revealing that training volume, particularly total distance, was the strongest predictor, outperforming intensity-based metrics such as high-speed running. Aggregated sRPE predictions were more accurate than RPE due to their reduced variability. Contextual factors, such as proximity to the match and psychological stress, also significantly influenced perceived exertion. Acute workloads were more strongly associated with RPE than chronic workloads or metrics such as the acute-to-chronic workload ratio. A strength of this study was the use of Recursive Feature Elimination with Cross-Validation (RFECV) for feature reduction, focussing on relevant predictors. However, the absence of correlation analysis may have inflated the importance of multicollinear variables, reducing the models robustness. The models also struggled with extreme RPE values, suggesting a need for methodological refinement and larger datasets. Although several ML methods were tested, including DT, Random Forest (RF), and Support Vector Regression, the results focused mainly on ordinal regression, limiting insight into model comparisons.

Marynowicz et al. (2021) emphasised model interpretability and individualisation, demonstrating the use of DT models in football. DT models provide clear vi-

sualisations and interpretable relationships between predictors and outcomes. They found that individualised models outperformed group-level models in predicting RPE, reflecting the variability in physiological and psychological responses among young athletes. The model prioritised high-speed running per minute as the most important predictor, contrasting with previous studies that emphasised volume metrics. However, the rationale for de-emphasising volume was not clearly justified, potentially overlooking its relevance in cumulative load and recovery, particularly in youth athletes.

Together, these studies highlight the evolution of ML applications in predicting RPE using GNSS data, highlighting advances and challenges. Early studies used NN to emphasise individualisation, but often excluded key variables and lacked integration of individual characteristics. Subsequent research incorporated a wider range of indicators and compared different ML models, finding that simpler models can outperform complex algorithms when effectively managed. However, gaps remain in the integration of relative load measures, contextual factors, and individual characteristics. Models often struggle with generalisability and predicting extreme RPE values due to overfitting and data set limitations. Building on these advancements, this study in Chapter 4 will address these limitations by developing predictive models tailored for elite Gaelic football. Absolute and relative GNSS measures along with individual, contextual, and perceived wellness variables will be incorporated to improve RPE prediction and optimise training load monitoring. In addition, a wide range of classification and regression models will be compared to identify the most effective algorithms for our specific context.

2.4.2 Machine Learning Applications in Predicting Recovery Status

Monitoring athlete fatigue and recovery is crucial to devising effective training schedules and optimising performance. The recovery process is complex, influenced by intrinsic factors such as training status, age, and gender, as well as extrinsic factors

such as match results and opponent quality. This complexity leads to significant variability in the recovery time between players (Nédélec et al., 2012, 2013). Despite the popularity of subjective wellness questionnaires, few studies have successfully integrated these with objective training load data for predictive purposes. ML approaches, by modelling these multidimensional datasets, can bridge this gap and support more individualised recovery monitoring.

A systematic review by Doeven et al. (2017) highlights the necessity of understanding recovery profiles in team ball sports such as soccer, rugby, handball, basketball and Australian rules football. These sports involve high-intensity, intermittent activities that place unique demands on athletes, including frequent jumps, sprints and collisions (Taylor et al., 2017). Consequently, specific recovery strategies are required to address the distinct physiological stresses of each sport. The review concludes that there is substantial variability in postmatch recovery time courses in different team ball sports, observed both within and between physical performance tests and biochemical markers (Doeven et al., 2017). Physical performance metrics such as Countermovement Movement Jump (CMJ) height and sprint times, recover more rapidly than biochemical markers such as Creatine Kinase (CK) levels. For example, players might be physically ready for subsequent training sessions or matches as early as 48 hours after the match based on performance recovery measures. However, impairments in sprinting and jumping abilities can persist up to 72 hours after a match (J. R. Silva et al., 2018).

Research in minor Gaelic football players demonstrated significant short-term changes in markers of muscle damage, soreness, and performance measures, such as CMJ height and sprint times of 5 m and 20 m after match play. CK levels, Delayed-Onset Muscle Soreness (DOMS) and sprint performance were transiently affected, muscle force and leukocyte counts showed significant fluctuations over a 60-hour period (Sheridan, 2021). These findings underscore the substantial physical demands of Gaelic football and the specific recovery requirements after a match. Insufficient recovery time can lead to both acute and chronic fatigue, potentially

leading to underperformance and an increased risk of injury. The challenge of optimising recovery is further compounded by considerable inter-individual variability in responses to training and competitive matches (Nédélec et al., 2012; J. R. Silva et al., 2018). This variability requires personalised recovery strategies to ensure that each athlete can return to optimal performance levels safely and effectively.

Determining which external load variables during soccer matches predict post-match acute and residual fatigue could help adjust training programmes to minimise injury risk and improve recovery. For example, running distance above 5.5 $\mathrm{m\,s^{-1}}$ was identified as the most sensitive variable for predicting biochemical and neuromuscular responses within the first 24 hours post-match. It showed a strong correlation between muscle damage markers and CMJ Peak Power Output (PPO). For every 100 metres run above 5.5 $\mathrm{m\,s^{-1}}$, CK activity increased by 30%, and CMJ PPO decreased by 0.5% 24 hours after the match. In contrast, the total distance covered did not show a significant relationship with fatigue-related markers, indicating that it is not a sensitive measure for fatigue monitoring (Hader et al., 2019). The analysis techniques in this study found no strong link between the tracking data and subjective markers of the players after the matches, although many soccer clubs use simple self-report tools to check the recovery and well-being of the players.

Subjective measures, such as self-reported recovery and wellness scores, are widely used due to their ease of implementation and low cost (Saw et al., 2016). These measures provide valuable information on the physical stress and recovery needs of athletes, helping inform training decisions and optimise performance. However, balancing exercise intensity and recovery remains a critical component of the training process to prevent performance declines and health issues (Bourdon et al., 2017; Halson, 2014). Despite their value, subjective scores are influenced by psychological and contextual factors, which may limit objectivity. Integrating these with physiological markerssuch as heart rate variability (HRV), electromyography (EMG), and skin temperaturecan improve the sensitivity of recovery models. However, these sensors are underutilised in team sport settings, likely due to cost, prac-

ticality, or validation constraints.

GNSS have revolutionised athlete monitoring by enabling the collection of extensive data on players' training and match workloads. This data provides a multidimensional perspective on players' physical demands, offering a more comprehensive understanding of their health and performance compared to traditional one-dimensional approaches (Rossi et al., 2022). Previous studies relying on single-variable analyses often oversimplified the complexity of training stimuli, missing intricate patterns within workloads (T. F. Gallo et al., 2016). ML models have the capacity to uncover latent patterns within complex, high-dimensional databridging subjective self-report tools and objective sensor data to produce more accurate, individualised recovery predictions.

ML techniques present a promising solution for predicting player wellness and recovery by analysing multidimensional training load data. These models enable sports scientists to simulate player responses to varying workloads, providing deeper insights into how training stimuli affect recovery and wellness. This data-driven approach facilitates the development of optimised training schedules, enhancing performance and recovery outcomes (Mandorino et al., 2022; Op De Beéck et al., 2019).

Op De Beéck et al. (2019) employed Gradient Boosting (GB) regression model to predict Future Perceived Wellness (FPW) in professional soccer players. Their model incorporated external load, internal load, and Presession Perceived Wellness (PPW), offering some individualisation. However, intrinsic factors such as age and fitness status were not explicitly accounted for, and contextual factors such as match schedules were excluded, potentially limiting the model's applicability. Despite these limitations, the study demonstrated how ML techniques can uncover complex relationships between training loads and wellness outcomes, paving the way for more personalised training programmes. Rossi et al. (2022) used XGBoost to predict FPW, achieving 74% accuracy in cross-validation and 87% in real-scenario validation. Chronic workload metrics, such as high metabolic load distance per minute and accelerations, were the most important predictors. However, intrinsic factors such

as age, training status and position, as well as contextual variables such as match results or opponent quality, were not included. This study highlighted the significant impact of specific external workload features on players' wellness and recovery status throughout the season. Mandorino et al. (2022) addressed previous gaps using ML to predict recovery status, measured as Total Quality Recovery (TQR) scale based on internal and external training load metrics. Their study demonstrated the utility of features such as previous TQR, age and decelerations for predicting recovery, but also noted challenges such as class imbalance and the absence of external validation. In future work, including underused physiological variables and contextual features such as travel, match pressure, or sleep quality may further improve model robustness.

Chapter 5.2.1 builds on these findings by integrating recovery status to detect early signs of fatigue and maximise players performance. This approach improves the integration of recovery metrics with training load analysis, offering a comprehensive method to optimise training, reduce injury risk, and support athlete well-being. The dual focus on recovery and training load establishes a robust framework to manage performance and ensure physical and psychological readiness (Selmi et al., 2018).

...

2.4.3 Improving Machine Learning Models with Feature Engineering of GNSS and IMU Data

Building on the importance of predicting athlete responses to training load, the next step involves enhancing the predictive power of ML models by using advanced data sources and feature engineering techniques. GNSS combined with IMU signals are widely used to track athletes' positions and physical demands during training and competition (Gómez-Carmona et al., 2019; Theodoropoulos et al., 2020). The analysis of data collected from these wearable devices provides valuable insights into the activities of players and their relationships with performance, enabling a better understanding and management of training demands (Claudino et al., 2019;

Mandorino et al., 2024).

Wearable GNSS devices typically produce summarised features such as total distance, high speed running, and average speed during a session or specific periods (Cardinale & Varley, 2017). Additionally, they can offer mechanical and energetic information, including accelerations, decelerations, and energy expenditure (Gray et al., 2018). However, the precision of these measurements can be influenced by technical factors, such as the devices sampling frequency, signal filtering, and satellite configuration. For example, GNSS devices often struggle with precision over short, high-speed movements or sharp turns due to low-velocity filtering (Varley et al., 2017). Moreover, the quality of the GNSS signals depends on the number and arrangement of satellites in the sky, which affects the accuracy and consistency (J. J. Malone et al., 2017). Despite these limitations, the integration of IMU technology with GNSS data has enhanced the capabilities of wearable devices. IMU provide additional data on accelerations, angular velocity, and orientation, complementing GNSS-based measurements and improving overall reliability. This integration addresses some limitations of GNSS, particularly in high-speed or multidirectional movements, where IMU can bridge data gaps (J. J. Malone et al., 2017).

Practitioners typically rely on summarised GNSS data for player monitoring, often neglecting the raw GNSS and IMU data. A study predicting Acceleration-Velocity (A-V) profiles in elite football players combined commercial GNSS features with novel features extracted from raw data. Using mean values and exponential weighting functions, features were engineered from both time and frequency domains, employing the discrete Fourier transform and the `TSfresh` Python module to identify 779 potential features. The 10 most relevant characteristics were selected on the basis of statistical significance (F statistic and p-value). Although GNSS features alone showed limited relevance in predicting A-V profiles, the study highlighted the potential of advanced signal processing and higher sampling rates to improve performance and injury modelling (Imbach et al., 2022). Kim et al. (2022) introduced a novel deep-learning model, *FatigueNet*, to predict RPE using raw move-

ment data rather than aggregated features. The raw GNSS data was transformed into local coordinates, down-sampled to reduce noise, and subjected to feature engineering to calculate feature vectors for each player. These vectors included both linear (velocity, acceleration, jerk) and angular (angular velocity, angular acceleration, angular jerk) components, better capturing the demands of intense team sports activities such as sprints and directional changes. *FatigueNet* demonstrated superior performance in RPE prediction, achieving a MAE of 0.8494 ± 0.0557 and a Root Mean Square Error (RMSE) of 1.2166 ± 0.0737, representing a 22.66% improvement in MAE and an 11.17% improvement in RMSE compared to the best GB regression model.

White et al. (2022) introduced a new methodological framework for identifying Sequential Movement Patterns (SMP) in athletes using GNSS data, improving the evaluation of external load structures in sports. This framework integrates spatiotemporal data, including geospatial coordinates (latitude and longitude), instantaneous velocity, and acceleration, to analyse athlete movement sequences in a comprehensive way. Movement units were formed by segmenting the data into distinct actions based on these variables. A key component was the computation of turning angles, defined as changes in direction between consecutive GNSS samples, capturing critical directional changes for field-based team sports. Antonini et al. (2024) introduced a multistep feature engineering framework to transform raw sequential data into optimised feature sets for ML applications. This framework addresses the challenges posed by raw data, which, while abundant, is not inherently suitable for direct analysis or ML tasks. The transformation involves multiple processing steps to refine and structure the data, making it applicable to both supervised and unsupervised ML models.

In addition to traditional time-domain features, researchers have explored frequency-domain and shape-based representations to improve the sensitivity of ML models to complex movement dynamics. Techniques such as Fast Fourier Transform (FFT) (Alemayoh et al., 2021) and discrete wavelet transform (DWT) (Mitchell et al.,

2013) enable the extraction of spectral and structural features from GNSS or IMU time series. These features can improve model performance in detecting fatigue, load patterns, and abnormal events. Such transformations are particularly useful for capturing periodicity, intensity fluctuations, or subtle variations in movement profiles that are not evident in basic kinematic data.

Recent studies have also introduced alternative machine learning architectures beyond tree-based models and standard neural networks. Hybrid models, such as convolutional LSTM (CNNLSTM) networks, and attention-based mechanisms have been proposed to capture spatiotemporal dependencies in athlete movement data (Chang et al., 2023). While these approaches show promise, their practical deployment in applied sport settings remains limited due to challenges related to interpretability, computational demands, and data requirements. In parallel, unsupervised approachessuch as clustering and dimensionality reductionare gaining traction for identifying movement signatures and recovery states without the need for labelled outcomes (J. Wang et al., 2019). Emerging alternatives such as unsupervised learning, frequency-domain features, and hybrid deep learning models offer potential for future work but remain limited in their current application due to deployment and interpretability challenges

In Chapter 5.2.1, these studies will be built on to develop a feature engineering process that allows more stable and accurate predictions of athlete responses to training. Using raw GNSS and IMU data, the *temporal* and *mechanical* aspects of movement, such as acceleration, deceleration, and directional changes will be examined. This approach provides a deeper understanding of how athletes respond to training loads, improves the reliability of ML models and supports better training and recovery strategies.

2.4.4 Machine Learning Applications in Predicting Fitness Status

The training loads in soccer are designed to improve the fitness of the players by tailoring the intensity and volume to individual needs throughout the season (Impellizzeri et al., 2005). However, group exercises can produce varied effects on performance, as they often do not take into account the unique fitness level of each player, which can lead to suboptimal improvements. The dose-response relationship in training describes how the training load influences physical and physiological adaptations (Impellizzeri et al., 2023). This relationship is shaped by factors such as player readiness, training suitability, season stage and exercise customisation. Furthermore, the methods used to measure training loads significantly affect this relationship, influencing the accuracy of adaptation assessments. Machine learning models offer an emerging avenue for estimating aerobic capacity and performance status without additional testing burden. Accurate modelling of athlete fitness using GNSS-derived data provides a means of assessing adaptations over time in a minimally intrusive way (Imbach et al., 2022). This invisible monitoring approach holds particular promise in congested team sport schedules, where traditional fitness testing may be impractical or disruptive to periodised plans.

Invisible monitoring leverages data collected during regular training sessions, avoiding additional testing burdens on players or coaches (S. W. West et al., 2021). This unobtrusive approach eliminates the need for traditional invasive testing procedures used to track physiological changes, making it an attractive alternative to monitor player adaptations over time (Lacome et al., 2018). Submaximal fitness tests have gained popularity for their efficiency in time and ease of administration (Shushan et al., 2023). These tests, categorised by exercise regimens and intensity progressions, typically rely on heart rate-derived indices to assess training effects, offering a practical method of tracking positive adaptations (Shushan et al., 2022).

Recent advances in machine learning have introduced innovative techniques for predicting fitness status. A new fitness index uses machine learning to compare

actual and predicted heart rates during training sessions, providing a novel non-invasive metric to assess player fitness. This fitness index holds promise for practical applications in soccer, such as personalised training adjustments and injury prevention. Integrating this index with traditional fitness tests could further enhance the accuracy of fitness evaluations and the effectiveness of individualised training programmes (Mandorino et al., 2024). Other wearable metrics, such as heart rate variability (HRV), skin temperature, or sweat-based sensors, may offer complementary physiological insights, but their adoption remains limited by issues with accuracy, validation, and practicality in applied team sport environments (Seshadri et al., 2019a).

The training loads in soccer are designed to improve the fitness of the players by tailoring the intensity and volume to individual needs throughout the season (Impellizzeri et al., 2005). However, group exercises can produce varied effects on performance, as they often do not take into account the unique fitness level of each player, which can lead to suboptimal improvements. The dose-response relationship in training describes how the training load influences physical and physiological adaptations (Impellizzeri et al., 2023). This relationship is shaped by factors such as player readiness, training suitability, season stage and exercise customisation. Furthermore, the methods used to measure training loads significantly affect this relationship, influencing the accuracy of adaptation assessments. Accurate modelling of athletic performance using GNSS data offers another approach to improve understanding and prediction of player performance (Imbach et al., 2022). This data-driven method provides information on player workloads without the need for additional testing, allowing more precise fitness and performance evaluations.

2.4.5 Estimating Oxygen Uptake from Wearable Sensor Data

Emerging methods for fitness assessment have advanced non-invasive monitoring techniques for evaluating player performance discussed above. Although these approaches provide valuable information, they often only offer estimates rather than

precise measurements of the physiological demands placed on athletes (Impellizzeri et al., 2005). Internal responses, stimulated by external loads during exercise, drive critical adaptations. However, these adaptations can only be directly measured by fitness testing (Clemente et al., 2019). Manari et al. (2016) highlight VO_2 max as a key metabolic parameter to determine the role of an athlete on the field and training needs in elite soccer, showing significant variability between positions and seasonal phases. In contrast, VO_2 at anaerobic threshold remains stable, which underscores its reliability in assessing overall performance.

Team sports require a high aerobic contribution, making it essential to track changes in aerobic fitness to optimise performance. Aerobic fitness significantly influences key performance metrics such as repeated sprint ability and fatigue resistance, both critical to maintaining high intensity performance in team sports (Gharbi et al., 2015; Kelly et al., 2018). For example, research in Gaelic football demonstrates that players with superior aerobic adaptations, measured by VO_2 max, cover greater distances during games, emphasising the importance of systematically monitoring and developing aerobic fitness to improve on-field performance (Daly et al., 2024).

Traditional fitness testing methods, such as spiroergometry or cardiopulmonary assessments, are often impractical during competitive periods due to scheduling constraints (Doeven et al., 2017). As a result, tests are typically limited to the preseason, leaving gaps in the monitoring of fitness levels and adaptations throughout the season (Altmann et al., 2020). To address this limitation, there is a growing need for non-obtrusive methods that allow for continuous monitoring of aerobic fitness. Recent advances in wearable technology have highlighted the potential of using external load metrics to estimate aerobic adaptations such as VO_2 max. Helwig et al. (2023) demonstrated that accumulated metrics such as PlayerLoadTM, total distances, and intensity-specific movement zones (e.g., high speed running and accelerations) strongly correlate with VO_2 max improvements over time (Clemente et al., 2019). Derived from MEMS, these wearable-based metrics offer practical,

noninvasive alternatives to traditional laboratory assessments, making them highly suitable for field-based sports environments.

Estimating energetic demands during team sports remains complex due to the dynamic nature of play, characterised by frequent accelerations, decelerations, and collisions (Taylor et al., 2017). GNSS-based metabolic power calculations have been proposed as a method to estimate instantaneous metabolic power and oxygen consumption (di Prampero & Osgnach, 2018; Osgnach & di Prampero, 2018). Although this approach provides valuable information on the energetic demands of team sports (S. Malone et al., 2017), its validity remains contested. Some studies report strong correlations with indirect calorimetry (Osgnach et al., 2016), while others highlight significant underestimations under certain conditions (Oxendale et al., 2017; Savoia et al., 2020).

To address these challenges, alternative approaches to quantifying loads for team sport athletes are required. IMU, which combine 3D accelerometers, 3D gyroscopes and 3D magnetometers, offer a promising solution. Captured at high frequencies (e.g. 100 Hz), the IMU data provides a more sensitive measure of high-intensity actions (Nedergaard et al., 2014). Custom accelerometer metrics derived from these data enable the quantification of three-dimensional movement and the estimation of oxygen consumption, although accuracy varies depending on activity (Gómez-Carmona et al., 2019; Vähä-Ypyä et al., 2023). Building on these advancements, ML has emerged as a transformative tool to model human physiological responses from sensor data, addressing the limitations of traditional statistical models. Unlike these models, which depend on significant predictors and often struggle with non-linear and complex relationships, ML techniques excel at uncovering intricate patterns within high-dimensional data (Zignoli et al., 2019). For example, ML models that use inputs such as heart rate, respiration rate and acceleration data from medical-grade wearables have significantly reduced errors in VO_2 estimation. In particular, during progressively intense exercises, such as the Bruce treadmill test, ML methods demonstrated superior accuracy compared to traditional heart rate-based models (Z.

Wang et al., 2022).

Recent advancements in VO_2 estimation during exercise have explored a variety of ML models, including random forests, LSTM networks, and Temporal Convolutional Networks (TCN). These models use various inputs, such as heart rate, work rate and accelerometer data, to predict ML dynamics at different intensity levels of exercise (Altini et al., 2016; Beltrame et al., 2016; Hedge & Hughson, 2020). Integrating data from IMU, GNSS, and heart rate sensors has further enhanced prediction accuracy during outdoor activities, highlighting the potential of neural networks for real-time energy expenditure assessments (Davidson et al., 2023). In particular, ML models have proven to be effective in estimating VO_2 dynamics during both steady-state and non-steady-state activities, including transitions between exercise intensities (Hedge et al., 2023). The development of personalised models marks a significant advancement in this domain. LSTM neural networks, trained on individual-specific data such as heart rate, mechanical power output, cadence, and respiratory rate have demonstrated high predictive accuracy in diverse exercise intensities (Zignoli et al., 2020). Such models facilitate the transition from laboratory-based assessments to field applications, enabling non-intrusive, real-time monitoring of aerobic fitness (Alzamer et al., 2021). By integrating wearable sensor data with ML algorithms, coaches and sports scientists can create personalised fitness models for athletes, continuously monitor cardiorespiratory fitness, and optimise training loads during both games and training. This approach represents a pivotal step in improving athletic performance and understanding physiological states in real world scenarios.

Despite these advancements, to date, no study has specifically examined the use of ML to predict VO_2 during team sports activities. This represents a unique challenge due to the intermittent nature of these sports and their frequent transitions between different intensity states. The dynamic and unpredictable nature of team sports introduces additional complexities in modelling VO_2 dynamics, requiring advanced ML techniques capable of handling non-linear relationships and real-time

variations in physiological responses. Addressing this gap presents an opportunity for innovative approaches accurately to monitor and predict oxygen uptake in such demanding environments. Chapter 6 will focus on investigating the performance of various ML models to predict individual VO_2 during simulated team sports circuits, using physiological test data collected during graded exercise tests in a laboratory setting. By integrating data from a combination of body-worn accelerometers, this study evaluate of the ability of IMU motion data ability to estimate VO_2. Personalised models developed from preseason test data allow for continuous monitoring of VO_2 demands during training, providing valuable insights into athletes' responses to exercise and supporting performance optimisation and load management.

2.5 Summary

This chapter explored the role of ML in enhancing athlete monitoring through wearable sensor data, identifying critical gaps and opportunities that shaped the direction of this research. The review began by detailing the frameworks that underpin athlete monitoring, distinguishing between internal and external loads. It highlighted the inherent challenge of managing internal loads, which are only measurable after training or competition. This challenge is further complicated by individual differences, which result in unique responses among athletes, and the reliance on subjective measures such as RPE.

The discussion then progressed to the advantages of applying ML to model the dose-response relationship between internal and external loads. Unlike traditional methods, ML models can integrate diverse data sources, providing a more comprehensive understanding of this relationship. These models have the potential to identify key factors that influence internal loads (for example, RPE and offer insight into the status of an athlete, allowing more informed and personalised training strategies.

The review then addressed the limitations of current research, which focusses primarily on aggregated sensor data, and examined the potential of using raw, high-

resolution wearable sensor data to build predictive models. Incorporating raw data can enhance the accuracy and reliability of predictions, offering a richer understanding of the relationship between training loads and athlete responses.

Finally, the chapter explored whether raw sensor data could be used to predict physiological responses, such as VO_2. This approach could overcome the reliance on subjective metrics like RPE by providing objective, non-intrusive methods for monitoring internal loads. Advancing this understanding could significantly improve the precision of training load management and expand the utility of wearable sensor data in sports science.

Through this progression, the chapter identified the research gaps addressed in this thesis: the need for integrating raw sensor data, leveraging advanced feature engineering, and applying ML to predict both perceptual and physiological responses. These foundations set the stage for the subsequent chapters, where these challenges are systematically tackled to advance the field of athlete monitoring.

Chapter 3

Data Acquisition and Machine Learning Methods

This chapter outlines the data collection hardware and methodologies employed in this research to generate the data required for subsequent machine learning experiments. Reusing existing data was not feasible due to the need for validated methodologies that ensured the accuracy, consistency, and relevance of the data for this specific context. Developing and adhering to a rigorous data acquisition process was considered a critical step to ensure the quality of the data and its suitability to address research objectives. Consequently, it was necessary to acquire or generate custom data sets tailored to the requirements of the three major studies presented in this thesis.

The chapter begins with a description of the devices and instruments in Section 3.1 used to generate data for each of the three studies that form the basis of this investigation. In Section 3.2 a detailed account of the data acquisition process for each experiment is provided. This includes descriptions of the hardware devices used, the methodologies implemented, and the adherence to established practices and standards during data collection. This comprehensive overview of the data and methods establishes a solid foundation for presenting the results in the following chapters.

3.1 Common Equipment, Instruments and Machine Learning Models

In this section, an overview of each sensor, together with their characteristics and capabilities is outlined. Some equipment is commonly used in sports- or activity-based research, while others are less common and are used in more specialised experiments.

Wearable tracking sensors, such as Inertial Measurement Unit (IMU) and Global Navigation Satellite Systems (GNSS), are widely used to monitor athletes' movement, energy expenditure and biomechanical and physiological parameters. These tools are key in tailoring training regimens to reduce the risk of soft tissue injuries (Seshadri et al., 2019b).

3.1.1 Global Navigation Satellite Systems (GNSS)

Originally developed for military applications, the GNSS technology consists of a satellite network that provides precise time and position data. A GNSS receiver calculates three-dimensional positions by determining the distance to multiple satellites, enabling accurate measurements of speed and position. Traditional methods for measuring human locomotion, often constrained by frame rate, viewing angle, and location suitability, have limitations. The introduction of GNSS in the 1990s provided a more flexible and accurate alternative, allowing improved measurement capabilities in field studies (Townshend et al., 2008). GNSS has shown promise in assessing speed during physical activities. However, improvements are necessary to increase accuracy, particularly at lower speeds. Future research should focus on the potential benefits of differential GNSS corrections to improve precision (Schutz & Chambaz, 1997). Non-differential GNSS refers to the standard method in which a single GNSS receiver determines its position, speed and other data solely based on signals from GNSS satellites. Speed measurements using non-differential GNSS, particularly through the Doppler shift method, demonstrate high accuracy and are strongly correlated with actual speeds, especially on straight paths, and maintained

reasonable accuracy even on curved paths. Speed data obtained by Doppler shift and positional changes over time were compared with actual speeds measured by chronometry, confirming the reliability of the method (Townshend et al., 2008).

In contrast, differential GNSS improves accuracy by using a network of stationary reference stations that have known positions. These stations calculate errors in satellite signals they receive and send correction data to nearby mobile GNSS receivers. The mobile receivers then adjust their calculated position and speed based on these corrections, significantly enhancing accuracy. differential GNSS is a reliable tool for assessing human locomotion speed, offering significant improvements over traditional methods and nondifferential GNSS. Studies show that differential GNSS improves accuracy by a factor of 10 compared to non-differential GNSS, making it a superior option for precise speed measurement in physical activity evaluations (Schutz & Herren, 2000). The utility of differential GNSS suggests that, in the future, it may combine with other biosignals to conduct comprehensive studies of exercise physiology in the field. It demonstrates the utility of differential GNSS in improving the accuracy and applicability of sport-specific field tests, contributing valuable insights to research on exercise physiology (Larsson, 2003).

3.1.2 Inertial Measurement Unit (IMU)

An IMU is a sensor device that measures and reports a specific force, angular rate and sometimes the magnetic field surrounding the device. IMU have gained popularity because of their compact size and ease of attachment to body segments. These wearable sensors typically contain a gyroscope, magnetometer and accelerometer to measure motion and orientation, making them practical and discreet. IMU can provide three-dimensional rotational data for each body segment by combining data from these components, allowing sufficiently accurate kinematic calculations outside laboratory settings (Morrow et al., 2017).

Accelerometer

Accelerometers measure linear acceleration on three perpendicular axes (X, Y, and Z). They detect changes in velocity along these axes, which can be used to infer the direction and magnitude of movement. They are widely used in various fields, including smartphones, vehicles, aerospace, and healthcare. Known for their high sensitivity and robust design, accelerometers can effectively capture motion data (Gauri Sawarkar et al., 2023). The basic operation of a Microelectromechanical Systems (MEMS) accelerometer relies on Newton's second law. There are two primary methods for measuring acceleration in MEMS accelerometers: displacement measurement, which tracks the displacement of a mass attached to a spring, and frequency measurement, which monitors changes in the frequency of a vibrating element as tension changes. Accelerometers are used to measure linear acceleration and, by integrating the acceleration signal twice, the position of an object can be determined (Alanen et al., 2021). Accelerometers equipped with triaxial sensors are increasingly used to measure external load by analysing movement acceleration across three axes. The PlayerLoad variable, derived from these measurements, has become a standard in sports such as basketball, rugby and soccer (Gómez-Carmona et al., 2020). In addition, physiological variables, such as HR and Maximal Oxygen Uptake (VO_2 max), play a significant role in the quantification of load. Muscle Oxygen Saturation (SmO_2) is also emerging as a reliable indicator, closely correlated with Oxygen Uptake (VO_2) (Gómez-Carmona et al., 2020).

Gyroscope

Gyroscopes are devices that sense the angular velocity, the rate of rotation around the three axes (roll, pitch, and yaw). They have been validated for use in various sports, such as tracking snow sports athletes' performance, including metrics like the number of jumps, flight time, and rotation rates. This information is valuable for managing training loads and analysing performance (Ross et al., 2020). MEMS gyroscopes use the Coriolis effect, which states that in a rotating frame of reference,

a moving mass experiences a force perpendicular to both its velocity and the axis of rotation. This effect is used to measure the angular velocity. The operation of a gyroscope involves detecting changes in capacitance. When the gyroscope rotates, variations in capacitance occur, allowing it to detect displacements and measure angular velocity. Mechanical gyroscopes consist of a spinning mass that rotates around its axis. When mounted on gimbals, which allow movement in three-dimensional space, the axis remains stable in its orientation despite changes in the environment. This stability results from physical phenomena like precession and nutation. By integrating the angular velocity signal, a gyroscope can determine the orientation of an object which is crucial for applications that require precise tracking of orientation and movement (Alanen et al., 2021).

3.1.3 Athlete Subjective Response Measures

Rating of Perceived Excretion (RPE)

The Rating of Perceived Exertion (RPE) is a widely used single item scale in exercise and sports science that serves as a surrogate measure of physiological stress or internal load, helping to quantify exercise intensity and overall training load (M. J. Chen et al., 2002). It has gained popularity as an alternative to objective measures such as heart rate and blood lactate levels, mainly due to its practicality and ease of use (Foster et al., 2021). The validity of RPE as an effective tool for quantifying internal training load has been established in various team sports, including soccer (A. J. Coutts et al., 2009; Impellizzeri et al., 2004), rugby (McLaren et al., 2017), and Australian football (T. J. Scott et al., 2013). However, it is important to note that RPE may not be sensitive to small changes in exercise intensity during brief, intermittent run bouts (T. J. Scott et al., 2013). Coaches are encouraged to incorporate additional subjective measures in conjunction with RPE to enhance the sensitivity and individualisation of athlete monitoring. This approach moves toward a more personalised system, allowing daily adjustments to the training schedule based on the athlete's actual status (Haddad et al., 2017). Despite some limitations,

the sRPE method remains inexpensive, simple, and practical for monitoring internal training load in sports settings (Eston, 2012). In this research project the RPE scale was used to collect data on athlete perceived exertion, acknowledging its widespread acceptance and validated effectiveness in capturing internal load responses during training and competition.

The Perceived Recovery Status Scale (PRS)

Monitoring recovery in team sports is crucial to optimise performance and reduce the risk of injury. Changes in perceived recovery have been associated with an increased risk of injury among athletes (van der Does et al., 2017) The Perceived Recovery Status (PRS) scale has emerged as a practical and reproducible tool to assess the recovery status of athletes and predict performance outcomes (Laurent et al., 2011; Paul et al., 2019; Tolusso et al., 2022). Studies have shown that PRS is strongly correlated with physical performance measures, suggesting that it may serve as a subjective marker of recovery following intense exercise protocols (Laurent et al., 2011). Furthermore, PRS has been effective in detecting changes in performance that were not captured by other physiological and psychological measures, highlighting its sensitivity and utility in athlete monitoring (Paul et al., 2019). Incorporating PRS assessments allows coaches to monitor players' psychometric recovery states before training sessions and matches, allowing the early detection of fatigue signs and the adjustment of training loads to improve performance while avoiding physical and technical declines (Selmi et al., 2022). The practicality and ease of use of the scale make it suitable for daily recovery evaluations, even in youth athletes, where it has been shown to be a highly reproducible tool (Paul et al., 2019). In general, the PRS scale provides valuable information on athletes' recovery levels, facilitating informed decision making in training prescription and contributing to enhanced performance and reduced injury risks.

3.1.4 Machine Learning Models Chapter 4 & Chapter 5

Logistic Regression

Logistic regression is a statistical and supervised machine learning algorithm used for binary and multiclass classification tasks. It uses a logistic function to model the relationship between the input features and the probability of a specific outcome. The algorithm estimates coefficients for each feature to maximise the likelihood of correctly predicting the target class. Logistic regression is simple, interpretable and works well for linearly separable data (Cox, 1958). Logistic regression provides a baseline model to evaluate classification tasks, such as predicting player RPE into categories such as low, moderate and high. Its interpretability allows practitioners to understand how individual predictors, such as training load metrics, wellness scores, or player characteristics, contribute to the classification. In addition, it serves as a valuable reference point for comparing the performance of more complex models.

Decision Tree (DT)

A DT is a supervised machine learning algorithm that is used for both classification and regression tasks. It works by splitting the data set into subsets based on feature values, creating a tree-like model of decisions. Each node represents a feature, branches indicate decision rules and leaves represent results. DTs are interpretable and easy to visualise but may overfit the data without pruning (Breiman et al., 2017). DTs are particularly valuable for their ability to provide information on the relationships between input features, such as training load metrics and the target variable, such as RPE or other performance indicators. Their interpretability makes them a suitable choice for practitioners who require models that can be easily translated into actionable decisions. For example, a DT can highlight thresholds or key factors that influence player responses, aiding in coaching decisions or training load adjustments. In addition, DTs serve as a foundational model to benchmark the performance of more complex algorithms.

Random Forest (RF)

RF is an ensemble learning method that builds multiple DTs during training and aggregates their predictions for classification or regression tasks. By introducing randomness in the feature selection and data sampling for each tree, it reduces overfitting and improves generalisation. RF is robust, handles missing data well and is effective for both small and large data sets (Breiman, 2001). RF is particularly suitable for analysing complex datasets with many features, such as training load metrics, wellness scores and player characteristics. Its ability to calculate feature importance provides valuable insight into the most influential predictors, helping in model interpretation and practical decision-making for athlete monitoring and performance optimisation.

Bootstrap Aggregating (Bagging)

Bootstrap Aggregating (Bagging) is an ensemble learning technique designed to improve the stability and accuracy of machine learning models, particularly for classification and regression. It works by creating multiple subsets of the training data through bootstrapping (random sampling with replacement). A separate model is trained on each subset, and their predictions are averaged (for regression) or voted on (for classification) to produce the final output. Bagging reduces variance and mitigates overfitting, making it particularly effective with high-variance models like DTs (Breiman, 1996). Bagging is useful for building stable models from noisy or variable data, such as GNSS or IMU metrics, where overfitting is a concern. It ensures that the model generalises well, even when the dataset contains anomalies or is limited in size.

Light Gradient Boosting Machine (LightGBM)

LightGBM is a highly efficient gradient boosting framework designed for large datasets and high-dimensional features. It splits the data using a leaf-wise strategy rather than the level-wise approach used in other gradient-boosting algorithms, which im-

proves its efficiency and reduces computation time. LightGBM is well suited for classification tasks, is able to handle categorical features and missing values effectively and supports advanced techniques such as feature importance and early stopping (Ke et al., 2017). LightGBM excels at handling large, structured datasets with many training load metrics or athlete characteristics. Its efficiency makes it ideal for rapid experimentation, while its ability to compute the importance of features aids in identifying key predictors of outcomes such as RPE.

Gradient Boosting (GB)

GB is an ensemble learning method that builds predictive models sequentially by combining weak learners, typically DTs. Each new tree corrects the errors of the previous ones by minimising a differentiable loss function. The classifier optimises performance by focussing more on incorrectly predicted instances in successive iterations, making it robust and highly accurate for classification tasks. However, it can be computationally intensive and prone to overfitting without proper regularisation (Friedman, 2001). GB is highly effective in sport science in capturing subtle nonlinear relationships between features and outcomes, such as the impact of combined training load metrics on player performance. Its robustness ensures high predictive accuracy, making it a preferred choice for complex data sets.

eXtreme Gradient Boosting (XGBoost)

XGBoost is an advanced gradient boost framework optimised for efficiency, scalability, and flexibility. It enhances traditional gradient enhancement by implementing features like regularisation (L1 and L2), parallel processing, and tree pruning, which improve model performance and prevent overfitting. XGBoost is highly effective for classification tasks in structured / tabular data and is known for its speed and predictive accuracy in machine learning competitions (T. Chen & Guestrin, 2016). XGBoost is well suited for data sets with many interrelated features, such as aggregated training data, wellness scores, and positional differences. Its efficiency and

predictive accuracy make it ideal for high-stakes applications such as predicting injury risk or performance outcomes.

Neural Networks (NN)

NN are computational models inspired by biological neural networks (McCulloch & Pitts, 1943). They consist of layers of interconnected nodes, or neurons, where each connection has an associated weight (Figure 3.1 . These weights are adjusted during training, enabling the network to learn patterns in the data (Ali et al., 2020).

Key Components of Neural Networks

- **Neurons:** Each neuron receives inputs, computes a weighted sum, applies an activation function (e.g., sigmoid, ReLU) and passes the result to the next layer.

- **Connections and Weights:** Connections between neurons are weighted links that determine the strength of the signal passed.

- **Layers:** NN typically have three types of layers:

  - *Input Layer:* Receives raw data or features.
  - *Hidden Layers:* Process and transform data through multiple neurons.
  - *Output Layer:* Produces the final prediction or classification.

NN are trained using algorithms such as backpropagation and gradient descent, where the weights are iteratively adjusted to minimise prediction error. This enables the network to model complex, non-linear relationships in data.

NN are widely used for tasks such as regression and classification because of their flexibility. However, their effectiveness depends on sufficient data and computational resources.

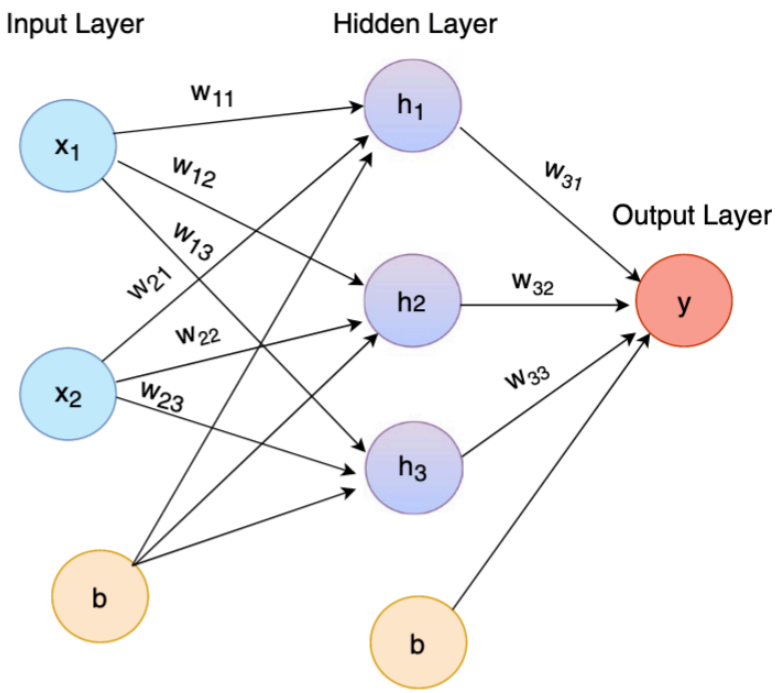


Figure 3.1: A simple artificial neural network with input, hidden and output layers.

3.1.5 Machine Learning Models Chapter 6

Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR) is a linear method for modelling the relationship between a dependent variable and one or more independent variables by fitting a linear function to the observed data (Bishop, 2006). Its simplicity and interpretability make it a common starting point for evaluating more complex models. The coefficients of the regression function are determined by minimising the residual error between the predicted and actual data values.

To enhance model efficiency and address potential multicollinearity, Principal Component Analysis (PCA) was applied prior to model training. PCA transforms the input feature space into orthogonal components, which can help reduce dimensionality, improve computational efficiency, and mitigate overfitting. In this study, MLR with PCA served as a baseline model for predicting breath-by-breath $\dot{V}O_2$, against which more complex, nonlinear models were compared.

Nonlinear Regression Model: XGBoost

Extreme Gradient Boosting (XGBoost) is a high-performance machine learning algorithm built on the Gradient Boosting Decision Trees (GBDT) framework, known for its superior performance in both regression and classification tasks (T. Chen & Guestrin, 2016). It effectively handles complex, nonlinear relationships and can manage missing data without requiring imputation.

XGBoost incorporates regularisation strategies, including both L1 (LASSO) and L2 (Ridge), which help control model complexity and reduce the risk of overfitting, thereby improving generalisability. It is optimised for speed and scalability, supporting large datasets and distributed training environments. These characteristics make it a leading model in modern machine learning pipelines.

For this study, the XGBoost regression model was implemented with various hyperparameter configurations to evaluate its performance. Model architectures tested included different numbers of trees (500, 1000, 3000, 5000) and maximum tree depths (3, 5, 7, 10). The final prediction is computed by summing the outputs of K decision trees, each contributing to the model's ability to represent complex, nonlinear mappings between input features and $\dot{V}O_2$.

Deep Learning Models

In addition to traditional machine learning models, Chapter 6 employed deep learning approaches to evaluate whether end-to-end temporal feature extraction from raw data could improve the estimation of VO_2. Deep learning architectures were selected due to their proven capacity to capture complex patterns and temporal dependencies in sequential dataan essential characteristic when modelling physiological responses in dynamic sports environments (Tunca et al., 2020).

Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP): MLP networks are adept at handling nonlinearity, internal randomness, and long-term unpredictability in time series data.

They transform high-dimensional input data into a manageable latent space to make accurate predictions (Rozos et al., 2021). **Architecture Details:** The activation function is ReLU; there is no dropout, and L2 regularisation is used with weight decay 1e-4 to reduce overfitting. Various configurations test different depths (from one to four layers) and widths (32 or 64 neurones per layer), examining how each configuration affects the performance of the model.

Long Short-Term Memory (LSTM)]

Long Short-Term Memory (LSTM): The LSTM model is well suited for time series predictions due to its ability to remember patterns based on previous time-step states, making it effective in capturing long-term dependencies within sequential data (Hochreiter & Schmidhuber, 1997). A bidirectional LSTM version, which processes its inputs in a bidirectional manner along the breath dimension to capture contextual information from past and future breaths, enhancing predictive accuracy was adpoted for this study. **Architecture Details:** In LSTM, we consider breath a time step, so we have a fixed length (7, 5, 3, 1 breath); no padding is needed. There is no dropout; regularisation with the weight decay 1e-4 to reduce overfitting are used. Various configurations, ranging from 1 to 4 hidden layers, each with 32 or 64 units are tested.

Convolutional Neural Network (CNN):

Convolutional Neural Network (CNN): A CNN is well suited for time series predictions due to its effectiveness in extracting local temporal features from sequential data (Reusch et al., 2022). The model uses a series of convolutional layers; each CNN layer has kernels to slide the breath dimension, computing and extracting temporal features at every single timestep. Like the LSTM, the CNN focusses on the output from the middle breath for final processing, ensuring relevance to the temporal centre of the data. **Architecture Details:** A fully connected layer initially processes each breath and generates latent vectors at the level of the breath

prepared for subsequent 1D convolution. The 1D convolution is performed along the breath dimension, using the'same' padding to maintain the original input shape by adding zero padding at the edges. A stride of 1 is used, ensuring that the convolution operation moves along the breath dimension one step at a time. The CNN configurations vary in depth, ranging from 3 to 5 convolutional layers and in the number of kernels, using either 32 or 64 kernels per layer. This variability allows the model to learn features at different levels of abstraction. A consistent kernel size of 3 is applied across all convolutional layers, effectively capturing local temporal patterns while maintaining broader contextual information.

3.2 Data Acquisition

3.2.1 Data Acquisition for Chapter 4

In this section the data acquisition procedures used in Chapter 4, which focusses on collecting and processing data to predict the RPE in elite Gaelic football players using external load metrics derived from GNSS wearable devices are outlined.

Participants

This study involved a cohort of 49 elite Gaelic football players (Mean ś SD, age: 25.59 ś 4.01 years; height: 1.82 ś 0.06 m; mass: 81.97 ś 7.10 kg) during the 2020, 2021 and 2022 intercounty seasons. The players competed in Division 2 and Division 3 of the National Football League (NFL) during the data collection period. Data from goalkeepers were excluded due to the distinct demands of their position. The study included players from the five outfield position groups: nine fullbacks, eleven halfbacks, eight midfielders, ten half-forwards and eight full-forwards.

Data were initially collected as part of routine performance monitoring in the candidates professional role as a sport scientist, prior to the commencement of the PhD and before formal ethical approval was in place. Ethical approval for the research use of these data was subsequently granted by the Dublin City University

Research Ethics Committee (REC Reference: DCUREC/2021/267), with notification received on 27th January 2022. From that point onward, all procedures adhered to the approved protocol in accordance with the Declaration of Helsinki.

Accordingly, the study design is best described as a **mixed retrospective-prospective design**, with early-season data analysed retrospectively and subsequent data collected prospectively under approved ethical oversight.

Study Design

Internal and external load data were collected during training sessions and matches in each season of an elite Gaelic football team over three seasons. All field training sessions were conducted on a full-length GAA grass field. Strength and recovery sessions were excluded to ensure greater consistency between training sessions. GNSS technology was used to measure the activity levels of elite Gaelic football players during training and match play in the three intercounty seasons. RPE was recorded from the players using a mobile application after the completion of each training session and match. Training sessions were limited to field sessions and were completed on a grass playing surface of approximately 140 m in length and 80 m in width. A total of 1616 GPS data files and concurrent measures of RPE were recorded. This included 96 training sessions and 44 matches, resulting in 1205 and 411 records, respectively. A total of 562, 575 and 479 records were collected during the 2020, 2021, and 2022 playing seasons, respectively. The median (range) number of observations per player was 29 (298) with a mean ś SD of 32.6 ś 22.9.

Data Acquisition Procedures

RPE was measured using the modified Borg CR10 scale (Foster et al., 2001). Ratings were recorded approximately 30 minutes after each training session using the Smartabase mobile app (Smartabase, v.6.8.08, Fusion Sport, Milton, Australia). The use of a mobile application allowed ratings to be recorded privately without the influence of peer presence or other related environmental factors (Minetti et

al., 2002). All participants were familiar with the use of the CR10 scale before participating in the present study.

External load metrics of activity were collected using GPS units sampling at a rate of 18 Hz (GPEXE LT, Exelio, Italy). During each training session and match, the players wore an individual GPS unit that was placed between the scapulae in a custom-made undergarment. These units have shown good to moderate validity (<10% typical error of estimate) and good (<5% coefficient of variation) reliability for distance covered over a range of movement speeds in a sport-specific team circuit (Hoppe et al., 2018; M. T. U. Scott et al., 2016). Following each training session and match, the data were downloaded to the manufacturer's proprietary software (GPEXE Bridge v.6.9.25) to remove values not related to the training session or match. Forty-two different movement indices related to speed and distance were extracted from each GNSS data file (Table A.1).

Perceived wellness was examined before each training session and matched using a 5-item questionnaire submitted through a mobile application (Smartabase, v.6.8.08, Fusion Sport, Milton, Australia). Each item was rated on a 10-point scale. Individual elements were then categorised as muscle soreness, sleep quality, or energy levels using the average score for each element in the section.

The stature was measured to the nearest 0.1 cm using a portable stadiometer (model 213, SECA, Hamburg, Germany). Body mass was measured to the nearest 0.1 kg using a portable digital scale (model 813, SECA, Hamburg, Germany). The lean mass index and body fat percentage were evaluated using Dual-energy X-ray Absorptiometry (DEXA). Age, body mass index, playing position and playing experience, classified as the number of years playing at the elite level of Gaelic football, were also included. MAS was determined using the time taken to complete a 1200 m Time Trial (TT). The TT was performed at the beginning of each playing season.

To track changes in workload over a given training period, 21 Acute Chronic Workload Ratio (ACWR) indices were generated from the GNSS data (Table 3.1), allowing effective monitoring of the dynamics of the training load.

ACWR Indices	Description
7-day total distance	Total distance in the previous 7 days (m)
14-day total distance	Total distance in the previous 14 days (m)
21-day total distance	Total distance in the previous 21 days (m)
28-day total distance	Total distance in the previous 28 days (m)
7-day average total distance	Average total distance across 7 days (m)
28-day average total distance	Average total distance across 28 days (m)
Total distance ACWR	Acute chronic ratio for total distance
7-day high-speed distance	High-speed distance in the previous 7 days (m)
14-day high-speed distance	High-speed distance in the previous 14 days (m)
21-day high-speed distance	High-speed distance in the previous 21 days (m)
28-day high-speed distance	High-speed distance in the previous 28 days (m)
7-day average high-speed distance	Average high-speed distance across 7 days (m)
28-day average high-speed distance	Average high-speed distance across 28 days (m)
High-speed ACWR	Acute chronic ratio for high-speed distance
7-day very high-speed distance	Very high-speed distance in the previous 7 days (m)
14-day very high-speed distance	Very high-speed distance in the previous 14 days (m)
21-day very high-speed distance	Very high-speed distance in the previous 21 days (m)
28-day very high-speed distance	Very high-speed distance in the previous 28 days (m)
7-day average very high-speed distance	Average very high-speed distance across 7 days (m)
28-day average very high-speed distance	Average very high-speed distance across 28 days (m)
Very high-speed ACR	Acute chronic ratio for very high-speed distance

Table 3.1: Acute Chronic Workload Ratio Indices.

A MySQL database was designed to manage the various data sources. Separate tables were created for the GNSS data, perceived wellness, RPE, and individual characteristics. Each table was linked through a unique user identification (UUID), forming a relational database structure. This setup allowed for seamless data integration and ensured the scalability of data processing. Automated routines were implemented to handle data entry, validation and merging across data sets. This section detailed the data acquisition procedures for Study 1, focusing on the collection of RPE scores, external load metrics via GNSS devices, wellness questionnaires and individual characteristics of elite Gaelic football players over three seasons.

Dataset Description

The dataset used for model development consisted of athletes who were highly adherent to the monitoring protocol and regularly submitted both GNSS and RPE data. As such, there is a potential selection bias toward more compliant and consistently available participants. This may limit the representativeness of the findings to broader team sport populations, particularly those with lower compliance or differ-

ent monitoring infrastructures. Furthermore, the cohort comprised elite intercounty Gaelic footballers, which may limit generalisability to other sports or levels of competition. The models developed should therefore be interpreted as context-specific to this population and should be validated externally before being applied more broadly.

Following all preprocessing stepsincluding cleaning, filtering for valid RPE labels, and alignment of input featuresthe final dataset used across Experiments 1 to 3 consisted of 1616 total records. This included 1205 training sessions and 411 match sessions collected between 2020 and 2022.

For the classification tasks, class imbalance was addressed through random undersampling to the size of the minority class. After an 80/20 train-test split resulted in a test set of 267 instances (89 per RPE category: Low, Moderate, High), as shown in Figure 4.7. This consistent subset was used across all attribute sets to enable direct model comparisons. Table 3.2 summarises the final instance counts and class balancing approach.

Table 3.2: Final data availability for classification and regression experiments after filtering and class balancing (Chapter 4)

Experiment	**Task Type**	**Total Records Used**	**Train Set (80%)**	**Test Set (20%)**
Experiment 1	Classification (Aggregated Metrics)	1616	1349	267 (balanced: 89/class)
Experiment 1	Regression (Aggregated Metrics)	1616	1349	267

3.2.2 Data Acquisition for Chapter 5

In this section the data acquisition procedures used in Chapter 5, which focusses on collecting and processing data to predict the RPE in elite Gaelic football players using raw data from the GNSS of wearable devices are outlined.

Participants

A cohort of 35 elite male Gaelic football players (mean ś standard deviation [SD]: age 24.94 ś 3.46 years; height 183.04 ś 3.82 cm; body mass 86.18 ś 5.00 kg) from one intercounty team participated in this study. The players competed in Division 2 of

the National Football League (NFL) during the data collection period. Goalkeepers were excluded due to the distinct demands of their position, resulting in a focus on outfield players in five positional groups: full back (10 players), half back (7 players), midfield (5 players), half forward (8 players), and full forward (5 players). All participants provided their informed written consent prior to participation. The study was approved by the Research Ethics Committee at Dublin City University in accordance with the Declaration of Helsinki (Reference: DCU/REC/2021/267).

Study Design

GNSS and IMU technology was used to measure the activity levels of elite Gaelic football players during training and match play in the 2024 season. RPE was recorded from the players on a one-to-one basis after the completion of each training session and match. All field training sessions and matches were conducted on a full-length grass pitch of the Gaelic Athletic Association (GAA) measuring approximately 140 metres in length and 80 meters in width. A total of 73 training sessions and matches were observed, comprising 48 field training sessions, 13 official matches, 9 challenge matches and 3 internal matches. A total of 1,031 GNSS raw data files and concomitant measures of RPE were collected. This included 644 training records, 200 official match records, 148 challenge match records, and 39 internal match records. The median number of observations (range) per player was 31 (24.50-33.4), with a mean ś standard deviation (SD) of 29.46 ś 9.51.

Data Acquisition Procedures

Stature was measured to the nearest 0.1 cm using a portable stadiometer (Model 213, SECA, Hamburg, Germany), and body mass was measured to the nearest 0.1 kg using a digital scale (Model 813, SECA, Hamburg, Germany). Age and playing position were also recorded. The maximum sprint speed (MSS) was determined using GNSS devices during 60-meter sprint trials conducted periodically throughout the season; the highest velocity recorded during these trials was used as the MSS

of each player. The maximum aerobic speed (MAS) was determined using the time taken to complete a 1,200-meter time trial. Due to more than 10% missing data for MAS, this variable was excluded from further analysis.

The recovery status of the players was estimated using the PRS prior to training sessions and matches. The 10-point PRS scale, a validated method for assessing recovery status in athletes, provided information on players' readiness and potential influences on their performance and perceived exertion (Laurent et al., 2011). Players individually completed a custom designed electronic survey using an iPad (Google LLC, n.d.) approximately 15 minutes before the warm-up of each training session or match. All players were familiar with the PRS scale prior to the study.

The RPE was recorded approximately 30 minutes after the completion of each training session and match. This timing allowed players to reflect on the session while minimising immediate fatigue effects on their perception. The Borg CR-10 Scale was used to quantify perceived exertion. Players provided their RPE using a custom-designed electronic survey on an iPad (Google LLC, n.d.), ensuring privacy and reducing potential biases from peer influence. All players were familiar with the RPE scale prior to the study.

External load data were collected using GNSS and IMU devices (STATSports Apex, Northern Ireland) sampling at a frequency of 10 Hz and 100 Hz, respectively. The STATSports Apex units have demonstrated good levels of accuracy (bias $< 5\%$) for measuring distance covered across various movement speeds in team sportspecific circuits (Beato et al., 2018). During each training session and match, players wore individual GNSS units placed between the scapulae using custom-made vests to minimise movement artefacts. Following each session, data were downloaded from GNSS units using the manufacturer's proprietary software (Sonra v4.5.19). Data not related to the training session or match were removed to ensure data integrity. Raw GNSS and IMU data for each player were downloaded and exported, enabling the design of features from the raw sensor data (Table 3.3).

To prepare GNSS and IMU data for modelling, we applied the following transfor-

Feature	Description	Data Type
PlayerID	Player identifier	String
Time	Timestamp for each recorded instance	String
Lat	Latitude position of the player	Float64
Lon	Longitude position of the player	Float64
Speed (m/s)	Speed of the player in meters per second	Float64
Heart Rate (bpm)	Heart rate in beats per minute	Float64
Hacc	Horizontal accuracy of the GNSS signal	Float64
Hdop	Horizontal dilution of precision	Float64
Quality of Signal	Quality of the GNSS signal	Float64
No. of Satellites	Number of satellites used to acquire the GNSS signal	Float64
Inst Accel Impulse	Instantaneous Acceleration Impulse value	Float64
Accl X	Acceleration in the x-axis	Float64
Accl Y	Acceleration in the y-axis	Float64
Accl Z	Acceleration in the z-axis	Float64
Gyro X	Gyroscope value in the x-axis	Float64
Gyro Y	Gyroscope value in the y-axis	Float64
Gyro Z	Gyroscope value in the z-axis	Float64

Table 3.3: Overview of Raw GNSS and IMU Dataset

mations: Using metadata from file names, we extracted session dates, drill names, and session labels, extracted metadata was integrated directly into the dataset to retain context, which aids in future filtering, grouping and analysis. A Butterworth Low-Pass Filter (Cutoff Frequency: 10 Hz; Sampling Rate: 100 Hz) was applied on accelerometer and gyroscope data to remove high-frequency noise. This smooths out the signal, improving the quality of movement metrics. The data were downsampled in two steps: first from 100 Hz to 10 Hz by averaging numeric columns and then from 10 Hz to 2 Hz. For nonnumeric columns, forward filling was applied after each step to maintain consistent metadata across rows. The distances between consecutive GNSS points was calculated using the Haversine formula 3.1, which provides an accurate movement distance between latitude and longitude points.

$$d = 2R \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \tag{3.1}$$

where:

$$d : \text{Great-circle distance between two points}$$

$$R : \text{Earth's radius (mean radius = 6,371,000 meters)}$$

$$\phi_1, \phi_2 : \text{Latitudes of the two points in radians}$$

$$\lambda_1, \lambda_2 : \text{Longitudes of the two points in radians}$$

$$\Delta\phi = \phi_2 - \phi_1$$

$$\Delta\lambda = \lambda_2 - \lambda_1$$

Anomaly Detection for Quality Control: Speed and distance data points were flagged as anomalous based on predefined thresholds for speed ($\geq$ 11 m/s) and positional changes ($\geq$11 m/s), helping to identify possible data collection errors. Time differences between consecutive points was computed, which is critical for detecting irregular time gaps and evaluating movement patterns. Instances where consecutive points exceed a specified time difference ($\geq$ 1 second) were flagged as anomalous, which could indicate gaps or irregularities in data collection. All anomaly types (speed, distance, time gap) were consolidated into a single anomaly column, simplifying the identification of data points that require further review or exclusion. These steps ensure that the data are cleaned, normalised, resampled and quality checked, ready for further analysis or model training with minimal noise and anomalies. These procedures ensured a clean, normalised dataset, ready for feature engineering and predictive modelling.

Dataset Description

The dataset used for model development consisted of athletes who were highly adherent to the monitoring protocol and regularly submitted both GNSS and RPE data. As such, there is a potential selection bias toward more compliant and consistently available participants. This may limit the representativeness of the findings to

broader team sport populations, particularly those with lower compliance or different monitoring infrastructures. Furthermore, the cohort comprised elite intercounty Gaelic footballers, which may limit generalisability to other sports or levels of competition. The models developed should therefore be interpreted as context-specific to this population and should be validated externally before being applied more broadly.

A separate dataset was used comprising 1,031 raw GNSS files matched with valid RPE responses. This dataset represented a refined time-series collection used solely for regression modelling and consisted of 644 training sessions, 200 official match sessions, 148 challenge match sessions, and 39 internal match sessions. Following standard preprocessing, feature extraction, and alignment procedures, all records were retained for model development and evaluation using an 80/20 train-test split. A breakdown of the final instance counts across all experiments is provided in Table 3.4.

Table 3.4: Final data availability per experiment after filtering and preprocessing

Experiment	**Input Type**	**Task**	**Total Records Used**	**Train/Test Split**
Experiment 1	Aggregated Metrics	Regression	1031	LOSOCV
Experiment 2	Raw GNSS Time-Series	Regression	1031	LOSOCV

3.2.3 Data Acquisition for Chapter 6

In this section the data acquisition procedures employed in Chapter 6 which focus on estimating individual oxygen uptake during simulated team sports activities using data form wearable devices are outlined. The data generation during a series of experiments is discussed. A description of the participants, how data was acquired and the types of sensors deployed is provided.

Participants

A total of eight healthy male team sports athletes (age: 28.63 ± 7.67 yr; height: 183.75 ± 3.49 cm; mass: 81.38 ± 9.81 kg; VO_2 max: 58.51 ± 4.57 $mL \cdot kg^{-1} \cdot min^{-1}$)

were recruited from November 2022 to July 2023 through social media advertisements. Participants were selected based on the following inclusion criteria: a minimum of four training sessions per week over two years, with at least three sessions being pitch-based training or games. Eligibility also required no self-reported history of metabolic, neurological, pulmonary, or cardiovascular disease and no significant lower extremity injuries for at least six months prior to the study. All participants provided their informed written consent in accordance with the Declaration of Helsinki. The study was approved by the local ethics committee of Dublin City University (DCUREC/2021/256).

The final dataset included six participants, due to one dropout and one athlete whose data were excluded because of sensor failure during testing. While the sample size is limited, this was a pilot study intended to evaluate proof-of-concept relationships between wearable sensor features and oxygen uptake. As such, the findings are not intended to be generalised, but rather to inform future larger-scale investigations. No imputation was performed on missing data; only complete datasets were retained for analysis.

Study Design

Each athlete participated in two sessions at Dublin City University (DCU), spaced at least 48 hours apart. The first laboratory visit consisted of three phases to assess VO_2: a resting phase, a submaximal exercise protocol, and a maximal Graded Exercise Test (GXT). During the resting phase, VO_2 and respiratory frequency were averaged over five minutes to establish baseline metabolic rates (Zignoli et al., 2020). The submaximal trial started at $9\,\mathrm{km\,h^{-1}}$, increasing by $1\,\mathrm{km\,h^{-1}}$ every six minutes until blood lactate levels reached ≥ 3 mmol/L, with intermittent 1-minute rest periods for lactate sampling (Garcia-Tabar et al., 2019). The treadmill gradient remained fixed at 1% to simulate the energetic cost of outdoor running (Jones & Doust, 1996). Following a 5-minute rest, the incremental maximal ramp test commenced, starting at a speed $1\,\mathrm{km\,h^{-1}}$ below the final submaximal speed

and increasing by $1\,\mathrm{km\,h^{-1}}$ each minute until it reaches $16\,\mathrm{km\,h^{-1}}$, followed by incremental increases in slope by 1% each minute until voluntary exhaustion. The second visit involved field tests on a synthetic pitch, incorporating a steady-state jog and an intermittent team sport simulated circuit developed from existing protocols (Wundersitz et al., 2015). Each circuit included three countermovement jumps, an eight-metre jog, an eight-meter change of direction agility section, two jumps for distance, a 10-m sprint, seven-metre walking and a tackle bag to be hit with force. These activities are designed to reflect the dynamic nature of team sports and lasted approximately 45 seconds, followed by 15 seconds of rest, repeated five times. All tests were carried out under similar conditions (20-21řC).

Data Acquisition Procedures

This section describes the data measured during laboratory and field visits and the features used for the dynamic oxygen prediction models. Figure 3.2 shows the measurement setup of wearable sensors worn by the athlete during the protocol. **Oxygen Uptake ($\mathbf{VO_2}$):** Pulmonary gas exchange data was captured using a Cosmed K5 breath-by-breath metabolic analyser, calibrated with a specific gas mixture and a flow metre before each session. VO_2 values were normalised by body mass, providing a detailed measure of aerobic capacity for each breath (Perez-Suarez et al. 2018). **Heart Rate (HR) & Breathing Rate (BR):** HR and BR were monitored using Zephyr Bio Harness 3.0, a validated tool for physiological monitoring in sports settings (Hailstone & Kilding, 2011). **Inertial Measurement Units (IMU):**IMU measured acceleration and angular velocity variations at five body locations, with data sampled at 250 Hz. Sensor calibration ensured accurate and synchronised data capture. Specific placements included the lower back, both tibiae and both wrists.

For each treadmill test, the VO_2 max of the runner was calculated as the maximum value of the rolling average of the VO_2 signal with a window length of 30 seconds. Recommendations for exercise physiologists have been published to adopt these data processing strategies to reduce variability in VO_2 measurements. A 15-

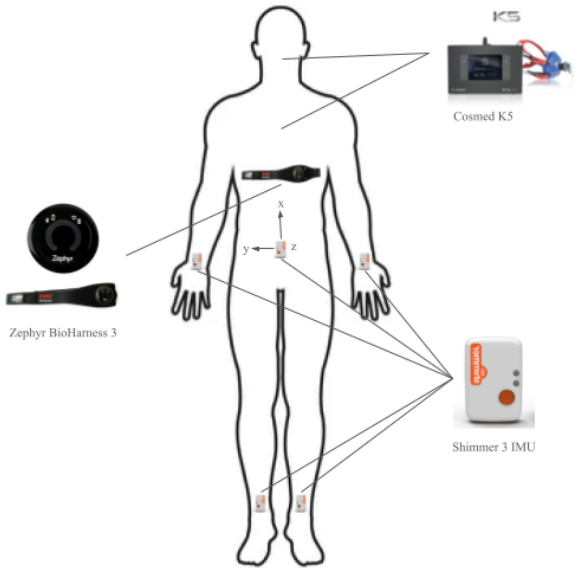
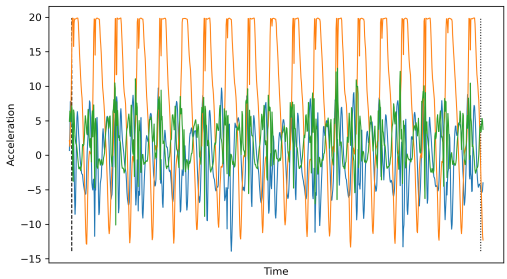
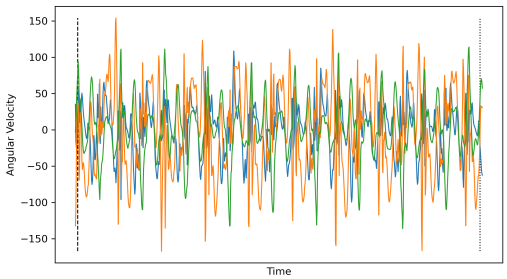


Figure 3.2: Measurement Setup: The Cosmed K5 portable Metabolic (the gas analyser is worn on the back and the face mask covers the mouth and nose), the Bio Harness 3 HR and BR device is worn under the shirt, and the Shimmer 3 IMU sensors are attached at five locations on the athlete's bodies.

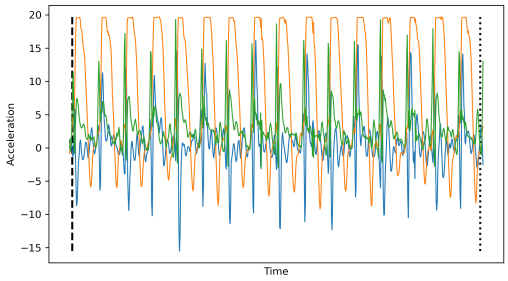
breath average can correct for a residual error in VO_2 data sets to within 10% of the raw variability (Robergs et al., 2010). Smoothing VO_2 with a 31-point moving average window to reduce interference noise was adopted for this exercise (Z. Wang et al., 2022). The calculation of the maximum value from the 31-point moving average window of the VO_2 signal was repeated for comparison. The treadmill speed ($km\,h^{-1}$) for each stage of the submaximal and maximal test was added to the VO_2 data for Visit 1; for the outdoor test, the GNSS speed ($km\,h^{-1}$) recorded on the Cosmed K5 device was used to determine the speed of outdoor running. The Activity Four class labels (Resting, Treadmill Running, Outdoor Running, Simulated Team Sports Circuit) to describe movement during the protocol were engineered and added to the VO_2 data. The physical characteristics of the subject, age (years), height (cm), weight (kg), resting oxygen uptake (VO_2 rest) and VO_2 max were included as features. The Zephyr data (1 Hz) were processed and merged into the breath-by-breath VO_2 data. The 5 Shimmer IMU data files were merged into one single file. To achieve this, the data was resampled from 250 Hz to 125 Hz to facilitate matching times. This raw IMU data was merged with the breath-by-breath



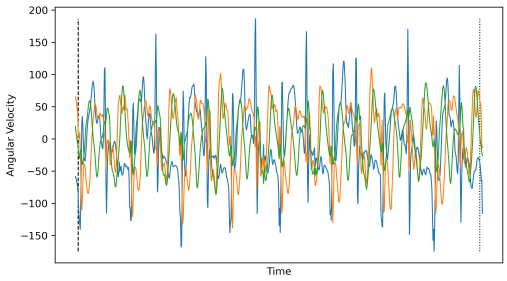
(a) Accelerometer Data TM Running



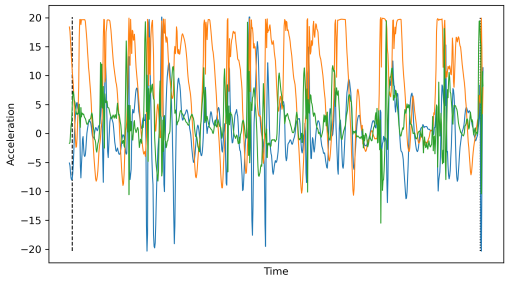
(b) Gyroscope Data TM Running



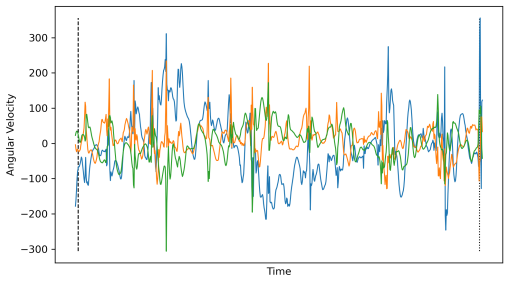
(c) Accelerometer Data Outdoor Running



(d) Gyroscopec Data Outdoor Running



(e) Accelerometer Data Simulated Circuit



(f) Gyroscope Data Simulated Circuit

Figure 3.3: Examples of IMU raw data from the Accelerometer and Gyroscope over a 3-breath window during three different activities: Treadmill Running (TM), Outdoor Running, and a Simulated Circuit.

VO_2 data to preserve its frequency. The IMU data are marked by windows of each breath recorded in the VO_2 data, and an example of the data can be seen in Figure 3.3. Due to issues with two sensors (right arm and left leg) during different sessions that failed to record during the experiment, these two of the five IMU sensors were removed from the final data. Data were used only from the torso, right

tibia, and left wrist.

3.3 Missing Data Patterns and Assumptions

Understanding the nature of missing data is essential in ensuring the validity of subsequent machine learning analyses. Three common mechanisms of missingness are typically considered in applied data science:

- **Missing Completely at Random (MCAR)**: The probability of missing data on a variable is unrelated to any observed or unobserved data.
- **Missing at Random (MAR)**: The probability of missingness is systematically related to observed data but not to the missing data itself.
- **Missing Not at Random (MNAR)**: The probability of missingness is related to the unobserved (missing) data, making it the most difficult to correct.

In this study, missing values primarily arose due to sensor dropout, file corruption, or missing concurrent RPE entries. To assess the pattern of missingness, exploratory analysis was performed using visual inspection (e.g., heatmaps of missing values) and statistical summaries. Most missing data were confined to a small subset of files, often clustered around match contexts or faulty sensor instances.

Given that missingness was not systematically associated with particular values of RPE, or player-level characteristics, the data is assumed to be **Missing at Random (MAR)**. For example, RPE was sometimes unrecorded after training sessions, but this was independent of the session's training load or performance metrics.

Where feasible, missing values were handled using the following strategies:

- Exclusion of incomplete records during feature selection and model training stages.
- For time-series derived features (e.g., in Experiment 2), sessions with partial signal loss were excluded to preserve integrity of derived metrics.

This assumption and handling strategy align with accepted practice in sports science research using wearable data (Van Buuren, 2018) and ensure that model results were not biased by imputation-based distortions or data leakage.

3.4 Summary

To support the machine learning experiments in this research, it was crucial to generate high-quality datasets under controlled conditions, applying rigorous methodologies to ensure the validity of the results and subsequent discussions. This chapter detailed the methodologies employed, including the selection of subjects, the design of data collection protocols, and the deployment of advanced hardware to capture relevant metrics with precision and consistency.

Key processes for cleaning and preparing the collected data were outlined, establishing a solid foundation for the feature extraction and analytical approaches presented in later chapters. The methodologies and sensor systems discussed here, including GNSS and IMU devices, informed the experimental designs that underpin the thesis, ensuring reliable insights into athlete monitoring.

By adhering to stringent data collection standards, this chapter provides the necessary groundwork for the predictive modelling and analyses explored in subsequent chapters, demonstrating a commitment to methodological rigour and data quality.

Chapter 4

Predicting Ratings of Perceived Exertion

4.1 Introduction

This chapter presents the first study of this thesis, which focusses on the prediction of Rating of Perceived Exertion (RPE) in elite Gaelic football players using traditional Machine Learning (ML) models. The study addresses key literature gaps, particularly the limited integration of relative External Load Indices (ELI), contextual factors, and individual characteristics into RPE prediction models. By establishing a baseline for RPE prediction, this chapter establishes the foundation for developing more advanced models in subsequent research. The primary objective of this study is to compare the predictive accuracy of absolute and relative ELI in various ML algorithms. Furthermore, whether incorporating variables such as personal characteristics (e.g., body mass, fitness level), perceived wellness scores (e.g., muscle soreness, sleep quality), and contextual factors (e.g., match schedules, recovery days) improve model performance is investigate. This comprehensive approach is designed to capture the complex interplay between external and internal training loads, improving the understanding of how athletes respond to planned workloads. A supervised ML approach is employed to identify relationships between predictor

variables and RPE outcomes, allowing predictions for new unseen data. In the context of elite Gaelic football, new data may be collected in real-time from live Global Navigation Satellite Systems (GNSS) and sensor input during training sessions. Two predictive modelling approaches are considered: regression, to estimate continuous RPE values, and classification, to categorise RPE into discrete levels. This dual approach allows for a more nuanced understanding of player responses to training loads and facilitates actionable insights for workload management.

The chapter is organised to provide a clear and systematic presentation of the study. Section 4.1 introduces the background and objectives, highlighting the importance of accurate prediction of RPE to optimise training load management. Section 4.2 details the experimental approach, feature engineering and the implementation of ML models. Section 4.3 presents the results of the experiments, highlighting key findings such as the comparative performance of absolute versus relative ELI and the impact of integrated variables on predictive precision. Section 4.4 interprets the results in the context of existing research, discusses practical implications for Gaelic football training, and addresses potential limitations. Section 4.5 concludes the chapter with a summary of key insights, contributions to the field, and recommendations for future research. Data acquisition procedures, including participant recruitment, GNSS data collection and preprocessing methods are detailed in Chapter 3, Section 3.2.1.

This study contributes to the growing field of sports analytics by providing practical tools to predict RPE in elite Gaelic football players. By comparing a variety of classification and regression models, this research identifies the most effective algorithms and features for RPE prediction. The findings aim to support coaches and sports scientists in designing individualised training interventions, optimising athlete performance and managing workloads more effectively. Furthermore, this study demonstrates how real-time sensor data can be integrated into predictive frameworks, paving the way for dynamic and responsive load management of training.

4.2 Methodology

Feature Engineering

Statistical summaries were generated for all features to identify potential inconsistencies, such as missing values or outliers. Only complete GNSS records with associated RPE data were included in the analysis to ensure consistency between all observations. The missing values for the performance of the Maximal Aerobic Speed (MAS) (4%) were imputed using the mean of the group. This approach was adopted to maintain sample size and ensure comparability between players. Regular audits were conducted throughout the data cleaning process to verify the accuracy and consistency of the data set before analysis. Correlations tests were performed on all GNSS indices before inclusion in the model. Figure 4.1 shows the correlation matrix of the GNSS metrics, with several indices showing high correlation values.

To mitigate multicollinearity, variables with a Variance Inflation Factor (VIF) greater than five were removed sequentially. Expert knowledge was used to retain features with the highest practical relevance (Hopkins et al., 2009). While VIF filtering supports interpretability and model stability, we acknowledge that it may result in the removal of informative but correlated features.

After this process, the number of GNSS indices was reduced from 42 to 4. The final retained indices included the total distance, the high-speed distance ($\geq 4.72\,\mathrm{m \cdot s^{-1}}$), and the number of accelerations and decelerations ($\geq 3\,\mathrm{m \cdot s^{-2}}$). These four GNSS indices were engineered to be expressed in relative terms (distance or events per unit of time) to allow direct comparison with their absolute values, as shown in Table 4.1. Following correlation tests of the 21 Acute Chronic Workload Ratio (ACWR) indices, four ACWR indices were retained for further analysis: total distance of 7 days, total distance of 28 days, total distance of ACWR, and high-speed distance of ACWR.

RPE is the target variable; this is reported by athletes on a scale from 0 to 10. For classification tasks, the RPE values were categorised into three intensity zones:

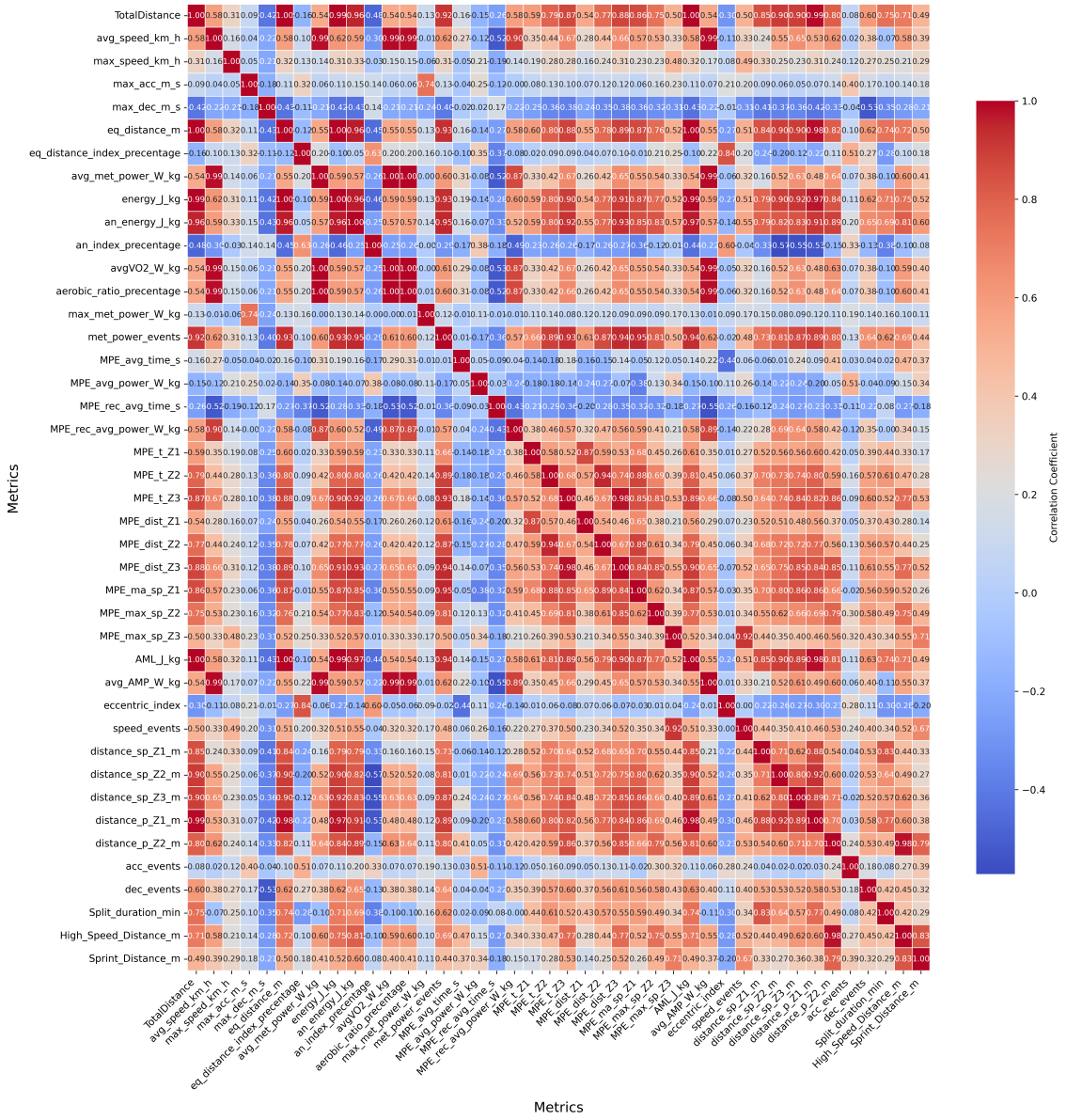


Figure 4.1: Correlation matrix of training load metrics. The heatmap shows pairwise correlations between GNSS data variables.

Absolute ELI	**Relative ELI**
Total Distance (m)	Relative Distance ($\mathrm{m} \cdot \mathrm{min}^{-1}$)
HSD ($\geq 4.72\,\mathrm{m} \cdot \mathrm{s}^{-1}$)	Relative HSD ($\geq 4.72\,\mathrm{m} \cdot \mathrm{s}^{-1}$)
Accelerations (n)	Relative Accelerations ($\mathrm{n} \cdot \mathrm{min}^{-1}$)
Decelerations (n)	Relative Decelerations ($\mathrm{n} \cdot \mathrm{min}^{-1}$)

Table 4.1: GNSS Absolute and Feature Engineered Relative External Load Indices (ELI). HSD: High-Speed Distance

low (0-5), moderate (6-7) and high (8-10) (Moreira et al., 2015). The target variable (RPE categories) was encoded using **LabelEncoder** to convert categorical labels (low, moderate, high) into numerical values.

Numerical features were normalised using the **StandardScaler** to ensure comparable scales, mainly because ML models are sensitive to feature scales. Given the imbalance in the RPE categories for the classification task, we under-sampled the larger classes to match the size of the smallest class. This ensured balanced training for the classification models and reduced bias towards the majority classes.

A set of contextual variables was generated and included to provide information on the data and time of each training session or match, session time (am or pm), day of the week, month and season. The type of activity (training, match) and the number of days to and from the next and previous match respectively were also included. Personal characteristics such as age, height, weight, body composition, fitness level, player position and playing experience were included. Any categorical variables were encoded using one-hot encoding to transform them into numerical features suitable for ML models. This ensured that categorical variables were represented without introducing ordinal relationships.

To investigate the performance of absolute and relative ELI in estimating RPE in various ML models, 4 combinations of predictor variables were tested for each modelling approach these are set our in Table 4.2. The combinations were chosen to investigate the predictive ability of absolute and relative ELI categories when used alone and with contextual variables.

Attribute Set	**Feature Set Categories**
1	Absolute ELI + Contextual variables
2	Relative ELI + Contextual variables
3	Absolute ELI
4	Relative ELI
5	Contextual variables

Table 4.2: Feature Sets for Predictive Models

4.2.1 Experimental Setup

A supervised ML approach was used to examine the performance of ELI, with and without the inclusion of contextual variables, in the prediction of RPE. These experiments use ML techniques to develop predictive models that estimate the RPE of an athlete. Given the complexity of athlete monitoring data, including physiological, positional, and contextual information, ML methods were selected to model and predict continuous RPE values (regression) and categorical RPE zones (classification). The evaluation of various ML models aims to identify the most effective approach to predicting RPE and understand the contribution of individual features to these predictions. In order to identify the most effective algorithms and features for RPE prediction from this first experiment, different ML functions were used.

The ML models used for this task are described in detail in Section 3.1.4 of the previous chapter. For classification tasks, which involve predicting categorical RPE zones, the models tested included Logistic Regression, Decision Tree (DT) Classifier, Random Forest (RF) Classifier, Bootstrap Aggregating (Bagging) Classifier, eXtreme Gradient Boosting (XGBoost) Classifier, Gradient Boosting (GB) Classifier and Light Gradient Boosting Machine (LightGBM) Classifier. For regression tasks, aimed at predicting continuous RPE values, the models included Linear Regression, DT Regressor, RF Regressor, XGBoost Regressor, GB Regressor and LightGBM Regressor. These models were evaluated to determine their ability effectively to handle the nuances of both tasks.

]

4.2.2 Model Training and Evaluation

The dataset was split into training and testing sets, with 80% of the data used for training and 20% reserved for testing. This ensured that model performance was evaluated on unseen data to prevent overfitting and to assess generalisability. **Feature Importance** was calculated for models that support this functionality (e.g., DT, RF, GB, XGBoost and LightGBM). This analysis revealed which features con-

tributed the most significantly to the predictions, providing insight into the most relevant factors influencing RPE. **Classification Metrics:** The models were evaluated using **accuracy**, **precision**, **recall**, and **F1-score** to assess their performance in correctly predicting RPE zones. **Regression Metrics:** The models were evaluated using **Mean Squared Error (MSE)**, **Mean Absolute Error (MAE)** and **Root Mean Square Error (RMSE)** which measure the precision of the model in predicting continuous RPE values.

Variable	**Low (Mean ś SD)**	**Moderate (Mean ś SD)**	**High (Mean ś SD)**	**ANOVA p-value**
Age (yr)	26.4 ś 4.3	25.7 ś 3.9	24.7 ś 3.7	<0.001
Lean mass index $(\mathrm{m} \cdot \mathrm{kg}^2)$	20.9 ś 1.3	20.8 ś 2.1	20.7 ś 1.9	0.348
Body fat percentage (%)	14.3 ś 4.1	14.2 ś 3.9	14.4 ś 4.1	0.804
Maximal aerobic speed $(\mathrm{m} \cdot \mathrm{s}^{-1})$	4.7 ś 0.2	4.7 ś 0.2	4.7 ś 0.2	0.004
Duration (min)	63.3 ś 20.9	73.4 ś 20.3	86.0 ś 25.3	<0.001
Total distance (m)	4958.8 ś 1454.7	6273.1 ś 1812.6	8178.0 ś 2966.2	<0.001
Relative distance $(\mathrm{m} \cdot \mathrm{min}^{-1})$	81.8 ś 19.4	88.8 ś 20.2	95.3 ś 23.4	<0.001
High-speed distance $(\geq 4.72\,\mathrm{m} \cdot \mathrm{s}^{-1})$	764.0 ś 453.2	962.0 ś 522.5	1255.5 ś 720.5	<0.001
Relative high-speed distance $(\geq 4.72\,\mathrm{m} \cdot \mathrm{s}^{-1})$	12.7 ś 7.3	14.1 ś 8.3	15.5 ś 9.1	<0.001
Accelerations (n)	9.4 ś 6.2	10.9 ś 7.8	11.7 ś 7.4	<0.001
Relative accelerations $(\mathrm{n} \cdot \mathrm{min}^{-1})$	0.2 ś 0.1	0.2 ś 0.2	0.2 ś 0.1	0.597
Decelerations (n)	6.0 ś 4.5	9.6 ś 6.0	13.3 ś 7.7	<0.001
Relative decelerations $(\mathrm{n} \cdot \mathrm{min}^{-1})$	0.1 ś 0.1	0.1 ś 0.1	0.2 ś 0.1	<0.001
Muscle soreness (AU)	4.8 ś 2.6	4.8 ś 2.7	4.9 ś 3.0	0.875
Sleep quality (AU)	7.9 ś 1.2	7.9 ś 1.1	8.0 ś 1.2	0.169
Energy level (AU)	7.7 ś 1.1	7.7 ś 1.1	8.0 ś 1.2	<0.001
7-d total distance (m)	15066.6 ś 6615.5	15653.5 ś 6223.2	16970.1 ś 6832.5	<0.001
28-d total distance (m)	47655.2 ś 22697.6	42589.2 ś 21329.8	41028.6 ś 23624.2	<0.001
ACWR total distance	1.0 ś 0.2	1.0 ś 0.2	1.0 ś 0.2	<0.001
ACWR high-speed distance	1.0 ś 0.3	1.0 ś 0.3	1.1 ś 0.3	<0.001
Days to last match	4.4 ś 4.8	2.6 ś 3.9	1.3 ś 3.7	<0.001
Days to next match	4.7 ś 4.0	7.8 ś 6.2	9.5 ś 6.0	<0.001

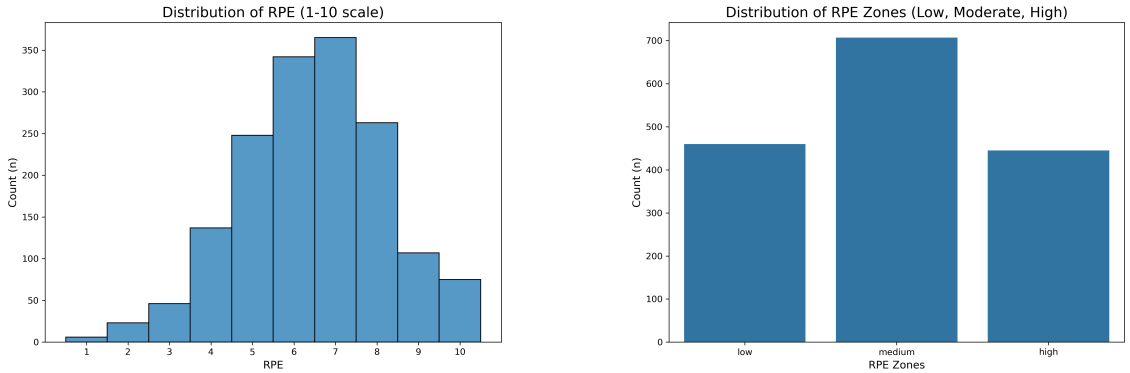
Table 4.3: Summary of variables with mean and standard deviation across low, moderate, and high categories, along with ANOVA p-values.

4.3 Results

Table 4.3 presents the mean ś SD of each model variable between the low, moderate and high RPE categories, along with the corresponding ANOVA p-values. This table illustrates the distribution and variation of key variables in the dataset used for modelling, highlighting significant differences between RPE categories. These differences provide insight into the potential relevance of these variables as predictors in the models.

Figure 4.2b shows the distribution of RPE scores on a scale of 1-10, providing an overview of how players rate their perceived exertion during training sessions. This distribution highlights the variability in perceived exertion across the dataset.

Figure 4.2a shows the distribution of the RPE zones classified as low, moderate and high, which are used for further analysis and modelling. These figures offer insight into the range and frequency of exertion levels,



(a) RPE (1-10 scale)

(b) RPE Zones (Low, Moderate, High)

Figure 4.2: Distributions of RPE and RPE Zones

4.3.1 Classification Models

The classification models to predict RPE were evaluated using accuracy, precision, recall, and F1 score metrics, as summarised in Table 4.4. The Bagging classifier demonstrated the best overall performance using attribute set 1, achieving an accuracy of 69%, a precision of 68.6%, a recall of 69%, and an F1 score of 68.7%. The XGBoost classifier, using the attribute set 2, was second with an accuracy of 67.8% and a precision of 67.4%. It had a comparable recall of 67.9% and an F1-score of 67.5%; its precision was slightly lower than the Bagging classifier. Among the models with moderate performance, such as LightGBM and RF classifiers, more balanced F1 scores were observed, though they did not surpass ensemble models such as Bagging classifier. The LightGBM classifier using attribute set 2 achieved an accuracy of 66.3% with a relatively high F1 score of 66.2%, indicating that it provided consistent predictions. However, it was not as strong as Bagging classifier in handling varying RPE categories. A deeper look into the other metrics revealed further insights for models with lower accuracy, such as Logistic Regression and DT. Logistic regression, for example, had the lowest accuracy 44.5% in attribute set 4,

but its precision (42.6%) and recall (44.6%) indicated poor prediction stability. The F1-score of 42.8% also highlighted that Logistic Regression struggled to make accurate predictions across different RPE zones. The DT model, which had an accuracy 46.4% in attribute set 3, had a relatively higher recall of 47.3%. This suggests that although the model was less precise in making accurate predictions, it captured a higher number of true positives, particularly in the high RPE range, but at the expense of increased false positives, as indicated by the lower precision. In general, the analysis of precision, recall, and F1-score highlights the superiority of ensemble methods like Bagging and XGBoost classifiers, which consistently outperformed individual models, such as Logistic Regression and DT classifier, particularly in their ability to maintain balanced precision and recall across different RPE categories. This balance is critical in athlete monitoring, where minimising false positives and negatives is essential for accurate training load management.

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	AUROC (Macro)	Attribute Set	Overall Rank
Bagging	0.6929	0.6910	0.6929	0.6917	0.7697	Attr_set2	1
Bagging	0.6891	0.6920	0.6891	0.6904	0.7669	Attr_set1	2
XGBoost Classifier	0.6854	0.6995	0.6854	0.6900	0.7640	Attr_set1	3
XGBoost Classifier	0.6854	0.6964	0.6854	0.6887	0.7640	Attr_set2	3
Gradient Boosting Classifier	0.6816	0.6896	0.6816	0.6844	0.7612	Attr_set1	4
LightGBM Classifier	0.6742	0.6837	0.6742	0.6771	0.7556	Attr_set1	5
LightGBM Classifier	0.6704	0.6698	0.6704	0.6700	0.7528	Attr_set2	6
Gradient Boosting Classifier	0.6667	0.6789	0.6667	0.6704	0.7500	Attr_set2	7
Bagging	0.6554	0.6514	0.6554	0.6527	0.7416	Attr_set5	8
Logistic Regression	0.6442	0.6393	0.6442	0.6406	0.7331	Attr_set2	9
Random Forest	0.6367	0.6340	0.6367	0.6334	0.7275	Attr_set2	10
Logistic Regression	0.6367	0.6307	0.6367	0.6312	0.7275	Attr_set1	11
Bagging	0.6330	0.6374	0.6330	0.6341	0.7247	Attr_set3	12
Gradient Boosting Classifier	0.6255	0.6304	0.6255	0.6275	0.7191	Attr_set5	13
Random Forest	0.6217	0.6229	0.6217	0.6219	0.7163	Attr_set1	14
Gradient Boosting Classifier	0.6180	0.6232	0.6180	0.6193	0.7135	Attr_set3	15
XGBoost Classifier	0.6142	0.6182	0.6142	0.6143	0.7107	Attr_set5	16
LightGBM Classifier	0.6105	0.6079	0.6105	0.6078	0.7079	Attr_set5	17
Logistic Regression	0.6067	0.6030	0.6067	0.6018	0.7051	Attr_set5	18
Logistic Regression	0.6030	0.6022	0.6030	0.5977	0.7022	Attr_set3	19
Random Forest	0.5918	0.5862	0.5918	0.5874	0.6938	Attr_set3	20
Random Forest	0.5805	0.5792	0.5805	0.5794	0.6854	Attr_set5	21
Decision Tree	0.5805	0.5897	0.5805	0.5836	0.6854	Attr_set5	22
Decision Tree	0.5543	0.5595	0.5543	0.5562	0.6657	Attr_set1	23
XGBoost Classifier	0.5506	0.5526	0.5506	0.5513	0.6629	Attr_set3	24
LightGBM Classifier	0.5506	0.5491	0.5506	0.5494	0.6629	Attr_set4	25
Random Forest	0.5506	0.5456	0.5506	0.5467	0.6629	Attr_set4	26
Decision Tree	0.5468	0.5452	0.5468	0.5450	0.6601	Attr_set2	27
Gradient Boosting Classifier	0.5468	0.5460	0.5468	0.5440	0.6601	Attr_set4	28
LightGBM Classifier	0.5393	0.5513	0.5393	0.5434	0.6545	Attr_set3	29
Bagging	0.5318	0.5281	0.5318	0.5270	0.6489	Attr_set4	30
Decision Tree	0.5131	0.5091	0.5131	0.5101	0.6348	Attr_set3	31
Logistic Regression	0.5131	0.5040	0.5131	0.4926	0.6348	Attr_set4	32
XGBoost Classifier	0.4944	0.4995	0.4944	0.4962	0.6208	Attr_set4	33
Decision Tree	0.4906	0.4911	0.4906	0.4900	0.6180	Attr_set4	34
Dummy Classifier	0.3333	0.1111	0.3333	0.1667	NaN	Attr_set5	34
Dummy Classifier	0.3333	0.1111	0.3333	0.1667	NaN	Attr_set3	34
Dummy Classifier	0.3333	0.1111	0.3333	0.1667	NaN	Attr_set4	34
Dummy Classifier	0.3333	0.1111	0.3333	0.1667	NaN	Attr_set2	34
Dummy Classifier	0.3333	0.1111	0.3333	0.1667	NaN	Attr_set1	34

Table 4.4: Evaluation metrics for classification models predicting RPE across different attribute sets. AUROC is macro-averaged.

4.3.2 Regression Models

The performance of the regression models was evaluated using the MSE, the MAE and the RMSE as summarised in Table 4.5, which provide insight into both the average magnitude of errors and the variance in predictions. The best-performing model was the LightGBM Regressor using attribute set 2, with an MSE of 1.69, RMSE 1.30 and an MAE of 0.94. Another LightGBM Regressor with attribute set 1 ranked second, with an MSE of 1.71, RMSE 1.31 and an MAE of 0.93. The XGBoost Regressor and GB Regressor models also performed well, with MAE values of 0.95 and 0.97, respectively. The XGBoost Regressor and GB Regressor models also performed well. While their RMSE values were higher than LightGBM's, they maintained competitive MAE scores, indicating that these ensemble-based methods were effective but introduced slightly more variability in prediction errors. Notably, the GB Regressor's MAE of 0.97 suggests that its prediction errors were somewhat larger on average than those of the LightGBM Regressor. Traditional models such as Linear Regression performed less favourably in this context. The Linear Regression model using attribute set 1 had an MSE of 2.12. an RMSE 1.46 and an MAE of 1.08, highlighting its difficulty in accurately predicting RPE values. DTs showed considerably higher error rates. The DT Regressor using attribute set 4 had an MSE of 4.69 and an MAE of 1.63, indicating that this model made substantial errors in its predictions. In summary, ensemble methods such as LightGBM and XGBoost Regressors demonstrated superior performance in predicting RPE, with lower MSE and MAE values. Traditional methods like linear regression exhibited higher error rates, particularly when handling the complex relationships between input features and RPE. The results suggest that advanced ML models, particularly those that utilise boost techniques, are more suitable to accurately predict RPE in this context.

4.3.3 Feature Importance

Combined Feature Importance Classification and Regression

In Table 4.6, the grouped and descriptively labeled importance scores for all classification and regression models provide an overview of which features were most influential in predicting RPE.

For classification models, Total Distance (m) remained the most important feature, with an average importance score of 0.272. Relative Total Distance ($\mathrm{m} \cdot \mathrm{min}^{-1}$) and Relative Decelerations ($\mathrm{n} \cdot \mathrm{min}^{-1}$) were also highly influential, with importance scores of 0.171 and 0.142, respectively. Accelerations (n), Decelerations (n), and their relative counterparts ($\mathrm{n} \cdot \mathrm{min}^{-1}$) also ranked highly.

For regression models, Total Distance (m) dominated the feature importance, with a grouped average score of 0.388, significantly higher than all other features. Relative Total Distance ($\mathrm{m} \cdot \mathrm{min}^{-1}$) ranked second with an importance score of 0.187. Relative Decelerations ($\mathrm{n} \cdot \mathrm{min}^{-1}$) and Relative Accelerations ($\mathrm{n} \cdot \mathrm{min}^{-1}$) followed, with importance scores of 0.144 and 0.132, respectively. Other variables such as High-Speed Distance ($\geq 4.72\,\mathrm{m} \cdot \mathrm{s}^{-1}$), Days Until Next Match, and Decelerations (n) contributed meaningfully, reflecting the influence of both load and contextual features.

These results reflect the aggregated model-based feature importances only. A more detailed comparison using SHAP values for regression models is presented in the next section.

Feature Importance Heat Map

Figure 4.3 is a heat map that illustrates the normalised importance scores for all classification models in different attribute sets. This visual representation highlights the key features that contributed to the predictive performance of the classification models for each set of attributes. Total Distance (m) continues to stand out as a dominant feature, particularly in Set 1, Set 4, and now Set 5, where its importance ranges between 0.20 and 0.40. Relative Accelerations ($\mathrm{n\mathring{u}min}^{-1}$) and Relative High-

Speed Distance ($\geq 4.72\,\mathrm{m} \cdot \mathrm{s}^{-1}$) also remain influential, with Set 4 showing peaks near 0.30, and Set 5 reinforcing their relevance with scores above 0.20.

Figure 4.4 displays a heat map of the combined normalised importance scores for regression models by attribute set. **Total Distance (m)** remains the most dominant feature, with values reaching 0.50 in Set 3 and exceeding 0.28 in Set 5. Notably, Relative Accelerations ($\mathrm{n\mathring{u}min}^{-1}$), Relative High-Speed Distance, and Decelerations (n) also show strong importance in Set 5, with values around 0.240.25, indicating their growing contribution in the more comprehensive feature combinations.

SHAP Feature Importance Regression Models

To support model-derived insights with a model-agnostic interpretability method, Figure 4.5 presents the combined SHAP importance values for all regression models. This provides a more nuanced view of which features consistently influenced RPE predictions across models. As shown, **Total Distance (m)** again emerges as the most impactful feature. Additional top-ranked features include **Days Until Next Match**, **Relative Total Distance**, **Decelerations**, and **Relative Decelerations**. Contextual variables such as **Age**, **Match Activity**, and **Lean Mass Index (LMI)** also contribute meaningfully, reaffirming the importance of both absolute load and contextual readiness indicators in estimating RPE. The strong alignment between SHAP-based and model-derived importance enhances confidence in the robustness and generalisability of the regression results.

Feature Importance Top Models

The normalised importance scores provide information on which features had the most significant influence on the model's prediction of RPE. Normalised importance scores for the best performing XGBoost classifier in Figure 4.6. Total Distance (m) was the most dominant feature. Match Activity, Days Until Next Match, and Days Since Last Match were considered highly important. Position: Midfield and Age (yrs) were prominent. AM Session and Decelerations (n) were significant as well.

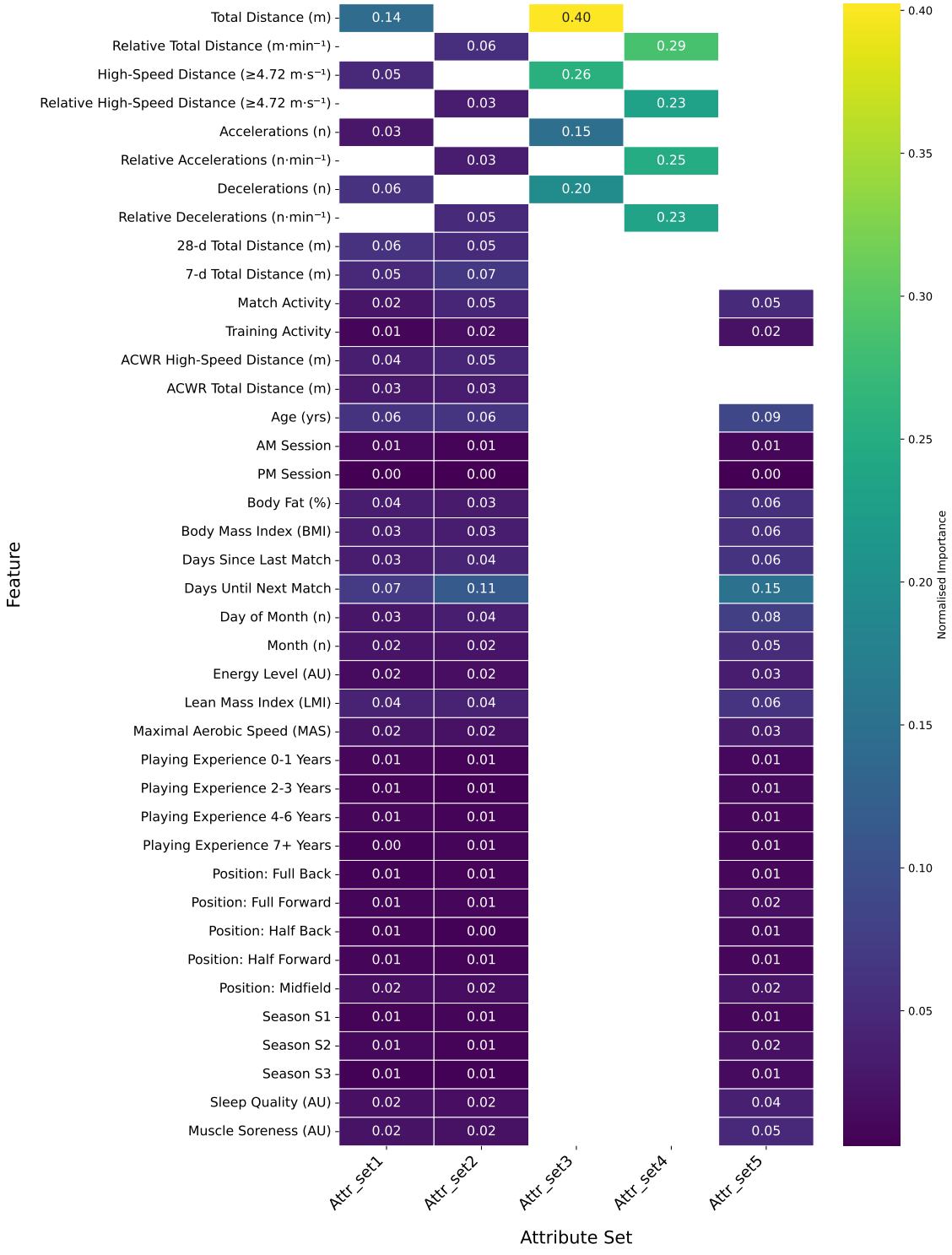


Figure 4.3: The normalised importance scores for all classification models across different attribute sets.

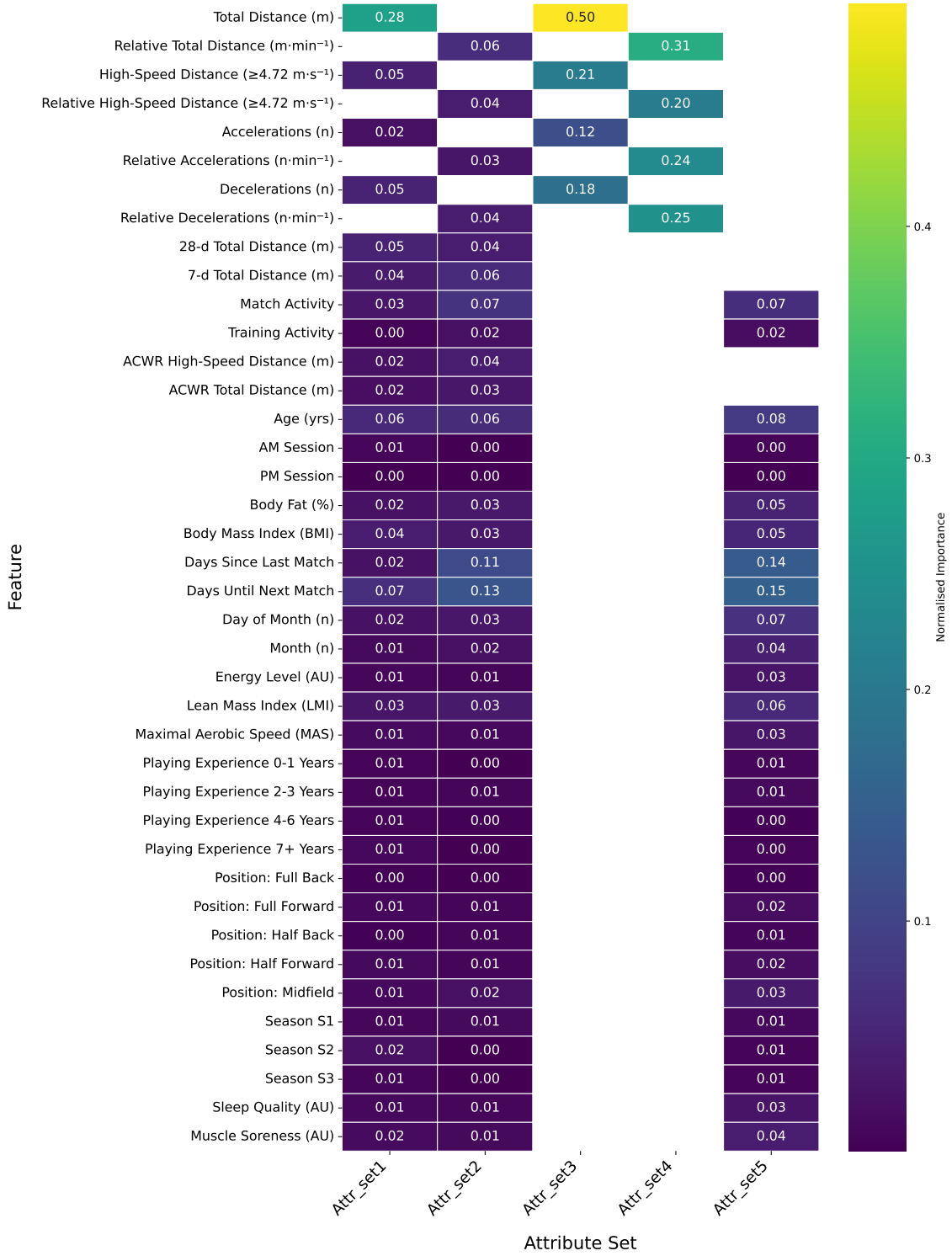


Figure 4.4: The normalised importance scores for all regression models across different attribute sets.

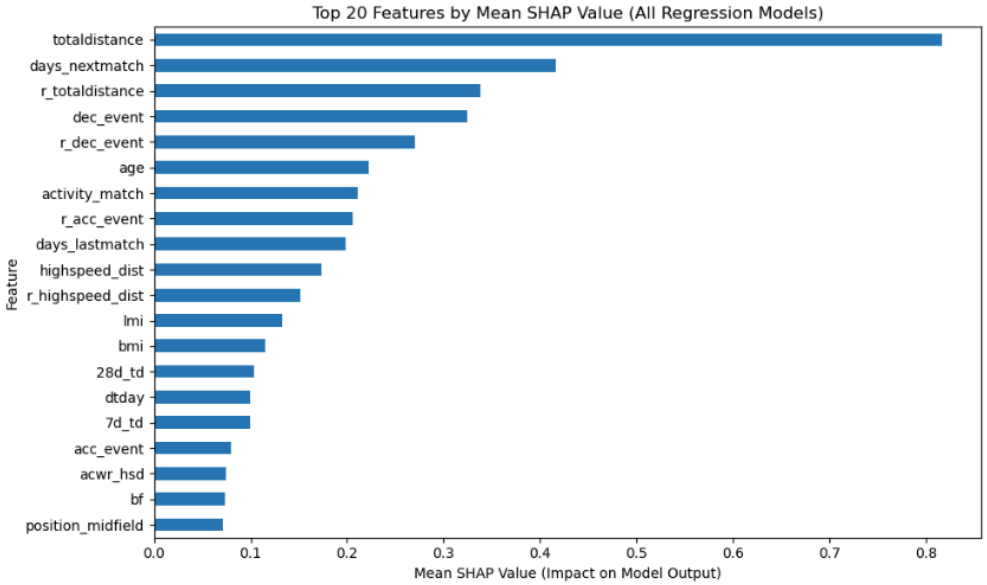


Figure 4.5: Top 20 features ranked by mean SHAP value across all regression models

Other variables such as High-Speed Distance ($\geq$ 4.72 můs^{-1}) and wellness metrics like energy level and sleep quality contributed to the model, albeit with lower impact.

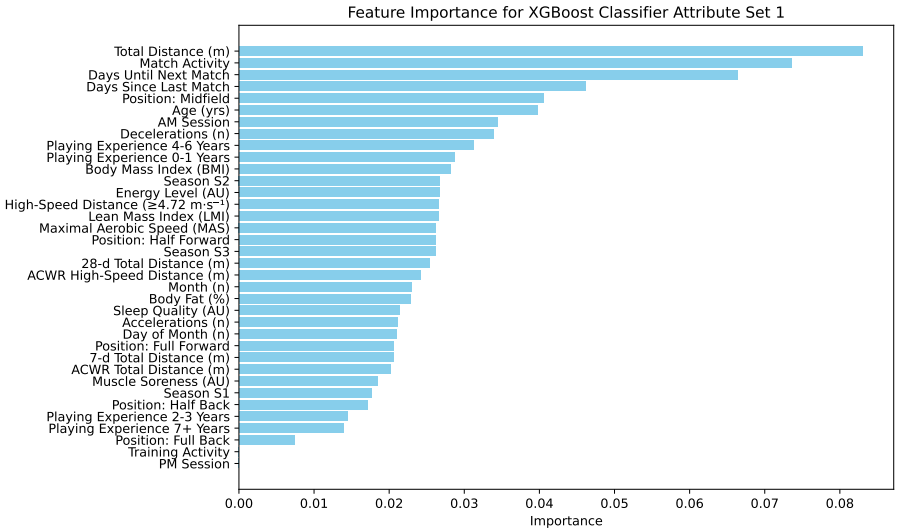


Figure 4.6: Normalised Importance scores for the best-performing XGBoost Classifier

The importance of features for the best performing LightGBM Regressor is presented in Figure 4.7. Total Distance (m) and 7-day Total Distance (m) consistently emerged as the most influential predictors of RPE response, with High-Speed Distance ($\geq$ 4.72 můs^{-1}) and 28-day Total Distance (m) also scoring highly. Age (yrs)

and total ACWR distance (m) were next in importance. Additional relevant variables included BMI, Lean Mass Index (LMI), Day of Month, and Decelerations (n).

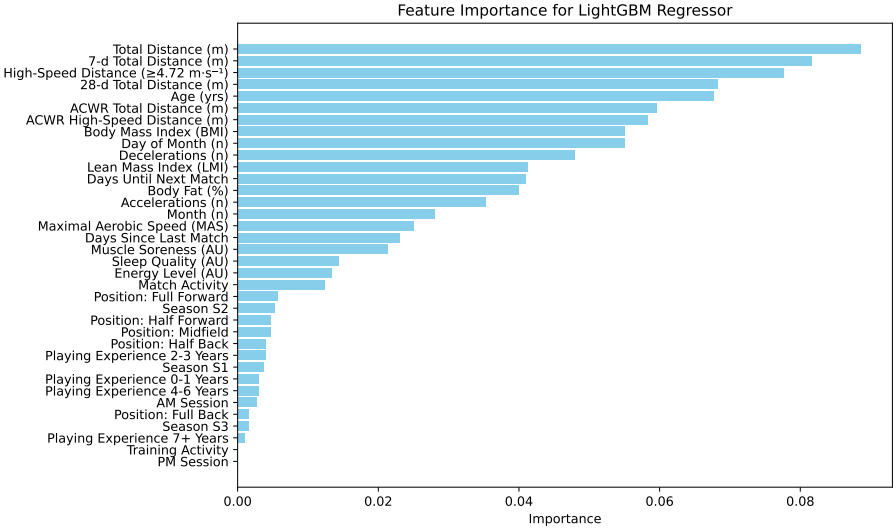


Figure 4.7: Normalised Importance scores for the best-performing LightGBM Regressor

To further validate the model-derived importance, SHAP values were computed for the best LightGBM regressor using Attribute Set 1 (Figure 4.8). The SHAP analysis closely aligned with model-derived rankings. Total Distance (m), 7-d Total Distance (m), and High-Speed Distance retained high SHAP values, confirming their predictive utility. Furthermore, variables like LMI, Day of Month, and Accelerations (n) showed moderate SHAP values, reinforcing their influence despite appearing lower in traditional importance rankings. This agreement strengthens the interpretability of the LightGBM model and supports the inclusion of both training load and contextual features.

Figure 4.9 illustrates the classification performance of the Bagging model across all five attribute sets. The model performed well in predicting **moderate RPE (67)**, with precision percentages consistently above 64%, peaking at **79.78%** in **Attribute Set 1** (absolute ELI + contextual variables). This indicates strong model sensitivity to moderate levels of perceived exertion, likely due to the broader data distribution in this zone. Similarly, the model showed solid performance for **low RPE (05)**, with **72.73%** accuracy in Set 1 and **69.66%** in Set 2 (relative

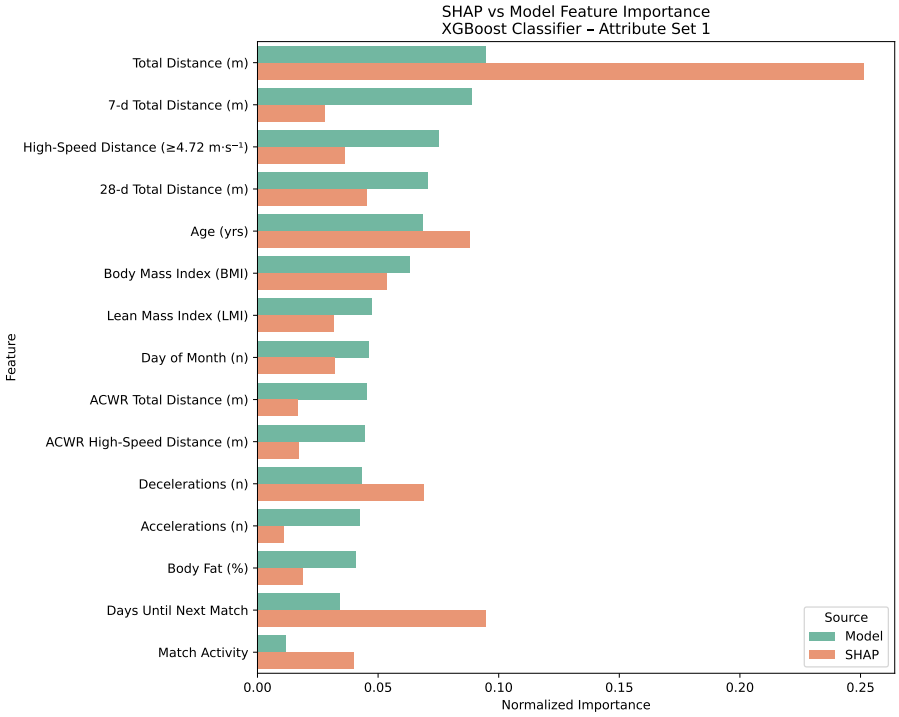


Figure 4.8: Comparison of SHAP values and model-derived feature importance for the best-performing LightGBM Regressor (Attribute Set 1)

ELI + contextual variables). Accuracy in predicting **high RPE (810)** was notably lower, falling to **50%** in **Attribute Set 4** and peaking at only **53.85%** in Set 1. Attribute Set 5 achieved moderate classification performance across all RPE zones, aligning closely with Set 3.

This suggests that while the Bagging model is generally strong in identifying moderate and low RPE, further refinement may be necessary to enhance high-RPE detection, potentially through additional contextual or physiological features.

In parallel, the scatter plot of the best-performing regression model (LightGBM with Attribute Set 1) demonstrates strong predictive alignment with actual RPE values. This indicates that the regression models are capable of capturing fine-grained variation in perceived exertion, particularly in the mid-to-high RPE ranges.

4.4 Discussion

For classification, Attribute Set 1 (Absolute ELI + contextual variables) is the top-performing configuration, achieving the highest accuracy (69%) and precision

(68.6%) when used with the Bagging model (Table 4.4). Attribute Set 5, all available contextual variables, achieved moderate performance, with its best model (Bagging) ranking 9th overall at 65.5% accuracy. Attribute Set 5 did not outperform the feature combinations such as Set 1 or Set 2. This suggests that absolute ELI (such as total distance and high-speed distance), combined with contextual variables (e.g., personal characteristics and wellness scores), are the most effective in predicting categorical RPE zones. Ensemble methods like Bagging and XGBoost outperformed traditional models, as they maintained balanced precision and recall, which is critical for effective load management.

For regression, Attribute Set 2 (Relative ELI + contextual variables) is the most effective, with the LightGBM Regressor achieving the lowest MSE of 1.69, RMSE of 1.30, and MAE of 0.94 (Table 4.5). Attribute Set 5s best model, also a LightGBM Regressor, ranked 10th overall with an MSE of 1.78, showing reasonable but not superior predictive power. SHAP analysis confirmed that key predictors such as Total Distance (m), 7-day Total Distance (m), and High-Speed Distance had the strongest influence, aligning with the model-derived rankings.

For both classification and regression tasks, advanced ML models such as LightGBM and XGBoost demonstrated superior performance in predicting RPE, especially when handling complex relationships between input features and RPE values. In contrast, traditional methods such as linear regression struggled to capture these complexities, leading to less accurate predictions.

Including a wide variety of variables related to athletic performance has been recommended to provide a more holistic approach (Geurkink et al., 2019; Rossi et al., 2019). Those chosen for inclusion in the present study were factors shown to influence RPE and activity performed during Gaelic football, such as body composition, player experience, physical fitness levels, cumulative training load, sleep quality and muscle soreness (Cullen et al., 2021; Daly et al., 2024; S. Malone et al., 2020) and those that have improved the predictive precision of previous models, such as individual characteristics and additional variables that contributed 4.533%

accuracy in soccer (Geurkink et al., 2019).

Contextual variables in conjunction with ELI significantly improved the predictive accuracy of all models in this study. Attribute Sets 1 and 2 consistently ranked among the best configurations. Attribute Set 5, designed to include only contextual variables, achieved moderate classification and regression performance. Its highest-performing classification model (Bagging) reached 65.5% accuracy, while the top regression model (LGBM) ranked 10th with an MSE of 1.78. This suggests that while contextual variables sets contribute to overall model robustness, they do not outperform combinations of ELI and contextual variables. SHAP analysis confirmed that features such as Total Distance (m), 7-day Total Distance (m), and High-Speed Distance were the most influential in the regression task, further validating the relevance of training load metrics.

These findings reinforce the importance of physiologically relevant variables over merely expanding the feature set. SHAP analysis further validated the dominance of key training load metricssuch as Total Distance, 7-day Total Distance, and High-Speed Distancehighlighting their central role in RPE prediction. These results underscore the significant improvement in predictive accuracy when incorporating personal characteristics, wellness scores, and training workloads alongside ELI and is in line with previous research (Geurkink et al., 2019; Rossi et al., 2019).

ML allows for the extraction of importance scores for the variables included in the models. The complexity of the algorithm and the number of variables considered directly influence these importance calculations. Understanding the contribution of each variable to the predictive accuracy of RPE can lead to more effective planning and control of training loads. Across both the classification (Figure 4.3) and regression (Figure 4.4) models, Total Distance (m) consistently emerged as the most influential feature. This was observed across all attribute sets, including the newly added Attribute Set 5. Notably, SHAP analysis further validated this finding by highlighting Total Distance (m), 7-day Total Distance, and High-Speed Distance as the most impactful predictors for the LightGBM Regressor (Figure 4.8). The volume

of training was found to have a greater impact on RPE than intensity, suggesting that the total amount of training is a stronger determinant of perceived exertion than the intensity of individual sessions.

This finding is consistent with previous research, in which total distance has been consistently identified as one of the strongest predictors of RPE when using ML models (Bartlett et al., 2017; Carey et al., 2016; Geurkink et al., 2019). Our results, supported by both traditional model-derived feature importance and SHAP-based interpretability, further support the view that the overall workloadas measured by the total distance covered in a session or matchis the strongest determinant of RPE. This is consistent with a recent meta-analysis, which found that the total distance had the strongest association with RPE (McLaren et al., 2018).

Features such as the Relative Total Distance ($\mathrm{m} \cdot \mathrm{min}^{-1}$) and Relative High-Speed Distance $\geq 4.72\,\mathrm{m} \cdot \mathrm{s}^{-1}$ consistently showed importance in multiple attribute sets. High-speed running ($\geq 4.0\,\mathrm{m} \cdot \mathrm{s}^{-1}$) (Bartlett et al., 2017) and relative high-speed running ($\geq 5.5\,\mathrm{m} \cdot \mathrm{s}^{-1}$) (Marynowicz et al., 2021) have been more influential in predicting RPE for some players. These differences may be due, in part, to variations in activity profiles between different team sports (Varley et al., 2014), as well as individual differences between players.

(Gaudino et al., 2015). A recent study examining the contribution of training intensity and duration to training load in rugby league and rugby union found that session duration represented a greater proportion of the total variance in training load compared to session intensity (Weaving et al., 2014). Additionally, the type of ML model and the high-speed running thresholds applied may also contribute to these observed differences.

In the present study, Relative ELI featuresincluding relative high-speed running and relative total distancewere consistently ranked among the top features in Attribute Sets 2 and 4. SHAP values confirmed that while absolute ELI features were dominant, several relative metrics contributed meaningfully to prediction. SHAP and model-based importance scores highlighted that Accelerations and Decelerations

also played a moderate but consistent role in predicting RPE, particularly in Sets 3 and 5. Decelerations have previously been identified as significant ELI contributing to the prediction of RPE (Jaspers et al., 2018).

The normalised importance scores for the best performing XGBoost classifier in Figure 4.6 highlight the role of contextual factors, such as the distance to official games and acute chronic workload features, which have been identified in previous research as key influences on RPE (Rossi et al., 2019; Vallance et al., 2023). Our analysis found that contextual match factors, playing position and age were also contributors, reflecting the impact of positional, individual, and contextual variables on the prediction of RPE, and well-being metrics playing a smaller role.

For the LightGBM regressor (Figure 4.7), 7-day and 28-day Total Distance (m) emerged as important features, reflecting the influence of both short- and long-term workloads on RPE. SHAP analysis (Figure 4.8) confirmed the model-based rankings, with Total Distance, High-Speed Distance, and 7-day load dominating the feature contributions. Variables such as BMI, LMI, Day of Month, and Accelerations were also highlighted by SHAP, supporting the inclusion of a diverse range of features in regression modelling. Attribute Set 5 included all available contextual and physiological variables, its regression models while competitive did not surpass the combinations in Sets 1 or 2.

These findings differ from previous studies that found acute workload features to have a greater impact on RPE (Rossi et al., 2019). A possible reason for these findings is that the study did not implement controls for multicollinearity, which can affect ELI (Bartlett et al., 2017; Jaspers et al., 2017). The analysis reveals a high degree of correlation between the ELI provided by the manufacturer's software as shown in Figure 4.1. While these metrics are useful for monitoring athlete workload, their inter-correlation can introduce multicollinearity, leading to biased or unstable predictions in ML models. To address this, both correlation analysis and VIF were applied to identify and mitigate multicollinearity, ensuring that the models provided accurate and interpretable RPE predictions without redundant information.

The findings of this study emphasise that, for both absolute and relative ELI, distance covered plays the most significant role in predicting RPE, while the contributions of high-speed running, accelerations, and decelerations are relatively similar. However, coaches and practitioners should be mindful of the differences in predictive power between absolute and relative ELI, as well as the shifting importance of individual variables when using these metrics to prescribe training and develop training programmes.

The improved accuracy observed when incorporating personal characteristics, wellness scores, and training workloads underscores that ELI alone do not fully capture or predict RPE. SHAP analysis provided additional validation by demonstrating that contextual and physiological factorssuch as BMI, Lean Mass Index, and training historycontributed meaningful predictive power.

These findings support the integration of machine learning interpretation techniques, such as SHAP, to better understand the role of both physical workload and individual characteristics in model predictions. This contributes to more informed and tailored approaches to load monitoring. A more holistic approachintegrating both physical and individual factors is therefore recommended for more accurate monitoring and load management.

In this study, RPE was divided into three categories, a method that has been used in previous research on intermittent team sports and endurance athletes (Lovell et al., 2013; Stellingwerf, 2012). The boundaries of these categories align with the first and second ventilatory thresholds (Seiler & Kjerland, 2006), which help distinguish between three distinct physiological exercise intensity domains (Jamnick et al., 2020). This classification is particularly practical for coaches, as the physiological adaptations vary across intensity domains, offering valuable insight into the athlete's response to training.

When using RPE to predict a player's response to training, a polarised training approach, consisting of low-volume, high-intensity efforts and high-volume, lo

4.5 Summary

In this chapter, a range of ML models were evaluated to predict RPE in a cohort of elite Gaelic football players. The findings demonstrated that absolute ELI were more effective predictors of RPE compared to relative ELI, highlighting the critical influence of training volume over intensity in determining perceived exertion.

Incorporating additional variables such as personal characteristics, wellness scores, and training workloads significantly enhanced the predictive accuracy of all models. The most effective performance was achieved using more targeted combinations, such as Attribute Sets 1 and 2, which combined contextual variables with absolute or relative external load indices. Feature importance and SHAP analyses consistently identified Total Distance, short- and long-term workloads, and select contextual variables as key contributors to accurate RPE prediction.

Among the models, the Bagging algorithm excelled in classifying RPE categories, effectively managing the responses of various athletes. For continuous prediction of RPE, the LightGBM model proved to be the most robust, capturing individual nuances in athlete load responses with high accuracy. Performance metrics, including R^2, AUROC, and baseline dummy model results, provided clearer benchmarking of model accuracy and discrimination. These additions helped contextualise the performance across models and attribute sets, reinforcing the consistency of the ensemble methods for both regression and classification tasks.

These models offer promising tools for coaches and practitioners to support the planning, monitoring, and evaluation of training and match demands. By integrating ML into athlete monitoring systems, practitioners may gain enhanced insight into factors influencing perceived exertion, which could inform training load decisions. However, the application of these models in real-world settings should be approached cautiously. Practical implementation challenges must be considered, such as athlete compliance with wearables, variability in data quality, and the need for further external validation. Moreover, while effective training load monitoring is an important component of athlete management, it does not directly reduce in-

jury risk or guarantee performance improvements, both of which are influenced by a broad range of contextual and individual factors.

However, reliance on aggregated ELI and metrics derived from proprietary software presents inherent limitations. These aggregated metrics, while convenient, may obscure important temporal patterns, nuanced changes in movement direction, and detailed spatial information that could provide deeper insight into athlete responses. Additionally, the use of proprietary processing pipelines often limits transparency and flexibility, restricting the ability to tailor analyses to specific research or practical questions. Furthermore, the multicollinearity among aggregated metrics adds complexity, challenging the interpretability and robustness of the models developed in this chapter.

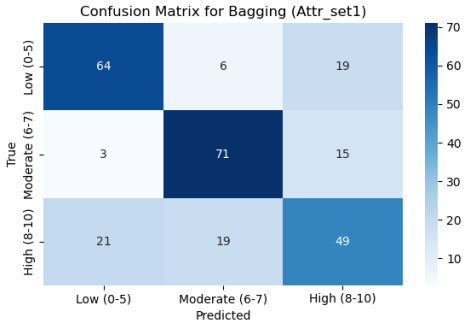
These insights shape the direction of the next chapter, where the focus is on leveraging raw GNSS data. Chapter 5 investigates whether the detailed temporal and spatial information embedded in raw data contains additional insights that can improve our ability to predict athletes' responses (RPE). By extracting a richer and more granular set of time-series features, the objective is to determine whether this detailed data offers greater predictive accuracy compared to aggregated metrics. Although time series features may not necessarily enhance model interpretability, they enable the identification of whether valuable information is captured within the temporal structure of the data. This exploration represents a critical step toward advancing the use of athlete monitoring data better to inform training and performance strategies.

Rank	Model	MSE	RMSE	MAE	Attribute Set
1	LightGBM Regressor	1.4044	1.1851	0.8708	Attr_set1
2	Gradient Boosting Regressor	1.4338	1.1974	0.9112	Attr_set1
3	Random Forest	1.5184	1.2322	0.9199	Attr_set1
4	XGBoost Regressor	1.5475	1.2440	0.9643	Attr_set2
5	XGBoost Regressor	1.5511	1.2454	0.9521	Attr_set1
6	LightGBM Regressor	1.5667	1.2517	0.9604	Attr_set2
7	Random Forest	1.5829	1.2581	0.9628	Attr_set2
8	Gradient Boosting Regressor	1.6281	1.2760	0.9667	Attr_set2
9	Linear Regression	1.7836	1.3355	1.0558	Attr_set1
10	LightGBM Regressor	1.7840	1.3357	0.9796	Attr_set5
11	Linear Regression	1.8915	1.3753	1.1080	Attr_set2
12	Gradient Boosting Regressor	1.9373	1.3919	1.0438	Attr_set5
13	Gradient Boosting Regressor	1.9608	1.4003	1.0998	Attr_set3
14	Random Forest	1.9816	1.4077	1.0500	Attr_set5
15	Linear Regression	1.9990	1.4139	1.1278	Attr_set3
16	XGBoost Regressor	2.1783	1.4759	1.1308	Attr_set5
17	Random Forest	2.1817	1.4771	1.1519	Attr_set3
18	LightGBM Regressor	2.2423	1.4974	1.1664	Attr_set3
19	Linear Regression	2.3278	1.5257	1.2173	Attr_set5
20	XGBoost Regressor	2.3630	1.5372	1.2187	Attr_set3
21	Decision Tree	2.5103	1.5844	1.1367	Attr_set2
22	Gradient Boosting Regressor	2.7016	1.6436	1.3337	Attr_set4
23	Decision Tree	2.8801	1.6971	1.2097	Attr_set1
24	Linear Regression	2.9027	1.7037	1.4046	Attr_set4
25	Random Forest	2.9597	1.7204	1.3449	Attr_set4
26	LightGBM Regressor	2.9849	1.7277	1.3359	Attr_set4
27	XGBoost Regressor	3.3116	1.8198	1.4260	Attr_set4
30	Dummy Regressor	3.6755	1.9172	1.6090	Attr_set4
30	Dummy Regressor	3.6755	1.9172	1.6090	Attr_set1
30	Dummy Regressor	3.6755	1.9172	1.6090	Attr_set3
30	Dummy Regressor	3.6755	1.9172	1.6090	Attr_set2
30	Dummy Regressor	3.6755	1.9172	1.6090	Attr_set5
31	Decision Tree	3.7079	1.9256	1.2734	Attr_set5
32	Decision Tree	4.0946	2.0235	1.4775	Attr_set3
33	Decision Tree	5.7575	2.3995	1.8146	Attr_set4

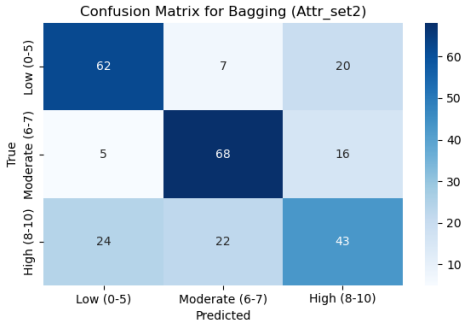
Table 4.5: Evaluation metrics for regression models predicting RPE across different attribute sets, including RMSE.

NI (Regression)	Rank	NI (Classification)	Feature
0.388	1	0.272	Total Distance (m)
0.187	2	0.171	Relative Total Distance (m$\cdot$min^{-1})
0.144	3	0.142	Relative Decelerations (n$\cdot$min^{-1})
0.132	4	0.141	Relative Accelerations (n$\cdot$min^{-1})
0.125	5	0.152	High-Speed Distance ($\geq$ 4.72 m$\cdot$s^{-1})
0.121	6	0.133	Relative High-Speed Distance ($\geq$ 4.72 m$\cdot$s^{-1})
0.116	7	0.113	Days Until Next Match
0.114	8	0.127	Decelerations (n)
0.092	9	0.046	Days Since Last Match
0.070	10	0.087	Accelerations (n)
0.068	11	0.071	Age (yrs)
0.056	12	0.040	Match Activity
0.048	13	0.059	7-d Total Distance (m)
0.043	14	0.055	28-d Total Distance (m)
0.041	15	0.041	Body Mass Index (BMI)
0.041	16	0.046	Day of Month (n)
0.040	17	0.048	Lean Mass Index (LMI)
0.034	18	0.042	Body Fat %
0.031	19	0.041	ACWR HSR
0.026	20	0.033	Month
0.025	21	0.031	ACWR TD
0.023	22	0.031	Soreness
0.023	23	0.019	Position: Midfield
0.018	24	0.024	Maximal Aerobic Speed (MAS)
0.016	25	0.025	Sleep
0.015	26	0.022	Energy
0.014	27	0.015	Training Activity
0.013	28	0.008	Position: Half Forward
0.012	29	0.012	Position: Full Forward
0.011	30	0.008	Season S1
0.010	31	0.012	Season S2
0.009	32	0.008	Playing Exp 2-3years
0.006	33	0.007	Playing Exp 7+years
0.006	34	0.008	Position: Half Back
0.006	35	0.009	Playing Exp 0-1years
0.005	36	0.008	Season S3
0.005	37	0.010	AM
0.004	38	0.009	Playing Exp 4-6years
0.003	39	0.007	Position: Full Back
0.001	40	0.003	PM

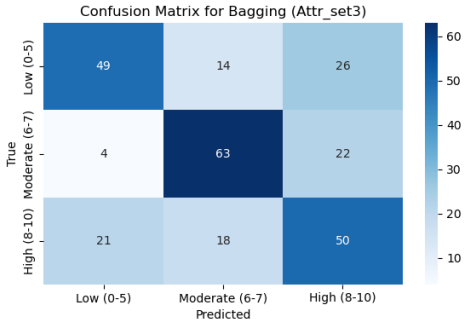
Table 4.6: Grouped Combined Normalised Feature Importance for Regression and Classification Models with Descriptive Feature Labels



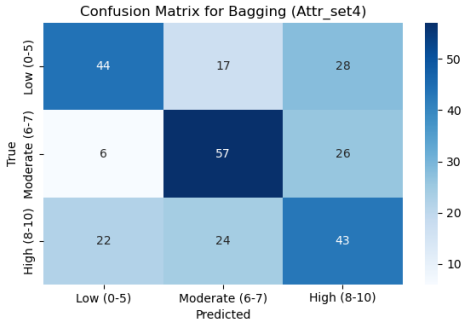
(a) Attribute Set 1 (Bagging)



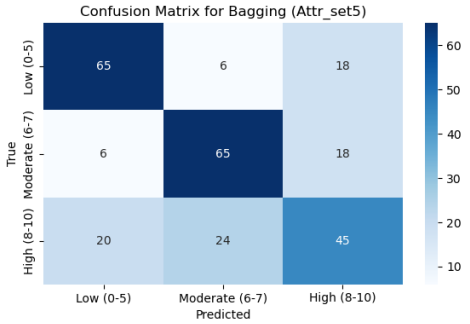
(b) Attribute Set 2 (Bagging)



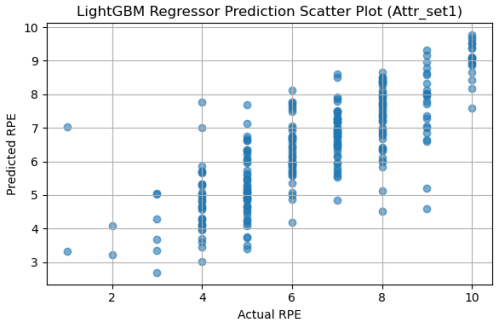
(c) Attribute Set 3 (Bagging)



(d) Attribute Set 4 (Bagging)



(e) Attribute Set 5 (Bagging)



(f) LightGBM Regressor Scatter Plot (Attr Set 1)

Figure 4.9: Classification confusion matrices for Bagging classifier across all attribute sets (left), and regression scatter plot for best-performing LightGBM Regressor (right).

Chapter 5

Evaluating Time Series Feature Sets for Predicting RPE

5.1 Introduction

This chapter investigates the predictive potential of time-series features derived from raw Global Navigation Satellite Systems (GNSS) data to estimate Rating of Perceived Exertion (RPE) in elite Gaelic football players. Building on the limitations of conventional approaches outlined in the previous chapter, this study shifts focus from aggregated metrics and proprietary External Load Indices (ELI) to the rich, detailed (raw) data provided by GNSS sensors. Specifically, the chapter evaluates whether transforming spatio-temporal data into time-series features can uncover additional insights and improve the accuracy of RPE predictions.

Raw GNSS data, represented by geospatial coordinates (latitude, longitude) and velocity, capture intricate multidirectional movement patterns inherent in team sports. These data are used to overcome the limitations of aggregated metrics, which can obscure key temporal patterns and introduce issues such as multicollinearity. Inspired by previous research, such as the work of Kim et al. (2022), which demonstrated the value of extracting angular and linear features to describe movement dynamics, this chapter expands on these methodologies by using automated

time-series feature extraction techniques. Specifically, tools like the `Time Series Feature Extraction (TSfresh)` library enable the quantification of temporal patterns and the identification of the most relevant attributes to predict RPE.

The primary objective of this chapter is to determine whether detailed time-series features extracted from raw GNSS data can enhance predictive accuracy compared to aggregated metrics. This includes addressing challenges such as information loss in aggregated data, multicollinearity and limited interpretability. Furthermore, the study explores whether integrating perceived recovery status, IC (e.g. age, body mass, and maximum sprint speed), and positional differences can further refine predictions. Through this analysis, the chapter aims to demonstrate how raw GNSS data can be transformed into actionable insights, advancing the understanding of athlete monitoring and training load management.

The chapter is organised as follows: Section 5.1 introduces the rationale and objectives of the study. Section 5.2 details the methods employed, including the feature engineering workflow and the Machine Learning (ML) modelling approach. Section 5.3 presents the results of the analyses, highlighting the impact of time-series features on predictive performance. Section 5.4 discusses the findings in the context of previous research and practical applications in athlete monitoring and coaching. Finally, Section 5.5 summarises the key contributions of the chapter and offers recommendations for future research. Details on data acquisition, participant recruitment, GNSS data collection, and preprocessing methods are provided in Chapter 3, Section 3.2.2.

5.2 Methodology

5.2.1 Feature Engineering

Feature engineering in this study focusses on evaluating time-series feature sets to predict RPE. The process involves transforming raw GNSS data into meaningful features through two key steps. This methodology employs a hybrid approach,

integrating techniques from White et al. (2022) and Kim et al. (2022).

1. **Movement Feature Extraction**: Compute linear and angular features from raw GNSS data (Time, Speed, Latitude, Longitude). This step is detailed in Section 5.2.2.

2. **Time Series Feature Extraction**: Generate time-series features using the `TSfresh` library. This process is described in Section 5.2.3.

Each step is motivated by the need to capture richer patterns and attributes in GNSS data for improved prediction of RPE, addressing the limitations of traditional aggregate metrics. The resulting feature sets and their characteristics are described at each stage of the process. The flow chart in Figure 5.1 illustrates the feature engineering process, highlighting the transformations of the data set at each step. The diagram shows how the initial dataset is expanded through the calculation of angular and linear features and then enriched with time-series features using `TSfresh`, resulting in data sets of varying dimensionality for analysis.

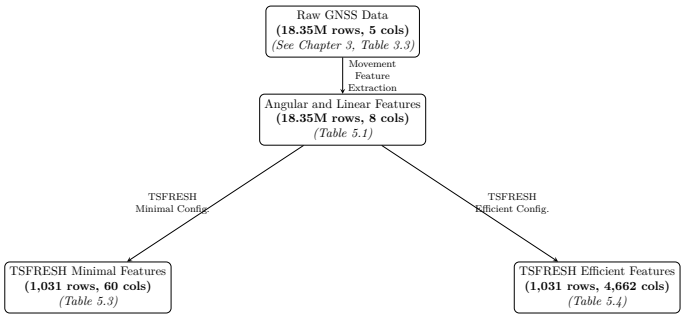


Figure 5.1: Feature Engineering Process and Dataset Evolution. Raw GNSS data was processed to derive angular and linear movement features. These were used to extract two distinct time-series feature sets using `TSfresh`: a compact Minimal configuration (Table 5.3) and a comprehensive Efficient configuration (Table 5.4).

5.2.2 Movement Feature Extraction from Raw GNSS Data

This step transforms raw GNSS data into linear and angular features, using methods proposed by White et al. (2022) and Kim et al. (2022). The raw dataset, described in Chapter 3 (Table 3.3), includes geospatial coordinates (latitude and longitude),

speed and additional metrics such as GNSS signal quality. This section focusses on the core GNSS features required for the extraction of movement features, Time, Latitude, Longitude, and Speed.

Bearing Calculation

To calculate angular features, we use latitude and longitude coordinates from the raw dataset without normalising them into Cartesian coordinates. Using the method described in White et al. (2022), the bearing (β) is calculated as the angle between the north-south line (Earth's meridian) and the line connecting two consecutive points in the geospatial time series. As illustrated in Figure 5.2, the bearing (β) indicates the direction change relative to the north connecting the target point (B) to the reference point (A).

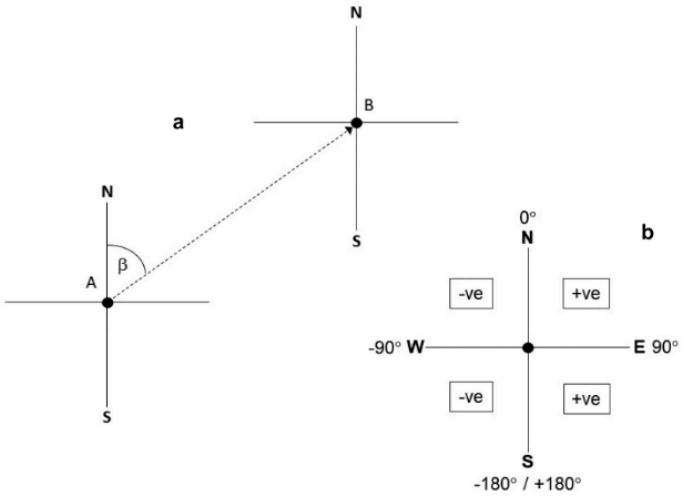


Figure 5.2: The bearing (β) relative to the NorthSouth line from point A to point B (a) and the compass quadrants (b).

The bearing (β) is calculated as:

$$X = \cos(\theta_B) \cdot \sin(\Delta L) \tag{5.1}$$

$$Y = \cos(\theta_A) \cdot \sin(\theta_B) - \sin(\theta_A) \cdot \cos(\theta_B) \cdot \cos(\Delta L) \tag{5.2}$$

$$\beta = \text{atan2}(X, Y) \tag{5.3}$$

where:

- θ_A, θ_B: Latitudes of points A and B, respectively (in radians),
- ΔL: Difference in longitude between A and B (in radians).

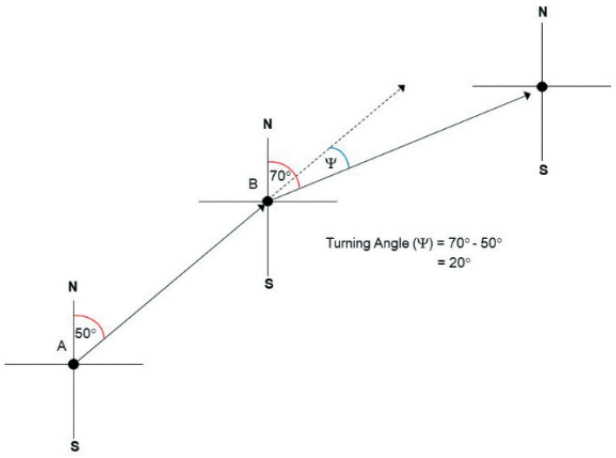


Figure 5.3: Example for computing the turning angle (Ψ) from consecutive GNSS sample bearing values.

Turning Angle Calculation

The turning angle (Ψ) between consecutive points, derived from the bearing (β), captures the directional changes in the movement of the player. These angular variations are crucial for analysing multidirectional movements typical of team sports, as shown in Figure 5.3.

The turning angle (Ψ) between two consecutive points is calculated as:

$$\Psi = \beta_i - \beta_{i-1} \tag{5.4}$$

The structure of the resulting dataset, including the calculated bearing and turning angle features, is summarised in Table 5.1. This table illustrates how the original latitude and longitude coordinates are replaced with derived features, providing a more interpretable representation of player movement dynamics.

Table 5.1: Feature Description After Bearing and Turning Angle Calculation

Feature	Description
PlayerID	A unique identifier for each player in the dataset.
Time	Timestamp indicating the exact time of the recorded data point.
Speed (m/s)	The player's velocity in meters per second at the corresponding timestamp.
Bearing (ř)	The player's directional heading in degrees, calculated from the GNSS data.
Turning Angle (ř)	The change in bearing between consecutive timestamps, representing the angle of directional change.

Linear and Angular Feature Calculation

Linear and angular features were calculated to capture both translational and rotational dynamics, enhancing the ability of the data set to describe the multidirectional movements characteristic of team sports (Kim et al., 2022).

Linear Features:

1. **Linear Velocity** ($v(t)$): Derived from the Doppler shift method for accuracy, as recommended by White et al. (2022).
2. **Linear Acceleration** ($a(t)$): First derivative of velocity.
3. **Linear Jerk** ($j(t)$): Second derivative of velocity.

Angular Features:

1. **Angular Velocity** ($\omega(t)$): Change in turning angle with time.
2. **Angular Acceleration** ($\alpha(t)$): First derivative of angular velocity.
3. **Angular Jerk** ($\zeta(t)$): Second derivative of angular velocity.

A detailed summary of the angular and linear features added to the dataset is presented in Table 5.2. These features provide a comprehensive representation of player movement, capturing both the directional changes and translational dynamics required for a detailed analysis of physical demands.

Table 5.2: Feature Description After Angular and Linear Feature Calculations

Feature	Description
PlayerID	A unique identifier for each player in the dataset.
Time	Timestamp indicating the exact time of the recorded data point.
Linear Velocity (m/s)	The player's velocity in a straight line, measured in meters per second.
Linear Acceleration (m/s^2)	The rate of change of linear velocity, representing the player's straight-line acceleration in meters per second squared.
Linear Jerk (m/s^3)	The rate of change of linear acceleration, indicating the smoothness or abruptness of linear movement in meters per second cubed.
Angular Velocity (ř/s)	The rate of change of the player's orientation, measured in degrees per second.
Angular Acceleration ($ř/s^2$)	The rate of change of angular velocity, representing rotational acceleration in degrees per second squared.
Angular Jerk ($ř/s^3$)	The rate of change of angular acceleration, representing the smoothness or abruptness of rotational movement in degrees per second cubed.

5.2.3 Time Series Feature Extraction with `TSfresh`

To capture the temporal dynamics of the GNSS-derived features, the `TSfresh` library was employed to extract two distinct sets of time-series features: a Minimal set and an Efficient set. Time-series feature extraction transforms raw temporal data into informative and interpretable attributes that are critical for downstream modelling. The `TSfresh` package is a recognised tool for automated time-series characterisation, computing an extensive feature set across statistical, temporal, and frequency domains (Christ et al., 2018).

Minimal Time-Series Feature Set: This configuration focused on essential statistical and temporal metrics, including mean, variance, and quantiles. It provided a concise representation of the data, generating 10 features per column. With six GNSS-derived features, this resulted in a total of 60 features. The features computed using the TSfresh Minimal configuration are summarised in Table 5.3, highlighting the 10 statistical and temporal metrics extracted for each of the six derived columns

(Angular Velocity, Angular Acceleration, Angular Jerk, Linear Velocity, Linear Acceleration, and Linear Jerk).

Efficient Time-Series Feature Set: In addition to the Minimal set, a more comprehensive Efficient configuration was also utilised. This configuration extracted a wide array of featuresincluding advanced statistical metrics, frequency-domain descriptors, autocorrelation structures, and temporal change patterns. This resulted in a total of 4,653 features, computed independently for each of the six GNSS-derived columns. The Efficient feature set enabled richer temporal representation and was especially beneficial for deeper model interpretation using SHAP values in later experiments.

While the Minimal feature set supported model benchmarking with a limited and interpretable subset, the Efficient set allowed for the exploration of high-dimensional, automated feature generation approaches. The results of models trained on these features are compared in later sections using both performance metrics and feature attribution methods.

Table 5.3: Minimal Time-Series Feature Set Computed Using `TSfresh`

General Feature Name	**Description (Applied to Each Column)**
Sum of Values (e.g., m/s)	Sum of values over the session
Median Value (e.g., m/s)	Median value
Mean Value (e.g., m/s)	Mean value
Number of Data Points	Number of data points
Standard Deviation (e.g., m/s)	Standard deviation of values
Variance (e.g., m^2/s^2)	Variance of values
Root Mean Square (e.g., m/s)	Root mean square of values
Maximum Value (e.g., m/s)	Maximum value
Absolute Maximum Value (e.g., m/s)	Absolute maximum value
Minimum Value (e.g., m/s)	Minimum value

Note: These features were computed independently for each of the six derived columns, resulting in 60 features.

Table 5.4: Efficient Time-Series Feature Set Computed Using `TSfresh`

Feature Category	**Examples of Computed Features**
Descriptive Statistics	Mean, Median, Standard Deviation, Variance, Skewness, Kurtosis
Frequency Domain Features	Fourier Coefficients (e.g., `fft_coefficient__attr_"real"__coeff_0`)
Autocorrelation Metrics	Autocorrelation (`autocorrelation__lag_1`), Partial Autocorrelation
Temporal Patterns	Energy Ratios, Peak Counts, Lempel-Ziv Complexity
Change Analysis	Mean Absolute Change, Longest Strike Below/Above Mean
Linear Trends	Slope, Intercept, and R-Value of Fitted Trends
Advanced Statistics	Benford Correlation, Quantile Values, Symmetry Statistics

Note: These features were computed independently for each of the six derived columns, resulting in 4,653 features.

Feature Selection and Dimensionality Reduction

To manage the high dimensionality introduced by the TSfresh feature setsespecially the efficient configuration, multiple post-extraction filtering and dimensionality reduction methods were employed. These aimed to retain the most predictive and non-redundant features for subsequent modelling.

F-statistics and p-value Filtering An initial statistical filtering step was applied using univariate feature selection. Each feature was evaluated against the target variable (RPE) using F-statistics, and features with a p-value below 0.0005 were retained. This reduced the Efficient feature set from 4,662 to 1,133 features, and the Minimal set from 60 to 39. These filtered features formed the input for the subsequent multicollinearity and dimensionality reduction steps.

Variance Inflation Factor (VIF) Filtering To address multicollinearity, VIF filtering was applied to the F-statistic-filtered feature sets. Features with a VIF score greater than 5 were iteratively removed until all remaining features met the threshold. This step reduced the Efficient feature set to 116 features and the Minimal set to 4 features.

Principal Component Analysis (PCA) As an alternative to VIF filtering, PCA was applied to the F-statistic-filtered features to reduce dimensionality while preserving 95% of the variance. This projection yielded 183 principal components

for the Efficient set and 10 components for the Minimal set.

5.2.4 Baseline Aggregated Features Extraction

To compare the performance of the newly engineered time-series features, conventional external workload metrics such as total duration, speed zone durations, total distance, speed zone distances, maximum speed, number of sprints, sprint distances, number of accelerations and distance of accelerations were computed as summarised in Table 5.5. These metrics, widely used by practitioners and researchers, have been extensively applied in previous studies to predict player responses (Bartlett et al., 2017; Jaspers et al., 2018; Kim et al., 2022; Rossi et al., 2019).

In the previous chapter, it was observed that these traditional aggregated workload variables are often linked, exhibiting high correlations that can lead to multicollinearity issues. Motivated by this observation, the relationships between these variables were assessed using a correlation matrix, shown in Figure 5.4. This matrix provides insight into how these baseline GNSS features relate to one another.

To ensure robust comparisons with the time-series feature sets, three variations of the baseline aggregated dataset were prepared:

- **All Features:** This version includes the full set of 21 commonly used aggregated GNSS-derived features, such as total duration, total distance, speed zone distances, sprint metrics, and acceleration/deceleration distances. A full description of these features and their thresholds is provided in Table 5.5.

- **VIF-Filtered Features:** This subset was created by sequentially removing features with a VIF > 5. Selection decisions were guided by expert knowledge to retain practically meaningful variables that reflect key physical workload demands.

- **PCA-Reduced Features:** In the third variation, Principal Component Analysis (PCA) was applied to the full 21-feature set, using `n_components=0.95` to

Feature Name	Description
`total_duration`	Total duration in a session.
`standing_duration`	Duration covered at speeds $0.00\,\mathrm{m/s} \leq$ Speed $< 0.19\,\mathrm{m/s}$.
`walking_duration`	Duration covered at speeds $0.19\,\mathrm{m/s} \leq$ Speed $< 2.00\,\mathrm{m/s}$.
`jogging_duration`	Duration covered at speeds $2.00\,\mathrm{m/s} \leq$ Speed $< 4.00\,\mathrm{m/s}$.
`running_duration`	Duration covered at speeds $4.00\,\mathrm{m/s} \leq$ Speed $< 5.50\,\mathrm{m/s}$.
`high_speed_running_duration`	Duration covered at speeds $5.50\,\mathrm{m/s} \leq$ Speed $< 7.00\,\mathrm{m/s}$.
`max_speed_running_duration`	Duration covered at speeds Speed $\geq 7.00\,\mathrm{m/s}$.
`total_distance`	Total distance covered in a session.
`standing_distance`	Distance covered at speeds $0.00\,\mathrm{m/s} \leq$ Speed $< 0.19\,\mathrm{m/s}$.
`walking_distance`	Distance covered at speeds $0.19\,\mathrm{m/s} \leq$ Speed $< 2.00\,\mathrm{m/s}$.
`jogging_distance`	Distance covered at speeds $2.00\,\mathrm{m/s} \leq$ Speed $< 4.00\,\mathrm{m/s}$.
`running_distance`	Distance covered at speeds $4.00\,\mathrm{m/s} \leq$ Speed $< 5.50\,\mathrm{m/s}$.
`high_speed_running_distance`	Distance covered at speeds $5.50\,\mathrm{m/s} \leq$ Speed $< 7.00\,\mathrm{m/s}$.
`max_speed_running_distance`	Distance covered at speeds Speed $\geq 7.00\,\mathrm{m/s}$.
`max_speed`	Maximum speed achieved in a session.
`number_of_sprints`	Number of times speed $\geq 7.00\,\mathrm{m/s}$ was maintained for more than 0.6 seconds in a session.
`distance_of_sprints`	Distance covered at speeds $\geq 7.00\,\mathrm{m/s}$ for more than 0.6 seconds in a session.
`number_of_accelerations`	Number of instances where acceleration $\geq 1.5\,\mathrm{m/s^2}$ was sustained for more than 0.5 seconds in a session.
`number_of_decelerations`	Number of instances where deceleration $\leq -1.5\,\mathrm{m/s^2}$ was sustained for more than 0.5 seconds in a session.
`distance_of_accelerations`	Distance covered during instances where acceleration $\geq 1.5\,\mathrm{m/s^2}$ was sustained for more than 0.5 seconds in a session.
`distance_of_decelerations`	Distance covered during instances where deceleration $\leq -1.5\,\mathrm{m/s^2}$ was sustained for more than 0.5 seconds in a session.

Table 5.5: Summary of Baseline Features with Speed Zones and Thresholds Chapter 5.

retain 95% of the dataset's variance. This transformation yielded a compact,ğ orthogonal feature set for efficient modelling.

After this process, the number of GNSS indices was reduced from 21 to 6 in the VIF-filtered version. The final retained indices, as shown in Table 5.6, included: *total duration*, *distance of accelerations* ($\geq 1.5\,\mathrm{m/s^2}$), *distance of decelerations* ($\leq$

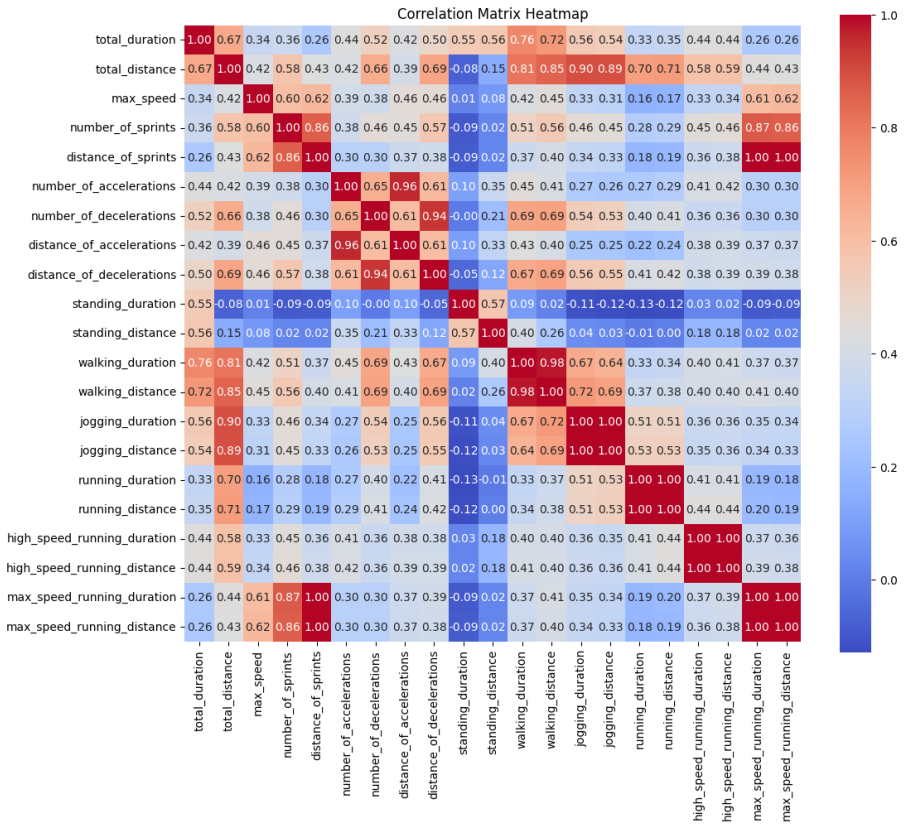


Figure 5.4: Correlation matrix of baseline GNSS features computed in this Chapter 5. Motivated by observations in the previous chapter that these variables are often linked, their relationships was tested to identify and mitigate potential multicollinearity issues. The correlation analysis informed the feature reduction process, which led to the selection of six key features for further analysis.

$-1.5\,\mathrm{m/s^2}$), *running distance* ($4.00\,\mathrm{m/s} \leq \text{Speed} < 5.50\,\mathrm{m/s}$), *high-speed running distance* ($5.50\,\mathrm{m/s} \leq \text{Speed} < 7.00\,\mathrm{m/s}$), and *max-speed running distance* ($\text{Speed} \geq 7.00\,\mathrm{m/s}$).

Table 5.6 provides a detailed description of the final retained features for the VIF-filtered version. In contrast, the PCA-reduced feature set offers an alternative compressed representation based on orthogonal components derived from the original 21 features. These three variants allowed for consistent comparison across all experiments.

5.2.5 Additional Features

Seven supplementary features were integrated into the GNSS movement dataset to enhance ML models. The player positions were categorised into five positional

Feature Name	Description
`total_duration`	Total duration in a session.
`distance_of_accelerations`	Distance covered during instances where acceleration $\geq 1.5\,\mathrm{m/s^2}$ was sustained for more than 0.5 seconds in a session.
`distance_of_decelerations`	Distance covered during instances where deceleration $\leq -1.5\,\mathrm{m/s^2}$ was sustained for more than 0.5 seconds in a session.
`running_distance`	Distance covered at speeds $4.00\,\mathrm{m/s} \leq$ Speed $< 5.50\,\mathrm{m/s}$ in a session.
`high_speed_running_distance`	Distance covered at speeds $5.50\,\mathrm{m/s} \leq$ Speed $< 7.00\,\mathrm{m/s}$ in a session.
`max_speed_running_distance`	Distance covered at speeds Speed $\geq 7.00\,\mathrm{m/s}$ in a session.

Table 5.6: Description of the reduced baseline feature set after correlation analysis and VIF testing. These six features were selected for their strong relationships with the target variable and their ability to minimise multicollinearity.

groups and included as a feature. Additional IC - age, height, weight, body mass index (BMI) and maximal sprint speed (MSS) - were also incorporated to account for individual variability (five features). Furthermore, the athletes evaluated their recovery status approximately 15 minutes before the warm-up of each training session or match using the Perceived Recovery Status (PRS) scale, a self-reported measure of readiness and recovery (one characteristic) (Laurent et al., 2011). Including PRS as a feature allowed the models to account for players' recovery levels and potential impacts on exertion. To prepare categorical variables for ML, one-hot encoding was applied. This ensured that non-numeric data, such as player positions, were represented as numerical features without introducing unintended ordinal relationships, making them compatible with model training.

5.2.6 Target Feature

The target variable in this study is RPE, reported by athletes on a scale of 1 to 10. RPE serves as a subjective measure of an athlete's perceived intensity during training or competition. To ensure that features were on comparable scales, all numerical variables were standardised using the **StandardScaler**. Standardisation

is crucial for models sensitive to scale differences as it improves model convergence and performance consistency.

5.2.7 Experimental Setup

The primary objective of this chapter is to assess whether detailed time series features extracted from raw GNSS data can improve predictive accuracy compared to traditional aggregated metrics. Furthermore, the study integrates IC (e.g., age, body mass, maximum sprint speed) and recovery metrics (e.g., Player Recovery Status), acknowledging their proven importance in predicting RPE, as discussed in Chapter 4. By incorporating these contextual factors, the analysis aims to enhance the precision and reliability of predictive models, tailoring them to individual-level variations.

To identify the most relevant predictors, Recursive Feature Elimination (RFE) was employed as part of the feature selection process in one of the experimental setups. This method iteratively ranks and removes less significant features, enabling the creation of optimised subsets tailored to each ML model. Stratified cross-validation K times (K = 3) was applied throughout to ensure robust evaluation and balanced representation of all RPE categories during training and validation.

Experimental Configurations

The experimental setup was designed to compare the predictive utility of baseline GNSS-derived metrics with advanced time-series feature sets for estimating ratings of perceived exertion (RPE). Two main pipelines were developed to support this comparison: one using conventional aggregated features, and another leveraging features extracted via `TSfresh`.

Baseline Aggregated Feature Experiment This experiment uses session-level summary statistics commonly applied in team sports to monitor external load. The aggregated features include total distance, session duration, speed zone distances

(e.g., walking, jogging, high-speed running), counts and distances of sprints, and acceleration and deceleration events. Three versions of this feature set were evaluated:

1. **Full Feature Set:** All 21 features as described in Table 5.5.
2. **VIF-Filtered Subset:** A reduced set of 6 features with variance inflation factor (VIF) < 5.
3. **PCA-Reduced Set:** A lower-dimensional representation retaining $\sim$95% of total variance via Principal Component Analysis (PCA).

Time-Series Feature Set Experiment Advanced feature sets were extracted from six derived GNSS signals (linear and angular velocity, acceleration, and jerk) using the `TSfresh` Python library. Two main feature configurations were evaluated:

- **Minimal Feature Set:** 10 core statistical features applied across six signals (total: 60 features).
- **Efficient Feature Set:** A high-dimensional feature set ($\sim$4,653 features) capturing statistical, frequency-domain, and complexity characteristics.

Each feature configuration was processed under three different dimensionality reduction strategies:

1. **Filtered:** Removal of features with non-significant univariate association to RPE (p-value > 0.05).
2. **VIF-Filtered:** Sequential removal of collinear features using a VIF threshold of 5.
3. **PCA-Reduced:** Principal Component Analysis retaining 95% explained variance.

In total, six time-series datasets were generated and modelled independently.

5.2.8 Regression Models

To predict continuous RPE values, six regression algorithms were implemented:

- Decision Tree (DT) Regressor
- Random Forest (RF) Regressor
- Gradient Boosting Machine (GBM) Regressor
- eXtreme Gradient Boosting (XGBoost) Regressor
- Light Gradient Boosting Machine (LightGBM) Regressor
- Dummy Regressor (mean baseline)

Models capable of interpreting feature importance (e.g., RF, GBM, XGBoost, LightGBM) were used to extract feature rankings. Additionally, XGBoost and LightGBM models were analysed using SHAP to provide model-agnostic explanations of feature impact.

5.2.9 Model Training and Evaluation

Model generalisability was assessed using Leave-One-Subject-Out Cross-Validation (LOSOCV), where each player served as a test fold once. The grouping variable was `Player_Display_Name`, ensuring subject-independent validation.

All features were standardised using z-score normalisation before training.

For each modeldataset combination, the following outputs were saved:

- Predictions vs. actual RPE values (with session metadata)
- Evaluation metrics (averaged over all LOSOCV folds)
- Feature importances (if applicable)
- SHAP values (for XGBoost and LightGBM models)

Evaluation Metrics

- **Coefficient of Determination (R^2):** Coefficient of determination, measuring variance explained.
- **Mean Absolute Error (MAE):** Mean Absolute Error, capturing average prediction error magnitude.
- **Mean Squared Error (MSE):** Mean Squared Error, penalising larger deviations more heavily.
- **Root Mean Square Error (RMSE):** Root Mean Squared Error, representing error in the original RPE scale.

This experimental structure enables a comprehensive comparison of conventional and time-series-based approaches to modelling perceived exertion, supporting both predictive performance and interpretability.

5.3 Results

5.3.1 Data Description

The data set provides detailed information on the distribution of RPE observations between players and positions. On average, each player has reported their RPE roughly 30 times, with a standard deviation of 9.51 observations per player. The median number of RPE observations per player is 31, indicating a slightly skewed distribution around the mean. The mean RPE value for all observations is 6.24, with a standard deviation of 1.89, indicating moderate variability in perceived exertion levels between players. The data set demonstrates a balanced number of observations per player with a moderate spread in RPE values, making it a good candidate for predictive modelling. As shown in Figure 5.5, the RPE distribution covers the 10 recorded values.

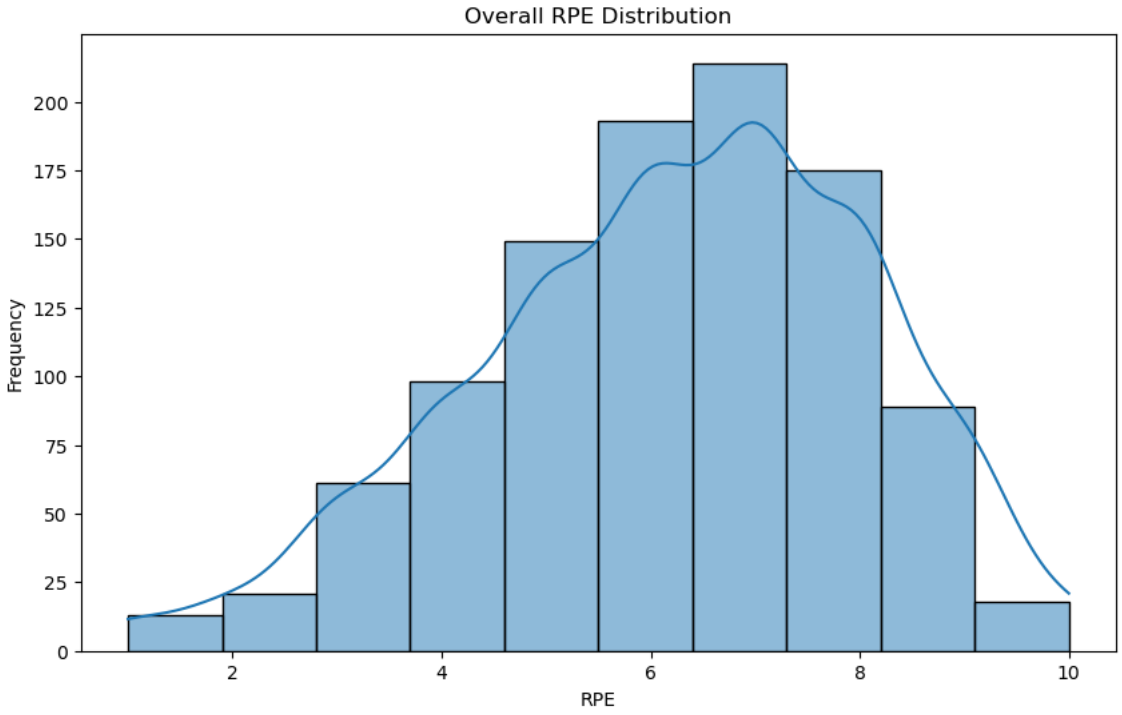


Figure 5.5: RPE Distribution. This figure illustrates the frequency distribution of RPE values on a scale from 0 to 10, reported by players. The histogram shows the overall spread of exertion levels, with a peak around the mid-range, indicating the most commonly reported RPE values.

The distribution of the RPE values by player position is shown in Figure 5.6. The boxplots highlight positional differences, with Full Forwards and Half Forwards tending to report higher RPE values compared to other positions. Midfielders, on the other hand, exhibited lower median RPE values, indicating possible differences in physical demands or perceived exertion between roles. Outliers, represented by individual points, underscore variability within positions.

The individual distributions of RPE values for each player are presented in Figure 5.7. The box plots illustrate substantial variability in RPE values between players, with some reporting a wider range of exertion levels than others. Outliers, marked as individual points, further highlight instances of extreme perceived exertion. This variability underscores the diverse training responses and physical loads experienced by players, reflecting both individual and positional characteristics within the dataset.

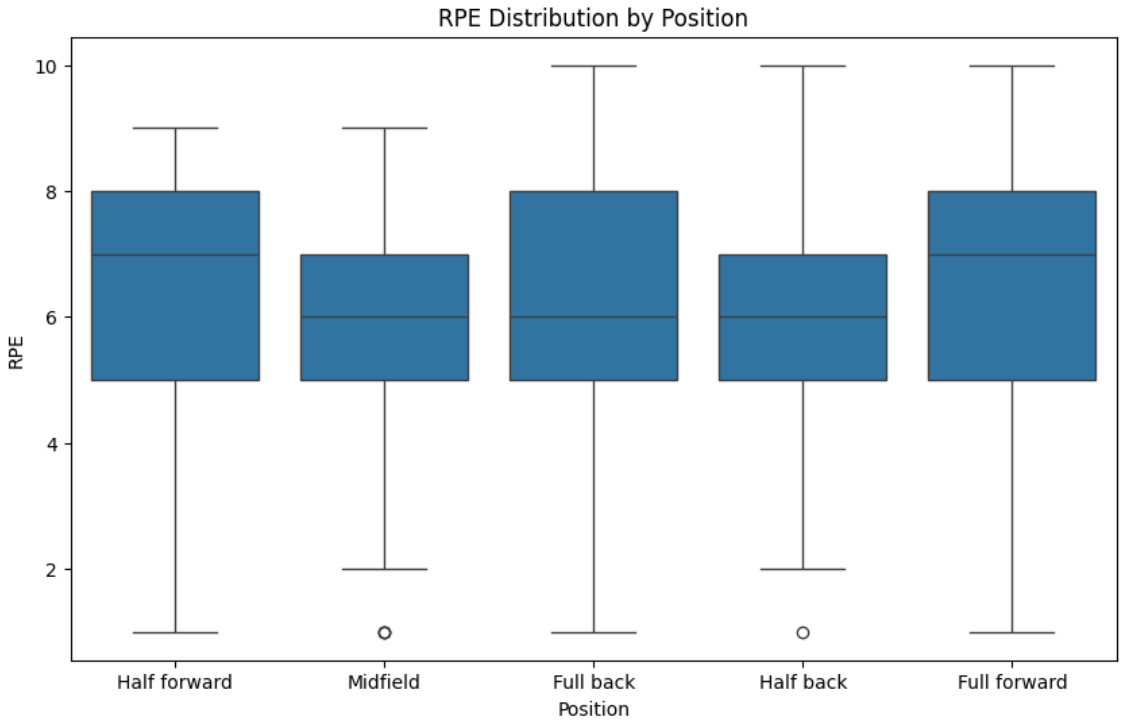


Figure 5.6: RPE Distribution by Position. This boxplot illustrates the variation in RPE values across different playing positions: Half Forward, Midfield, Full Back, Half Back, and Full Forward. The figure highlights positional differences in perceived exertion, with median values and ranges varying across roles. Outliers are shown as individual points.

5.3.2 Baseline Experiment Results

The baseline performance of ML models using aggregated GNSS features was re-evaluated across three feature configurations: full feature set, VIF-filtered subset, and PCA-reduced set. These modelsDT, RF, GBM, XGBoost, and LightGBMwere assessed using R^2, MSE, MAE, and RMSE. Among them, the RF Regressor achieved the highest R^2 (0.495), the lowest MAE (1.08), and the lowest RMSE (1.34) when using the full aggregated feature set. GBM and LightGBM also performed strongly, with consistent results across all three feature configurations. PCA-based features resulted in lower model performance, particularly for the DT, which saw a negative R^2 and elevated error values (e.g., RMSE = 2.12). As expected, the DummyRegressor baseline yielded negative R^2 and high error metrics. These findings establish a robust benchmark using traditional session-level GNSS summaries for comparison with time-series-derived feature sets.

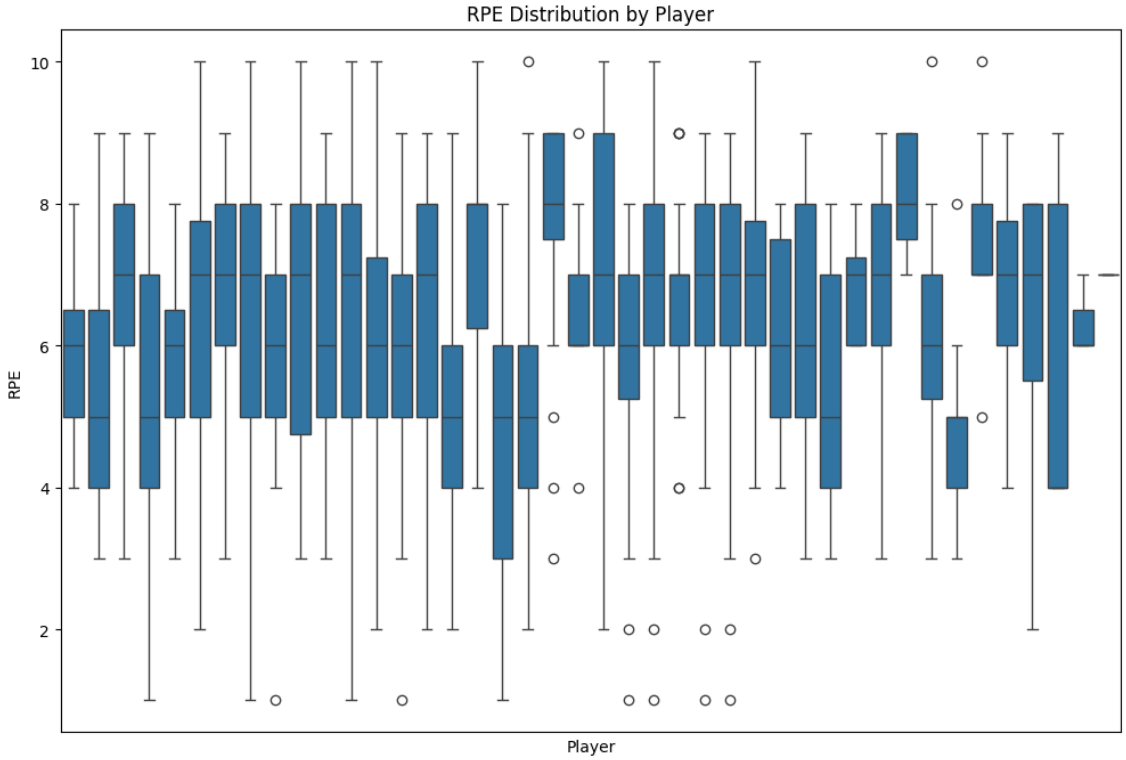


Figure 5.7: RPE Distribution by Player. This boxplot shows the range, median, and interquartile range of RPE values for individual players. Outliers are indicated by individual points, highlighting variability in perceived exertion across different players.

Feature Importance of Aggregated Features

To interpret the predictions made by the machine learning models using full aggregated GNSS feature set, SHAP analysis was conducted on the two best-performing models: XGBoost and LightGBM. SHAP (SHapley Additive exPlanations) values provide insight into how much each feature contributed to the final RPE prediction for each observation.

The SHAP summary plots in Figure 5.8 and Figure 5.9 illustrate the top contributing features for each model across the full feature set. For both models, `total_distance`, `running_duration`, and `running_distance` emerged as the most influential predictors. These findings align with prior research emphasizing the role of total load and high-intensity efforts in perceived exertion.

A detailed comparison of model-based feature importances and average SHAP values across RF, XGBoost, and LightGBM is provided in Table 5.11. Notably, `age`,

Table 5.7: R^2 Performance of Regression Models on Aggregated GNSS Feature Sets

Model	All	VIF	PCA
RandomForest	0.495	0.450	0.374
GradientBoosting	0.470	0.435	0.376
XGBoost	0.414	0.390	0.292
LGBMRegressor	0.450	0.434	0.369
DecisionTree	0.034	-0.023	-0.260
DummyRegressor	-0.008	-0.008	-0.008

Table 5.8: Mean Absolute Error (MAE) for Aggregated GNSS Feature Sets

Model	All	VIF	PCA
RandomForest	1.081	1.122	1.182
GradientBoosting	1.099	1.138	1.174
XGBoost	1.161	1.181	1.256
LGBMRegressor	1.121	1.151	1.193
DecisionTree	1.426	1.461	1.623
DummyRegressor	1.542	1.541	1.541

`height_cm`, and `mss_ms` also showed moderate influence in both models, highlighting the role of individual physical characteristics in shaping RPE responses.

5.3.3 Time-Series Feature Set Results

To evaluate the predictive performance of time-series features extracted with `TSfresh`, six distinct feature sets were generated: minimal and efficient versions, each filtered, VIF-filtered, and PCA-reduced. These feature sets were used to train multiple ML models using Leave-One-Subject-Out Cross-Validation (LOSOCV). Model performance was assessed using R^2, MAE, MSE, and RMSE. Figure 5.10 compares the R^2 scores across all TSFRESH feature sets and models. The **efficient filtered** feature set consistently outperformed other variants, with the RF and GBM regressors achieving R^2 scores of 0.509 and 0.502 respectively. These results exceeded the performance of models using aggregated features, highlighting the value of rich temporal feature representations. The best RMSE (1.318) was achieved by the GBM regressor on the efficient PCA-reduced feature set, followed closely by the RF and LightGBM regressors. In contrast, Decision Tree and DummyRegressor models showed poor

Table 5.9: Mean Squared Error (MSE) for Aggregated GNSS Feature Sets

Model	**All**	**VIF**	**PCA**
RandomForest	1.798	1.959	2.232
GradientBoosting	1.891	2.012	2.224
XGBoost	2.088	2.172	2.523
LGBMRegressor	1.962	2.016	2.248
DecisionTree	3.443	3.643	4.487
DummyRegressor	3.593	3.589	3.589

Table 5.10: Root Mean Squared Error (RMSE) for Aggregated GNSS Feature Sets

Model	**All**	**VIF**	**PCA**
RandomForest	1.341	1.400	1.494
GradientBoosting	1.375	1.418	1.491
XGBoost	1.445	1.474	1.589
LGBMRegressor	1.401	1.420	1.499
DecisionTree	1.856	1.909	2.118
DummyRegressor	1.895	1.895	1.895

Table 5.11: Top SHAP and Model-Based Feature Importances from Aggregated Feature Models

Feature	**RF Imp.**	**XGB SHAP**	**LGBM SHAP**	**XGB Imp.**
total_distance	0.225	**0.435**	**0.435**	0.099
running_duration	0.083	0.224	0.088	0.178
running_distance	0.211	0.196	0.281	0.261
standing_duration	0.045	0.169	0.141	0.031
age	0.027	0.142	0.155	0.035
height_cm	0.025	0.133	0.149	0.025
walking_distance	0.024	0.116	0.082	0.034
high_speed_running_distance	0.024	0.114	0.084	0.030
mss_ms	0.023	0.105	0.103	0.035
max_speed	0.024	0.093	0.086	0.012

Note: Table shows top 10 most important features across models, including Random Forest (RF) importance and SHAP values for XGBoost and LightGBM.

performance across all feature sets, with negative or near-zero R^2 values. These results demonstrate that models trained on high-dimensional time-series features, especially the efficient filtered subset provided consistently improved RPE prediction accuracy compared to baseline aggregated metrics. Full results are summarised in Table 5.5.

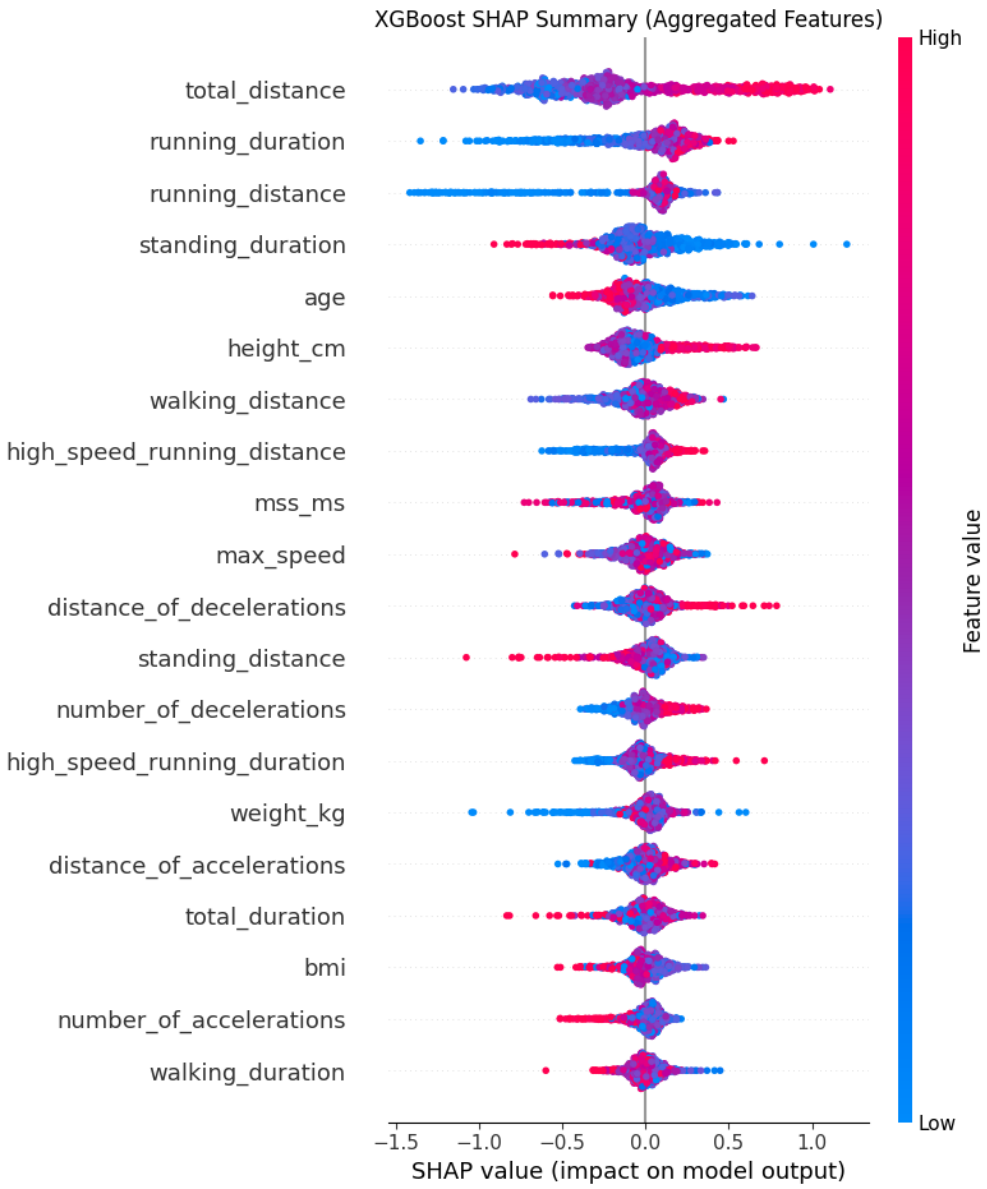


Figure 5.8: SHAP summary plot for XGBoost model using aggregated GNSS features.

Feature Importance from TSFRESH-Derived Features

To interpret model predictions using TSFRESH-derived time-series features, SHAP analysis was applied to the best-performing models XGBoost and LightGBM using the **efficient PCA-filtered feature set**. SHAP (SHapley Additive exPlanations) values quantify the contribution of each feature to model output, offering a consistent method to understand model behaviour across complex, transformed inputs.

Figures 5.11 and 5.12 display the SHAP summary plots for XGBoost and Light-

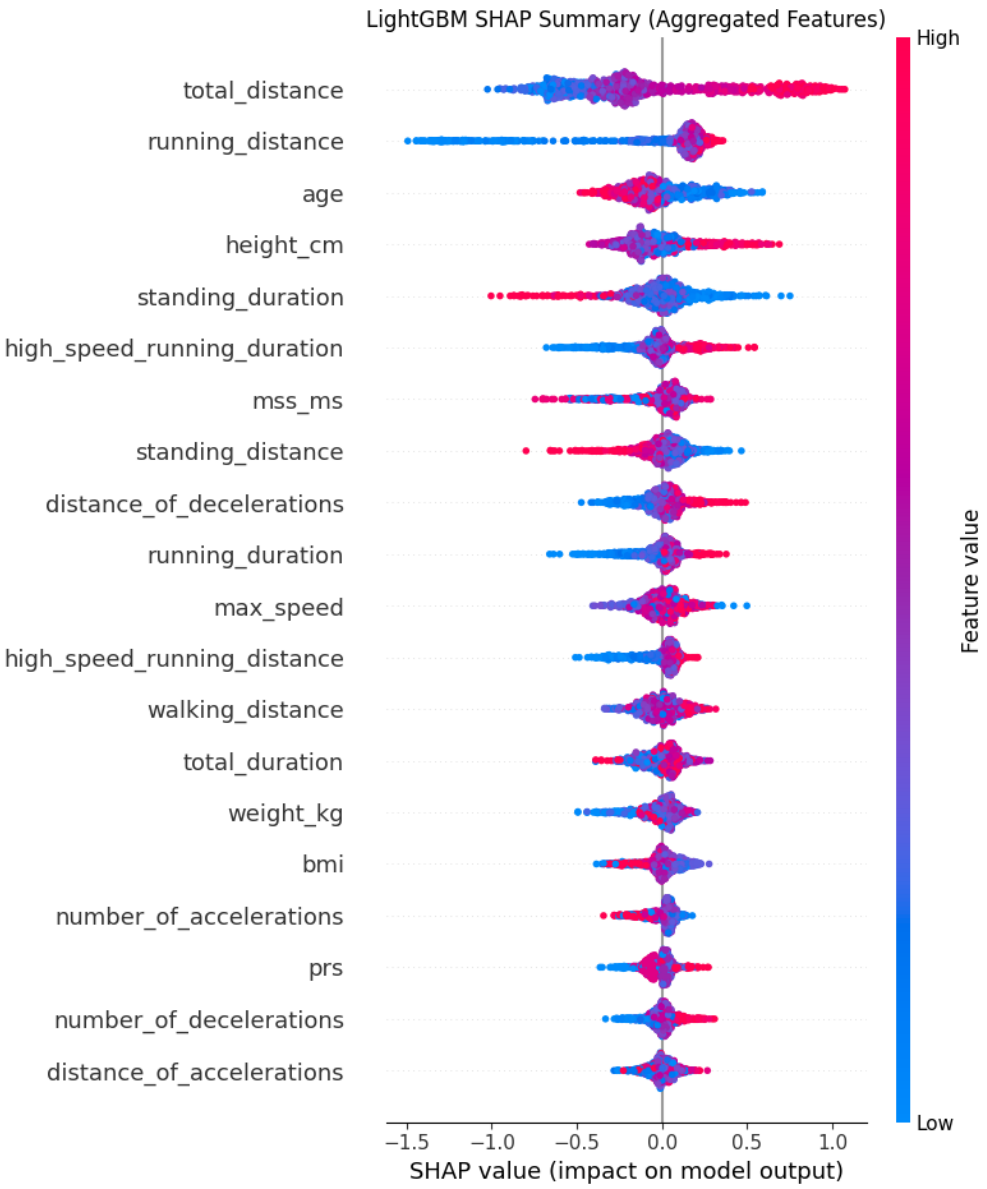


Figure 5.9: SHAP summary plot for LightGBM model using aggregated GNSS features.

GBM. The results clearly show that `pca_1` and `pca_2` dominate the prediction landscape, alongside key contextual features such as `age`, `height_cm`, and `weight_kg`. This suggests that both dynamic signal-derived components and static player characteristics significantly influence perceived exertion.

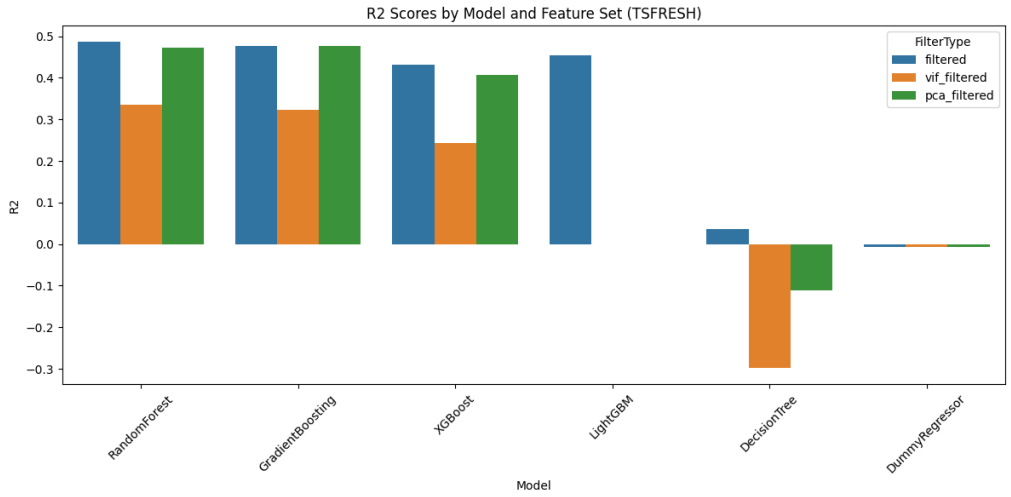


Figure 5.10: R^2 scores by model and feature version for TSFRESH-derived time-series features.

Table 5.12: Top R^2 Scores from TSFRESH Feature Sets

Model	**Feature Set**	**Filter Type**	**R^2**
RandomForest	efficient	filtered	0.509
GradientBoosting	efficient	pca_filtered	0.512
GradientBoosting	efficient	filtered	0.502
RandomForest	efficient	pca_filtered	0.490
GradientBoosting	efficient	vif_filtered	0.489
LightGBM	efficient	filtered	0.489
RandomForest	minimal	filtered	0.463
XGBoost	efficient	filtered	0.463

5.3.4 Comparison of Time-Series Feature Sets to Baseline Aggregated Metrics

This section compares the predictive performance, dimensionality, and interpretability of traditional aggregated GNSS features and TSFRESH-derived time-series features.

Table 5.13: Lowest MAE Values from TSFRESH Feature Sets

Model	Feature Set	Filter Type	MAE
GradientBoosting	efficient	pca_filtered	1.052
RandomForest	efficient	filtered	1.061
GradientBoosting	efficient	filtered	1.065
LightGBM	efficient	filtered	1.085
RandomForest	efficient	pca_filtered	1.079
GradientBoosting	efficient	vif_filtered	1.080
RandomForest	minimal	filtered	1.112
XGBoost	efficient	filtered	1.114

Table 5.14: Lowest MSE Values from TSFRESH Feature Sets

Model	Feature Set	Filter Type	MSE
GradientBoosting	efficient	pca_filtered	1.738
RandomForest	efficient	filtered	1.749
GradientBoosting	efficient	filtered	1.772
LightGBM	efficient	filtered	1.821
RandomForest	efficient	pca_filtered	1.817
GradientBoosting	efficient	vif_filtered	1.819
XGBoost	efficient	filtered	1.914
RandomForest	minimal	filtered	1.912

Predictive Performance

Figure 5.10 shows R^2 scores across models using both aggregated and TSFRESH features. TSFRESH-derived features (efficient PCA-filtered) achieved the highest predictive performance, with Gradient Boosting yielding an R^2 of 0.512 compared to the best aggregated model (Random Forest) at 0.495. Table 5.17 summarises these top-performing models across feature types. While the margin of improvement was modest, it demonstrates the added predictive value of leveraging time-series feature extraction.

Table 5.15: Lowest RMSE Values from TSFRESH Feature Sets

Model	Feature Set	Filter Type	RMSE
GradientBoosting	efficient	pca_filtered	1.318
RandomForest	efficient	filtered	1.323
GradientBoosting	efficient	filtered	1.331
LightGBM	efficient	filtered	1.350
RandomForest	efficient	pca_filtered	1.348
GradientBoosting	efficient	vif_filtered	1.349
XGBoost	efficient	filtered	1.383
RandomForest	minimal	filtered	1.383

Table 5.16: Top SHAP and Model-Based Feature Importances from TSFRESH Efficient PCA-Filtered Models

Feature	XGBoost Imp.	XGBoost SHAP	LightGBM Imp.	LightGBM SHAP
pca_1	0.129	**0.837**	83	**0.851**
pca_2	0.053	0.332	65	0.324
age	0.005	0.102	30	0.124
pca_17	0.020	0.087	25	0.079
pca_5	0.014	0.079	38	0.069
height_cm	0.004	0.078	34	0.089
pca_12	0.007	0.061	32	0.079
weight_kg	0.006	0.055	20	0.055
pca_23	0.013	0.053	21	0.040
pca_19	0.017	0.050	30	0.049

Note: Top 10 features ranked by SHAP values in XGBoost, with corresponding importance across both models.

Model Complexity

While aggregated feature sets required only 621 handcrafted features, the TSFRESH-based PCA feature set relied on over 90 principal components to achieve slightly improved performance. This highlights a trade-off between complexity and accuracy, where the richer time-series feature extraction comes at the cost of increased dimensionality and reduced interpretability.

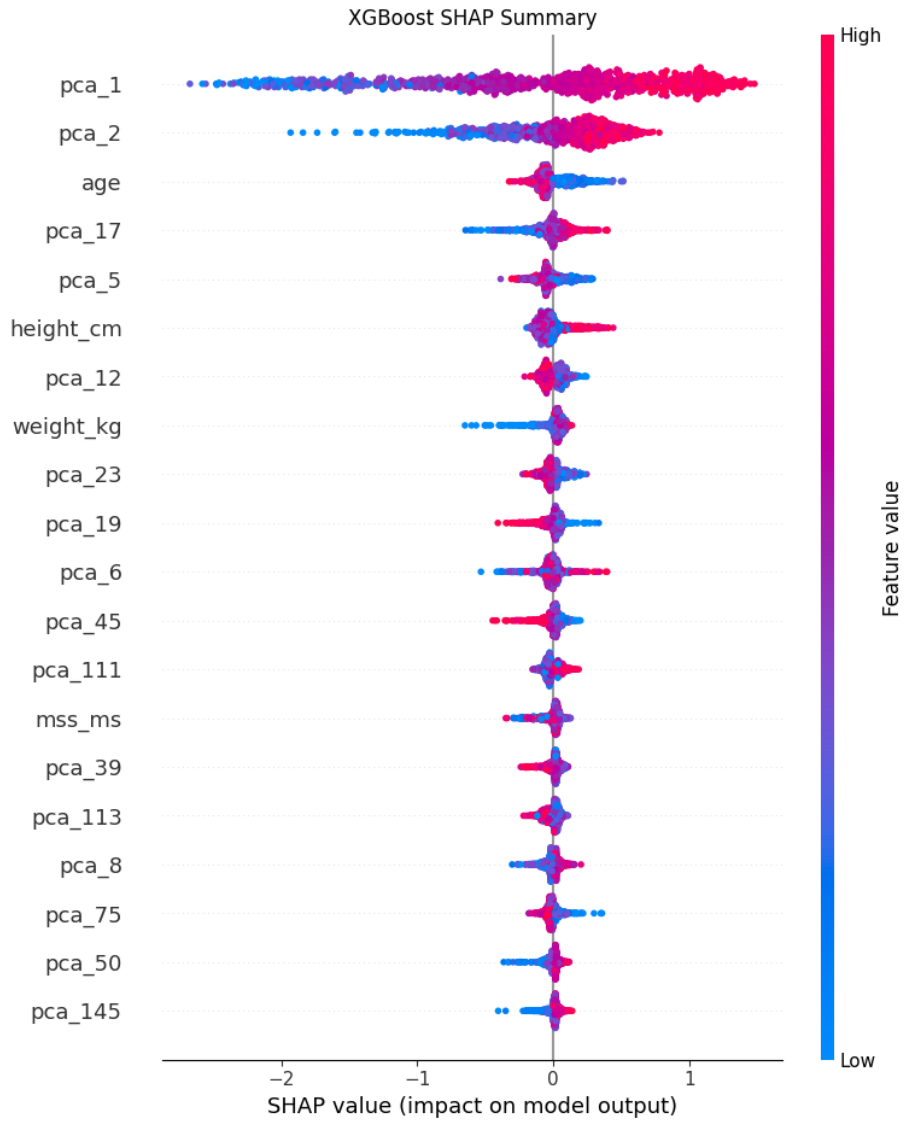


Figure 5.11: SHAP summary plot for XGBoost using TSFRESH efficient PCA-filtered features.

Interpretability

SHAP analysis revealed aggregated models relied heavily on interpretable features such as total distance and running time, whereas TSFRESH models derived importance primarily from latent PCA components. Despite reduced interpretability, the richer time-series structure may enable better generalisability.

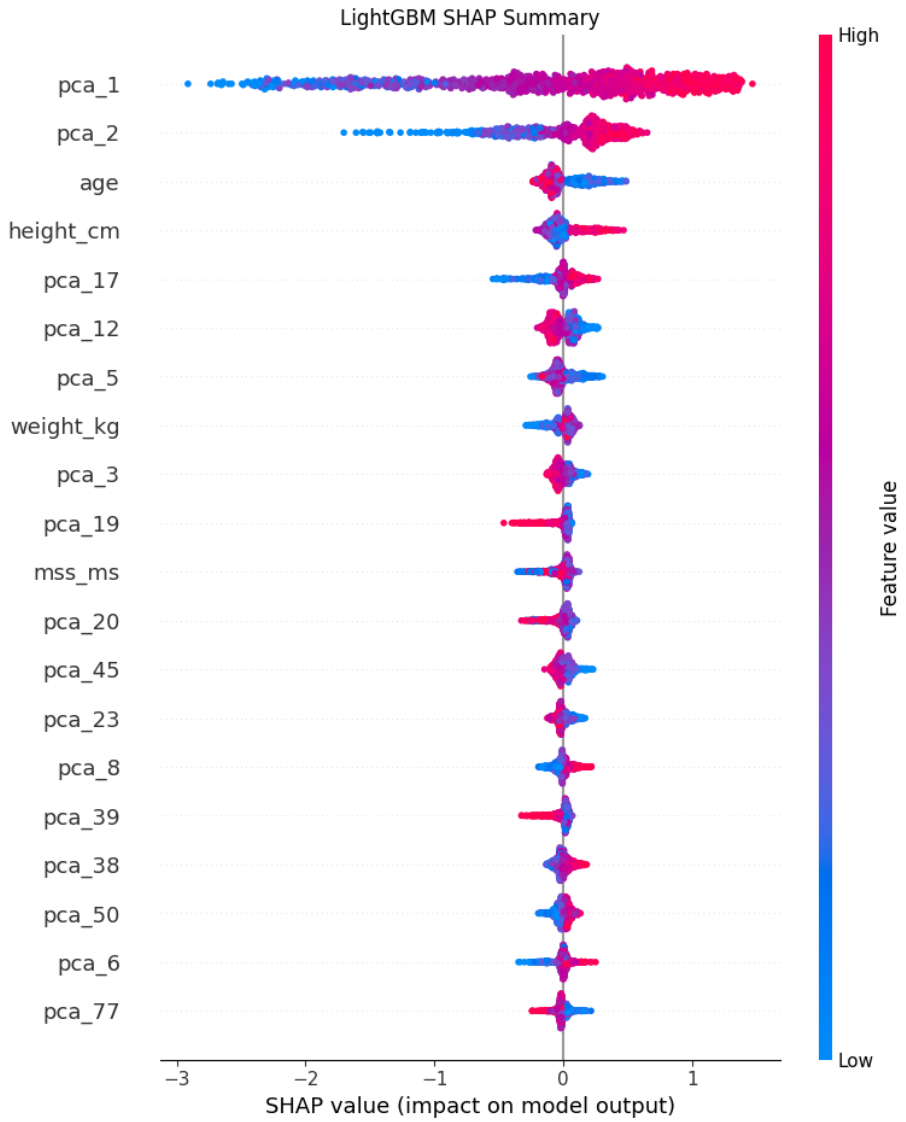


Figure 5.12: SHAP summary plot for LightGBM using TSFRESH efficient PCA-filtered features.

5.4 Discussion

The objective of these experiments was to determine whether the extraction of detailed time-series features from raw GNSS data could enhance the accuracy of ML models in predicting RPE. This study leverages the Python library `TSfresh` to extract a wide range of time and frequency domain features from raw positional and velocity data sampled at 2Hz.

By comparing models built on both traditional aggregated features and multiple TSFRESH-based time-series feature sets, this research addresses key challenges such

Table 5.17: Comparison of Best-Performing Models using Aggregated and TSFRESH Feature Sets

Model	Feature Set	R^2	MAE	RMSE
Random Forest	Aggregated (All Features)	0.495	1.081	1.341
Gradient Boosting	TSFRESH (Efficient + PCA)	**0.512**	**1.052**	**1.318**

as multicollinearity and the loss of temporal structure in aggregated representations.

By capturing trends and fluctuations that aggregated metrics may overlook, this approach provides a more nuanced understanding of athlete workloads.

This is the first study in team sports to apply time-series feature extraction techniques to GNSS data using multiple TSFRESH configurations, including efficient and minimal feature sets, to predict RPE. It addresses a significant gap in the literature and proposes a novel approach for leveraging raw sensor data to better explain subjective athlete responses.

To establish a benchmark, models were trained using conventional aggregated features widely used in sports science. These included total distance, high-speed running distance, and acceleration counts, which served as a practical and interpretable baseline.

The experimental design systematically explored three time-series feature configurations: *(i)* a statistically filtered subset of features selected via F-statistics and p-values, *(ii)* VIF-filtered features to address multicollinearity, and *(iii)* PCA-transformed features retaining 90% of the variance.

The experimental setup systematically explores the impact of time-series feature engineering and custom feature selection on predictive performance.

The experimental results demonstrate that models trained on TSFRESH-derived time-series features consistently outperformed those using traditional aggregated GNSS metrics across multiple evaluation metrics. Notably, the Gradient Boosting model using the efficient PCA-reduced TSFRESH feature set achieved the highest R^2 score (0.512), the lowest RMSE (1.318), and the lowest MAE (1.052). In comparison, the best aggregated modelRandom Forest using the full feature setreached an R^2 of

0.495 and RMSE of 1.341. Although the performance margin appears modest, this consistent improvement across models underscores the added value of preserving temporal structure when modelling perceived exertion.

The observed improvements in predictive accuracy with TSFRESH-derived features came at the cost of increased model complexity. Aggregated feature sets typically required only 621 handcrafted variables, offering straightforward interpretability and minimal computational overhead. In contrast, the TSFRESH PCA-reduced feature sets relied on more than 90 principal components to achieve slightly superior performance. This increase in dimensionality introduces additional computational burden and reduces the transparency of model decision-making. While richer in temporal detail, these models are inherently more complex, presenting a trade-off that must be considered in applied athlete monitoring systems.

Performance trends across models highlighted the clear advantage of ensemble methods in handling high-dimensional time-series data. Gradient Boosting and Random Forest consistently achieved the highest R^2 scores (0.512 and 0.509, respectively) and the lowest RMSE values (1.318 and 1.323) across TSFRESH feature sets, particularly when paired with efficient or PCA-reduced representations. In contrast, Decision Tree regressors showed poor performance under all feature conditions, with negative R^2 values when using PCA-reduced aggregated metrics. This reinforces the limitation of shallow learners in capturing complex temporal patterns and multivariate dependencies. Interestingly, XGBoost showed only moderate performance despite its reputation for robustness, suggesting it may not generalise as well under LOSOCV or when exposed to high-dimensional but noisy feature spaces. LightGBM consistently performed well but did not outperform Gradient Boosting or Random Forest in any configuration. These trends suggest that, for subjective outcome prediction like RPE, model stability and tolerance to multicollinearity are more important than raw complexity or speed.

To ensure generalisability, all models were evaluated using Leave-One-Subject-Out Cross-Validation (LOSOCV). This approach guards against overfitting and en-

sures that predictive performance reflects the ability to generalise across different athletes a critical factor in real-world deployment of RPE prediction models. Consistent improvements in R^2 and RMSE across LOSOCV folds further confirmed the robustness of the time-series feature sets.

To enhance the interpretability of model predictions, SHAP (SHapley Additive exPlanations) values were used to quantify feature contributions across models. In the aggregated feature models (Figure 5.8 and 5.9), `total_distance`, `running_duration`, and `running_distance` emerged as the most influential predictors. These are intuitive and widely accepted markers of training load, aligning well with practitioner expectations. Physical characteristics such as `age`, `height_cm`, and `mss_ms` also showed moderate importance, suggesting that RPE is shaped by both workload and individual context.

In contrast, SHAP analysis of the TSFRESH PCA-based models (Figure 5.11 and 5.12) revealed that predictive power was concentrated in the first few principal components (e.g., `pca_1`, `pca_2`), which capture latent temporal patterns in the raw data. While this dimensionality reduction enabled strong predictive performance, it also reduced feature transparency. Nonetheless, contextual variables such as `age`, `height_cm`, and `weight_kg` remained influential even in PCA-transformed models, highlighting the persistent role of athlete-specific characteristics across feature engineering pipelines.

Together, these findings underscore the trade-off between interpretability and predictive precision: aggregated features support practitioner-facing insights, while TSFRESH-derived components offer a more powerful but opaque data-driven representation.

Existing studies, such as Imbach et al. (2022), have highlighted the untapped potential of raw GNSS data for modelling performance and injury. However, these efforts primarily focused on mechanical indicators such as force, speed, and velocity-derived metrics. In contrast, this study extends the application of raw GNSS data to predict subjective internal responses, specifically RPE, through time-series feature

extraction. By incorporating contextual variablesPRS, IC (e.g., age, body composition, playing position)this study provides a more holistic approach to athlete monitoring.

FatigueNet, a deep learning approach proposed by Kim et al. (2022), demonstrated notable gains over traditional methods by leveraging multidimensional raw GNSS data directly. Their model achieved a **19.39% reduction in MAE** and a **10.75% reduction in RMSE** compared to a random forest baseline, illustrating the value of end-to-end temporal learning. By extracting complex temporal and directional information across full-length sequences, the model enhances RPE predictions but at the cost of interpretability and increased computational burden.

The current study employed a time-series feature engineering approach using `TSfresh` and dimensionality reduction (e.g., VIF filtering and PCA) to capture signal dynamics at 2Hz while maintaining model interpretability. While the R^2 improvement was modestfrom 0.495 with Random Forest (aggregated) to 0.512 with Gradient Boosting (TSFRESH PCA)the Root Mean Squared Error decreased from 1.341 to 1.318. This reduction in RMSE, though only 1.71%, corresponds to more precise RPE predictions on average. Given the bounded and subjective nature of RPE, even small improvements in prediction accuracy are valuable in applied contexts. When such gains are achieved using interpretable, handcrafted features, they represent a meaningful step toward practical, deployable models in athlete monitoring.

While mid-range RPE values were predicted accurately by TSFRESH-based models, performance at extreme values remained less stablean area where deep learning methods like FatigueNet showed an advantage. This suggests that handcrafted features may underrepresent rare, high-exertion patterns present in the full time series.

These results support the view that while deep learning models can leverage raw data more comprehensively, handcrafted time-series features remain valuable for their transparency, lower computational cost, and practical utility in applied

settings. Future work could explore hybrid pipelines that combine interpretable engineered features with deep learning representations to enhance both accuracy and usability.

While the R^2 improvement was modestfrom 0.495 with Random Forest (aggregated) to 0.512 with Gradient Boosting (TSFRESH PCA)the Root Mean Squared Error decreased from 1.341 to 1.318. This reduction in RMSE, though only 1.71%, corresponds to more precise RPE predictions on average. Given the bounded scale of RPE and its relevance in applied contexts, this improvement reflects a meaningful gain in prediction accuracy, especially when achieved through interpretable, handcrafted features.

Interestingly, the aggregated baselines in this studywhich integrated contextual variables such as PRS, IC, and positional labelsoutperformed the baseline models reported in Kim et al. (2022), reaffirming the importance of integrating non-GNSS information to explain RPE variability (Sheridan et al., 2024).

As also noted by Imbach et al. (2022), raw GNSS data presents considerable potential for performance modelling, but introduces complexity in the form of signal noise and high dimensionality. This study used `TSfresh` to address those challenges, extracting meaningful temporal patterns while reducing the need for deep model architectures.

By bridging the gap between summary statistics and full deep learning models, time-series feature engineering offers a practical and interpretable solution. This approach is particularly suited for applied environments with limited data or compute infrastructure, enabling insights to be drawn from wearable data without full-scale model training.

5.5 Summary

This chapter investigated time-series feature sets and modelling approaches to predict RPE using raw GNSS data in elite Gaelic football players. The study focused on the extraction of features with `TSfresh`, comprising core time-series features,

and used custom-tailored feature selection methods to optimise model performance. These approaches aimed to balance feature richness, model complexity, and predictive accuracy. GBM consistently emerged as the most effective model, leveraging its capability as both a feature selector and predictor. Interestingly, the full aggregated feature set without VIF or PCA filtering produced the best baseline performance. This suggests that, despite potential multicollinearity, the ensemble-based models used (e.g., RF, GBM) were sufficiently robust to maintain accuracy. Nevertheless, this highlights a trade-off: while dimensionality reduction improves model generalisability in some cases, retaining domain-specific features can offer better practical performance when interpretability and completeness are priorities.

The best-performing configurationGradient Boosting with PCA-reduced TSFRESH efficient featuresachieved an RMSE of 1.318 and an R^2 of 0.512, outperforming all models trained on aggregated metrics. The importance of model-specific feature selection was highlighted, with algorithms like XGBoost and GBM achieving significant gains when tailored RFE was applied. Key findings demonstrated that these time-series features provided a richer temporal and multidirectional representation of player movements, potentially benefiting advanced algorithms capable of handling intricate data, such as GBM and XGBoost, while simpler models like DT were less effective.

The results indicate the potential of time-series features for athlete monitoring. Using raw GNSS data and applying sophisticated time-series feature extraction techniques, it was demonstrated that the predictive accuracy for RPE could be improved. Although the performance gain in RMSE (1.71%) over the best aggregated model was small, it was achieved with generalisable models using LOSOCV and enhanced feature attribution through SHAP analysis. The selection of model-specific features emerged as a critical factor, allowing algorithms to harness the richness of the `TSfresh` feature set fully. Advanced models such as GBM and XGBoost excelled in handling complex data, achieving the highest accuracy across experimental setups. SHAP-based interpretability further revealed that while aggregated models relied on

intuitive metrics like total distance and running duration, TSFRESH-based models drew predictive power from latent principal components and contextual features such as age and height. These findings support the potential use of raw GNSS data in combination with time-series features as a promising approach to optimise performance monitoring systems in team sports. Overall, highlight the value of aligning time-series feature extraction and model selection strategies with the strengths of predictive algorithms.

In the first two experimental chapters, the focus was on predicting athletes' internal load through subjective perceptual responses, specifically RPE. Using external load metrics derived from GNSS and additional contextual features, the goal was to model the intensity of the workout and improve the precision of the training prescriptions. These chapters highlighted how subjective measures can provide valuable insight into an athlete's immediate response to training stimuli, informing decision-making and tailored load management strategies. In the next chapter, the focus shifts to a more objective physiological measure: Oxygen Uptake (VO_2). Unlike RPE, which reflects an athlete's perception of effort, oxygen uptake provides a direct and quantifiable metric of internal load and aerobic capacity. This transition allows for exploration of the predictive power of raw data in estimating VO_2, addressing long-term adaptations and improvements in fitness, such as VO_2 max. By investigating the physiological response, a deeper understanding of how consistent training enhances aerobic capacity will be developed, bridging the gap between subjective and objective monitoring in athlete performance evaluation.

Chapter 6

Estimating VO_2 Using Wearable Sensors

6.1 Introduction

This chapter investigates the estimation of Oxygen Uptake (VO_2) during team sports activities using wearable sensor data collected during load monitoring. Accurate VO_2 estimation is critical for understanding an athlete's physiological demands, to help optimise performance and effectively manage training load. Given the dynamic and intermittent nature of team sports, predicting VO_2 poses unique challenges that require innovative solutions. To address these challenges, this study conducts a comprehensive evaluation of traditional Machine Learning (ML) models alongside advanced deep learning approaches, such as Long-Short-Term Memory (LSTM) networks. The chapter focuses on assessing the effectiveness of different data representations and configurations of body-worn accelerometers in predicting VO_2. Using data from a Inertial Measurement Unit (IMU), we aim to identify optimal methods to provide an accurate real-time estimation of VO_2 during training and competition. This work represents a significant step toward the development of improved athlete monitoring systems. By bridging the gap between wearable sensor data and physiological metrics, the findings contribute to the design of practical tools that improve

training precision and athlete performance in team sports.

The hypothesis is that raw IMU data represent an underused source of information, offering greater sensitivity for capturing high-intensity actions and rapid movement transitions compared to traditional metrics. This precision enables a more accurate assessment of the physiological demands placed on athletes. Based on this hypothesis, a study was developed to compare the predictive precision of various deep learning models in estimating individual VO_2 during outdoor jogging and simulated team sports activities. By examining raw IMU data and multiple sensors configurations, this study aims to identify the optimal input method for ML models to estimate VO_2 during high-intensity team sports activities. This comparative approach evaluates the effectiveness of different data representations and modelling techniques, offering valuable insight into the complex relationship between external workload and internal physiological responses. Ultimately, this chapter seeks to optimise athlete monitoring by leveraging advanced ML techniques and diverse sensor inputs, paving the way for noninvasive, real-time VO_2 monitoring. Such advancements have the potential to revolutionise the management of training load, enabling coaches and sports scientists better to understand athlete responses, reduce injury risk and improve performance during high-demand scenarios.

6.2 Methodology

6.2.1 Feature Engineering

In this study, the influence of different representations of IMU data on model accuracy is investigated by comparing features derived directly from raw data (RAW) with those derived through feature engineering, specifically using the Mean Absolute Deviation (MAD). This approach aims to determine the most effective representation of the data set to estimate VO_2 during team sports activities.

To evaluate the impact of data set representation, the raw time series data from the 6-axis IMU sensor, consisting of 3-axis accelerations and 3-axis angular

velocities, were transformed into engineered features. The MAD values specific to the axes were calculated for each sensor axis (x, y, z), and a resultant MAD value (MAD_{xyz}) was calculated using Equation 6.1:

$$\text{MAD}_{xyz} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})^2 + \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{Y})^2 + \frac{1}{N}\sum_{i=1}^{N}(z_i - \overline{Z})^2} \quad (6.1)$$

This transformation was applied to windows of IMU data segmented between breaths, compressing the data into a single row per breath. The resultant MAD (MAD_{xyz}) is sensitive to axis inclination and movement, offering a more nuanced representation of activity than axis-specific MAD alone (Vähä-Ypyä et al., 2023). The same procedure was applied to the acceleration and angular velocity data, generating two MAD_{xyz} features (Accel xyz and Gyro xyz) per IMU sensor. For three IMUs, this process yielded six engineered features, which were subsequently used in the modelling experiments.The input data structure was also optimised to improve deep learning performance. Four different input window sizes (1, 3, 5, and 7 breaths) were evaluated to determine the best temporal representation for predicting VO_2. This exploration ensured that the input format was well-adapted to the modelling approaches, balancing data compression with temporal context.

Sensor Configurations

The study also examined the influence of various sensor placements on the predictive accuracy of VO_2 estimation. By systematically combining different IMU placements and physiological data inputs, four different input configurations were tested:

- **Input Set A:** Heart Rate (HR) + Breathing Rate (BR) + IMU Torso
- **Input Set B:** HR + BR + IMU Torso + IMU Arm
- **Input Set C:** HR + BR + IMU Torso + IMU Leg
- **Input Set D:** HR + BR + IMU Torso + IMU Arm + IMU Leg

These configurations were designed to assess the relative importance of different sensor placements and combinations in capturing the physiological and biomechanical demands associated with VO_2. By testing incremental additions of sensors, the study aimed to determine the trade-off between model complexity and predictive accuracy.

6.3 Machine Learning Approach

Deep learning models that have shown potential in predicting VO_2 kinetics during intermittent transitions (Amelard et al., 2021; Davidson et al., 2023; Hedge et al., 2023; Zignoli et al., 2020) against a baseline Multiple Linear Regression (MLR) model were used.

A supervised ML regression approach was used, employing a modified Leave-One-Subject-Out (LOSO) cross-validation strategy. In this approach, test subject data from the first visit is included in the training phase before testing on data from the second visit. A portion of the training data is set aside as a validation set to ensure the model's reliability and accuracy. This validation set is used to select hyperparameters by minimising the Root Mean Square Error (RMSE). A grid search technique is applied to identify the optimal hyperparameters that achieve the lowest RMSE. The number of layers and filters are hyperparameters and are selected in the same way for all models.

Model Evaluation

Subsequently, a residual analysis was performed to assess the prediction capacity of the model. This involves calculating residuals as the difference between the measured VO_2 values and the predicted VO_2 values. To quantify the accuracy of the predictions, Mean Absolute Error (MAE) and the RMSE for the residuals are calculated. In addition, a regression analysis of these residuals yields the Pearson correlation coefficient (r) and explains the proportion of variance (R^2) for which each model accounts. Furthermore, a Bland-Altman analysis is employed to assess

the agreement between the measured and predicted VO_2 values, calculating both the mean bias and the limits of agreement at a 95% confidence interval (twice the standard deviation).

6.4 Experimental Results

From Table 6.1, the LSTM model with RAW data representation and Set C input configuration achieved the best performance (RMSE: 4.976, MAE: 3.698 $mL \cdot kg^{-1} \cdot min^{-1}$). This indicates that the LSTM model had the lowest prediction error among all tested models and configurations. Other models, such as CNN and MLP, also showed competitive performance. For example: CNN with RAW and Set C: MAE of 4.174 ($mL \cdot kg^{-1} \cdot min^{-1}$); MLP with RAW and Set C: MAE of 4.326 ($mL \cdot kg^{-1} \cdot min^{-1}$). The results indicated that deep learning models, particularly LSTM and CNN, performed best with RAW data representations. MLR models showed robust performance on both RAW and MAD data representations. For example: LSTM model with RAW data and Set C input: Test MAE of 3.698 ($mL \cdot kg^{-1} \cdot min^{-1}$); MLR model with MAD data and Set B input: Test MAE of 3.797 ($mL \cdot kg^{-1} \cdot min^{-1}$). The analysis of different sensor configurations revealed that using a multisensor setup composed of one sensor on the torso and additional sensors on the arm (Set B) or leg (Set C) provided the best performance for estimating oxygen consumption. A single sensor on the torso (Input A) also provided good prediction accuracy.

Figure 6.1a shows the relationship between the predicted VO_2 and the measured VO_2 for the LSTM model with RAW data and the Set C input configuration. The value of R^2 is 0.87, indicating a strong correlation between the predicted and measured values of VO_2. This high value R^2 suggests that the LSTM model explains 87% the variability in the measured VO_2 data. There is a consistent trend where the predicted values closely follow the measured values, indicating good model performance. Figure 6.1b is a Bland-Altman plot, which shows the difference between the measured and predicted VO_2 values versus the average of the two. This plot helps to identify any systematic bias and the limits of agreement (LoA) for the predictions.

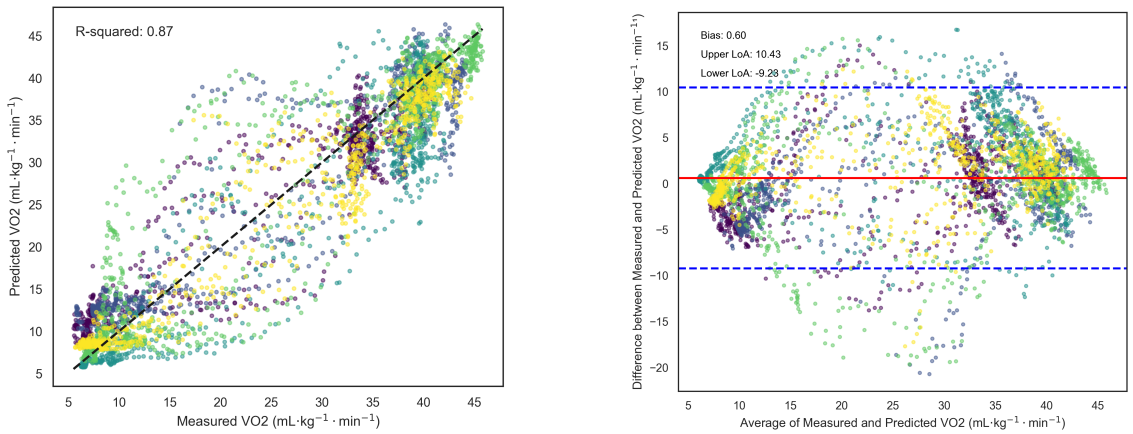
Table 6.1: **Performance metrics of ML models.**

Model	Data	Input	Valid		Test	
			RMSE	MAE	RMSE	MAE
LSTM	RAW	Set C	4.1113	3.2988	4.9763	3.6979
MLR	MAD	Set B	7.2175	6.3824	5.0101	3.7977
MLR	MAD	Set D	7.2620	6.4323	5.0861	3.8283
MLR	RAW	Set A	6.2807	5.2638	4.9406	3.9537
MLR	MAD	Set A	7.1939	6.3416	5.0850	3.9756
MLR	RAW	Set C	5.7539	4.7465	5.2296	4.0882
CNN	RAW	Set C	4.5602	3.7178	5.6945	4.1747
MLR	MAD	Set C	7.2695	6.4395	5.4021	4.3061
MLP	RAW	Set C	4.2032	3.3455	5.6987	4.3268
MLP	RAW	Set A	4.1333	3.0814	5.7361	4.3335
LSTM	RAW	Set D	3.9768	3.0669	5.9422	4.3835
XGBoost	RAW	Set D	4.4614	3.4353	5.8601	4.5239
MLP	RAW	Set D	4.4614	3.4353	5.8601	4.5239
CNN	RAW	Set A	4.1312	3.2282	6.0970	4.5933
LSTM	RAW	Set A	4.3833	3.6292	6.5239	4.7223

Table notes: This table shows the performance metrics of different ML models, including RMSE and MAE ($mL \cdot kg^{-1} \cdot min^{-1}$) for validation and test sets.

The bias is 0.50 ($mL \cdot kg^{-1} \cdot min^{-1}$), indicating a slight overprediction on average. The upper limit of agreement (upper LoA) is 10.24 ($mL \cdot kg^{-1} \cdot min^{-1}$), and the lower limit of agreement (lower LoA) is -9.23 ($mL \cdot kg^{-1} \cdot min^{-1}$). This range indicates the spread of the differences between the measured and predicted values. Most data points are within the limits of agreement, suggesting that the model predictions are generally accurate.

Figure 6.2 illustrates the residuals (predicted VO_2 minus measured VO_2) under different exercise conditions for the LSTM model using RAW data and the set C input configuration. The exercise conditions include baseline, jogging, recovery1, circuit1, recovery2, circuit2, and recovery3. The box plot reveals that the LSTM model with RAW data and input set C input configuration generally has a median residual of around 0 under different exercise conditions, indicating that there is no significant prediction bias. However, there is noticeable variability in the residuals, particularly during the recovery phases. The model predicts during circuit activities more consistently than in the baseline and recovery phases.



(a) Linear Correlation plot

(b) Bland-Altman plot

Figure 6.1: The linear correlation plot showing the relationship between predicted VO_2 and measured VO_2 for the LSTM model using RAW data and Set C input configuration, with an R^2 value of 0.87. The Bland-Altman plot illustrating the difference between measured and predicted VO_2 values against the average of the two for all subjects combined.

Figure 6.3 illustrates the breath-by-breath VO_2 predictions (blue line) versus the measured VO_2 values (green line) for the LSTM model using RAW data and Set C input configuration. Although the predicted VO_2 values (blue line) generally follow the trend of the measured VO_2 values (green line), there are noticeable deviations during the exercise and recovery phases. The model captures the general pattern of VO_2 changes but struggles with precise tracking. The model shows a significant variation from the measured values of VO_2 during high-intensity exercise phases. The predicted values tend to either overshoot or undershoot the peaks of the measured values. This indicates that the model has difficulty accurately estimating VO_2 during increased physical activity. In the recovery phases, the model predictions do not closely track the rapid decrease in VO_2 seen in the measured values. It appears to lag or lead in response to the actual changes. The discrepancies suggest that the model struggles to accurately capture the oxygen kinetics during recovery. Finally, the model exhibits inconsistencies in tracking rapid changes, particularly during transitions between exercise and rest.

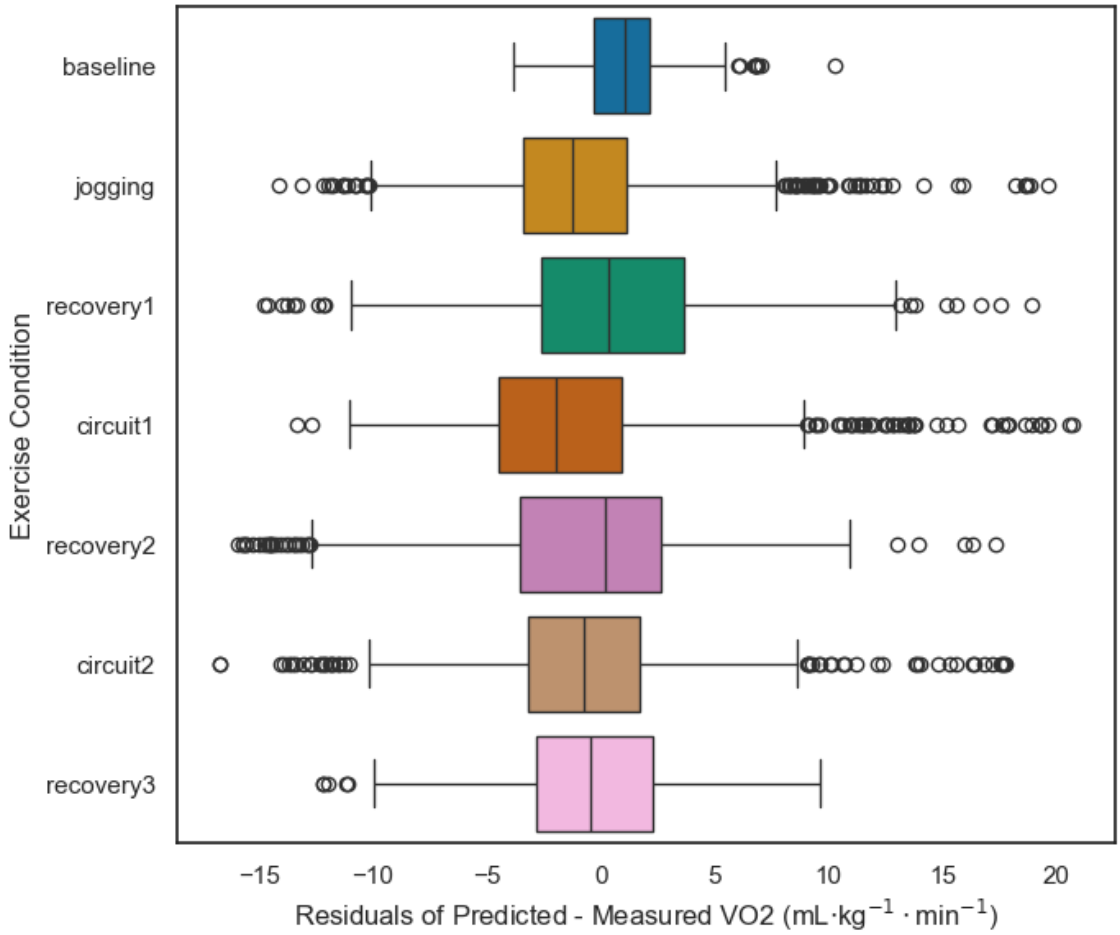
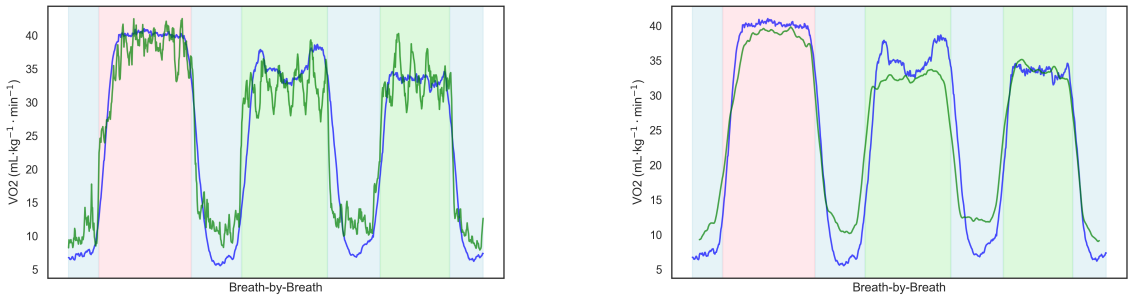


Figure 6.2: The box plot illustrates the residuals (predicted VO_2 minus measured VO_2) across different exercise conditions for the LSTM model using RAW data and Set C input configuration. The exercise conditions include baseline, jogging, recovery1, circuit1, recovery2, circuit2, and recovery3.

6.5 Discussion

In these experiments, the ability of ML models to estimate individual VO_2 during simulated team sports activities using wearable sensor data was investigated. The comparative analysis revealed a marginal advantage for the deep learning LSTM model over the baseline MLR model in predictive power. The MLR model with the best performance achieved an MAE of 3.80 ($mL \cdot kg^{-1} \cdot min^{-1}$), slightly behind the LSTM model, which achieved an MAE of 3.70 ($mL \cdot kg^{-1} \cdot min^{-1}$) in the test set, as shown in Table 6.1. This result demonstrates that while deep learning models provide strong performance, traditional linear models such as MLR can still deliver highly competitive predictions.



(a) Unsmoothed predictions

(b) Smoothed predictions

Figure 6.3: The graph compares breath-by-breath VO_2 predictions (blue line) with measured VO_2 values (green line) for the LSTM model using RAW data and Set C input configuration for Subject 2. The left plot shows unsmoothed predictions with an MAE of 3.374 ($mL \cdot kg^{-1} \cdot min^{-1}$), while the right plot shows smoothed predictions with an MAE of 2.902 ($mL \cdot kg^{-1} \cdot min^{-1}$). The plot includes different exercise and recovery phases, shaded as follows: baseline and recovery phases (light blue), jogging (light grey), and simulated soccer circuit (light green).

The performance of the model across two data representations was computed: raw sensor data (RAW) and feature-engineered data (MAD). Deep learning models, such as LSTM and CNN, exhibited strong performance when trained on RAW data. However, MLR models were more effective with MAD data representations, indicating the value of feature engineering for simpler models. In particular, the performance gap between the RAW and MAD data representations was less pronounced for the MLR models, highlighting the robustness of this approach.

The choice of sensor configurations significantly influenced the accuracy of the model. Configurations incorporating multiple sensors, such as the torso and leg (Set C) or the torso and arm (Set B), yielded the most accurate predictions, with Set C emerging as the most effective across all models. A single torso-mounted sensor (Set A) also provided reasonably good predictive performance, offering a simpler alternative for practical applications.

As this research represents one of the first applications of ML models on wearable sensors data collected during load monitoring training in team sports to predict VO_2 during simulated team sports activities, direct comparisons with existing studies remain limited. However, some parallels can be drawn. For example, an LSTM model using heart rate, mechanical power output, pedalling cadence, and respiratory

frequency as input was used to estimate VO_2 during variable high-intensity cycling exercises, achieving an MAE of approximately 3.50 ($mL \cdot kg^{-1} \cdot min^{-1}$) and an R^2 value of 0.89 (Zignoli et al., 2020). In the present study, the LSTM model achieved an R^2 value of 0.87 (Figure 6.1a), closely aligning with these findings.

Similarly to Zignoli et al. (2020), the LSTM model in this study was trained using data from a graded exercise test, with two arbitrary protocols of varying intensities used to evaluate its predictive performance. Zignoli et al. (2020) also compared the LSTM model with two baseline analytical models, which achieved R^2 values of 0.83 and 0.90, respectively. These results indicate a comparable performance between the LSTM and the baseline models. The present study's findings align with this observation, as the baseline MLR model performed on a par with the LSTM model, achieving a R^2 value of 0.87. These results suggest that while deep learning models such as LSTM offer effective tools for VO_2 prediction, simpler approaches like MLR remain highly competitive. This underscores the importance of considering model complexity and data representation when selecting an approach for real-world applications.

Compared to a study that used an LSTM model with motion features from GNSS and IMU data during unconstrained outdoor walking and running, the results in this study highlight notable differences in performance and experimental design. Their reported MAE of 1.36 ($mL \cdot kg^{-1} \cdot min^{-1}$) significantly outperformed the best LSTM model by 2.33 ($mL \cdot kg^{-1} \cdot min^{-1}$) (Davidson et al., 2023). The superior performance observed in their study can likely be attributed to differences in the experimental protocol. Their protocol involved four different three-minute exercise conditions, consisting of two walking and two running sessions, which likely facilitated steady-state activity. This contrasts with the protocol in this present study, which aimed to simulate the intermittent and variable intensity nature of team sports activities. The steady-state conditions in their study may have simplified the physiological dynamics, allowing the LSTM model to achieve greater accuracy.

In addition to supporting this, their LSTM model that uses only Heart Rate (HR)

as input achieved an MAE of 2.52 ($mL \cdot kg^{-1} \cdot min^{-1}$), still outperforming the best LSTM configuration. However, this improvement may reflect the more predictable physiological responses during steady-state activities, rather than a fundamental advantage of the model. Their best-performing LSTM model utilised 93,151 total parameters and was trained for more than 8,000 epochs. While such a high parameter count demonstrates the model's capacity to learn complex relationships, it also raises concerns about overfitting, particularly when applied to smaller datasets. As highlighted in Shirdel et al. (2021), increasing model complexity can lead to diminishing returns, and marginal improvements over simpler models often fail to justify the additional computational cost and risk of overfitting. In contrast, the present study maintained a balance between model complexity and generalisability. Despite using fewer parameters and a less resource-intensive training process, the deep learning models performed comparably to simpler approaches such as MLR. This suggests that for tasks involving intermittent and variable intensity activities, simpler models can offer a more competitive and practical alternative to deep learning.

One possible explanation for the observed results is the tendency of deep learning models to overfit the training data, as evidenced by the consistently higher RMSE and MAE values on the validation sets compared to the test sets (Table 6.1). This discrepancy highlights the models' limited ability to generalise to unseen data. This pattern where test error appears lower than validation error is likely a result of fold-dependent variability during validation. Validation performance was estimated using a modified LOSO cross-validation on the training data, where some folds may have included particularly difficult subject cases or imbalanced input distributions. In contrast, the test set, held out entirely during training, may have had a more typical or representative distribution. No leakage occurred in the pipeline, and all test results reflect models trained and validated independently. However, these differences suggest that model performance estimates from validation alone may not fully capture real-world generalisation performance. Future studies should consider nested cross-validation or more comprehensive folds to mitigate this effect. Among

the deep learning models, the most pronounced overfitting was observed in the MLP, followed by the CNN and LSTM models. Although the LSTM showed the best generalisation among deep learning approaches, even it exhibited signs of overfitting, indicating a broader challenge in achieving robust performance on test data.

To mitigate overfitting, several techniques were applied, including L2 regularisation to constrain model complexity, cross-validation using a modified LOSO strategy to ensure robust evaluation, and early stopping to terminate training when validation performance plateaued. Despite these efforts, the results reveal that further optimisation is required to enhance the generalisability of deep learning models. Additional strategies, such as incorporating dropout layers, exploring simpler architectures or increasing the size of the data set, may help address these challenges.

These findings underscore a key limitation of deep learning in this context: while these models excel at capturing complex patterns in the data, their reliance on extensive training data and their susceptibility to overfitting may limit their utility for predicting VO_2 in highly variable and dynamic activities. For practical applications, this raises the question of whether the added complexity of deep learning is justified, particularly when simpler models like MLR demonstrate comparable performance. Future research should prioritise developing lightweight, generalisable architectures and refining training strategies better to balance model complexity with robust real-world performance.

A study conducted in a simulated futsal setting demonstrated the limitations of VO_2 estimation using simple linear regression equations derived from treadmill test HR data. While this method successfully matched measured VO_2 at a group level (p-value = 0.38), it failed to provide reliable predictions at an individual level. This was evident in the weak correlations and significant bias reported. A Bland-Altman analysis revealed a bias of -2.8 ($mL \cdot kg^{-1} \cdot min^{-1}$), with errors reaching up to 19 ($mL \cdot kg^{-1} \cdot min^{-1}$) (P. Silva et al., 2018). In contrast, the model in this study achieved a markedly reduced bias of 0.51 ($mL \cdot kg^{-1} \cdot min^{-1}$), with narrower limits of agreement (ś9 $mL \cdot kg^{-1} \cdot min^{-1}$), indicating superior accuracy in capturing

individual physiological responses.

Another study used a mixed effects non-penalised linear regression model to predict VO_2 max using HR and accelerometer data during submaximal running, achieving a MAE of 2.33 ($mL \cdot kg^{-1} \cdot min^{-1}$) (Brabandere et al., 2018). Although this performance is notable, this study's baseline MLR models demonstrated consistent accuracy in all six configurations tested in this study (Table 6.1). These findings highlight the potential of MLR models when combined with data fusion from wearable sensors, offering a balance of simplicity, robustness, and interpretability. In particular, linear models are less prone to overfitting, making them particularly suitable for studies with small sample sizes (Allen & Tkatchenko, 2022).

Both the studies by P. Silva et al. (2018) and Brady et al. (2022) relied on data collected from incremental fitness tests to develop VO_2 estimation models. These traditional approaches used linear regression analysis of HR and VO_2 data derived from treadmill tests (Achten & Jeukendrup, 2003). Although effective in controlled environments, such methods often struggle to account for the variability and complexity of VO_2 responses in dynamic, real-world activities. In contrast, the present study extends these approaches by integrating wearable sensor data and exploring both linear and non-linear models, demonstrating the advantages of data fusion and modern ML techniques in enhancing the accuracy of VO_2 estimation for individual athletes.

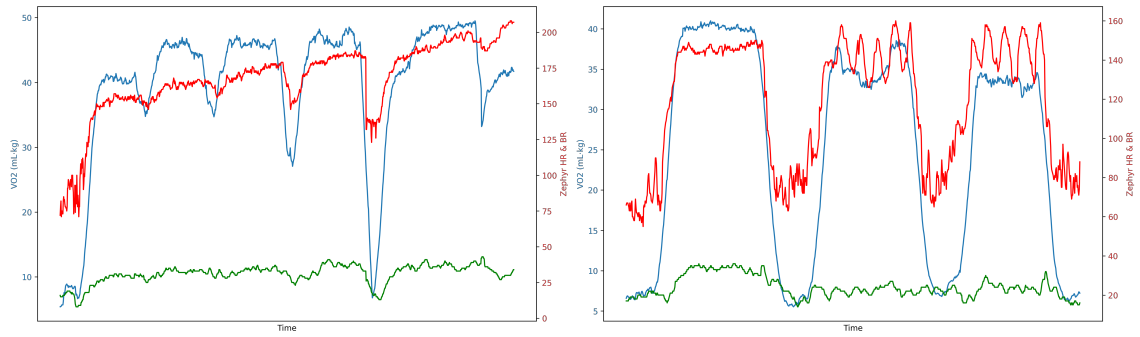


Figure 6.4: Examples of oxygen uptake (VO_2) measurements (blue) and the easy-to-obtain input physiological variables are HR measurements (red) and Breathing Rate (BR) (green) during the first (left) and second (right) visits for Subject 2.

As shown in Figure 6.4, the incremental fitness test data in this study demon-

strated a clear linear relationship between HR and VO_2. However, this relationship became less consistent during the outdoor simulated circuit test, where significant variations in HR were observed despite steady VO_2 levels. This highlights the complexity of VO_2 estimation during dynamic unstructured activities compared to controlled laboratory conditions.

One of the key strengths of deep learning models is their ability to capture transitions and patterns in dynamic and variable contexts (Amelard et al., 2021). However, adequate and diverse training data are essential for these models to generalise effectively. One of the research questions in this study aimed to determine whether data from structured laboratory fitness tests could be used to build ML models for predicting VO_2 during team sports activities. Although the incremental fitness test provided a controlled and structured dataset, its protocol may not fully represent the unstructured and intermittent activities typical of team sports.

Figure 6.2 illustrates the variability in the performance of the LSTM model with RAW data and the set C input configuration under different exercise conditions. The box plot shows that, overall, the model demonstrates a median residual close to zero in all conditions, indicating no significant prediction bias. However, greater variability in residuals is evident during recovery phases compared to circuit activities, where the model predictions were more consistent. This suggests that the structured, high-intensity nature of circuit activities aligns more closely with the training data than the lower-intensity, irregular recovery periods.

Simulated team sports activities, such as those used in this study, involve variable-speed locomotion and high-intensity actions such as directional changes, jumping, and sprinting (Taylor et al., 2017). These activities pose unique challenges for VO_2 estimation due to their dynamic and intermittent nature. Figure 3.3 illustrates how RAW 3D accelerometer and gyroscope data can capture nuanced differences between treadmill running, outdoor running, and team sport simulated circuits over a three-breath window. These high-frequency data are essential for detecting intense movement, which are critical components of team sports.

Each circuit in this study's protocol included eight distinct movements repeated five times, significantly increasing the number of transitions. Prior research has demonstrated the potential of IMU signals to detect high-intensity sports movements (Wundersitz et al., 2015). The ability of IMU data to reflect these transitions underscores its value as a data source for ML models, particularly for activities with rapid movement changes and high physiological demands.

Figure 6.3 illustrates the predicted unsmoothed VO_2 output across five individual simulated circuits, while Figure 6.3a highlights the effects of applying a smoothing function to the VO_2 input signal. Although smoothing is commonly recommended for steady-state VO_2 measurements, its application to intermittent activities may remove critical dynamic information. This loss of detail likely impairs the model's ability to learn VO_2 kinetics, contributing to the lag observed in its performance. As noted by Altini et al. (2016) neglecting transitional periods can reduce accuracy in VO_2 estimation. For example, during recovery phases, the model used here consistently overestimates VO_2 demands, as shown in Figure 6.3a. When smoothing was also applied to the VO_2 output (Figure 6.3b), the model underestimated demands during jogging and circuit 1, although predictions for circuit 2 aligned well. While the model captured overall trends, substantial variations and inaccuracies remain, particularly during recovery, suggesting the need for further refinement to improve its reliability in tracking VO_2 dynamics in team sports.

6.6 Summary

In this chapter, our final study demonstrated the effectiveness of using ML models to estimate VO_2 during simulated team sports activities. For these experiments, wearable sensors were required to regularly collect data during training load monitoring. This study was specifically chosen as the type of problem was one where we believed ML could play a major contribution. While it is accepted that this particular study may have been a challenge too far in terms of the scale of ambition, nevertheless, it is suggested that the learnings were sufficiently positive to continue

with this direction of research. Having now demonstrated three quite diverse applications to evaluate the application of ML in sport, the dissertation now moves to the final chapter for a summary and discussion around suitable topics to extend the work presented in this dissertation.

Chapter 7

Conclusions

The primary objective of this thesis was to investigate the use of Machine Learning (ML) models and wearable sensor data to enhance athlete monitoring. Specifically, this research focused on predicting perceptual metrics (Rating of Perceived Exertion (RPE)) and physiological metrics (Oxygen Uptake (VO_2)) to optimise training load management. By addressing critical research questions, this work aimed to demonstrate how ML can go beyond traditional descriptive analytics and aggregated metrics to provide actionable insights. This chapter summarises the key findings of the research, highlights the contributions made, and proposes future directions to advance this field.

7.1 Summary of Findings

This research addressed four key research questions, exploring the potential of ML and wearable sensor data in athlete monitoring through three interlinked studies. The findings underscore several contributions to the field.

Machine learning has demonstrated its utility in enhancing athlete monitoring by identifying the variables most influential in determining internal load responses, as measured by RPE. In Chapter 4, predictive experiments confirmed that absolute external load indices (e.g., total distance) provide more reliable signals of perceived exertion than relative intensity-based measures. Additional analyses incorporating

contextual-only models and dummy regressors/classifiers established clear performance baselines, strengthening the interpretability of results.

ML models proved especially effective when integrating multiple data modalitiesincluding personal characteristics, perceived wellness, and training contextreflecting the multidimensional nature of athlete monitoring. Through feature importance comparisons and SHAP-based explanations, this study highlighted the consistency and relevance of certain workload and contextual variables across model types. These capabilities support the application of ML not only as a predictive tool, but as a means of informing individualised training decisions grounded in objective and subjective data streams.

A key challenge in athlete monitoring is the reliance on manufacturer-provided aggregated metrics, which can obscure critical temporal and spatial information. This research demonstrated that raw wearable sensor data, particularly from Global Navigation Satellite Systems (GNSS), offers rich, high-frequency information that can improve the predictive accuracy of ML models for RPE. Chapter 5 applied advanced time-series feature extraction techniques using `Time Series Feature Extraction (TSfresh)`, revealing previously untapped temporal dynamics that enhanced model performance over traditional aggregated metrics. The best-performing configurationGradient Boosting with PCA-reduced time-series featuresachieved an RMSE of 1.318 and R^2 of 0.512, surpassing all baseline models. These improvements, though modest, were consistent across folds using Leave-One-Subject-Out Cross-Validation, reinforcing their generalisability.

Importantly, the study also found that the full aggregated feature set performed best when left unfiltered, suggesting that well-designed handcrafted features retain strong predictive value, even when multicollinearity may be present. Ensemble models such as Random Forest (RF) and Gradient Boosting Machine (GBM) proved robust to this, although multicollinearity remains a consideration in more interpretable or linear settings.

This research conducted an extensive evaluation of ML models, identifying en-

semble and tree-based models (e.g., RF, GBM, eXtreme Gradient Boosting (XGBoost)) as the most effective for RPE prediction tasks. These models consistently outperformed simpler methods, demonstrating their ability to handle the complexity and variability inherent in athlete monitoring data. The inclusion of SHAP analysis further enhanced the interpretability of both aggregated and time-series models, offering insight into which features, whether handcrafted or derivedcontributed most to predictions. This supports practical implementation by enabling coaches and practitioners to trace model decisions to meaningful training indicators.

This analysis provides valuable guidance for sports scientists selecting ML models for practical applications.

Chapter 6 explored the feasibility of using raw Inertial Measurement Unit (IMU) data to predict physiological responses such as VO_2. The findings showed that multi-sensor configurations, particularly those involving the torso and leg, provided the most accurate predictions. Although deep learning models such as Long-Short-Term Memory (LSTM) demonstrated some improvements, traditional methods such as. Multiple Linear Regression (MLR) offered comparable accuracy with lower computational demands. This research represents an important step toward continuous monitoring of athletes' aerobic fitness, leveraging existing wearable data without the need for additional testing.

7.2 Areas for Further Research

The findings of this research address several gaps identified in Chapter 2, particularly the need for more diverse datasets and advanced analytical techniques capable of integrating contextual and physiological variables. However, further research is necessary to build upon these contributions and overcome the limitations identified in this work.

For Chapters 4 and 5, the studies were limited to data collected from a single elite Gaelic football teamone study spanning three seasons and the other focused on a single season. Although these datasets provided high-resolution insight into elite-

level training, their homogeneity constrained the generalisability of the findings. Future research should expand datasets to include athletes from various sports, age groups, competitive tiers, and cultural contexts to improve model robustness. Collaborations between institutions and sports organisations to enable cross-sport validation of models would also advance the field.

Another area for future work involves the continued integration of contextual variablessuch as weekly schedule, recovery status, and travel demandsand the inclusion of objective well-being data from wearables, such as sleep quality, HRV, and heart rate recovery. These variables may provide valuable insight into individual responses and readiness, complementing subjective measures like RPE and PRS.

The use of raw sensor data opens opportunities for deeper exploration of feature sets that can better represent the temporal and spatial aspects of athlete movement profiles.

The use of raw sensor data opens opportunities for deeper exploration of feature sets that can better represent the temporal and spatial aspects of athlete movement profiles.

This study used TSFRESH to extract a diverse set of interpretable time- and frequency-domain features. However, future work could explore additional feature generation methodsincluding wavelet transforms, signal energy, and recurrence based metrics to capture more nuanced dynamics of training load. While `tsfresh` offers scalable and interpretable summary statistics, future studies should also explore deep learning-based approaches such as Long Short-Term Memory (LSTM) networks. LSTMs have the potential to learn latent temporal dependencies in the data directly and could serve as a strong comparative approach for time series regression or classification tasks. Additionally, evaluating the trade-off between feature richness and multicollinearity remains important. This study found that full aggregated feature sets performed best without VIF filtering, suggesting tree-based models can tolerate correlated inputs. Future work should explore how multicollinearity affects model interpretability, particularly in linear models or SHAP-based explanations.

Chapter 6 highlighted the challenges of predicting VO_2 using small, homogeneous sample sizes and controlled testing protocols. Future research should focus on capturing VO_2 in field-based settings that reflect the dynamic, intermittent nature of team sports. This would improve ecological validity and bridge the gap between lab-based physiological assessment and real-world training. Expanding the sample size to include athletes with diverse aerobic profiles would also improve model generalisability.

More experimentation is needed to optimise models for VO_2 estimation. Hybrid pipelinescombining handcrafted features with deep learning representationsmay offer a way to balance interpretability and predictive power. Models trained with LOSOCV and regularised architectures could mitigate overfitting, particularly when working with small samples. Future research should also explore semi-supervised or transfer learning methods to enhance model robustness when labelled physiological data are limited.

Practical considerations regarding multi-sensor configurations must also be addressed to enhance real-world applicability. Future systems should focus on lightweight, cost-effective sensor configurations that minimise intrusion during training while maximising data fidelity. Combining GNSS, IMU, and physiological data streams (e.g., HRV, heart rate recovery) could yield richer insights and enable real-time monitoring at scale.

In conclusion, future research should continue to refine modelling pipelines that link wearable sensor data with athlete responsesboth subjective and physiologicalacross broader populations, sports contexts, and use cases. Advancing explainability, model robustness, and ecological validity will be key to operationalising these tools in daily practice and long-term athlete development.

Appendix A

Supplementary Data

Category	**Indices**	**Description**
Duration indices	Total time	Duration of the session
Kinematic indices (9)		
Distance indices	Total distance	Total distance (m) covered during the session
	Standing and Walking	Distance (m) covered between 01.92 m$\mathring{u}$s^{-1} during the session
	Jogging	Distance (m) covered between 1.963.33 m$\mathring{u}$s^{-1} during the session
	Cruising/Striding	Distance (m) covered between 3.334.72 m$\mathring{u}$s^{-1} during the session
	High-Speed Running	Distance (m) covered $>$ 4.72 m$\mathring{u}$s^{-1} during the session
	Sprinting	Distance (m) covered $>$ 7.22 m$\mathring{u}$s^{-1} during the session
Speed indices	Average speed	Average speed for a selected time interval (m$\mathring{u}$s^{-1})
	Maximum speed	Highest speed achieved in the session (m$\mathring{u}$s^{-1})
	Very High-Speed effort (n)	Number of speed occurrences $>$ 9.72 m$\mathring{u}$s^{-1} during the session
Mechanical indices (7)		
	Max acceleration	A change $>$ +3 m$\mathring{u}$s^{-2} for 0.5 sec minimum.
	Acceleration events	Number of events above +3 m$\mathring{u}$s^{-2} lasting 0.5 sec
	Max deceleration	A change $<$ $-$3 m$\mathring{u}$s^{-2} for 0.5 sec minimum
	Deceleration events	Number of events below $-$3 m$\mathring{u}$s^{-2} lasting 0.5 sec
	Active muscle load	Total work done by the athlete during the session (J)
	Average active muscle power	Average power (W$\mathring{u}$kg^{-1}) during the session
	Eccentric Index	The ratio between active muscle power and mechanical power
Metabolic indices (24)		
	EQ distance	Distance covered at constant speed using total energy consumed (m)
	EQ distance index percentage	Ratio between EQ distance and total distance (%)
	Average metabolic power	Average metabolic power sustained during the session (W$\mathring{u}$kg^{-1})
	Energy	Total energy spent by the athlete (kcal)
	Anaerobic energy	Ratio between anaerobic and total energy (%)
	Average VO2	Average aerobic power sustained by the athlete (L$\mathring{u}$min^{-1})
	Aerobic ratio percentage	Ratio between average metabolic power and VO2max (%)
	Max metabolic power	Peak metabolic power attained (W$\mathring{u}$kg^{-1})
	Met Power Events	Number of high-energy demand actions during a session
	MPE average time	Average time (s) in the selected time interval
	MPE average power	Average power (W$\mathring{u}$kg^{-1}) in the selected time interval
	MPE average recovery time	Average recovery time (s) in the selected time interval
	MPE average recovery power	Average recovery power (W$\mathring{u}$kg^{-1}) in the selected time interval
	MPE time Z1	Number of events lasting $<$ 2.99 sec
	MPE time Z2	Number of events lasting between 3.00-5.99 sec
	MPE time Z3	Number of events lasting $>$ 6.00 sec
	MEP distance Z1	Number of events performed over a distance $>$ 9.99 m
	MEP distance Z2	Number of events performed over a distance between 10.00-19.99 m
	MEP distance Z3	Number of events performed over a distance $>$ 20.00 m
	MEP max speed Z1	Number of events with peak speed $<$ 5.50 m$\mathring{u}$s^{-1}
	MEP max speed Z2	Number of events with peak speed between 5.567.00 m$\mathring{u}$s^{-1}
	MEP max speed Z3	Number of events with peak speed $>$ 7.00 m$\mathring{u}$s^{-1}
	MEP power distance Z1	Distance covered $<$ 24.99 W$\mathring{u}$kg^{-1}
	MEP power distance Z2	Distance covered $>$ 25 W$\mathring{u}$kg^{-1}

Table A.1: External Load Indices (ELIs) and Descriptions form Chapter 3.2.1.

Bibliography

Abbiss, C. R., Peiffer, J. J., Meeusen, R., & Skorski, S. (2015). Role of Ratings of Perceived Exertion during Self-Paced Exercise: What are We Actually Measuring? *Sports Medicine (Auckland, N.Z.)*, *45*(9), 1235–1243. https://doi.org/10.1007/s40279-015-0344-5

Achten, J., & Jeukendrup, A. E. (2003). Heart Rate Monitoring. *Sports Medicine*, *33*(7), 517–538. https://doi.org/10.2165/00007256-200333070-00004

Alanen, A., Räisänen, A., Benson, L., & Pasanen, K. (2021). The use of inertial measurement units for analyzing change of direction movement in sports: A scoping review [Number: 6 Publisher: SAGE Publications]. *International Journal of Sports Science & Coaching*, *16*(6), 1332–1353. https://doi.org/10.1177/17479541211003064

Alemayoh, T. T., Lee, J. H., & Okamoto, S. (2021). New Sensor Data Structuring for Deeper Feature Extraction in Human Activity Recognition. *Sensors (Basel, Switzerland)*, *21*(8), 2814. https://doi.org/10.3390/s21082814

Ali, Z., Qi, G., Kefalas, P., Abro, W., & Ali, B. (2020). A graph-based taxonomy of citation recommendation models. *Artificial Intelligence Review*, *53*. https://doi.org/10.1007/s10462-020-09819-4

Allen, A. E. A., & Tkatchenko, A. (2022). Machine learning of material properties: Predictive and interpretable multilinear models. *Science Advances*, *8*(18), eabm7185. https://doi.org/10.1126/sciadv.abm7185

Altini, M., Penders, J., & Amft, O. (2016). Estimating Oxygen Uptake During Nonsteady-State Activities and Transitions Using Wearable Sensors. *IEEE*

Journal of Biomedical and Health Informatics, *20*(2), 469–475. https://doi.org/10.1109/JBHI.2015.2390493

Altmann, S., Neumann, R., Woll, A., & Härtel, S. (2020). Endurance Capacities in Professional Soccer Players: Are Performance Profiles Position Specific? *Frontiers in Sports and Active Living*, *2*. Retrieved September 11, 2022, from https://www.frontiersin.org/articles/10.3389/fspor.2020.549897

Alzamer, H., Abuhmed, T., & Hamad, K. (2021). A Short Review on the Machine Learning-Guided Oxygen Uptake Prediction for Sport Science Applications. *Electronics*, *10*(16), 1956. https://doi.org/10.3390/electronics10161956

Amelard, R., Hedge, E. T., & Hughson, R. L. (2021). Temporal convolutional networks predict dynamic oxygen uptake response from wearable sensors across exercise intensities. *npj Digital Medicine*, *4*(1), 1–8. https://doi.org/10.1038/s41746-021-00531-3

Antonini, V., Mileo, A., & Roantree, M. (2024). Engineering Features from Raw Sensor Data to Analyse Player Movements during Competition. *Sensors*, *24*(4), 1308. https://doi.org/10.3390/s24041308

Bartlett, J. D., O'Connor, F., Pitchford, N., Torres-Ronda, L., & Robertson, S. J. (2017). Relationships Between Internal and External Training Load in Team-Sport Athletes: Evidence for an Individualized Approach. *International Journal of Sports Physiology and Performance*, *12*(2), 230–234. https://doi.org/10.1123/ijspp.2015-0791

Bastiaansen, B. J., Wilmes, E., Brink, M. S., De Ruiter, C. J., Savelsbergh, G. J., Steijlen, A., Jansen, K. M., Van Der Helm, F. C., Goedhart, E. A., Van Der Laan, D., Vegter, R. J., & Lemmink, K. A. (2020). An Inertial Measurement Unit Based Method to Estimate Hip and Knee Joint Kinematics in Team Sport Athletes on the Field. *Journal of Visualized Experiments*, (159), 60857. https://doi.org/10.3791/60857

Beato, M., Coratella, G., Stiff, A., & Iacono, A. D. (2018). The Validity and Between-Unit Variability of GNSS Units (STATSports Apex 10 and 18 Hz) for Mea-

suring Distance and Peak Speed in Team Sports. *Frontiers in Physiology*, *9*, 1288. https://doi.org/10.3389/fphys.2018.01288

Beltrame, T., Amelard, R., Villar, R., Shafiee, M. J., Wong, A., & Hughson, R. L. (2016). Estimating oxygen uptake and energy expenditure during treadmill walking by neural network analysis of easy-to-obtain inputs. *Journal of Applied Physiology*, *121*(5), 1226–1233. https://doi.org/10.1152/japplphysiol.00600.2016

Bennett, T., Marshall, P., Barrett, S., Malone, J. J., Simpson, A., Bray, J., Christopherson, C., Nickolay, T., Metcalfe, J., & Towlson, C. (2024). Validation of field-based running tests to determine maximal aerobic speed in professional rugby league (C. F. Buzzachera, Ed.). *PLOS ONE*, *19*(7), e0306062. https://doi.org/10.1371/journal.pone.0306062

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(33), 8689–8692. https://doi.org/10.1073/pnas.1702076114

Bodemer, O. (n.d.). Enhancing Individual Sports Training through Artificial Intelligence: A Comprehensive Review. Retrieved November 21, 2024, from https://www.authorea.com/users/684875/articles/682878-enhancing-individual-sports-training-through-artificial-intelligence-a-comprehensive-review

Borg, G. (1998, July). *Borg's Perceived Exertion And Pain Scales*.

Bourdon, P. C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M. C., Gabbett, T. J., Coutts, A. J., Burgess, D. J., Gregson, W., & Cable, N. T. (2017). Monitoring Athlete Training Loads: Consensus Statement [Number: s2]. *International Journal of Sports Physiology and Performance*, *12*(s2), S2–161–S2–170. https://doi.org/10.1123/IJSPP.2017-0208

Brabandere, A. D., Beéck, T. O. D., Schütte, K. H., Meert, W., Vanwanseele, B., & Davis, J. (2018). Data fusion of body-worn accelerometers and heart rate to

predict VO2max during submaximal running. *PLOS ONE*, *13*(6), e0199509. https://doi.org/10.1371/journal.pone.0199509

Brady, A. J., Moyna, N. M., Scriney, M., & McCarren, A. (2022). Activity profile of elite Gaelic football referees during competitive match play [Publisher: Routledge _eprint: https://doi.org/10.1080/24733938.2022.2049456]. *Science and Medicine in Football*, *0*(0), 1–7. https://doi.org/10.1080/24733938.2022.2049456

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (2017, October). *Classification and Regression Trees*. Chapman; Hall/CRC.

Brink, M. S., Frencken, W. G., Jordet, G., & Lemmink, K. A. (2014). Coaches and Players Perceptions of Training Dose: Not a Perfect Match [Number: 3]. *International Journal of Sports Physiology and Performance*, *9*(3), 497–502. https://doi.org/10.1123/ijspp.2013-0009

Buchheit, M. (2012). Repeated-sprint performance in team sport players: Associations with measures of aerobic fitness, metabolic control and locomotor function. *International Journal of Sports Medicine*, *33*(3), 230–239. https://doi.org/10.1055/s-0031-1291364

Cabrera Hernández, M. A., Tafur Tascon, L. J., Cohen, D. D., García-Corzo, S. A., Quiñonez Sánchez, A., Povea Combariza, C., & Tejada Rojas, C. X. (2018). Concordance between the indirect VO2max value estimated through the distance in Yo-Yo intermittent recovery test level 1 and the direct measurement during a treadmill protocol test in elite youth soccer players. *Journal of Human Sport and Exercise - 2018 - Spring Conferences of Sports Science*. https://doi.org/10.14198/jhse.2018.13.Proc2.24

Cardinale, M., & Varley, M. C. (2017). Wearable Training-Monitoring Technology: Applications, Challenges, and Opportunities. *International Journal of Sports Physiology and Performance*, *12*(s2), S2–55–S2–62. https://doi.org/10.1123/ijspp.2016-0423

Carey, D. L., Ong, K., Morris, M. E., Crow, J., & Crossley, K. M. (2016). Predicting ratings of perceived exertion in Australian football players: Methods for live estimation. *International Journal of Computer Science in Sport*, *15*(2), 64–77. https://doi.org/10.1515/ijcss-2016-0005

Carrier, B., Helm, M. M., Cruz, K., Barrios, B., & Navalta, J. W. (2023). Validation of Aerobic Capacity (VO2max) and Lactate Threshold in Wearable Technology for Athletic Populations. *Technologies*, *11*(3), 71. https://doi.org/10.3390/technologies11030071

Casamichana, D., Castellano, J., Calleja-Gonzalez, J., San Román, J., & Castagna, C. (2013). Relationship Between Indicators of Training Load in Soccer Players [Number: 2]. *Journal of Strength and Conditioning Research*, *27*(2), 369–374. https://doi.org/10.1519/JSC.0b013e3182548af1

Caserman, P., Yum, S., Göbel, S., Reif, A., & Matura, S. (2024). Assessing the Accuracy of Smartwatch-Based Estimation of Maximum Oxygen Uptake Using the Apple Watch Series 7: Validation Study. *JMIR Biomedical Engineering*, *9*, e59459. https://doi.org/10.2196/59459

Castagna, C., Impellizzeri, F. M., Belardinelli, R., Abt, G., Coutts, A., Chamari, K., & D'Ottavio, S. (2006). Cardiorespiratory responses to Yo-yo Intermittent Endurance Test in nonelite youth soccer players. *Journal of Strength and Conditioning Research*, *20*(2), 326–330. https://doi.org/10.1519/R-17144.1

Chang, P., Wang, C., Chen, Y., Wang, G., & Lu, A. (2023). Identification of runner fatigue stages based on inertial sensors and deep learning. *Frontiers in Bioengineering and Biotechnology*, *11*. Retrieved December 13, 2023, from https://www.frontiersin.org/articles/10.3389/fbioe.2023.1302911

Chen, M. J., Fan, X., & Moe, S. T. (2002). Criterion-related validity of the Borg ratings of perceived exertion scale in healthy individuals: A meta-analysis [Number: 11]. *Journal of Sports Sciences*, *20*(11), 873–899. https://doi.org/10.1080/026404102320761787

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chmait, N., & Westerbeek, H. (2021). Artificial Intelligence and Machine Learning in Sport Research: An Introduction for Non-data Scientists. *Frontiers in Sports and Active Living*, *3*. Retrieved March 14, 2022, from https://www.frontiersin.org/article/10.3389/fspor.2021.682287

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh A Python package). *Neurocomputing*, *307*, 72–77. https://doi.org/10.1016/j.neucom.2018.03.067

Claudino, J. G., Capanema, D. d. O., de Souza, T. V., Serrão, J. C., Machado Pereira, A. C., & Nassis, G. P. (2019). Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: A Systematic Review. *Sports Medicine - Open*, *5*(1), 28. https://doi.org/10.1186/s40798-019-0202-3

Clemente, F. M., Nikolaidis, P. T., Rosemann, T., & Knechtle, B. (2019). Dose-Response Relationship Between External Load Variables, Body Composition, and Fitness Variables in Professional Soccer Players. *Frontiers in Physiology*, *10*, 443. https://doi.org/10.3389/fphys.2019.00443

Coutts, A. (2019). Dodging Silver Bullets and Opening Black Boxes: Recommendations for Developing Athlete Monitoring Systems. *Journal of Science and Medicine in Sport*, *22*, S2. https://doi.org/10.1016/j.jsams.2019.08.026

Coutts, A. J., Crowcroft, S., & Kempton, T. (2018). Developing athlete monitoring systems Theoretical basis and practical applications. In *Recovery and Well-being in Sport and Exercise: Interdisciplinary Insights*. Routledge.

Coutts, A. J., Rampinini, E., Marcora, S. M., Castagna, C., & Impellizzeri, F. M. (2009). Heart rate and blood lactate correlates of perceived exertion during small-sided soccer games [Number: 1]. *Journal of Science and Medicine in Sport*, *12*(1), 79–84. https://doi.org/10.1016/j.jsams.2007.08.005

Cox, D. R. (1958). The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, *20*(2), 215–242. Retrieved December 4, 2024, from https://www.jstor.org/stable/2983890

Cullen, B. D., McCarren, A. L., & Malone, S. (2021). Ecological validity of self-reported wellness measures to assess pre-training and pre-competition preparedness within elite Gaelic football. *Sport Sciences for Health*, *17*(1), 163–172. https://doi.org/10.1007/s11332-020-00667-x

Daly, L. S., Catháin, C. Ó., & Kelly, D. T. (2024). Do players with superior physiological attributes outwork their less-conditioned counterparts? A study in Gaelic football. *Biology of Sport*. https://doi.org/10.5114/biolsport.2024.129479

Davidson, P., Trinh, H., Vekki, S., & Müller, P. (2023). Surrogate Modelling for Oxygen Uptake Prediction Using LSTM Neural Network. *Sensors*, *23*(4), 2249. https://doi.org/10.3390/s23042249

Delaney, J. A., Duthie, G. M., Thornton, H. R., & Pyne, D. B. (2018). Quantifying the relationship between internal and external work in team sports: Development of a novel training efficiency index. *Science and Medicine in Football*, *2*(2), 149–156. https://doi.org/10.1080/24733938.2018.1432885

di Prampero, P., & Osgnach, C. (2018). Metabolic Power in Team Sports - Part 1: An Update [Number: 08]. *International Journal of Sports Medicine*, *39*(08), 581–587. https://doi.org/10.1055/a-0592-7660

Doeven, S. H., Brink, M. S., Frencken, W. G. P., & Lemmink, K. A. P. M. (2017). Impaired Player-Coach Perceptions of Exertion and Recovery During Match Congestion [Number: 9]. *International Journal of Sports Physiology & Performance*, *12*(9), 1151–1156. Retrieved May 31, 2022, from https://dcu.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=126616067&site=ehost-live&scope=site

Doherty, C., Baldwin, M., Keogh, A., Caulfield, B., & Argent, R. (2024). Keeping Pace with Wearables: A Living Umbrella Review of Systematic Reviews Evaluating the Accuracy of Consumer Wearable Technologies in Health Measurement. *Sports Medicine*, *54*(11), 2907–2926. https://doi.org/10.1007/s40279-024-02077-2

Düking, P., Ruf, L., Altmann, S., Thron, M., Kunz, P., & Sperlich, B. (2024). Assessment of Maximum Oxygen Uptake in Elite Youth Soccer Players: A Comparative Analysis of Smartwatch Technology, Yoyo Intermittent Recovery Test 2, and Respiratory Gas Analysis. *Journal of Sports Science & Medicine*, *23*(2), 351–357. https://doi.org/10.52082/jssm.2024.351

Eston, R. (2012). Use of Ratings of Perceived Exertion in Sports [Number: 2]. *International Journal of Sports Physiology and Performance*, *7*(2), 175–182. https://doi.org/10.1123/ijspp.7.2.175

Ferretti, G. (2015). *Energetics of Muscular Exercise*. Springer International Publishing. https://doi.org/10.1007/978-3-319-05636-4

Foster, C., Boullosa, D., McGuigan, M., Fusco, A., Cortis, C., Arney, B. E., Orton, B., Dodge, C., Jaime, S., Radtke, K., van Erp, T., de Koning, J. J., Bok, D., Rodriguez-Marroyo, J. A., & Porcari, J. P. (2021). 25 Years of Session Rating of Perceived Exertion: Historical Perspective and Development. *International Journal of Sports Physiology & Performance*, *16*(5), 612–621. Retrieved June 7, 2022, from https://dcu.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=149971188&site=ehost-live&scope=site

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., Doleshal, P., & Dodge, C. (2001). A New Approach to Monitoring Exercise Training: [Number: 1]. *Journal of Strength and Conditioning Research*, *15*(1), 109–115. https://doi.org/10.1519/00124278-200102000-00019

Foster, C., Rodriguez-Marroyo, J. A., & de Koning, J. J. (2017). Monitoring Training Loads: The Past, the Present, and the Future. *International Journal of Sports Physiology and Performance*, *12*(s2), S2–2–S2–8. https://doi.org/10.1123/IJSPP.2016-0388

Freeberg, K. A., Baughman, B. R., Vickey, T., Sullivan, J. A., & Sawyer, B. J. (2019). Assessing the ability of the Fitbit Charge 2 to accurately predict VO2max. *mHealth*, *5*, 39. https://doi.org/10.21037/mhealth.2019.09.07

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gabbett, T. J., Nassis, G. P., Oetter, E., Pretorius, J., Johnston, N., Medina, D., Rodas, G., Myslinski, T., Howells, D., Beard, A., & Ryan, A. (2017). The athlete monitoring cycle: A practical guide to interpreting and applying training monitoring data. *British Journal of Sports Medicine*, *51*(20), 1451–1452. https://doi.org/10.1136/bjsports-2016-097298

Gallo, T., Cormack, S., Gabbett, T., Williams, M., & Lorenzen, C. (2015). Characteristics impacting on session rating of perceived exertion training load in Australian footballers [Number: 5]. *Journal of Sports Sciences*, *33*(5), 467–475. https://doi.org/10.1080/02640414.2014.947311

Gallo, T. F., Cormack, S. J., Gabbett, T. J., & Lorenzen, C. H. (2016). Pre-training perceived wellness impacts training output in Australian football players [Number: 15]. *Journal of Sports Sciences*, *34*(15), 1445–1451. https://doi.org/10.1080/02640414.2015.1119295

Garcia-Tabar, I., Rampinini, E., & Gorostiaga, E. M. (2019). Lactate Equivalent for Maximal Lactate Steady State Determination in Soccer [Number: 4 Pub-

lisher: Routledge _eprint: https://doi.org/10.1080/02701367.2019.1643446]. *Research Quarterly for Exercise and Sport*, *90*(4), 678–689. https://doi.org/10.1080/02701367.2019.1643446

Gaudino, P., Iaia, F. M., Strudwick, A. J., Hawkins, R. D., Alberti, G., Atkinson, G., & Gregson, W. (2015). Factors Influencing Perception of Effort (Session Rating of Perceived Exertion) during Elite Soccer Training [Number: 7]. *International Journal of Sports Physiology and Performance*, *10*(7), 860–864. https://doi.org/10.1123/ijspp.2014-0518

Gauri Sawarkar, Maithali Kulkarni, Sayali Marathe, Sayali Marathe, & Prof. Sagar Padiya. (2023). Analysis of Accelerometer Applications. *International Journal of Advanced Research in Science, Communication and Technology*, 426–430. https://doi.org/10.48175/IJARSCT-9439

Geurkink, Y., Vandewiele, G., Lievens, M., de Turck, F., Ongenae, F., Matthys, S. P. J., Boone, J., & Bourgois, J. G. (2019). Modeling the Prediction of the Session Rating of Perceived Exertion in Soccer: Unraveling the Puzzle of Predictive Indicators. *International Journal of Sports Physiology and Performance*, *14*(6), 841–846. https://doi.org/10.1123/ijspp.2018-0698

Gharbi, Z., Dardouri, W., Haj-Sassi, R., Chamari, K., & Souissi, N. (2015). Aerobic and anaerobic determinants of repeated sprint ability in team sports athletes. *Biology of Sport*, *32*(3), 207–212. https://doi.org/10.5604/20831862.1150302

Gómez-Carmona, C. D., Bastida-Castillo, A., González-Custodio, A., Olcina, G., & Pino-Ortega, J. (2020). Using an Inertial Device (WIMU PRO) to Quantify Neuromuscular Load in Running: Reliability, Convergent Validity, and Influence of Type of Surface and Device Location. *Journal of Strength and Conditioning Research*, *34*(2), 365–373. https://doi.org/10.1519/JSC.0000000000003106

Gómez-Carmona, C. D., Pino-Ortega, J., Sánchez-Ureña, B., Ibáñez, S. J., & Rojas-Valverde, D. (2019). Accelerometry-Based External Load Indicators in Sport: Too Many Options, Same Practical Outcome? *International Journal of En-*

vironmental Research and Public Health, *16*(24), 5101. https://doi.org/10.3390/ijerph16245101

Gray, A. J., Shorter, K., Cummins, C., Murphy, A., & Waldron, M. (2018). Modelling Movement Energetics Using Global Positioning System Devices in Contact Team Sports: Limitations and Solutions [Number: 6]. *Sports Medicine*, *48*(6), 1357–1368. https://doi.org/10.1007/s40279-018-0899-z

Guazzi, M., Bandera, F., Ozemek, C., Systrom, D., & Arena, R. (2017). Cardiopulmonary Exercise Testing: What Is its Value? *Journal of the American College of Cardiology*, *70*(13), 1618–1636. https://doi.org/10.1016/j.jacc.2017.08.012

Haddad, M., Stylianides, G., Djaoui, L., Dellal, A., & Chamari, K. (2017). Session-RPE Method for Training Load Monitoring: Validity, Ecological Usefulness, and Influencing Factors. *Frontiers in Neuroscience*, *11*, 612. https://doi.org/10.3389/fnins.2017.00612

Hader, K., Rumpf, M. C., Hertzog, M., Kilduff, L. P., Girard, O., & Silva, J. R. (2019). Monitoring the Athlete Match Response: Can External Load Variables Predict Post-match Acute and Residual Fatigue in Soccer? A Systematic Review with Meta-analysis. *Sports Medicine - Open*, *5*(1), 48. https://doi.org/10.1186/s40798-019-0219-7

Hailstone, J., & Kilding, A. E. (2011). Reliability and Validity of the Zephyr BioHarness to Measure Respiratory Responses to Exercise. *Measurement in Physical Education and Exercise Science*, *15*(4), 293–300. https://doi.org/10.1080/1091367X.2011.615671

Halperin, I., & Emanuel, A. (2020). Rating of Perceived Effort: Methodological Concerns and Future Directions. *Sports Medicine (Auckland, N.Z.)*, *50*(4), 679–687. https://doi.org/10.1007/s40279-019-01229-z

Halson, S. L. (2014). Monitoring Training Load to Understand Fatigue in Athletes [Number: S2]. *Sports Medicine*, *44*(S2), 139–147. https://doi.org/10.1007/s40279-014-0253-z

Hedge, E. T., Amelard, R., & Hughson, R. L. (2023). Prediction of oxygen uptake kinetics during heavy-intensity cycling exercise by machine-learning analysis. *Journal of Applied Physiology*. https://doi.org/10.1152/japplphysiol.00148.2023

Hedge, E. T., & Hughson, R. L. (2020). Frequency domain analysis to extract dynamic response characteristics for oxygen uptake during transitions to moderate- and heavy-intensity exercises. *Journal of Applied Physiology*, *129*(6), 1422–1430. https://doi.org/10.1152/japplphysiol.00503.2020

Helwig, J., Diels, J., Röll, M., Mahler, H., Gollhofer, A., Roecker, K., & Willwacher, S. (2023). Relationships between External, Wearable Sensor-Based, and Internal Parameters: A Systematic Review [Number: 2 Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, *23*(2), 827. https://doi.org/10.3390/s23020827

Hernandez, B., Roberts, B., Jamro-Comer, E., Kodidhi, A., Roelle, L., Miller, N., Littell, L., Ybarra, A., Silva, J., Silva, J., & Orr, W. (2023). Validity Study Comparing Polar Ignite Estimated VO2max to Traditional Cardiopulmonary Exercise Testing in Normal Cardiac Anatomy Fontan Pediatric Patients. *American Journal of Pediatrics*, *9*(2), 79–84. https://doi.org/10.11648/j.ajp.20230902.15

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, *41*(1), 3–13. https://doi.org/10.1249/MSS.0b013e31818cb278

Hoppe, M. W., Baumgart, C., Polglaze, T., & Freiwald, J. (2018). Validity and reliability of GPS and LPS for measuring distances covered and sprint mechanical properties in team sports (L. P. Ardigò, Ed.) [Number: 2]. *PLOS ONE*, *13*(2), e0192708. https://doi.org/10.1371/journal.pone.0192708

Houtmeyers, K. C., Jaspers, A., & Figueiredo, P. (2021). Managing the Training Process in Elite Sports: From Descriptive to Prescriptive Data Analytics. *International Journal of Sports Physiology and Performance*, *16*(11), 1719–1723. https://doi.org/10.1123/ijspp.2020-0958

Imbach, F., Ragheb, W., Leveau, V., Chailan, R., Candau, R., & Perrey, S. (2022). Using global navigation satellite systems for modeling athletic performances in elite football players. *Scientific Reports*, *12*(1), 15229. https://doi.org/10.1038/s41598-022-19484-y

Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J. (2019). Internal and External Training Load: 15 Years On. *International Journal of Sports Physiology and Performance*, *14*(2), 270–273. https://doi.org/10.1123/ijspp.2018-0935

Impellizzeri, F. M., Rampinini, E., Coutts, A. J., Sassi, A., & Marcora, S. M. (2004). Use of RPE-Based Training Load in Soccer: [Number: 6]. *Medicine & Science in Sports & Exercise*, *36*(6), 1042–1047. https://doi.org/10.1249/01.MSS.0000128199.23901.2F

Impellizzeri, F. M., Rampinini, E., & Marcora, S. M. (2005). Physiological assessment of aerobic training in soccer [Number: 6]. *Journal of Sports Sciences*, *23*(6), 583–592. https://doi.org/10.1080/02640410400021278

Impellizzeri, F. M., Shrier, I., McLaren, S. J., Coutts, A. J., McCall, A., Slattery, K., Jeffries, A. C., & Kalkhoven, J. T. (2023). Understanding Training Load as Exposure and Dose. *Sports Medicine (Auckland, N.Z.)*, *53*(9), 1667–1679. https://doi.org/10.1007/s40279-023-01833-0

Inoue, A., dos Santos Bunn, P., do Carmo, E. C., Lattari, E., & da Silva, E. B. (2022). Internal Training Load Perceived by Athletes and Planned by Coaches: A Systematic Review and Meta-Analysis [Number: 1]. *Sports Medicine - Open*, *8*(1), 35. https://doi.org/10.1186/s40798-022-00420-3

Jamnick, N. A., Pettitt, R. W., Granata, C., Pyne, D. B., & Bishop, D. J. (2020). An Examination and Critique of Current Methods to Determine Exercise

Intensity. *Sports Medicine*, *50*(10), 1729–1756. https://doi.org/10.1007/s40279-020-01322-8

Jaspers, A., Brink, M. S., Probst, S. G. M., Frencken, W. G. P., & Helsen, W. F. (2017). Relationships Between Training Load Indicators and Training Outcomes in Professional Soccer. *Sports Medicine*, *47*(3), 533–544. https://doi.org/10.1007/s40279-016-0591-0

Jaspers, A., De Beéck, T. O., Brink, M. S., Frencken, W. G., Staes, F., Davis, J. J., & Helsen, W. F. (2018). Relationships Between the External and Internal Training Load in Professional Soccer: What Can We Learn From Machine Learning? *International Journal of Sports Physiology and Performance*, *13*(5), 625–630. https://doi.org/10.1123/ijspp.2017-0299

Jones, A. M., & Doust, J. H. (1996). A 1% treadmill grade most accurately reflects the energetic cost of outdoor running [Publisher: Routledge _eprint: https://doi.org/10.1080/02640419608727717]. *Journal of Sports Sciences*, *14*(4), 321–327. https://doi.org/10.1080/02640419608727717

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, *30*. Retrieved December 4, 2024, from https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

Kelly, D. T., Tobin, C., Egan, B., McCarren, A., O'Connor, P. L., McCaffrey, N., & Moyna, N. M. (2018). Comparison of Sprint Interval and Endurance Training in Team Sport Athletes [Number: 11]. *Journal of Strength and Conditioning Research*, *32*(11), 3051–3058. https://doi.org/10.1519/JSC.0000000000002374

Kim, J., Kim, H., Lee, J., Lee, J., Yoon, J., & Ko, S.-K. (2022). A Deep Learning Approach for Fatigue Prediction in Sports Using GPS Data and Rate of Perceived Exertion [Conference Name: IEEE Access]. *IEEE Access*, *10*, 103056–103064. https://doi.org/10.1109/ACCESS.2022.3205112

Lacome, M., Simpson, B., & Buchheit, M. (2018). 2018 Monitoring training status with player-tracking technology. Still on the road to Rome. Part 1. *7*, 55–63.

Larsson, P. (2003). Global Positioning System and Sport-Specific Testing. *Sports Medicine*, *33*(15), 1093–1101. https://doi.org/10.2165/00007256-200333150-00002

Laurent, C. M., Green, J. M., Bishop, P. A., Sjökvist, J., Schumacker, R. E., Richardson, M. T., & Curtner-Smith, M. (2011). A Practical Approach to Monitoring Recovery: Development of a Perceived Recovery Status Scale. *The Journal of Strength & Conditioning Research*, *25*(3), 620. https://doi.org/10.1519/JSC.0b013e3181c69ec6

Lopes, T. R., Pereira, H. M., & Silva, B. M. (2022). Perceived Exertion: Revisiting the History and Updating the Neurophysiology and the Practical Applications. *International Journal of Environmental Research and Public Health*, *19*(21), 14439. https://doi.org/10.3390/ijerph192114439

Lovell, T. W., Sirotic, A. C., Impellizzeri, F. M., & Coutts, A. J. (2013). Factors Affecting Perception of Effort (Session Rating of Perceived Exertion) During Rugby League Training [Number: 1]. *International Journal of Sports Physiology & Performance*, *8*(1), 62–69. Retrieved November 22, 2021, from https://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=85039188&site=ehost-live

Malone, J. J., Lovell, R., Varley, M. C., & Coutts, A. J. (2017). Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. *International Journal of Sports Physiology and Performance*, *12*(s2), S2–18–S2–26. https://doi.org/10.1123/ijspp.2016-0236

Malone, S., Hughes, B., Roe, M., Mangan, S., & Collins, K. (2020). Factors that Influence Session-Rating of Perceived Exertion in Elite Gaelic Football [Number: 4]. *Journal of Strength and Conditioning Research*, *34*(4), 1176–1183. https://doi.org/10.1519/JSC.0000000000002192

Malone, S., Solan, B., Collins, K., & Doran, D. (2017). The metabolic power and energetic demands of elite Gaelic football match play [Number: 5]. *The Journal of Sports Medicine and Physical Fitness*, *57*(5), 7.

Manari, D., Manara, M., Zurini, A., Tortorella, G., Vaccarezza, M., Prandelli, N., Ancelotti, D., Vitale, M., Mirandola, P., & Galli, D. (2016). VO2Max and VO2AT: Athletic performance and field role of elite soccer players. *Sport Sciences for Health*, *12*(2), 221–226. https://doi.org/10.1007/s11332-016-0278-9

Mandorino, M., Figueiredo, A., Cima, G., & Tessitore, A. (2022). Analysis of Relationship between Training Load and Recovery Status in Adult Soccer Players: A Machine Learning Approach. *International Journal of Computer Science in Sport*, *21*(2), 1–16. https://doi.org/10.2478/ijcss-2022-0007

Mandorino, M., Clubb, J., & Lacome, M. (2024). Predicting Soccer Players' Fitness Status Through a Machine-Learning Approach. *International Journal of Sports Physiology and Performance*, *19*(5), 443–453. https://doi.org/10.1123/ijspp.2023-0444

Marcora, S. (2010, January). Effort: Perception of.

Marcora, S. (2019). Psychobiology of fatigue during endurance exercise. In *Endurance performance in sport* (pp. 15–34). Routledge. Retrieved August 7, 2024, from https://www.taylorfrancis.com/chapters/edit/10.4324/9781315167312-2/psychobiology-fatigue-endurance-exercise-samuele-marcora

Martínez-Lagunas, V., & Hartmann, U. (2014). Validity of the Yo-Yo Intermittent Recovery Test Level 1 for direct measurement or indirect estimation of maximal oxygen uptake in female soccer players. *International Journal of Sports Physiology and Performance*, *9*(5), 825–831. https://doi.org/10.1123/ijspp.2013-0313

Marynowicz, J., Lango, M., Horna, D., Kikut, K., & Andrzejewski, M. (2021). Predicting ratings of perceived exertion in youth soccer using decision tree mod-

els. *Biology of Sport*, *39*(2), 245–252. https://doi.org/10.5114/biolsport.2022.103723

McCall, A., Fanchini, M., & Coutts, A. J. (2017). Prediction: The Modern-Day Sport-Science and Sports-Medicine Quest for the Holy Grail. *International Journal of Sports Physiology and Performance*, *12*(5), 704–706. https://doi.org/10.1123/ijspp.2017-0137

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

McGuigan, H., Hassmén, P., Rosic, N., & Stevens, C. J. (2020). Training monitoring methods used in the field by coaches and practitioners: A systematic review. *International Journal of Sports Science & Coaching*, *15*(3), 439–451. https://doi.org/10.1177/1747954120913172

McLaren, S. J., Macpherson, T. W., Coutts, A. J., Hurst, C., Spears, I. R., & Weston, M. (2018). The Relationships Between Internal and External Measures of Training Load and Intensity in Team Sports: A Meta-Analysis [Number: 3]. *Sports Medicine*, *48*(3), 641–658. https://doi.org/10.1007/s40279-017-0830-z

McLaren, S. J., Smith, A., Spears, I. R., & Weston, M. (2017). A detailed quantification of differential ratings of perceived exertion during team-sport training. *Journal of Science and Medicine in Sport*, *20*(3), 290–295. https://doi.org/10.1016/j.jsams.2016.06.011

Miguel, M., Oliveira, R., Loureiro, N., García-Rubio, J., & Ibáñez, S. J. (2021). Load Measures in Training/Match Monitoring in Soccer: A Systematic Review. *International Journal of Environmental Research and Public Health*, *18*(5), 2721. https://doi.org/10.3390/ijerph18052721

Minetti, A. E., Moia, C., Roi, G. S., Susta, D., & Ferretti, G. (2002). Energy cost of walking and running at extreme uphill and downhill slopes [Number: 3]. *Journal of Applied Physiology*, *93*(3), 1039–1046. https://doi.org/10.1152/japplphysiol.01177.2001

Mitchell, E., Monaghan, D., & O'Connor, N. (2013). Classification of Sporting Activities Using Smartphone Accelerometers. *Sensors*, *13*(4), 5317–5337. https://doi.org/10.3390/s130405317

Molina-Garcia, P., Notbohm, H. L., Schumann, M., Argent, R., Hetherington-Rauth, M., Stang, J., Bloch, W., Cheng, S., Ekelund, U., Sardinha, L. B., Caulfield, B., Brønd, J. C., Grøntved, A., & Ortega, F. B. (2022). Validity of Estimating the Maximal Oxygen Consumption by Consumer Wearables: A Systematic Review with Meta-analysis and Expert Statement of the INTER-LIVE Network. *Sports Medicine (Auckland, N.Z.)*, *52*(7), 1577–1597. https://doi.org/10.1007/s40279-021-01639-y

Moreira, A., Bilsborough, J. C., Sullivan, C. J., Cianciosi, M., Saldanha Aoki, M., & Coutts, A. J. (2015). Training Periodization of Professional Australian Football Players During an Entire Australian Football League Season [Number: 5]. *International Journal of Sports Physiology & Performance*, *10*(5), 566–571. Retrieved May 27, 2022, from https://dcu.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=103585255&site=ehost-live&scope=site

Morrow, M. M. B., Lowndes, B., Fortune, E., Kaufman, K. R., & Hallbeck, M. S. (2017). Validation of Inertial Measurement Units for Upper Body Kinematics. *Journal of Applied Biomechanics*, *33*(3), 227–232. https://doi.org/10.1123/jab.2016-0120

Nédélec, M., McCall, A., Carling, C., Le Gall, F., Berthoin, S., & Dupont, G. (2013). Physical performance and subjective ratings after a soccer-specific exercise simulation: Comparison of natural grass versus artificial turf. *Journal of Sports Sciences*, *31*(5), 529–536. https://doi.org/10.1080/02640414.2012.738923

Nédélec, M., McCall, A., Carling, C., Legall, F., Berthoin, S., & Dupont, G. (2012). Recovery in Soccer: Part I Post-Match Fatigue and Time Course of Recovery.

Sports Medicine, *42*(12), 997–1015. https://doi.org/10.2165/11635270-000000000-00000

Nedergaard, N. J., Kersting, U., & Lake, M. (2014). Using accelerometry to quantify deceleration during a high-intensity soccer turning manoeuvre. *Journal of Sports Sciences*, *32*(20), 1897–1905. https://doi.org/10.1080/02640414.2014.965190

ODriscoll, R., Turicchi, J., Beaulieu, K., Scott, S., Matu, J., Deighton, K., Finlayson, G., & Stubbs, J. (2018). How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *British Journal of Sports Medicine*, bjsports–2018–099643. https://doi.org/10.1136/bjsports-2018-099643

Op De Beéck, T., Jaspers, A., Brink, M. S., Frencken, W. G., Staes, F., Davis, J. J., & Helsen, W. F. (2019). Predicting Future Perceived Wellness in Professional Soccer: The Role of Preceding Load and Wellness. *International Journal of Sports Physiology and Performance*, *14*(8), 1074–1080. https://doi.org/10.1123/ijspp.2017-0864

Osgnach, C., Paolini, E., Roberti, V., Vettor, M., & di Prampero, P. E. (2016). Metabolic Power and Oxygen Consumption in Team Sports: A Brief Response to Buchheit et al. *International Journal of Sports Medicine*, *37*(1), 77–81. https://doi.org/10.1055/s-0035-1569321

Osgnach, C., & di Prampero, P. (2018). Metabolic Power in Team Sports - Part 2: Aerobic and Anaerobic Energy Yields [Number: 08]. *International Journal of Sports Medicine*, *39*(08), 588–595. https://doi.org/10.1055/a-0592-7219

Oxendale, C. L., Highton, J., & Twist, C. (2017). Energy expenditure, metabolic power and high speed activity during linear and multi-directional running [Number: 10]. *Journal of Science and Medicine in Sport*, *20*(10), 957–961. https://doi.org/10.1016/j.jsams.2017.03.013

Passfield, L., & Hopker, J. G. (2017). A Mine of Information: Can Sports Analytics Provide Wisdom From Your Data? *International Journal of Sports Physiology and Performance*, *12*(7), 851–855. https://doi.org/10.1123/ijspp.2016-0644

Paul, D. J., Tomazoli, G., & Nassis, G. P. (2019). Match-Related Time Course of Perceived Recovery in Youth Football Players. *International Journal of Sports Physiology and Performance*, *14*(3), 339–342. https://doi.org/10.1123/ijspp.2017-0521

Rana, M., & Mittal, V. (2021). Wearable Sensors for Real-Time Kinematics Analysis in Sports: A Review. *IEEE Sensors Journal*, *21*(2), 1187–1207. https://doi.org/10.1109/JSEN.2020.3019016

Ravé, G., Granacher, U., Boullosa, D., Hackney, A. C., & Zouhal, H. (2020). How to Use Global Positioning Systems (GPS) Data to Monitor Training Load in the Real World of Elite Soccer. *Frontiers in Physiology*, *11*. https://doi.org/10.3389/fphys.2020.00944

Reusch, R. S., Juracy, L. R., & Moraes, F. G. (2022). Assessment and Optimization of 1D CNN Model for Human Activity Recognition. *2022 XII Brazilian Symposium on Computing Systems Engineering (SBESC)*, 1–7. https://doi.org/10.1109/SBESC56799.2022.9964520

Robergs, R. A., Dwyer, D., & Astorino, T. (2010). Recommendations for Improved Data Processing from Expired Gas Analysis Indirect Calorimetry. *Sports Medicine*, *40*(2), 95–111. https://doi.org/10.2165/11319670-000000000-00000

Ross, C., Lambs, P., Mcalpine, P., Kennedy, G., & Button, C. (2020). Validation of gyroscope sensors for snow sports performance monitoring. *International Journal of Computer Science in Sport*, *19*(1), 51–59. https://doi.org/10.2478/ijcss-2020-0004

Rossi, A., Perri, E., Pappalardo, L., Cintia, P., Alberti, G., Norman, D., & Iaia, F. M. (2022). Wellness Forecasting by External and Internal Workloads in Elite Soccer Players: A Machine Learning Approach. *Frontiers in Physiology*,

13. Retrieved December 14, 2023, from https://www.frontiersin.org/articles/10.3389/fphys.2022.896928

Rossi, A., Perri, E., Pappalardo, L., Cintia, P., & Iaia, F. M. (2019). Relationship between External and Internal Workloads in Elite Soccer Players: Comparison between Rate of Perceived Exertion and Training Load. *Applied Sciences*, *9*(23), 5174. https://doi.org/10.3390/app9235174

Rozos, E., Dimitriadis, P., Mazi, K., & Koussis, A. D. (2021). A Multilayer Perceptron Model for Stochastic Synthesis [Number: 2 Publisher: Multidisciplinary Digital Publishing Institute]. *Hydrology*, *8*(2), 67. https://doi.org/10.3390/hydrology8020067

Ryan, S., Kempton, T., Impellizzeri, F., & Coutts, A. (2019). Training monitoring in professional Australian football: Theoretical basis and recommendations for coaches and scientists. *Science and Medicine in Football*, *4*, 1–7. https://doi.org/10.1080/24733938.2019.1641212

Savoia, C., Padulo, J., Colli, R., Marra, E., McRobert, A., Chester, N., Azzone, V., Pullinger, S. A., & Doran, D. A. (2020). The Validity of an Updated Metabolic Power Algorithm Based upon di Pramperos Theoretical Model in Elite Soccer Players [Number: 24]. *International Journal of Environmental Research and Public Health*, *17*(24), 9554. https://doi.org/10.3390/ijerph17249554

Saw, A. E., Main, L. C., & Gastin, P. B. (2016). Monitoring the athlete training response: Subjective self-reported measures trump commonly used objective measures: A systematic review. *British Journal of Sports Medicine*, *50*(5), 281–291. https://doi.org/10.1136/bjsports-2015-094758

Schutz, Y., & Chambaz, A. (1997). Could a satellite-based navigation system (GPS) be used to assess the physical activity of individuals on earth? *European Journal of Clinical Nutrition*, *51*(5), 338–339. https://doi.org/10.1038/sj.ejcn.1600403

Schutz, Y., & Herren, R. (2000). Assessment of speed of human locomotion using a differential satellite global positioning system. *Medicine and Science in Sports and Exercise*, *32*(3), 642–646. https://doi.org/10.1097/00005768-200003000-00014

Scott, M. T. U., Scott, T. J., & Kelly, V. G. (2016). The Validity and Reliability of Global Positioning Systems in Team Sport: A Brief Review. *The Journal of Strength & Conditioning Research*, *30*(5), 1470–1490. https://doi.org/10.1519/JSC.0000000000001221

Scott, T. J., Black, C. R., Quinn, J., & Coutts, A. J. (2013). Validity and Reliability of the Session-RPE Method for Quantifying Training in Australian Football: A Comparison of the CR10 and CR100 Scales. *Journal of Strength and Conditioning Research*, *27*(1), 270–276. https://doi.org/10.1519/JSC.0b013e3182541d2e

Seiler, K. S., & Kjerland, G. Ø. (2006). Quantifying training intensity distribution in elite endurance athletes: Is there evidence for an "optimal" distribution? *Scandinavian Journal of Medicine & Science in Sports*, *16*(1), 49–56. https://doi.org/10.1111/j.1600-0838.2004.00418.x

Selmi, O., Marzouki, H., Ouergui, I., BenKhalifa, W., & Bouassida, A. (2018). Influence of intense training cycle and psychometric status on technical and physiological aspects performed during the small-sided games in soccer players [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15438627.2018.1492398]. *Research in Sports Medicine*, *26*(4), 401–412. https://doi.org/10.1080/15438627.2018.1492398

Selmi, O., Ouergui, I., Muscella, A., My, G., Marsigliante, S., Nobari, H., Suzuki, K., & Bouassida, A. (2022). Monitoring Psychometric States of Recovery to Improve Performance in Soccer Players: A Brief Review. *International Journal of Environmental Research and Public Health*, *19*(15), 9385. https://doi.org/10.3390/ijerph19159385

Seshadri, D. R., Li, R. T., Voos, J. E., Rowbottom, J. R., Alfes, C. M., Zorman, C. A., & Drummond, C. K. (2019a). Wearable sensors for monitoring the physiological and biochemical profile of the athlete. *npj Digital Medicine*, *2*(1), 1–16. https://doi.org/10.1038/s41746-019-0150-9

Seshadri, D. R., Li, R. T., Voos, J. E., Rowbottom, J. R., Alfes, C. M., Zorman, C. A., & Drummond, C. K. (2019b). Wearable sensors for monitoring the internal and external workload of the athlete. *npj Digital Medicine*, *2*(1), 71. https://doi.org/10.1038/s41746-019-0149-2

Sheridan, D. (2021). *Effects of minor Gaelic football match play on markers of muscle damage, delayed onset muscle soreness and muscle function* [Master's thesis, Dublin City University]. https://doras.dcu.ie/25230/1/Effects%20of%20Minor%20Gaelic%20Football%20Match%20Play%20on%20Markers%20of%20Muscle%20Damage%2C%20Delayed%20Onset%20Muscle%20Soreness%20and%20Muscle%20Function%20MSc%20-Dermot%20Sheridan%20%282021%29%20%281%29.pdf

Sheridan, D., Brady, A. J., Nie, D., & Roantree, M. (2024). Predictive analysis of ratings of perceived exertion in elite Gaelic football. *Biology of Sport*, *41*(4), 61–68. https://doi.org/10.5114/biolsport.2024.134753

Shirdel, M., Asadi, R., Do, D., & Hintlian, M. (2021, September). Deep Learning with Kernel Flow Regularization for Time Series Forecasting [arXiv:2109.11649 [cs]]. https://doi.org/10.48550/arXiv.2109.11649

Shushan, T., Lovell, R., McLaren, S. J., Barrett, S., Buchheit, M., Scott, T. J., & Norris, D. (2023). A Methodological Comparison of Protocols and Analytical Techniques to Assess the Outcome Measures of Submaximal Fitness Tests. *International Journal of Sports Physiology and Performance*, *-1*(aop), 1–13. https://doi.org/10.1123/ijspp.2023-0164

Shushan, T., McLaren, S. J., Buchheit, M., Scott, T. J., Barrett, S., & Lovell, R. (2022). Submaximal Fitness Tests in Team Sports: A Theoretical Frame-

work for Evaluating Physiological State. *Sports Medicine*, *52*(11), 2605–2626. https://doi.org/10.1007/s40279-022-01712-0

Silva, J. R., Rumpf, M. C., Hertzog, M., Castagna, C., Farooq, A., Girard, O., & Hader, K. (2018). Acute and Residual Soccer Match-Related Fatigue: A Systematic Review and Meta-analysis. *Sports Medicine*, *48*(3), 539–583. https://doi.org/10.1007/s40279-017-0798-8

Silva, P., Santos, E. D., Grishin, M., & Rocha, J. M. (2018). Validity of Heart Rate-Based Indices to Measure Training Load and Intensity in Elite Football Players. *Journal of Strength and Conditioning Research*, *32*(8), 2340–2347. https://doi.org/10.1519/JSC.0000000000002057

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. Retrieved November 21, 2024, from https://www.semanticscholar.org/paper/A-review-of-supervised-machine-learning-algorithms-Singh-Thakur/53b55682222692323a3a0d546d9e1a3de29454f0

Srivastava, S., Tamrakar, S., Nallathambi, N., Vrindavanam, S. A., Prasad, R., & Kothari, R. (2024). Assessment of Maximal Oxygen Uptake (VO2 Max) in Athletes and Nonathletes Assessed in Sports Physiology Laboratory. *Cureus*, *16*(5), e61124. https://doi.org/10.7759/cureus.61124

Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J., Kosub, S., Schreck, T., Keim, D., & Grossniklaus, M. (2017). How to Make Sense of Team Sport Data: From Acquisition to Data Modeling and Research Aspects. *MDPI*, *2*(1), 2. https://doi.org/10.3390/data2010002

Stellingwerf, T. (2012). Case study: Nutrition and training periodization in three elite marathon runners. *International Journal of Sport Nutrition and Exercise Metabolism*, *22*(5), 392–400. https://doi.org/10.1123/ijsnem.22.5.392

Sulaiman, N., Mahat, A., Adnan, R., Kassim, R.-N., & Sulaiman, M. (2012, December). *Validity of selected aerobic capacity field-based test among rugby players*.

Taylor, J. B., Wright, A. A., Dischiavi, S. L., Townsend, M. A., & Marmon, A. R. (2017). Activity Demands During Multi-Directional Team Sports: A Systematic Review. *Sports Medicine*, *47*(12), 2533–2551. https://doi.org/10.1007/s40279-017-0772-5

Theodoropoulos, J. S., Bettle, J., & Kosy, J. D. (2020). The use of GPS and inertial devices for player monitoring in team sports: A review of current and future applications. *Orthopedic Reviews*, *12*(1). https://doi.org/10.4081/or.2020.7863

Thornton, H. R., Delaney, J. A., Duthie, G. M., & Dascombe, B. J. (2019). Developing Athlete Monitoring Systems in Team Sports: Data Analysis and Visualization. *International Journal of Sports Physiology and Performance*, *14*(6), 698–705. https://doi.org/10.1123/ijspp.2018-0169

Timmerman, W. P., Abbiss, C. R., Lawler, N. G., Stanley, M., & Raynor, A. J. (2024). Athlete monitoring perspectives of sports coaches and support staff: A scoping review. *International Journal of Sports Science & Coaching*, *19*(4), 1813–1832. https://doi.org/10.1177/17479541241247131

Tolusso, D. V., Dobbs, W. C., MacDonald, H. V., Winchester, L. J., Laurent, C. M., Fedewa, M. V., & Esco, M. R. (2022). The Validity of Perceived Recovery Status as a Marker of Daily Recovery Following a High-Volume Back-Squat Protocol. *International Journal of Sports Physiology and Performance*, *17*(6), 886–892. https://doi.org/10.1123/ijspp.2021-0360

Townshend, A. D., Worringham, C. J., & Stewart, I. B. (2008). Assessment of Speed and Position during Human Locomotion Using Nondifferential GPS. *Medicine & Science in Sports & Exercise*, *40*(1), 124–132. https://doi.org/10.1249/mss.0b013e3181590bc2

Tunca, C., Salur, G., & Ersoy, C. (2020). Deep Learning for Fall Risk Assessment With Inertial Sensors: Utilizing Domain Knowledge in Spatio-Temporal Gait Parameters. *IEEE journal of biomedical and health informatics*, *24*(7), 1994–2005. https://doi.org/10.1109/JBHI.2019.2958879

Vähä-Ypyä, H., Bretterhofer, J., Husu, P., Windhaber, J., Vasankari, T., Titze, S., & Sievänen, H. (2023). Performance of Different Accelerometry-Based Metrics to Estimate Oxygen Consumption during Track and Treadmill Locomotion over a Wide Intensity Range. *Sensors*, *23*(11), 5073. https://doi.org/10.3390/s23115073

Vallance, E., Sutton-Charani, N., Guyot, P., & Perrey, S. (2023). Predictive modeling of the ratings of perceived exertion during training and competition in professional soccer players. *Journal of Science and Medicine in Sport*, *26*(6), 322–327. https://doi.org/10.1016/j.jsams.2023.05.001

Van Buuren, S. (2018). *Flexible Imputation of Missing Data*. Chapman; Hall/CRC. https://stefvanbuuren.name/fimd/

van der Does, H. T. D., Brink, M. S., Otter, R. T. A., Visscher, C., & Lemmink, K. A. P. M. (2017). Injury Risk Is Increased by Changes in Perceived Recovery of Team Sport Players. *Clinical Journal of Sport Medicine*, *27*(1), 46. https://doi.org/10.1097/JSM.0000000000000306

Varley, M. C., Gabbett, T., & Aughey, R. J. (2014). Activity profiles of professional soccer, rugby league and Australian football match play [Number: 20]. *Journal of Sports Sciences*, *32*(20), 1858–1866. https://doi.org/10.1080/02640414.2013.823227

Varley, M. C., Jaspers, A., Helsen, W. F., & Malone, J. J. (2017). Methodological Considerations When Quantifying High-Intensity Efforts in Team Sport Using Global Positioning System Technology. *International Journal of Sports Physiology & Performance*, *12*(8), 1059–1068. Retrieved April 22, 2022, from https://dcu.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=126023934&site=ehost-live&scope=site

Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, *119*, 3–11. https://doi.org/10.1016/j.patrec.2018.02.010

Wang, Z., Zhang, Q., Lan, K., Yang, Z., Gao, X., Wu, A., Xin, Y., & Zhang, Z. (2022). Enhancing instantaneous oxygen uptake estimation by non-linear model using cardio-pulmonary physiological and motion signals. *Frontiers in Physiology*, *13*. https://doi.org/10.3389/fphys.2022.897412

Weaving, D., Jones, B., Till, K., Abt, G., & Beggs, C. (2017). The Case for Adopting a Multivariate Approach to Optimize Training Load Quantification in Team Sports. *Frontiers in Physiology*, *8*, 1024. https://doi.org/10.3389/fphys.2017.01024

Weaving, D., Marshall, P., Earle, K., Nevill, A., & Abt, G. (2014). Combining internal- and external-training-load measures in professional rugby league. *International Journal of Sports Physiology and Performance*, *9*(6), 905–912. https://doi.org/10.1123/ijspp.2013-0444

West, S., Clubb, J., Blake, T., Fern, J., Bowles, H., & Dalen-Lorentsen, T. (2024). Big data. Big potential. Big problems? *BMJ Open Sport & Exercise Medicine*, *10*, e001994. https://doi.org/10.1136/bmjsem-2024-001994

West, S. W., Clubb, J., Torres-Ronda, L., Howells, D., Leng, E., Vescovi, J. D., Carmody, S., Posthumus, M., Dalen-Lorentsen, T., & Windt, J. (2021). More than a Metric: How Training Load is Used in Elite Sport for Athlete Management. *International Journal of Sports Medicine*, *42*(04), 300–306. https://doi.org/10.1055/a-1268-8791

White, R., Palczewska, A., Weaving, D., Collins, N., & Jones, B. (2022). Sequential movement pattern-mining (SMP) in field-based team-sport: A framework for quantifying spatiotemporal data and improve training specificity? *Journal of Sports Sciences*, *40*(2), 164–174. https://doi.org/10.1080/02640414.2021.1982484

Wiig, H., Andersen, T. E., Luteberget, L. S., & Spencer, M. (2020). Individual Response to External Training Load in Elite Football Players [Publisher: Human Kinetics Section: International Journal of Sports Physiology and Perfor-

mance]. *International Journal of Sports Physiology and Performance*, *15*(5), 696–704. https://doi.org/10.1123/ijspp.2019-0453

Wundersitz, D. W., Josman, C., Gupta, R., Netto, K. J., Gastin, P. B., & Robertson, S. (2015). Classification of team sport activities using a single wearable tracking device [Number: 15]. *Journal of Biomechanics*, *48*(15), 3975–3981. https://doi.org/10.1016/j.jbiomech.2015.09.015

Zignoli, A., Fornasiero, A., Bertolazzi, E., Pellegrini, B., Schena, F., Biral, F., & Laursen, P. B. (2019). State-of-the art concepts and future directions in modelling oxygen consumption and lactate concentration in cycling exercise [Number: 2]. *Sport Sciences for Health*, *15*(2), 295–310. https://doi.org/10.1007/s11332-019-00557-x

Zignoli, A., Fornasiero, A., Ragni, M., Pellegrini, B., Schena, F., Biral, F., & Laursen, P. B. (2020). Estimating an individuals oxygen uptake during cycling exercise with a recurrent neural network trained from easy-to-obtain inputs: A pilot study. *PLOS ONE*, *15*(3), e0229466. https://doi.org/10.1371/journal.pone.0229466