Faculty of Health and Social Sciences

Bournemouth University

Injury risk and performance: Towards a better understanding of the complexities and intricacies of load monitoring within an elite football club

A thesis submitted in partial fulfilment for the degree of

Doctor of Philosophy

Doctoral Thesis of

Aritra Majumdar

Supervisory Team:

Prof. Tim Rees

Dr. Rashid Bakirov

March 2024

Abstract

Load monitoring has emerged as a pivotal aspect of contemporary sports science, particularly in the context of athlete training and competition. This thesis delves into the dynamic landscape of load monitoring with a particular focus on soccer, a sport of unparalleled global popularity, boasting 200,000 professional and 240 million amateur players. The prevalence of soccer-related injuries, surpassing those in other sports, has underscored the imperative for effective load monitoring strategies to optimize training adaptations, evaluate fatigue and recovery, and mitigate injury risks. Professional sports teams, cognizant of the multifaceted implications of inadequate load management, have invested significantly in this domain. In the realm of soccer, where injuries can lead to prolonged player absences, impacting team performance and incurring substantial financial costs, the need for comprehensive load monitoring becomes even more apparent. Notably, English Premier League soccer clubs bore an approximate financial burden of £45 million per season due to injuries from 2012-2013 through to the 2016-2017 season. In response to the pressing demand for a nuanced understanding of the intricate relationship between training load and soccer injuries, this thesis integrates insights from machine learning. Building upon existing research, we explore how machine learning techniques contribute to the refinement of load monitoring strategies in soccer, offering a promising avenue for enhancing injury prevention protocols. By bridging the gap between traditional sports science methodologies and cutting-edge machine learning applications, this research seeks to provide a comprehensive framework for optimizing athlete performance and well-being in the dynamic context of soccer with the help of Machine learning.

This thesis undertook three comprehensive investigations aimed at advancing the understanding of the relationship between training load and soccer injuries through the application of machine learning methodologies.

The initial inquiry critically examined recent research endeavours in football that incorporated machine learning techniques. This exploration highlighted the profound implications of football injuries, which not only result in prolonged player absences affecting team performance but also entail considerable financial ramifications. Despite the burgeoning interest in the relationship between training load and injuries, prevailing models and statistical approaches were found to inadequately capture the intricate nuances of this association. The lack of consensus on variables for analysis posed a significant challenge, hindering the effective utilization of existing studies in guiding the selection of key training load variables. (Chapter – 2)[1]

Subsequently, the second investigation employed machine learning to scrutinize the connection between training load and soccer injuries, utilizing a multi-season dataset from an English Premier League club. A pioneering aspect of this chapter was the application of Artificial Neural Networks, marking the first instance of employing such a method on a multi-season dataset for injury prediction. The results indicated a promising capability to predict injuries with high recall, identifying a majority of injury cases. However, precision suffered due to the prevalent class imbalance, emphasizing the need for further refinement in this methodology. Despite these challenges, the chapter provided valuable insights for soccer organizations and practitioners engaged in load injury monitoring. (Chapter – 3)[2]

The third and final investigation contributed a pioneering analysis of online continual and adaptive learning methodologies for soccer injury prediction, utilizing a distinctive multi-season dataset from Elite Premier League players. Noteworthy findings demonstrated the

[1] The second chapter is published as a journal paper entitled "Machine Learning for Understanding and Predicting Injuries in Football" in the Sports Medicine – Open Journal (SMOA).

[2] The third chapter is published as a journal paper entitled "A Multi-Season Machine Learning Approach to Examine the Training Load and Injury relationship in Professional Soccer" in the Journal of Sports Analytics (JSA).

superiority of these adaptive learning approaches over static learning, with cumulative training identified as a critical factor enhancing model adaptability and performance. The practical applications extended to injury prevention and player well-being management in professional soccer. The research's forward-looking stance emphasized the necessity for future exploration into advanced continual learning frameworks and real-time injury prediction systems to refine and enhance the efficacy of injury prevention strategies. (Chapter – 4)[3]

[3] The fourth chapter, entitled as "A Multi-Season Continual Machine Learning Approach to Examine the Training Load and Injury relationship in Professional Soccer" is ready for submission.

Acknowledgements

I would like to express my sincere gratitude to my esteemed supervisors, Professor Tim Rees and Dr. Rashid Bakirov, whose unwavering support, guidance, and expertise have been instrumental throughout the entire journey of my thesis. Their dedication to fostering academic excellence, coupled with their insightful feedback and encouragement, significantly enriched the quality and depth of this research. I am truly fortunate to have had the privilege of working under their mentorship. In addition to their academic guidance, Professor Tim Rees and Dr. Rashid Bakirov have been incredible pillars of support on a personal level, often going above and beyond to ensure not just the success of my research, but also my well-being and growth as a scholar and individual. Their genuine care and mentorship have illuminated my path with wisdom and kindness, teaching me invaluable lessons that extend far beyond the realm of academic inquiry. Their influence has shaped not only this thesis but also my future endeavors, for which I am eternally grateful. Their spirit of generosity and mentorship will always be a beacon of inspiration for me.

In addition, I would like to express my gratitude to Mrs. Tina Ikin, the Course Coordinator of the Faculty of Health and Social Sciences. Mrs. Ikin played a pivotal role not only in overseeing the academic aspects of my research but also in providing crucial assistance in navigating and completing the various academic norms and requirements. Her meticulous attention to detail, coupled with her consistent reminders and gentle encouragement, greatly contributed to the successful completion of these academic milestones. Despite my occasional challenges in meeting deadlines, Mrs. Ikin exhibited remarkable patience and understanding, offering unwavering support and guidance. Her commitment to fostering an environment conducive to academic achievement has left an indelible mark on my educational journey, and I am sincerely thankful for her indispensable role in ensuring the fulfilment of academic requirements.

I would also like to extend my heartfelt thanks to my friends and family, who provided unwavering support and encouragement throughout this challenging yet rewarding endeavour. Their understanding, patience, and belief in my capabilities were constant sources of motivation. In particular, I am deeply grateful to my parents - Mr. Partha Sarathi Majumdar and Mrs. Maitrayee Sinharay, for their endless support, sacrifices, and belief in my potential. Their encouragement has been the cornerstone of my academic pursuits, and I am profoundly thankful for their unwavering faith in my abilities.

Lastly, I must acknowledge my own perseverance and insatiable curiosity. These past years have been not just a journey of academic enrichment in the field of Workload and Injury but also a period of personal growth. Beyond the confines of my research, I have eagerly explored the realm of football analytics, a passion that has both broadened my horizons and deepened my appreciation for the intricate balance between data and the beautiful game. I have cultivated a dynamic online presence through my analytics channel (EPL Analytics) on Twitter (handle: X). Here, I engage with large datasets, employing advanced scraping techniques to extract meaningful insights that I then meticulously analyze and present to a diverse audience, gained more than 13k Followers.

This thesis is a testament to the collaborative efforts and support of these remarkable individuals, each of whom has played an indispensable role in the realization of this academic achievement. I am deeply appreciative of the guidance, encouragement, and inspiration they have provided, and I look forward to continuing to build upon the foundation they have helped me establish.

Table of Contents

List of Tables

List of Figures

Chapter 1

1. Introduction

1.1 Background and motivation

Football clubs are increasingly adopting data-driven approaches as technological advancements in data collection and storage evolve. Sophisticated methods such as multi-camera systems, electronic performance tracking, wearable sensors, and detailed questionnaires now enable the systematic collection of nuanced physical, technical, and psychological data from players (De Silva et al., 2018). Initially utilised for scouting, performance analysis, and tactical planning, these datasets now play a pivotal role in understanding injury aetiology. Injuries not only impact team performance but also carry substantial financial repercussions, particularly in elite leagues such as the English Premier League (Eliakim et al., 2020). Thus, the relationship between training load and injuries has become a central focus in sports science.

Despite data-driven methodologies being widely applied in sectors like healthcare and autonomous vehicles, their adoption in football injury prediction remains limited (Claudino et al., 2019). Existing research primarily focuses on a restricted set of training load variables, neglecting the broader potential of multivariate and machine learning (ML) approaches (Rein and Memmert, 2016). This gap underscores the need for research harnessing ML to enhance prognostic insights into football injuries, thereby contributing to sports science and performance optimization.

Machine learning offers a novel perspective on analysing the interplay between training load and injuries. As a field of chapter, ML leverages mathematical and statistical models to enable computers to learn from data and improve decision-making. Its applications span various domains, including image detection, cancer diagnostics, stock market predictions, and customer behavior analysis (Guyon et al., 2002). In sports, the effective use of ML is still emerging (Oliver et al., 2020).

Training load, a key component in injury risk analysis, is categorized into internal and external loads. Internal loads include biological and psychological factors like heart rate, blood lactate levels, session-RPE, and well-being metrics, recorded via wearable sensors and questionnaires (Halson, 2014). External loads measure physical activity, such as velocity, acceleration, total workload, and high-speed distances, captured using GPS and electronic performance tracking systems (Akenhead and Nassis, 2016). A comprehensive assessment of these loads provides insights into players' physiological and psychological stress during training and competition.

Athletes subjected to training loads experience both fitness and fatigue responses. While fitness adaptations enhance resilience to injuries, fatigue increases injury susceptibility. Evidence links higher workloads to increased injury risk, leading to recommendations for careful workload management (Gabbett, 2016). However, the "Workload-Injury Paradox" suggests that high workloads might also promote injury resilience (Windt and Gabbett, 2017).

Models like the Acute Chronic Workload Ratio (ACWR) illustrate this paradox. ACWR compares the rolling average of acute (5-10 days) and chronic (4-6 weeks) workloads, with values above 1.5 indicating a higher injury risk (Hulin et al., 2014). Innovations like the Exponential Weighted Moving Average (EWMA) improve the sensitivity of ACWR, emphasizing the importance of internal and external workload balance (Murray et al., 2017).

Further, methodologies such as monotony and strain analysis have been employed to quantify workload impact. High monotony—the ratio of mean to standard deviation of training loads—

has been associated with increased injury risk (Brink et al., 2010). Studies in sports like basketball and rugby highlight these metrics' predictive power, underscoring their relevance in soccer (Anderson et al., 2003).

Advances in ML now allow deeper insights into the workload-injury relationship. By incorporating multiple explanatory variables, ML provides a holistic view, transcending traditional models like ACWR (Majumdar et al., 2022). Despite its promise, ML in soccer injury prediction faces challenges, including class imbalance and dynamic data shifts (concept drift), which necessitate adaptive learning frameworks (Lundberg and Lee, 2017). Concept drift—the temporal alteration of data distributions—poses significant challenges for static ML models, as it renders earlier assumptions obsolete. Addressing this requires ensemble methods, online learning algorithms, and adaptive modeling techniques (Hussain et al., 2021).

Continual learning, also known as lifelong learning, represents a paradigm shift in ML. Unlike static models, it allows systems to learn from both historical and incoming data, adapting to evolving conditions while mitigating catastrophic forgetting (Disabato and Roveri, 2022). This dynamic framework is crucial for applications like soccer injury prediction, where training regimes and player conditions vary seasonally.

This thesis aims to deepen the understanding of the intricate relationship between training load and soccer injuries through three interconnected investigations, each contributing uniquely to the field. First, it evaluates the current state of research, identifying gaps and limitations in existing injury prediction models (Majumdar et al., 2022). Second, it employs machine learning on a multi-season dataset from the English Premier League to uncover patterns and improve predictive accuracy (Majumdar et al., 2024). Finally, it pioneers the use of continual and adaptive learning methodologies to address the dynamic nature of football data, offering significant advancements in injury prevention strategies. The first evaluates existing research, highlighting inadequacies in current models, such as the limited scope of training load variables considered, the lack of consensus on key predictors, and the reliance on static machine learning models that fail to account for the dynamic nature of football data. The second employs ML to analyze multi-season data from an English Premier League club, implementing a carefully pre-processed pipeline that addresses data imbalances through resampling techniques. The chapter utilized an Artificial Neural Network (ANN) to identify patterns in training load and injury data, achieving high recall by effectively capturing injury cases despite the inherent class imbalance. This approach yielded insights into the predictive value of load variables, informing targeted injury prevention strategies. The third explores online continual and adaptive learning methodologies, showing their superiority in handling dynamic data. For instance, the chapter demonstrated that adaptive learning models achieved a 15% higher precision in injury prediction compared to static models when applied to a multi-season dataset. This improvement underscores the ability of adaptive approaches to adjust to evolving data distributions, offering practical advantages in real-world injury prevention scenarios. Collectively, these studies advocate for advanced frameworks to enhance injury prevention and athlete well-being.

1.2 Overview of Thesis Contributions

This thesis presents three major contributions that advance the understanding of training load and injury prediction in professional soccer:

Comprehensive Literature Review and Identification of Research Gaps:

The first contribution systematically reviews the existing literature on the relationship between training load and injuries, particularly in soccer. It identifies significant gaps, such as the lack of multivariate analysis and limited application of machine learning methodologies, establishing a strong foundation for further investigation.

Application of Machine Learning on Multi-Seasonal Data:

The second contribution involves the innovative use of machine learning, particularly Artificial Neural Networks (ANNs), to analyze a longitudinal dataset from an English Premier League club. This research addresses challenges such as class imbalance and uncovers valuable insights into the predictive capabilities of specific training load variables. The results contribute to the development of precise and actionable injury prevention strategies.

Integration of Continual and Adaptive Learning Techniques:

The third and most novel contribution introduces continual and adaptive learning methodologies. These approaches address the limitations of static models by adapting to evolving data distributions (concept drift). By comparing static, continual, and adaptive models, the chapter highlights the superior performance of adaptive techniques in maintaining high predictive accuracy, ensuring the long-term relevance of injury prediction models in dynamic sports environments.

Collectively, these contributions not only enhance injury prediction frameworks but also provide actionable insights for practitioners, paving the way for more effective injury prevention strategies in professional soccer.

1.3 Literature Review

1.3.1 General Overview

The English Premier League (EPL) is renowned as the most affluent, popular, and competitive football league worldwide (Ekstrand et al., 2016; Eliakim et al., 2020). Achievements within the league are diverse; elite teams compete for the championship and European Champions League qualifications, mid-tier teams strive for entry into the UEFA Cup, and others battle to avoid relegation (Ekstrand et al., 2016). These varying degrees of success correlate directly with financial gains, which are crucial for clubs to attract and retain top talent, thereby enhancing their competitive status (English Football League, 2018). As a result, the importance of load monitoring in athletes has gained considerable attention in the realm of sports science, leading professional sports teams to dedicate significant resources towards understanding and optimizing athlete workloads (Halson, 2014; Akenhead and Nassis, 2016). This process is crucial for evaluating how athletes adapt to training, gauging fatigue and recovery levels, and reducing the likelihood of injuries and illnesses (Soligard et al., 2016; Owoeye et al., 2020).

Soccer, as the world's most widely played sport with a vast number of professional and amateur participants, experiences a higher incidence of injuries compared to other sports. These injuries not only restrict players to play for extended durations, negatively affecting team performance, but also carry substantial financial costs. In the context of the English Premier League, the financial toll of player injuries has been substantial, with costs reaching around £45 million annually over a five-year period (Eliakim et al., 2020). This underscores the critical need for effective load monitoring strategies to mitigate injury risks and their associated impacts in soccer.

The availability of players has been identified as a crucial factor influencing the success of football clubs across Europe (Ekstrand et al., 2016). Clubs that experienced lower injury rates and higher player availability compared to the previous season observed notable improvements in their average points per match and overall league standing. Essentially, having a fuller roster of available talent directly correlates with enhanced team performance.

Over an extended period, muscle and tendon injuries, particularly in the hamstring and groin, along with ligament and joint injuries to the knee and ankle, emerged as the most significant contributors to injury burden, adversely affecting team outcomes (Bahr and Holme, 2003; Gabbett, 2016). A substantial portion of these injuries, which negatively impact team performance, are considered preventable. Despite ongoing injury prevention efforts, elite football players average two injuries each season, leading to a considerable number of injuries within a standard team (Ekstrand et al., 2016). Hamstring injuries alone account for 12% of these, resulting in significant time lost from training and matches (Ekstrand et al., 2016). Notably, while overall muscle injury rates have remained steady, hamstring injuries have seen an annual increase, exacerbated by the heightened intensity of competitive leagues like the EPL.

The financial implications of injuries are substantial, with millions paid in wages to injured players each season, not including the additional costs for treatment. This financial strain is compounded by potential losses in club income due to diminished team performance and lower league placements. Clubs in the top half of the league, typically with larger squads and more depth, are less impacted by injuries to key players compared to those in the lower half, where the threat of relegation looms larger, potentially leading to significant decreases in revenue from various sources (Bourdon et al., 2017; Bowen et al., 2019).

Injuries not only impact the competitive edge and financial health of clubs but also highlight the importance of effective injury prevention measures. Without such strategies, clubs risk falling into a detrimental cycle where reduced player availability leads to poorer performance and, consequently, less financial capacity to invest in talent, perpetuating the cycle of injury and underperformance. This underscores the critical need for clubs to prioritize and refine their approach to injury prevention to safeguard both their athletic and financial futures.

1.3.2 Training Loads and Injury

Injuries in sports, while complex and resulting from various factors, invariably occur under the duress of training or competition workloads (Gabbett, 2016). Essentially, a sport injury manifests as structural damage when the applied physical forces surpass the body's resilience. This underscores the imperative for sport scientists to delineate workloads that stretch athletes' capabilities without breaching their physiological limits (Bahr and Holme, 2003; Carey et al., 2017). Accordingly, the meticulous monitoring and tailoring of football players' training regimens are pivotal to optimizing workload, enhancing adaptability, and diminishing injury rates.

Injuries within football are broadly characterized into three categories: any physical complaint derived from football activities, injuries necessitating medical intervention, and injuries causing absence from training or matches (Ekstrand et al., 2019; Gabbett and Ullah, 2012). The latter, known for its significant impact on performance, facilitates a practical approach to injury data collection. Furthermore, injuries are classified based on duration out of action—ranging from minimal to severe—and detailed by factors like location, type, and whether they represent new or recurrent issues. Distinctions between training versus match-induced injuries, and contact versus non-contact injuries, offer additional insights into their nature (Bahr, 2009; Windt and Gabbett, 2017).

Incorporating workloads into injury prevention frameworks demands a nuanced comprehension of how workloads interact with various injury risk factors. A developed Workload-Injury Aetiology Model clarifies the influence of workload on injury occurrence, an area not fully addressed in earlier models. This model outlines three primary pathways through which workload contributes to injury risk: exposure to external risk factors and potential injury events, the induction of negative adaptations such as fatigue, and the facilitation of positive adaptations like improved fitness (Soligard et al., 2016; Hulin et al., 2014). Workloads, therefore, play a dual role in both predisposing athletes to injuries by exposing them to external risks and modifying their injury risk through the body's adaptations to physical stress.

The relationship between athlete workloads and injury prevention is intricate, where workloads act as a means through which athletes encounter potentially injurious scenarios rather than being direct causes of injury (Brink et al., 2010; Drew et al., 2016). High workloads have been associated with increased injury risks across various sports, emphasizing the need for careful management. Initial studies in elite rugby highlighted strong correlations between training intensity, duration, and perceived exertion with injury occurrence, supporting the theory that higher workloads lead to increased injury risks. This concept was further explored in Australian football, revealing that cumulative workloads over weeks correlated with injury risks, particularly rapid increases in high-speed running which heightened the chance of injuries like hamstring strains (Carey et al., 2017; Murray et al., 2017).

In soccer, the relationship between internal workload, as measured by perceived exertion, and non-contact injuries has been documented, indicating a higher probability of injury following periods of high workload intensity over several weeks (Halson, 2014; Owoeye et al., 2020). However, the measurement of workload and its impact on injury risk is multifaceted. Subjective measures like perceived exertion can be influenced by various factors including the athlete's personality and environmental conditions, suggesting that both internal and external workload measurements should be considered for a more holistic understanding (Akenhead and Nassis, 2016; Gabbett and Ullah, 2012).

Research incorporating both perceived exertion and GPS-derived external workloads in elite soccer has identified external workloads, such as total distance covered and decelerations over weeks, as significant injury risk factors (Rossi et al., 2018; Oliver et al., 2020). Moreover, while reducing workloads might lower injury risks, it could also inhibit positive physical adaptations essential for performance enhancement and injury tolerance (Bourdon et al., 2017; Bowen et al., 2019). Studies have shown that athletes with higher physical capacities can tolerate greater workloads and exhibit a decreased relative risk of injury, highlighting the bidirectional relationship between workload exposure and physical conditioning. Thus, a balanced approach to training, which avoids excessive workloads while fostering physical development, is crucial for minimizing injury risks and enhancing athlete performance (Gabbett, 2016; Windt and Gabbett, 2017).

The English Premier League (EPL) has seen significant increases in the speed, intensity, and competitive nature of the game, alongside a rise in physical and technical demands. This evolution, coupled with a packed schedule, necessitates that players consistently perform under high workloads. Consequently, training regimens that do not adequately prepare players can lead to decreased fitness levels and a reduced ability to handle physical stress, thus increasing the risk of injury. It's been suggested that there's a U-shaped relationship between workload and injury risk, implying that both insufficient and excessive workloads can elevate injury risk. Previous chronic exposure to workloads can also influence an athlete's current injury risk (Ekstrand et al., 2016; Carey et al., 2017).

The concept of relative workload monitoring, specifically the acute chronic workload ratio (ACWR), has gained prominence (Hulin et al., 2014; Soligard et al., 2016). This involves comparing the workload of a recent week (acute workload) against the average workload over the past four weeks (chronic workload), offering insights into whether current workloads are above, equal to, or below what an athlete has been conditioned for (Murray et al., 2017; Gabbett, 2016). Chronic workload represents the athlete's fitness level, while acute workload is indicative of fatigue. A high chronic workload with a lower acute workload suggests an athlete is well-prepared, whereas a sudden increase in acute workload compared to chronic workload indicates excessive fatigue, potentially harming performance and elevating injury risk. Initial research in elite sports, like cricket and rugby league (Hulin et al., 2014), demonstrated that acute workloads significantly higher than the athlete's chronic workload were linked to an increased injury risk. Conversely, higher chronic workloads were associated with lower injury risks, attributed to positive training adaptations.

Guidelines derived from this research suggest that ACWRs above 1.5 indicate a high risk of injury, while ratios between 0.8 to 1.3 are considered optimal for minimizing injury risk. Workloads below this optimal range or significantly above it are associated with increased injury risks. Thus, maintaining workloads within this 'sweet spot' can help balance the risks of under- and over-training, ensuring athletes are adequately prepared for competition demands while minimizing the likelihood of injury and performance decline (Bourdon et al., 2017; Bowen et al., 2019).

The calculation of the Acute Chronic Workload Ratio (ACWR), a pivotal metric in sports science, primarily employs rolling averages, summarizing workloads over predefined periods to gauge chronic workload. This traditional method, however, overlooks the nuances of training stimuli variations and their timing within these intervals. It treats all stimuli equally, regardless of whether they occurred recently or several weeks prior, thus potentially misrepresenting their true impact on an athlete's preparedness and risk of injury.

As a response to these limitations, an alternative approach has been suggested, utilizing an exponentially weighted moving average (EWMA). This model assigns progressively lesser significance to older workload data, more accurately reflecting the diminishing influence of past training stimuli on the athlete's current state. This refinement aims to offer a more precise assessment of workload impacts, enhancing the applicability of the ACWR in monitoring and managing athlete training (Murray et al., 2017; Hulin et al., 2014).

Workload management emerges as a strategic measure within the injury prevention framework, which is divided into primary, secondary, tertiary, universal, selective, and indicated prevention strategies (Soligard et al., 2016; Gabbett, 2016). Primary prevention focuses on avoiding potential injury risk factors by balancing workloads appropriately. Secondary prevention seeks to identify and mitigate early signs of injury through workload adjustments. Tertiary prevention aims to facilitate a safe return to activity post-injury, minimizing the risk of recurrence.

Universal prevention considers general risk factors across sports disciplines, including mental health and physical activity, among others. Given the correlation between workload and injury risk across various sports, workload is regarded as a universal risk factor (Hulin et al., 2014; Soligard et al., 2016). Selective prevention targets specific demographic and physiological attributes that may influence an individual's injury risk, emphasizing personalized training programs. Lastly, indicated prevention concentrates on athletes with a heightened injury risk, necessitating vigilant workload management to prevent injury occurrence.

This holistic approach underscores the critical role of workload management in injury prevention, advocating for a nuanced understanding of each athlete's unique risk profile and adaptive capacity. In exploring workload management within sports, particularly Australian football, studies have contrasted two methods for calculating the Acute Chronic Workload Ratio (ACWR): the traditional rolling averages and the exponentially weighted moving average (EWMA) (Murray et al., 2017; Carey et al., 2017). While both methods link very high ACWRs to increased injury risk, the EWMA has shown greater sensitivity to these risks. Despite this, the practical application of rolling averages in planning training workloads seems more straightforward, allowing for easier adjustments to manage player load effectively. Conversely, the EWMA's complexity, due to its weighting system, poses challenges in applying specific workload adjustments in a dynamic training schedule (Bourdon et al., 2017; Bowen et al., 2019).

The concept of ACWR serves as a foundational framework for injury prevention, emphasising that chronic workload builds an athlete's capacity to handle acute loads. However, individual factors such as age, training history, and physical fitness play a crucial role in moderating injury risk, making some athletes more resilient than others. Well-developed physical qualities are known to enhance an athlete's tolerance to higher workloads, highlighting the reciprocal relationship between workload exposure and physical conditioning. Effective workload management, which gradually increases chronic workload while avoiding abrupt spikes, is key to enhancing physical capacities and, by extension, workload tolerance.

Research across various sports has investigated the predictive power of workload measures, including ACWR, for injury risk, often finding limited predictive accuracy. This limitation is attributed to the complex, multifactorial nature of injury occurrence and the inherently low probability of injury in certain sports like football. Nonetheless, the association between workload and injury underscores the value of workload monitoring as a preventive tool, despite challenges in prediction. The overall evidence suggests that understanding and managing workloads can contribute to injury prevention strategies, yet the direct impact of such interventions on reducing injuries remains to be fully determined (Gabbett, 2016; Windt and Gabbett, 2017).

1.3.3 Machine Learning

Machine Learning (ML) is a foundational discipline within artificial intelligence (AI) that empowers computers to execute tasks without explicit programming. By leveraging sophisticated algorithms and statistical models, ML systems extract patterns from data, make predictions, and continually improve their accuracy (Lundberg and Lee, 2017; Guyon et al., 2002). This ability to learn from data and adapt to changing conditions makes ML indispensable across numerous domains, including finance, healthcare, and sports science.

The success of any ML model is contingent upon several critical components. Data serves as the cornerstone of ML, providing the raw material from which insights are drawn. Data can be unstructured, such as images or text, or structured, as in databases and spreadsheets. High-quality, comprehensive datasets are essential to ensure robust model performance (Halson, 2014; Mehlig, 2019). Algorithms form the backbone of the learning process, enabling systems to process data and uncover underlying patterns. Linear regression, logistic regression, decision trees, and neural networks are commonly used algorithms, each suited to specific types of problems and data structures (Chen and Guestrin, 2016; Krawczyk et al., 2017). Features, representing measurable attributes of the phenomenon under chapter, are integral to model training. Proper feature selection and engineering enhance a model’s predictive capability by focusing on the most relevant variables, ensuring that the model generalizes well to unseen data (Lundberg and Lee, 2017).

Training and inference are pivotal stages in ML. During training, the model learns from historical data by iteratively refining its parameters to minimize prediction errors. Once trained, the model enters the inference stage, where it applies its learned knowledge to new, unseen data, enabling real-time predictions and decision-making (Oliver et al., 2020). The evaluation of model performance involves metrics such as accuracy, precision, recall, and F1-score, which provide insights into the model's strengths and areas for improvement (Guyon et al., 2002; Emmert-Streib et al., 2020).

ML paradigms are broadly classified into four categories: supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised learning involves training a model on labeled data, where the relationship between inputs and outputs is known. This paradigm is widely used for tasks like regression and classification (Loyola-Gonzalez, 2019). Unsupervised learning, in contrast, focuses on identifying hidden patterns within unlabeled data, making it valuable for clustering and anomaly detection (Rossi et al., 2021). Semi-supervised learning strikes a balance by leveraging a small amount of labeled data alongside a larger volume of unlabeled data, improving model accuracy without the need for extensive labeling efforts (Bourdon et al., 2017). Reinforcement learning enables models to learn optimal strategies by interacting with an environment and receiving feedback in the form of rewards or penalties, making it particularly effective in robotics, gaming, and autonomous systems (Tang et al., 2010).

The versatility of ML is evident in its wide-ranging applications. In finance, ML enhances credit scoring, fraud detection, and algorithmic trading (Belle and Papantonis, 2020). Healthcare relies on ML for disease diagnosis, treatment personalization, drug discovery, and medical imaging (Bohr and Memarzadeh, 2020). In e-commerce, ML powers recommendation systems and optimizes supply chain management, while in the automotive industry, it drives the development of autonomous vehicles and advanced safety systems (Chen and Guestrin, 2016). In the domain of sports science, ML facilitates performance analysis, injury risk prediction, and the refinement of scouting and recruitment strategies (Majumdar et al., 2022). Furthermore, in cybersecurity, ML strengthens malware detection, threat prediction, and network defense mechanisms (Sáez et al., 2019).

Despite its transformative potential, ML presents several challenges. Ensuring data privacy and security is paramount, particularly in sensitive domains such as healthcare and finance. Bias and fairness are critical concerns, as biased training data can lead to discriminatory outcomes and perpetuate social inequalities (Lundberg and Lee, 2017). The interpretability of complex models, especially those involving deep learning, remains a significant hurdle, as stakeholders demand transparent and explainable decision-making processes (Johnson and Khoshgoftaar, 2019). Finally, the acquisition of high-quality, diverse datasets is often resource-intensive but essential for building reliable and unbiased models (Emmert-Streib et al., 2020).

In sports science, ML is revolutionising the management of training loads and injury prevention. By analysing both historical and real-time data, ML models provide actionable insights to optimize athletic performance and reduce injury risks. Predictive models for injury risk, which integrate physiological, psychological, and biomechanical factors, surpass traditional statistical approaches by capturing complex, non-linear relationships (Oliver et al., 2020). Wearable technologies, equipped with sensors, further enhance ML's capabilities by enabling continuous monitoring of training loads, recovery, and player well-being. These real-time insights allow coaches and medical staff to make data-driven decisions, tailoring training regimens to individual athletes and minimizing the likelihood of overtraining and injury (Majumdar et al., 2022).

Moreover, ML supports rehabilitation by predicting optimal recovery pathways and tracking an athlete's progress post-injury. By personalizing rehabilitation programs, ML ensures that athletes regain peak performance levels safely and efficiently, reducing the risk of re-injury. Additionally, ML-driven performance analytics offer a competitive edge by identifying strengths and weaknesses in team dynamics and individual player contributions, enabling strategic adjustments that maximize success (Bohr and Memarzadeh, 2020; Rossi et al., 2021).

In conclusion, ML represents a paradigm shift in sports science and beyond, providing innovative, data-driven solutions to complex challenges. By leveraging the power of ML, practitioners can deliver personalized, adaptive strategies that enhance performance, mitigate risks, and advance the broader understanding of athletic and physiological dynamics. As ML continues to evolve, its integration into sports science will undoubtedly drive further breakthroughs, solidifying its role as a cornerstone of modern innovation.

1.3.4 Online Learning and Concept Drift: Significance and Strategies

In the realm of machine learning and data analysis, concept drift refers to the temporal alteration of data distributions over time, posing a significant challenge for static predictive models (Wang et al., 2013; Krawczyk et al., 2017). Traditional machine learning models are often built under the assumption of a stationary data distribution. However, in dynamic environments such as football, where factors like coaching strategies, player performance, and training regimes evolve, these models may become obsolete or suboptimal without frequent retraining (Hussain et al., 2021; Zenisek et al., 2019).

Concept drift manifests in various forms, including sudden, gradual, or recurring shifts in data patterns. Sudden drift occurs when a major change happens abruptly in the data distribution, such as an injury crisis leading to altered training strategies. Gradual drift involves a slow change over time, like a team's progressive adaptation to a new tactical approach. Recurring drift, on the other hand, refers to patterns that re-emerge periodically, such as seasonal variations in player workload (Lundberg and Lee, 2017; Gama et al., 2014).

Addressing concept drift is crucial for ensuring that injury prediction models remain accurate and relevant over time. Adaptive learning frameworks and continual learning methodologies help mitigate its effects by enabling models to incrementally update their knowledge base as new data becomes available (Disabato and Roveri, 2022).

Various techniques exist to handle concept drift effectively. Online learning algorithms continuously update model parameters with incoming data, ensuring real-time adaptability (Goel and Batra, 2021). Ensemble methods combine multiple models trained on different data segments to enhance robustness against drift (Minku et al., 2010). Periodic retraining involves re-training models at scheduled intervals to capture recent trends. Additionally, drift detection methods like Drift Detection Method (DDM) and Early Drift Detection Method (EDDM) monitor changes in error rates to signal significant distributional shifts, prompting necessary model updates (Hussain et al., 2021; Gama et al., 2014).

Online learning is a core approach to tackling dynamic data environments, especially in sports science. Unlike batch learning, where models are trained on entire datasets at once, online learning updates the model incrementally with each new data point. This ensures the model remains adaptable to real-time changes, a necessity in high-paced domains like football (Rossi et al., 2021). For instance, as new injury data is recorded daily, online learning allows immediate model refinement without the computational overhead of retraining from scratch.

Online learning methods are particularly effective when combined with drift detection techniques. They ensure that the model evolves in response to concept drift while maintaining robustness against overfitting to transient changes. Moreover, these methods support continual learning, enabling the integration of historical and incoming data without catastrophic forgetting—a common challenge in traditional retraining approaches (Disabato and Roveri, 2022; Zenisek et al., 2019).

In the context of this thesis, addressing concept drift and incorporating online learning are paramount for enhancing the reliability and applicability of injury prediction systems. By leveraging these techniques, the research ensures that the developed models remain resilient in the face of evolving football data, thereby improving both short-term and long-term injury prevention strategies. Ultimately, understanding and managing concept drift, coupled with robust online learning frameworks, not only enriches the scientific discourse on sports science but also contributes to the development of adaptive and effective predictive systems for real-world applications.

1.4 Aims and Objectives of the PhD

This thesis presents three interconnected studies that utilize machine learning methodologies to deepen the understanding of the complex relationship between training load and soccer injuries. The first chapter provides a comprehensive review of existing literature, the second applies machine learning techniques to analyse multi-season data and develop predictive models, and the third further enhances these predictive models using adaptive machine learning techniques.

Aim 1: Developing a clearer understanding of what we currently know and think about player load and injury, across sports specifically in soccer and the application of Machine Learning to unpick the relationship between training load and injury.

Objectives:

1. In pursuit of the primary aim, a comprehensive literature review was undertaken, encompassing empirical studies, reviews, unpublished documents ('grey' literature), and anecdotal evidence pertaining to player load and injury.
2. Additionally, an exploration of machine learning methodologies for injury prediction was conducted, alongside an examination of pertinent works from well-established application domains, notably industrial faults prediction. The fulfilment of these objectives establishes a foundational framework for subsequent research endeavours concerning machine learning applications in injury prediction, as expounded further within this thesis and potentially extending to broader academic discourse.

Aim 2: To examine, employing sophisticated machine learning methodologies, the association between individual and multiple load variables and injury, utilizing a longitudinal dataset derived from the first team of AFCB, with the aim of constructing an interpretable machine learning model.

Objectives:

1. Employ exploratory methodologies, leveraging proficiency in data science and machine learning, to analyse a comprehensive 6-year longitudinal dataset encompassing player load and injury records curated by AFC Bournemouth. The dataset comprises diverse variables pertaining to GPS and on-field activities, general off-field and well-being indicators, gym-related metrics, readiness assessments, and altitude-related data—encompassing technical, physical, psychological load aspects, and personal information.
2. Employing a distinctive pre-processing methodology involving data sampling, feature engineering, and feature selection techniques to construct a machine learning model.

The primary objective is to discern between instances of injuries and non-injuries, optimizing precision and recall metrics within the model.

Aim 3: To enhance the prognostication of soccer injuries, integrating online continual and adaptive learning methodologies leveraging a longitudinal dataset sourced from the first team of AFC Bournemouth, contributing to the refinement and advancement of injury prediction models within the realm of professional soccer.

Objectives:

1. Focusing on static, continual, and adaptive learning, with the aim to construct continual and adaptive drift retraining models.
2. Executing diverse adaptive machine learning methodologies with the objective of augmenting the predictive efficacy of the models beyond that achievable by static machine learning models.

1.5 Original contributions

1. **Developing a clearer understanding between Training load and Injuries and prediction of injuries using Machine learning.** Aligning with Aim 1, a comprehensive literature review is presented in Chapter 2 that identifies the key strengths and weaknesses of the current state-of-the-art approaches in training load and football injury relationship using Machine learning.
2. **Investigation of the relationship between training load and Injuries using Machine Learning**. Addressing the challenges highlighted in Aim 1, and aligning with Aim 2, A Multi-Season Machine Learning Approach to Examine the Training Load and Injury Relationship in Professional Soccer is presented in Chapter 3.

   a. To date, this is the first chapter that investigates the training load and injury relationship using machine learning with data from English Premier League.
   b. To date, this is the first chapter that uses multi-seasonal training load injury data to predict injuries.
   c. To date, this is the first chapter that uses Artificial Neural Network (ANN) to predict injuries.
3. **Investigation of the relationship between training load and Injuries using Continual and Adaptive Machine Learning**. Addressing the challenges highlighted in Aim 2, and aligning with Aim 3, A Multi-Season Continual Machine Learning Approach to Examine the Training Load and Injury Relationship in Professional Soccer is presented in Chapter 4.
   a. To date, this is the first chapter that uses continual and adaptive machine learning technologies to predict injuries.
   b. To date, this is the first chapter that compares static, continual and adaptive machine learning technologies to predict injuries.

1.6 List of resulting Publications

The following publications are a result of this work:

1. Majumdar, A., Bakirov, R., Hodges, D., et al. (2022) 'Machine learning for understanding and predicting injuries in football', *Sports Medicine - Open*, 8(1).
2. Majumdar, A., Bakirov, R., Hodges, D., McCullagh, S. and Rees, T. (2024) ‘A multi-season machine learning approach to examine the training load and injury relationship in professional soccer’, *Journal of Sports Analytics*, 1 January, pp. 47–65.
3. Majumdar, A., Bakirov, R., Hodges, D., McCullagh, S. and Rees, T. (2024) ‘A multi-season continual machine learning approach to examine the training load and injury

relationship in professional soccer', is ready for submission at *Journal of Sports Science and Medicine*.

1.7 Structure of the Thesis

This thesis is structured as a compilation of interconnected research papers, each presented as an individual chapter, collectively constituting the three primary studies conducted. As a result, some content overlap is inherent across different chapters.

Chapter One serves as the introduction, providing an overview of the subject areas under consideration. It aims to establish a comprehensive understanding of the major issues and limitations within these domains, offering a rationale that guides the research direction throughout the thesis.

Chapter Two focuses on elucidating the existing knowledge and perspectives on player load and injury, particularly in the context of soccer. Additionally, it explores the application of Machine Learning to unravel the intricate relationship between training load and injury.

Chapter Three employs sophisticated machine learning methodologies to investigate the association between individual and multiple load variables and injury. Utilizing a longitudinal dataset extracted from the first team of AFC Bournemouth, the objective is to construct an interpretable machine learning model, enhancing our understanding of the complex relationship between training load and injury occurrence.

Chapter Four aims to advance the prognostication of soccer injuries by integrating online continual and adaptive learning methodologies. Leveraging a longitudinal dataset sourced from the first team of AFC Bournemouth, this chapter contributes to refining and advancing injury prediction models within the professional soccer domain.

Chapter Five encapsulates an overarching discussion of the entire thesis. It synthesizes the findings from the preceding chapters, delves into the limitations encountered, and outlines implications for future research and applied practices based on these insights.

Chapter 2

2. Abstract

Attempts to better understand the relationship between training and competition load and injury in football are essential for helping to understand adaptation to training programmes, assessing fatigue and recovery, and minimizing the risk of injury and illness. To this end, technological advancements have enabled the collection of multiple points of data for use in analysis and injury prediction. The full breadth of available data has, however, only recently begun to be explored using suitable statistical methods. Advances in automatic and interactive data analysis with the help of machine learning are now being used to better establish the intricacies of the player load and injury relationship. In this article, we examine this recent research, describing the analyses and algorithms used, reporting the key findings, and comparing model fit. To date, the vast array of variables used in analysis as proxy indicators of player load, alongside differences in approach to key aspects of data treatment—such as response to data imbalance, model fitting, and a lack of multi-season data—limit a systematic evaluation of findings and the drawing of a unified conclusion. If, however, the limitations of current studies can be addressed, machine learning has much to offer the field and could in future provide solutions to the training load and injury paradox through enhanced and systematic analysis of athlete data.

Key Points

Football injuries can lead to extended periods of absence from competition, with associated impacts on team performance, as well as financial implications. The relationship between training load and injuries is now a key research and applied focus, but current models and statistical approaches to data analysis fail to sufficiently capture the nuances of this relationship.

The application of machine learning to the training load and injury relationship is a new but fast growing research area, but there is a lack of consensus regarding which variables to consider for analysis, let alone those subsequently proving to be key in predicting players' injuries, making it difficult at this time to draw on those studies when choosing which training load variables upon which to focus.

Although questions remain as to the current utility of machine learning for real-world application, the use of machine learning has great potential to unearth new insights into the workload and injury relationship, if research is expanded to examine multiple seasons' data, accounts for data imbalance, and uses explainable artificial intelligence.

2.1 Introduction

With technological developments in data collection and storage, football clubs are increasingly data-driven (De Silva et al., 2018; Rein and Memmert, 2016). The multi-camera method and electronic performance and tracking systems, alongside wearable sensors and use of questionnaires, has allowed practitioners to collect more detailed physical, technical, and psychological data from players (De Silva et al., 2018; Rein and Memmert, 2016). These data can be used to inform scouting, performance analysis, and tactics (De Silva et al., 2018; Anderson and Sally, 2014), but increasingly they are being used to better understand the aetiology of injuries (Bourdon et al., 2017). Injuries can lead to extended periods of absence from matches, with associated impacts on team performance, as well as financial implications (De Silva et al., 2018; Bourdon et al., 2017). As such, the relationship between training load and injuries is now a key focus in football (as it is in all sports). In contrast to other data-centric contexts (e.g., health care; autonomous vehicles), however, comparatively little effort has been invested in understanding football injuries and their prediction using machine learning. Indeed, much of the existing injury research has tended to focus on a limited number of training load variables, while the application of multivariate statistical and machine learning methods—despite their obvious utility for understanding complex, multi-dimensional, problems—has been largely ignored (Claudino et al., 2019). The few studies that have used machine learning techniques to understand and predict football injuries show its potential. The timeliness of using machine learning for sports injury prediction is also highlighted by recent reviews (Van Eetvelde et al., 2021; Rossi et al., 2022). We complement this work via close examination of research on injury prediction in football, providing details of the approaches employed, along with comparison of methods, data, and results, and by providing recommendations for practitioners. Before closer examination of these studies in football, we first briefly highlight current approaches to understanding the training load and injury relationship, and then introduce machine learning and techniques from machine learning with application to understanding the prediction of football injuries. Thus, as well as highlighting the specifics of those studies on football injury, this article should serve to aid readers both from sport science and machine learning communities in their understanding of sports injury articles employing machine learning.

2.2 Training Loads and Injuries

Monitoring the load placed on athletes in training (and competition) is a current “hot topic” (Halson, 2014) in sport science, with professional sports teams investing substantial resources to this end (Bourdon et al., 2017). Load monitoring is essential for determining adaptation to training programmes, understanding responses to training, assessing fatigue and recovery, and minimizing the risk of injury and illness (Halson, 2014; Gabbett, 2016). Load can be broadly classified into two types: internal and external. Internal training load includes physiological (e.g., heart rate, blood lactate, oxygen consumption) and psychological (e.g., RPE—ratings of perceived exertion, stress, well-being) markers, collected via wearable sensors and questionnaires; external training load includes variables collected via electronic performance and tracking systems (EPTS)—e.g., velocity, acceleration, deceleration, average speed, top speed—as well as numerous other variables, such as power output and weight lifted.

Although accumulated evidence that higher training workloads may be associated with greater injury risk has led to the recommendation that workloads should be reduced to minimise injury risk (Gabbett, 2016; Windt and Gabbett, 2017; Drew et al., 2016; Soligard et al., 2016), the “Workload-Injury Paradox” (Gabbett, 2016; Windt and Gabbett, 2017) describes the phenomenon whereby intense workloads may also be associated with injury resilience. Indeed, for sport scientists working full-time in the field, any instruction to reduce workloads for currently healthy players will frequently prove to be unpopular. In seeking to better understand

and unpick the key features and components of training load and associated injury risk, several methods have been developed. Banister and colleagues (Banister et al., 1976) described differences between a positive fitness function and a negative fatigue function. The "10% rule" (Buist et al., 2007) describes protection from injury to the extent that week-to-week workload changes do not exceed 10%. The Acute Chronic Workload Ratio (denoted ACWR), developed by Hulin and colleagues (Hulin et al., 2014), is the most popular and well-researched model of the injury process (despite known limitations; Impellizzeri et al., 2020), describing the ratio of acute (i.e., rolling average of training load completed in the past week) to chronic (i.e., rolling average of training load completed in the past 4-6 weeks) workload. ACWR values exceeding 1.5 have been shown to lead to a 2-4 times greater injury risk over the following week, with an optimal range for ACWR suggested as between 0.85 and 1.35. Session load (Foster, 1998) is the product of RPE of training sessions and the duration of those sessions. "Overtraining syndrome" occurs when session loads exceed a player's ability to fully recover (Foster, 1998), and the related concept of monotony (i.e., the ratio of the mean and standard deviation of training loads—the sum of all session loads—recorded each week) has been noted as a strong risk factor for injury in studies of skating, basketball, and football (Anderson et al., 2003) (Brink et al., 2010). Finally, the ratio of internal (e.g., physiological and psychological factors) and external (e.g., data collected via GPS) workload variables (Bourdon et al., 2017; Bowen et al., 2019) has been demonstrated to be important as a predictor of injury.

2.3 Machine Learning

Machine learning is the scientific chapter of mathematics and statistical models to enable computers to use data to automatically learn and make better decisions from experience (Hastie et al., 2009). Machine learning has been applied to many areas of science, health care, and finance industries, such as for image detection, cancer detection, stock market prediction, and customer churn prediction (Claudino et al., 2019; Hastie et al., 2009). In some areas, such as sport, the effective use of machine learning is in its infancy (Claudino et al., 2019; Ruddy et al., 2019).

The algorithms (the 'rules' to be followed in calculations) used in machine learning are termed supervised learning methods (e.g., regression and classification) and unsupervised learning methods (e.g., clustering) (Hastie et al., 2009). Supervised learning methods are based on labelled input and output data (i.e., every piece of input data has a corresponding output—in the case of injury prediction, training load variables would be considered input data; and injury occurrence as output data); unsupervised learning methods are based only on unlabeled input data (i.e., the input data do not have corresponding outputs) (Hastie et al., 2009; Ruddy et al., 2019). The focus in this paper is on supervised algorithms, especially classification—predicting classes or categories as opposed to continuous values—because injury prediction is commonly based on clearly labelled training data and player injuries. In its simplest form, the task of any machine learning model is to correctly predict injuries (a positive class) and non-injuries (a negative class). Common supervised machine learning algorithms are linear and logistic regression, decision tree, random forest, k-nearest neighbors (often denoted KNN), support vector machine (often denoted SVM), artificial neural networks (often denoted ANN or NN), and 'ensemble methods' (e.g., bagging; and boosting) (Ruddy et al., 2019). Of these machine learning algorithms, some are termed white-box algorithms (e.g., linear regression, logistic regression, k-nearest neighbors, decision tree); some are termed black-box algorithms (e.g., ensemble methods, random forest, artificial neural networks, support vector machine) (Belle and Papantonis, 2020). White-box algorithms are known as interpretable approaches, which are useful, because they present a clear mapping from inputs to outputs, clarifying how analysis decisions are made—and potentially aiding practitioners and clinicians in deriving applied implications from such research. With black-box algorithms, however, this mapping from inputs to outputs is opaque. Thus, with the latter algorithms, additional post-hoc methods are needed to interpret and understand their results (Belle and Papantonis, 2020; Loyola-Gonzalez, 2019). The key point to note from the above is that all these terms are simply various algorithms that may be used, each of which may perform better or worse under different conditions.

Many real-life machine learning tasks, including injury prediction, are based on imbalanced datasets. Imbalanced datasets include a far higher number of negative examples (i.e., non-injuries) than positive examples (i.e., injuries). A problem for machine learning models can then arise, because they tend to learn from those data points present in the highest numbers (in this case, the non-injuries) and subsequently predict those non-injuries well, but fail to predict injuries (Krawczyk, 2016; Leevy et al., 2018). To improve the performance of models with such imbalanced data, studies can employ balancing techniques such as oversampling (e.g., to artificially create more injury data points) or undersampling (e.g., to remove non-injury data points), resulting in datasets with a more even balance of non-injuries and injuries. Although each approach has its drawbacks, such a process should lead to machine learning models which favour neither prediction of injury nor non-injury (Krawczyk, 2016; Leevy et al., 2018).

Classification machine learning models are typically evaluated via a number of fit metrics, some of which, such as accuracy and area under the curve (AUC) are expressed as a single value, while others, such as precision, recall, and specificity can have different values depending on the choice of the positive class (Hastie et al., 2009; Ruddy et al., 2019). Assuming injuries are considered as the positive class and non-injuries as the negative class, accuracy is the ratio of correctly predicted injuries and non-injuries to the total number of observed injuries and non-injuries; precision is the ratio of either the correctly predicted injuries to the total number of correctly and incorrectly predicted injuries; recall is the ratio of correctly predicted injuries to the total observed injuries (often described as the true positive rate or sensitivity); specificity is the ratio of correctly predicted non-injuries to the total of observed non-injuries (often described as the true negative rate); and F1-score is the “harmonic” mean (compared to a simple average, this helps to protect against any extreme values) of precision and recall (as such, this metric is sometimes considered an optimal blend of precision and recall). These metrics are often expressed in percentages. AUC is the probability curve of the true positive rate and false positive rate, with scores close to 1 indicating the best-fitting models (Ruddy et al., 2019).

Often the per-class metrics (precision, recall, specificity, and F1-score) are calculated for each class (e.g., injuries and non-injuries) separately and averaged to provide a single overall score. Although this can be reasonable in some instances, the overall score can also be misleading with imbalanced datasets, such as is often the case with soccer injury data. This is because this overall score tends not to reflect how well the model performs on what is termed the “minority class” (in this case, the injury data, because there tend to be far fewer injury than non-injury data points)—our principal focus of interest. Thus, in the latter case, recall and F1-score of just the injury class would be considered particularly useful metrics, while at the same time precision and specificity of both the injury and non-injury data help to protect against drawing conclusions which may then be biased towards the prediction of injuries. Finally, although AUC is often regarded as a very useful evaluation metric, it has also been noted to be misleading with imbalanced data (Saito and Rehmsmeier, 2015). Studies (including those highlighted in the present article) do not use these metrics in a uniform manner—that is, studies employ some but not all of, and not the same, metrics—as such, comparing studies is far from a simple process.

Extending the above, a typical machine learning chapter would proceed as follows: data collection, data pre-processing, application of machine learning algorithms (i.e., model training), and model evaluation (Kamiri and Mariga, 2021; Gibert et al., 2016). Following data collection, data pre-processing can include data cleaning (e.g., missing values imputation, handling outliers, anomaly detection), data transformation (including data normalization), feature selection (where only a subset of the original data are used in the model), and feature extraction (where new features are created from the original raw data, to perform better within the machine learning algorithm) (Gibert et al., 2016; Kotsiantis et al., 2007). This pre-processing stage generally enhances the performance of the machine learning algorithms than if they were fed with the original raw data (Gibert et al., 2016; Kotsiantis et al., 2007). Following data pre-processing, there are two main approaches to evaluate the performance of machine learning models. In the first approach, the dataset is divided into two parts—training data (c. 70%-80% of the dataset) and validation data (c. 20%-30% of the dataset). This process is termed train-validation split (although it is also frequently termed train-test split). The training data are fed into a machine learning algorithm (e.g., decision tree, support vector machine, or artificial neural network), resulting in a trained model. The predictive performance of this trained model is then subsequently assessed with the validation data. In the second approach, a machine learning model is trained on different subsets of the data and then assessed with further (validation) subsets of the same data. This process is termed cross-validation. Regardless of approach, some researchers also set aside a final portion of the dataset as “test” data—here, after validation, the models are applied to the test data to provide a final unbiased estimate of the models’ performance (Kamiri and Mariga, 2021; Gibert et al., 2016; Kotsiantis et al., 2007). How well the trained model performs with the (validation or) test data is then assessed by means of the evaluation metrics noted above (i.e., accuracy, precision, recall, specificity, F1-score, and AUC) (Kamiri and Mariga, 2021). The purpose of these validation and test processes is to try to reduce overfitting—a phenomenon whereby a model is biased towards the data it has been trained on, but has poor predictive performance when applied to previously unseen validation/test data. Machine learning is usually an iterative and cyclical process, such that, depending on the model’s performance, analysts return to earlier stages of the process, to change feature selection, to modify the settings (often called hyperparameters) of their machine learning algorithm (a process termed hyperparameter optimisation), or to try an alternative machine learning algorithm. This entire iterative and cyclic process occurs during the training and validation phases (Kamiri and Mariga, 2021; Gibert et al., 2016; Kotsiantis et al., 2007). A key point to note from the above discussion is that pre-processing techniques are applied to the training, validation, and test data, but balancing techniques are only applied to the training data. Indeed, balancing of the validation or test data would be undesirable, because assessment of the trained model would not reflect its application and performance with real-world (and unbalanced) data. In following all the preceding steps, the prediction performance of the machine learning model is often assessed and compared against what is termed a baseline model. Baseline models may be simple machine learning algorithms or dummy classifiers which use simple heuristics such as predicting the most frequent class (i.e., in our case non-injuries). With regard to feature selection, baselines normally include the most basic set of features. These base classifiers vary across studies and are set by the researchers (i.e., there are no fixed baseline criteria that must be adhered to). Ordinarily, researchers also attempt to compare their results with those from similar previous studies, a challenging process with football injury prediction, given the infancy of the area, and (as we note below), the differences in load variables used and evaluation methods employed across these studies. Ultimately, the goal is to derive a model with the best evaluation metrics with the test data. For non-experts, understanding this process is useful when trying to glean the key message from research using machine learning.

2.4 The Application of Machine Learning for Injury Prediction in Football

In section 4.1 we highlight research applying machine learning techniques to football injury prediction, describing the type of injury, the machine learning algorithms employed, the machine learning methodology, and, if mentioned, the important injury predictors (it is worthy of note that not all studies are explicit with regard to the key predictors in their models); section 4.2 (and Tables 1-3) provides a summary.

2.4.1 Existing research

Rossi and colleagues (Rossi et al., 2018) examined non-contact injuries. The authors collected 954 data recordings (each data record held information about players' daily training load) from 80 training sessions, using 18 training load variables. To account for data imbalance, they employed the "ADASYN" (He et al., 2008) oversampling technique. The authors used decision trees as the machine learning algorithm, employed both train-test split and cross-validation approaches, and constructed four baseline models with different combinations of training loads and machine learning models (i.e., logistic regression and random forest). The classification models examined in this chapter included ACWR, the ratio of mean and standard deviation (MSWR), and the exponentially moving average (EWMA) of each external training load variable (i.e., training load variables collected via GPS) individually, as well as with all training load variables simultaneously. The results demonstrated that a model including all load variables produced the best evaluation metrics when compared with standalone models for ACWR, MSWR and EWMA. In this model including all load variables, EWMA of previous injuries, EWMA of high-speed running distance, and MSWR of total distance monotony appeared to be the key predictors.

Naglah and colleagues (2018) examined non-contact football injuries caused by what they termed high-intensity workouts (more detailed information is not presented). The authors initially implemented the k-means classification (an unsupervised classification algorithm) and k-nearest neighbors algorithm on each of 65 training load variables individually using a cross-validation approach, albeit no baseline models are explicitly noted. Subsequently, using those training load variables which were significant in the initial approach simultaneously, and with support vector machine, they reported an accuracy score of 83.5%; for comparison, k-means classification with each load variable individually generated accuracy of between 40% and 75%. Overall, a model including all 65 training load variables appeared the most optimal, but further specifics on which individual variables might be most important are lacking.

López-Valenciano and colleagues (2018) and Ayala and colleagues (Ayala et al., 2019) examined lower limb muscle injuries (López-Valenciano et al., 2018) and hamstring strains (Ayala et al., 2019), comparing a range of machine learning models using 151 and 229 training load variables respectively. To account for data imbalance, both studies employed several balancing techniques: random oversampling, random undersampling, and synthetic minority oversampling (SMOTE) (Chawla et al., 2002). Bagging and boosting machine learning algorithms were tested, in order to select the best performing machine learning model for injury prediction, with both studies using a cross-validation approach and the ADTree machine learning algorithm as a baseline model. The SmoteBoost (i.e., a combination of SMOTE and boosting) technique provided the best machine learning model (with 52 (López-Valenciano et al., 2018) and 66 (Ayala et al., 2019) of the load variables). Of the 52 variables found to be important for predicting injury in López-Valenciano and colleagues' chapter, three key ones were history of lower extremity muscle injury in the last season, peak torque knee flexor concentric 300-degree dominant leg, and sport devaluation (an aspect of burnout). Of the 66 variables found to be important for predicting injury in Ayala and colleagues' chapter, history of hamstring strain injury last season, sleep quality, reduced sense of accomplishment, and range of motion-passive hip flexion with the knee extended-dominant leg appeared to be key variables.

Rommers and colleagues (Rommers et al., 2020) examined both the prediction of (a) total injuries, and (b) acute versus overuse injuries. The authors used the XGBoost algorithm to build their machine learning models, employed both train-test (on the whole dataset) and cross-validation (on the training data only) approaches, alongside grid-search (a type of hyperparameter optimisation process) as the hyperparameter optimisation process. The authors did not, however, mention any baseline model. This chapter is notable for being interpretable (see following section on black-box models), because Shapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017) was used for interpretation and visualization. SHAP demonstrated that, of the 29 training load variables examined, the five most important for predicting injuries were age at peak high velocity, body height, leg length, percent body fat, and standing broad jump. For classifying injuries as either acute or overuse, the five most important variables were age at peak high velocity, moving sideways, sitting height, 20-metre sprint, and T-test left (a specific form of sprint test, involving movements forwards and sideways).

Oliver and colleagues (Oliver et al., 2020) examined non-contact lower limb injuries based on "neuromuscular" training loads (using 20 variables). The authors examined the relationship of continuous and categorical training load variables with injuries individually, then used those variables significantly associated with injuries as inputs for multivariate logistic regression. In the latter analysis, only single leg counter movement jump (SLCMJ) peak vertical ground reaction force asymmetry remained a significant contributor to injury. The authors also implemented different ensemble (e.g., bagging, boosting) machine learning algorithms. To account for the data imbalance inherent in this dataset, the authors employed four unspecified balancing techniques. The authors used a cross-validation approach, with the J48 machine learning algorithm as a baseline model. A total of 57 machine learning models were generated, with the bagging machine learning algorithm leading to the best performing model. Across all models SLCMJ asymmetry figured prominently, attesting to its importance. Single leg hop for distance asymmetry, hop and stick (75% hop) asymmetry, knee valgus on the left leg, age, body mass, height, and leg length were also (albeit less so) prominent.

Vallance and colleagues (Vallance et al., 2020) examined non-contact injuries, with data from 245 training sessions, using 27 training load variables. The authors ran analyses with a focus on (a) the up-coming week; and (b) the following month, using machine learning with five different sets of training load variables (termed “feature sets”)—each set contained a combination of personal information, plus GPS, physical, and psychological data. The authors used a cross-validation approach, alongside Bayesian optimization (a type of hyperparameter optimization process) as the hyperparameter optimization process, with a baseline model which predicted only non-injuries. Across all analyses, k-nearest neighbors, random forest, decision tree, and XGBoost achieved the best results. The inclusion of personal, GPS, and psychological data to a baseline model (which considered past injuries only) resulted in the most accurate models. For the up-coming week, the best results were achieved using decision tree and random forest, with the following psychological features being the key predictors: RPE, pleasure, and satisfaction. For the subsequent month, the best results were achieved using XGBoost, with the following features being key predictors: RPE, pleasure, satisfaction, pain, physical shape, worry, fatigue, and maximum velocity. The presence of data imbalance in this chapter was likely somewhat alleviated by the increased number of positive cases (i.e., injuries) occurring with the focus on the upcoming week/month.

Finally, Venturelli and colleagues (Venturelli et al., 2011) examined thigh muscle strains in young players using a survival probability model (i.e., evaluation of the time—from the first training load assessment date—players “survived” without injury until occurrence of a first injury) with univariate and multivariate Cox regression on 26 variables. In their multivariate analysis, previous injuries, height, and percentage difference between two kinds of jumps (countermovement jump and squat jump) were found to be significant injury predictors. Further, in an unpublished PhD thesis (Kampakis, 2016), using various machine learning models with 69 training load variables, supervised principal components analysis outperformed all other machine learning models for injury prediction, but model fits were quite poor.

2.4.2 Summary of the research

In sum, Rossi et al. (2018), López-Valenciano et al. (2018), Ayala et al. (2019), Oliver et al. (2020) and Vallance et al. (2020) implemented various white-box, tree-based machine learning algorithms in their models. Naglah et al. (2018), Vallance et al. (2020), Venturelli et al. (2011), and Kampakis (2016) applied black-box machine learning algorithms (support vector machine, artificial neural networks, Cox regression). Rommers et al. (2020) also used a black-box model, but to counter the problem of interpretability, employed SHAP to interpret and visualise their results. The majority of articles used techniques such as SMOTE, random undersampling, and random oversampling to counter data imbalance. Further, all articles used cross-validation, although note that Rossi et al. (2018) used a prequential evaluation approach (common in stream data classification—also noted below), whereby their model was repeatedly tested on incoming (in their case, weekly) small data batches, which were then added to the training data—this approach of evaluation and up-dating with new data may more closely mirror the real-world experience of practitioners using all available data to predict injuries in real time. Table 1 (below) gives basic descriptive information about each chapter, including players' ages, types of injury, and time-frame—each of these factors could be important in determining which features are selected during machine learning as the most prominent injury predictors.

Table 1 Descriptive data for the highlighted papers

	No. of players	No. of injuries	Age group (years)	Injury type	Dataset time span
Rossi et al. (2018)	26	21	20-30	Every non-contact	23 weeks
Naglah et al. (2018)	21	36	Unreported	Every non-contact	16 months
López-Valenciano et al. (2018)	132	32	Unreported	Lower leg muscle	Pre-season+ 1 Season
Ayala et al. (2019)	96	18	Unreported	Hamstring strain	Pre-season+ 1 season
Rommers et al. (2020)	734	368	10-15	Acute and overuse	Pre-season+ 1 season
Oliver et al. (2020)	400	99	10-18	Non-contact lower leg	Pre-season+ 1 season
Vallance et al. (2020)	40	142	23.6-35.2	Every non-contact	Pre-season+ 1 season
Venturelli et al. (2011)	84	27	14-18	Thigh muscle strain	Pre-season+ 1 season
Kampakis (2016)	Unreported	Unreported	Unreported	Not specified	Unreported

Note. Only Oliver et al. and Vallance et al. specifically reported using “male” players. The other papers noted the following: young football players, elite football players, youth players, and/or professional football players.

Table 2 (below) lists the training load variables considered as input features in the studies. Despite some consistency, there is also wide variability in features, meaning that drawing conclusions across studies is complex. Thus, the lack of consensus regarding which features to consider for analysis, let alone those subsequently proving to be key in predicting players' injuries, makes it difficult at this time for practitioners to rely on these studies when choosing which training load features upon which to focus.

Table 2 Training load features in the highlighted papers

	[1]	[2]	[3, 4]*	[5]	[6]	[7]	[8]	[9]
External Load								
Exposure							X	
Jumps		X					X	
Distance	X	X				X		X
Accelerations and decelerations	X	X				X		X
DSL (Total weighted impacts above 2g)	X							
Duration		X				X		
Player Load		X				X		X
Speed and velocity		X				X		X
Meterage per minute		X						
Total efforts		X						
High Inertial Movement Analysis’s		X						
Average Metabolic Power								X
Dynamic Stress Load								X
Impacts								X
Energy Expenditure								X
Step Balance								X
Dribbling				X				
Sprint				X				
Jumping, moving and balancing				X				
Internal Load – Physical Data								
Body Mass Index	X		X	X	X	X	X	
Fat percentage				X			X	
Step yo-yo test				X			X	
Heart rate		X						
Ratings of perceived exertion (RPE)						X		
Internal Load – Psychological Data								
Sleep quality			X			X		
Physical exhaustion			X					
Reduced sense of exhaustion			X					
Sport devaluation			X					
Fatigue, shape, pain, pleasure, worry, satisfaction						X		
Personal Information								
Height and weight			X	X	X	X	X	

Age	X		X	X	X	X	X	X
Role of the player (Position)/ field position	X		X			X	X	X
Previous injury	X		X			X	X	
Minutes played in previous games	X							
Number of games played before each training session	X							
Dominant leg	X		X					
Current level of play			X					
Injury details								X
Season stage								X
Activity								X
Phase of play								X
Footwear								X
Surface condition								X
Sitting height, curl-ups, leg length				X				
75% Hop, SLCMJ, SLHD, Y-balance, TJ Knee					X			
ACWR and MSWR of training loads	X							
Neuromuscular training loads				X	X			
Total training load features	55	65	151, 229	29	20	27	18	18

Note. Neuromuscular training loads is an over-arching "feature" which includes multiple variables not explicitly mentioned here. *These two papers included 151 and 229 training load variables, under eight over-arching topics (with the most important ones noted here).

[1] denotes Rossi et al. (2018), [2] denotes Naglah et al. (2018), [3] denotes López-Valenciano et al. (2018), [4] denotes Ayala et al. (2019), [5] denotes Rommers et al. (2020), [6] denotes Oliver et al. (2020), [7] denotes Vallance et al. (2020), [8] denotes Venturelli et al. (2011) and [9] denotes Kampakis (2016).

The above notwithstanding, the evaluation metrics in Table 3 (below) appear to demonstrate that overall, the best models for injury prediction are those reported by Rossi et al. (2018), Ayala et al. (2019), Rommers et al. (2020), and Vallance et al. (2020) The work of Rommers et al. (2020) and Vallance et al (2020). considered a far greater number of injuries than the other studies, potentially improving prediction. Ayala et al. (2019), Rommers et al (2020)., and Vallance et al. (2020) used boosting-based algorithms, which thus appear to work well in this context. Both Rossi et al. (2018) and Ayala et al. (2019) used data oversampling, while Rommers et al. (2020) and Vallance et al. (2020) did not use any data balancing techniques, presumably because of their larger datasets and greater number of positive cases (i.e., injuries).

Table 3 Model fit for the best-fitting models from each paper

	Machine Learning Algorithms	Pre-processing techniques	Accuracy (%)	Precision (%)	AUC	Recall (%)	F1 score (%)	Specificity (%)
Rossi et al. (2018)	Decision Tree	Feature selection, Oversampling – SMOTE	-	50	0.76	80	64	-
Naglah et al. (2018)	SVM	Data Normalization	83.50	-	-	-	-	-
López-Valenciano et al. (2018)	SmoteBoost	Oversampling - SMOTE	-	-	0.75	65.90	-	79.10
Ayala et al. (2019)	SmoteBoost	Oversampling - SMOTE	-	-	0.84	77.80	-	83.80
Rommers et al.(a) (2020)	XGBoost	Unmentioned	-	85	-	85	85	-
Rommers et al. (b) (2020)			-	78	-	78	78	-

Oliver et al. (2020)	Decision Tree	Various balancing techniques	-	-	0.66	55.60	-	74.20
Vallance et al. (a)* (2020)	Random Forest	Missing values imputation	95.5	92.2	0.92	94.5	-	-
Vallance et al. (b)* (2020)	XGBoost		97	97	0.97	97	-	-
Venturelli et al. (2011)	Cox Regression	Unmentioned	-	-	-	-	-	-
Kampakis (a) (2016)	Supervised Principal	Unmentioned	88.80	55	-	33	-	-
Kampakis (b) (2016)	Componen ts Analysis		97.07	19	-	20	-	-

Note. Each paper used a different overall set of model fit metrics. In papers Rommers et al., Vallance et al. and Kampakis, two key differential approaches (denoted *a* and *b*) were used. *This article did not explicitly mention evaluation metrics—we approximated these values from the article's presented boxplots.

Overall, although the research highlighted in this article demonstrates the potential of machine learning for bringing new insight to our understanding of injury prediction in football, as readers might observe, there is considerable variability in chapter design and analysis. More generally, a major concern (and a future research issue) is that the studies examined here are based on data collected across a single season. An important future direction would be to test and refine the developed models on subsequent seasons' data, with their inherent changes in players, coaches, training, and matches. Indeed, in addition to the above, might a consideration of aspects such as the workload-injury paradox, ACWR, and overtraining syndrome aid in the design of research and analysis plans to make the most of the predictive ability of machine learning models? The paper from Rossi and colleagues (Rossi et al., 2018) is the only one to take the latter approach, drawing on ACWR, MSWR and EWMA in their machine learning analysis.

Building from the above, although the machine learning techniques employed in the research highlighted above are quite sound, greater detail regarding the machine learning approaches employed would help any objective assessment of their contribution towards better understanding the workload-injury relationship. For example, greater clarity with regard to whether the reported evaluation metrics are "per-class" or "averaged" would be important—only Rossi and colleagues (2018) explicitly mentioned recall and precision of their models for injury and non-injury data separately. Further, as injury datasets likely have large amounts of missing and unclean data, greater detail regarding the various pre-processing techniques employed (e.g., in relation to any missing values, different data imputation techniques required, balancing, and clarity regarding all types of demographic data, and internal and external load variables) would be important for drawing conclusions and guiding future work. Here, only López-Valenciano et al. (2018) and Ayala et al. (2019) gave a complete account of all the various pre-processing techniques they used in their research.

With the above said, researchers would be well advised to consider several key points before employing machine learning models. The first is to clearly define the task—often drawing on the needs and preferences of football practitioners. For example, is the interest simply in raw predictions, probabilities, or in examining specific features impacting injuries? Second, with regard to data compilation and pre-processing, practitioners at football clubs are likely to have varied sources of data, often in unique formats, such that great care should be taken to avoid errors when compiling such data into one final dataset. Third, we would recommend ensuring that input data are examined in relation to injuries occurring after collection of those input data—i.e., such that the model may predict injuries in the future (e.g., in one day's time or in seven days' time). That is, any training and input data from the same day that an injury has occurred should be disregarded, because such data may be confounded by the injury occurrence. Fourth, with regard to data pre-processing, given the longitudinal nature of football injury datasets, it would make more sense to replace missing values on a player-by-player basis, rather than across the whole dataset, as well as using interpolation for this purpose for some features. Similarly, data balancing might also be conducted on a player-by-player and/or season-by-season basis.). Finally, as noted above, changes of coaches, managers, players, and training regimes across seasons mean that the underlying distribution and quality of data will vary from season to season. Traditional machine learning algorithms assume that the underlying distribution of the data is the same. To counter this problem, a focus on what is termed stream learning may help to better understand and interpret machine learning models with multi-season data. What the preceding lines suggest is that future machine learning research in this area could be well served by drawing from current expertise, insight, and understanding from sport science and sport practitioners.

2.5 Conclusion

Machine learning for football injury prediction is a new but fast growing research area. Machine learning approaches can help expand the focus from univariate models, to create a better understanding of the relative influence of various (physical and psychological) aspects of training load on injury risk. In this article, part of our aim was to highlight (and to an extent de-mystify) the machine learning process. Machine learning is simply an analytical technique, but its power lies in its ability to work so eloquently with such a vast array of load variables. Although this can offer greater flexibility over analysis with more simplified models (e.g., using ACWR), the myriad ways machine learning can be employed can also lead to difficulty in synthesising the current research evidence into an overall, unified, conclusion. Indeed, there remain questions as to the utility of these models for real-world application. The use of white-box machine learning algorithms in a number of the present articles should aid understanding and application. Black-box models may, however, offer better predictive performance, despite being difficult to interpret and understand. The latter issue of interpretability can be addressed using explainable artificial intelligence approaches, like SHAP (Lundberg and Lee, 2017), Local Interpretable Model-agnostic Explanations (Ribeiro et al., 2016), and partial dependency plots (Goldstein et al., 2014; Friedman, 2001). Despite its infancy, coupled with the limitations we have noted, machine learning for understanding the workload-injury relationship in football is clearly a method whose time has come. By expanding the focus to multiple seasons' data, accounting for data imbalance, and using explainable artificial intelligence, machine learning should help to unlock new insights into the workload-injury relationship.

Chapter 3

3. Abstract

The purpose of this chapter was to use machine learning to examine the relationship between training load and soccer injury with a multi-season dataset from one English Premier League club. Participants were 35 male professional soccer players (aged 25.79±3.75 years, range 18-37 years; height 1.80±0.07 m, range 1.63-1.95 m; weight 80.70±6.78 kg, range 66.03-93.70 kg), with data collected from the 2014-2015 season until the 2018-2019 season. A total of 106 training loads variables (40 GPS data, 6 personal information, 14 physical data, 4 psychological data and 14 ACWR, 14 MSWR and 14 EWMA data) were examined in relation to 133 non-contact injuries, with a high imbalance ratio of 0.013. XGBoost and Artificial Neural Network were implemented to train the machine learning models using four and a half seasons' data, with the developed models subsequently tested on the following half season's data. During the first four and a half seasons, there were 341 injuries; during the next half season there were 37 injuries. To interpret and visualize the output of each model and the contribution of each feature (i.e., training load) towards the model, we used the Shapley Additive Explanations (SHAP) approach. Of 37 injuries, XGBoost correctly predicted 26 injuries, with recall and precision of 73% and 10% respectively. Artificial Neural Network correctly predicted 28 injuries, with recall and precision of 77% and 13% respectively. In the model using Artificial Neural Network (the relatively more accurate model), last injury area and weight appeared to be the most important features contributing to the prediction of injury. This was the first chapter of its kind to use Artificial Neural Network and a multi-season dataset for injury prediction. Our results demonstrate the potential to predict injuries with high recall, thereby identifying most of the injury cases, albeit, due to high class imbalance, precision suffered. This approach to

using machine learning provides potentially valuable insights for soccer organizations and practitioners when monitoring load injuries.

3.1 Introduction

Monitoring the load placed on athletes in training and competition is a current "hot topic" (Kalkhoven et al., 2021) in sport science, with professional sports teams investing substantial resources to this end (Bourdon et al., 2017). Load monitoring is essential for determining adaptation to training programs, assessing fatigue and recovery, and minimizing the risk of injury and illness (Kalkhoven et al., 2021; Halson, 2014). As the most popular global sport, with 200,000 professional and 240 million amateur players, and with injury incidence higher than any other sport (Rahnama, 2011; Owoeye et al., 2020; Jones et al., 2019), soccer has become a key focus for research into load monitoring and injury. Soccer injuries can lead to extended periods of absence from matches, with associated impacts on team performance, as well as financial implications (Rahnama, 2011; Owoeye et al., 2020; Jones et al., 2019; Ibrahimović et al., 2021). Indeed, from 2012-2013 through to the 2016-2017 season, injuries cost English Premier League soccer clubs approximately £45 million per season (Eliakim et al., 2020). In attempting to better understand the relationship between training load and soccer injury, recent research has begun to draw on techniques from machine learning (for a review, see Majumdar et al., 2022). In the present chapter, we employed a multi-dimensional and multi-season interpretable machine learning approach to examine the relationship between training load and soccer injury using data from one English Premier League club.

The timeliness of using machine learning for sports injury prediction is highlighted by recent reviews (Van Eetvelde et al., 2021; Rossi et al., 2021). Machine learning approaches can help expand the focus from more simplified models of the injury process—such as when using the Acute Chronic Workload Ratio (ACWR) (Hulin et al., 2013), the most popular and well-researched model of the injury process—to create a better understanding of the relative influence of various (physical and psychological) aspects of training load on injury risk. The original research into the ACWR (Hulin et al., 2013) in the sport of cricket suggested an optimal ACWR range of between 0.85 and 1.5, with ACWR values exceeding 1.5 leading to a 2-4 times greater injury risk. But there have been recent methodological and theoretical criticisms of ACWR (e.g., see Impellizzeri et al., 2021). Further, although tests of the ACWR with data from the English Premier League (Bowen et al., 2019) have shown that if the ACWR exceeds a value of 2 when chronic load is low, there is 5-7 times greater risk of injury, other work within Italian professional soccer (Rossi et al., 2018) has not observed any training sessions with ACWR values exceeding 2, finding that the highest injury risks occur when the ACWR is less than 1. These sorts of concerns and equivocal results have led to recent machine learning research examining soccer injury with a greater number or explanatory load variables (Rossi et al., 2018; Vallance et al., 2020; Naglah et al., 2018; Lopez-Valenciano et al., 2018; Ayala et al., 2019; Rommers et al., 2020; Oliver et al., 2020; Venturelli et al., 2011; Kampakis, 2016). The above notwithstanding, however, there are a number of limitations in this research that have been noted (Majumdar et al., 2022). These include, though are not limited to, a need for (a) greater clarity with regard to the reported evaluation metrics (e.g., recall and precision), and whether they are "per-class" of injury or "averaged" across injury and non-injury data; (b) greater detail regarding the various pre-processing techniques employed (e.g., in relation to any missing values, different data imputation techniques required, balancing, and clarity regarding all types of demographic data, and internal and external load variables); and (c) studies over more than one season, wherein models are tested and refined on subsequent seasons' data, with their inherent changes in players, coaches, training, and matches.

In the present chapter we addressed each of these limitations, examining the relationship between training load and soccer injury with a multi-season dataset from one English Premier League club. The latter point is important, because previous research has, with the exception of the work of Rossi and colleagues (2018), tended to focus on developing models with just one season's data, using cross-validation and train-validation split, leaving questions as to how accurate such models would be in predicting "unseen" data (such as from a subsequent season). Specifically, then, our novel approach was to train models on data collected across four and a half soccer seasons, and then to test those models on the next unseen half season's data. Alongside addressing the known limitations of previous papers, we also sought to examine multiple forms of data (e.g., Global Positioning System data, physical data, psychological data, and demographic data)—something only Vallance et al. (2020) had previously reported.

To provide the best opportunity to then unearth insights with our training load input data and injury output data, we drew upon state-of-the-art processes from machine learning (such as using the XGBoost algorithm: Chen and Guestrin, 2016), but also drew upon deep learning, wherein the employed algorithms are inspired by the structure and functions of biological neural networks—often called Artificial Neural Networks (or ANNs) (Mehlig, 2019). Finally, we should note another criticism of previous papers examining load monitoring and soccer injury—the lack of clarity with regard to the key variables underpinning the machine learning models developed. This is important, because if machine learning is to become a key part of a practitioner's toolkit in understanding injury risk, machine learning models need to provide clarity with regard to the causes of (or key risks for) injury—i.e., the importance of "interpretability" (Belle and Papantonis, 2020). In this context, white-box models use algorithms (e.g., linear regression, logistic regression, k-nearest neighbors, decision tree) that are interpretable, presenting a clear mapping from inputs to outputs, clarifying how analysis decisions are made—and potentially aiding practitioners and clinicians in deriving applied implications from such research (Loyola-Gonzalez, 2019). On the other hand, black-box models use algorithms (e.g., ensemble methods, random forest, artificial neural networks, support vector machine) that are not easily interpretable, but may be more powerful. In the latter examples, the mapping from inputs to outputs is opaque, but additional post-hoc methods can then be used to interpret and understand the results (Loyola-Gonzalez, 2019). In the present chapter, we employed black-box methods, and thus to aid interpretability, we employed the Shapley Additive exPlanations (SHAP) approach (Lundberg and Lee, 2017)—an explainability framework based on game theory, which can be used to unpick the key predictors of machine learning models by computing the contribution of each feature to prediction.

Overall, then, in the first chapter of its type, we report a novel approach which can address gaps in existing research and produce a practical solution for soccer injury prediction. Through comprehensive analysis of a unique multi-season dataset of Elite Premier League soccer players, we aimed to develop a multi-dimensional predictive machine learning model to assess injury risk among players in the following seven days.

3.2 Materials and Methods

3.2.1 Data collection and Feature creation

Participants were 35 male professional soccer players (aged 25.79±3.75 years, range 18-37 years; height 1.80±0.07 m, range 1.63-1.95 m; weight 80.70±6.78 kg, range 66.03-93.70 kg) from one English Premier League club, with data collected from the 2014-2015 season until the 2018-2019 season. Players' positions were recorded as follows: eight full-backs, nine center-backs, seven central mid-fielders, eight wing-forwards, and three strikers. Data were provided to the research team by the club's first team sports science department, having been collected as part of the club's day-to-day data collection processes, and with all permissions in place. The dataset contained 343 injury data points, of which our focus was the 133 non-contact injuries. Of these 133 non-contact injuries, there were 43 thigh injuries, 29 knee injuries, 24 hip injuries, 19 ankle injuries, and 18 'lower leg' injuries. Across injuries, eight players were injured once, nine players were injured two times, four players were injured three times, two players were injured four times, four players were injured five times, two players were injured six times, four players were injured seven times, one player was injured 11 times, and one player was injured 16 times. Overall, there were 11 injuries recorded in the 2014-2015 season, six in the 2015-2016 season (the club's first in the English Premier League), 28 in the 2016-2017 season, 41 in the 2017-2018 season, and 47 in the 2018-2019 season.

The available 'load' data included Global Positioning System (termed GPS) data, physical (e.g., various skinfold measurements, bodyfat percentage) data, psychological (e.g., RPE) data, and demographic information. Feature selection first focused on removing features with more than 60% missing values. Please note, when players missed training sessions, their absence of training load data is not noted in the dataset and is thus not treated as missing data. Subsequently, different missing values imputation methods were used across the features. We also created two additional features within the dataset: "last injury area" and "days since last injury". Below, Table 4 lists all training load variables considered as input features in the present chapter, along with their description, source, method of collection, frequency of data collection (e.g., GPS and psychological data are collected daily; physical data are collected every two weeks), and missing values imputation techniques.

Table 4 Training Load Variables, Variable Descriptions, Missing Value Imputation Techniques, Method of Data Collection, and Data Collection Frequency

Variable Name	Variable Description	Missing Value Imputation	Method of data collection	Frequency of data collection
GPS measures/ External Load				
Total Duration	Total time in minutes an athlete is in activity	knn	Time taken from activity on GPS device	Every pitch session and game
Total Distance (m)* (TDM)	Distance in meters covered during the activity	knn	GPS device	Every pitch session and game
Meterage Per Minute* (MPM)	Distance in meters covered during the activity per minute	knn	GPS device	Every pitch session and game
Sprint Efforts* (SE)	Number of efforts above 7 m/s	knn	GPS device	Every pitch session and game

Sprint Distance (m)	Distance in meters covered above 7 m/s	knn	GPS device	Every pitch session and game
High Speed Distance (m)* (HSD)	Distance in meters covered above 5.5 m/s	knn	GPS device	Every pitch session and game
High Speed Distance Per Minute (m/min) (m)	Distance in meters covered above 5.5 m/s per minute of activity	knn	GPS device	Every pitch session and game
Maximum Velocity (m/s) * (MV)	Maximum velocity reached in activity	knn	GPS device	Every pitch session and game
Velocity Band 7 Total Effort Count	Number of efforts above 90% of players maximum velocity	knn	GPS device	Every pitch session and game
Velocity Band 7 Total Distance (m)	Distance in meters covered above 90% of players maximum velocity	knn	GPS device	Every pitch session and game
Total Player Load* (TPL)	Sum of the accelerations across all axes of the internal tri-axial accelerometer during movement. It considers instantaneous rate of change of acceleration and divides it by a scaling factor (divided by 100).	knn	GPS device	Every pitch session and game
Accels* (ACC)	Number of accelerations above 0.5 m/s^2	knn	GPS device	Every pitch session and game
Decels* (DCC)	Number of decelerations above -0.5 m/s^2	knn	GPS device	Every pitch session and game
Perceived Exertion* (PE)	The Borg Rating of Perceived Exertion (RPE) using Borg CR10 Scale	knn	Questionnaire	Every pitch session and game
Workload* (WD)	Perceived Exertion x Total Duration	knn	Calculation of Total Duration x RPE	Every pitch session and game
Meta Energy (KJ/kg) * (ME)	Estimated energy expenditure, based on GPS acceleration	knn	GPS device	Every pitch session and game
Velocity Work/Rest Ratio	Time working divided by Time resting where work	knn	GPS device	Every pitch session and game

	and rest are defined by velocity thresholds			
Work/Rest Ratio	The amount of time spent above the work velocity threshold divided by the amount of time spent below the rest velocity threshold	knn	GPS device	Every pitch session and game
Relative Intensity	(High Speed Distance (m) / Total Distance (m)) * 100	knn	GPS device	Every pitch session and game
Mean Heart Rate	Average heart rate (beats per minute) in activity	knn	GPS device	Every pitch session and game
Maximum Heart Rate	Maximum heart rate (beats per minute) in activity	knn	GPS device	Every pitch session and game
Player Load Per Minute	Average Player Load accumulated per minute of activity	knn	GPS device	Every pitch session and game
Player Load (1D Fwd)	Player Load accumulated in the sagittal plane	knn	GPS device	Every pitch session and game
Player Load (1D Side)	Player Load accumulated in the frontal plane	knn	GPS device	Every pitch session and game
Player Load (1D Up)	Player Load accumulated in the sagittal plane	knn	GPS device	Every pitch session and game
Player Load (2D)	Player Load accumulated in the frontal and sagittal planes	knn	GPS device	Every pitch session and game
RHIE Total Bouts	The total occurrences of Repeated High Intensity Effort (RHIE) events	knn	GPS device	Every pitch session and game
RHIE Effort Duration - Mean	The average duration of a RHIE event	knn	GPS device	Every pitch session and game
RHIE Effort Duration - Min	The shortest duration of a RHIE event	knn	GPS device	Every pitch session and game
RHIE Effort Duration - Max	The longest duration of a RHIE event	knn	GPS device	Every pitch session and game

RHIE Bout Recovery - Mean	The average amount of time between RHIE events	knn	GPS device	Every pitch session and game
RHIE Bout Recovery - Min	The shortest time between RHIE events	knn	GPS device	Every pitch session and game
RHIE Bout Recovery - Max	The longest amount of time between RHIE events	knn	GPS device	Every pitch session and game
IMA Jump Count Low Band	The total number of jumps registered 0-20 cm	knn	GPS device	Every pitch session and game
IMA Jump Count Med Band	The total number of jumps registered 20-40 cm	knn	GPS device	Every pitch session and game
IMA Jump Count High Band	The total number of jumps registered >40 cm	knn	GPS device	Every pitch session and game
HMLD*	Distance in meters covered by a player where his/her Metabolic Power is >25.5 W/kg	knn	GPS device	Every pitch session and game
HML Distance Per Minute (m/min) (m)	Distance in meters covered by a player where his/her Metabolic Power is >25.5 W/kg per minute	knn	GPS device	Every pitch session and game
Explosive Efforts* (EE)	IMA Accel High + IMA Decel High + IMA CoD Left High + IMA CoD Right High + IMA Accel Medium + IMA Decel Medium + IMA CoD Left Medium + IMA CoD Right Medium	knn	GPS device	Every pitch session and game
Explosive Efforts per Min* (EEM)	EE/ minute	knn	Calculation	Every pitch session and game
Personal Information				
Age	Age of player			
BMI	Body Mass Index; ratio between weight (in kg) and the square of height (in meters)	None	Calculation	
Height	Player’s height in centimetres	None	Measurement from Sadiometer	Pre-season

Weight	Players weight in kilograms	Linear Interpolation	Measurement from Secca Scales	Fortnightly
Last Injury Area	Last injury area	None		
Days since last injury		None		
Internal Load – Physical data				
TRICEP	Triceps' skinfold measurement	Linear Interpolation	Skinfold measurement taken with Harpenden Callipers	Fortnightly
SUBSCAP	Subscapular skinfold measurement	Linear Interpolation	Skinfold measurement taken with Harpenden Callipers	Fortnightly
BICEP	Bicep skinfold measurement	Linear Interpolation	Skinfold measurement taken with Harpenden Callipers	Fortnightly
ILIAC	Iliac Crest skinfold measurement	Linear Interpolation	Skinfold measurement taken with Harpenden Callipers	Fortnightly
SUPRA	Supraspinal skinfold measurement	Linear Interpolation	Skinfold measurement taken with Harpenden Callipers	Fortnightly
ABDOM	Abdominal skinfold measurement	Linear Interpolation	Skinfold measurement taken with Harpenden Callipers	Fortnightly
THIGH	Thigh skinfold measurement	Linear Interpolation	Skinfold measurement taken with Harpenden Callipers	Fortnightly
CALF	Calf skinfold measurement	Linear Interpolation	Skinfold measurement taken with Harpenden Callipers	Fortnightly
Skinfolds	Sum of 8 site skinfold measurements	Linear Interpolation	Calculation	Fortnightly
% Bodyfat (Yuhasz).	(0.1051 x sum of triceps, subscapular, su praspinal, abdominal, thi gh, calf) + 2.585	Linear Interpolation	Calculation	Fortnightly
% Bodyfat (Jackson)	(0.29288 x sum of skinfolds) – (0.0005 x square of the sum of	Linear Interpolation	Calculation	Fortnightly

	skinfolds) + (0.15845 x age) – 5.76377			
Fat Mass	(Weight/100) * % Bodyfat (Jackson)	Linear Interpolation	Calculation	Fortnightly
Lean Mass	Weight - Fat Mass	Linear Interpolation	Calculation	Fortnightly
Relative Lean Mass	Lean Mass/Weight	Linear Interpolation	Calculation	Fortnightly
Internal Load – Psychological data				
Sleep	Previous night's sleep quality	Forward fill and back fill	Questionnaire	Every training day
Fatigue	Fatigue level	Forward fill and back fill	Questionnaire	Every training day
Ext. Stress	Stress level	Forward fill and back fill	Questionnaire	Every training day
Soreness	Muscle Soreness	Forward fill and back fill	Questionnaire	Every training day
ACWR, MSWR and EWMA				
ACWR of 14 daily GPS features*	Given a training load feature, the Acute Chronic Workload Ratio (ACWR) is the ratio of acute (i.e., rolling average of training load completed in the past week) to chronic (i.e., rolling average of training load completed in the past 4 weeks) workload.	knn	Calculation	
MSWR of 14 daily GPS features*	Monotony of a player. Given a training load feature, MSWR is calculated by taking the ratio of the mean and standard deviation of the values of the training load in the past 1 week/ 7 days.	knn	Calculation	
EWMA of 14 daily GPS features*	Exponential weighted moving average puts greater weight and significance to the most recent training loads (i.e., data points). It follows a	knn	Calculation	

decay rule of $\alpha = \frac{2}{Span+1}$
where the span is set to 7.

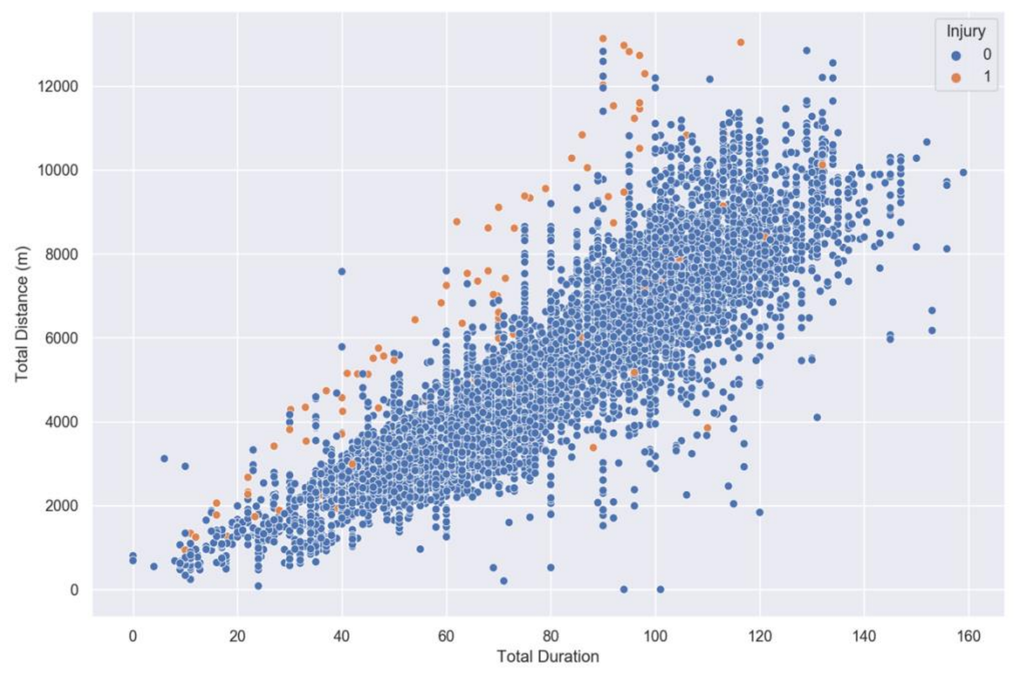
Note. *These training load variables are used in the calculation of ACWR, MSWR and EWMA.

3.2.2 Dataset construction

We constructed a multi-dimensional load-injury prediction model to forecast whether a player would become injured in the next seven-day window. This seven-day window was chosen to mirror the standard frequency of English Premier League match occurrence—i.e., a match is played approximately every seven days (and generally at the weekend). A similar approach was employed by Vallance et al. (2020). There are generally between three and four training sessions each week, with training loads reaching their peak towards the end of each week.

To accomplish the task of constructing an injury prediction model, we initially built a master dataset consisting of 106 training load variables (see Table 4): 40 GPS data variables, six personal information variables, 14 physical data variables, four psychological data variables, 14 ACWR, 14 MSWR, and 14 EWMA data variables (mentioned in Table 4), one injury label (indicating 1 if the player is injured and 0 if not), and 10653 data points (i.e., each data point is a row which describes the training information and personal information for each player). In this master dataset, there were 10,520 non-injury data points and 133 injury data points, indicating a high imbalance ratio of 0.013. Importantly, in this master dataset, the injury label was assigned to the original injuries that happened on the same day (i.e., injuries that were recorded on the day of occurring), but our aim was to predict injuries in the next seven-day window. To achieve the latter focus, we thus assigned the previous data points (i.e., each data point or row that came before the original data points) present in the past seven days of the original injury data point to 1 and removed the original injury data points. The assumption behind removing the original data points is that the injury occurring on a specific day is caused by the training loads of the previous days. As a result, our seven-day injury prediction model is based on a revised dataset containing 10,520 data points, of which there are 10,142 non-injury data points and 378 injury data points, giving an imbalance ratio of 0.037. Figure 1 presents the injury and non-injury distribution in the original and seven-day injury prediction dataset (denoted D) respectively. In the seven-day injury prediction dataset (D) the injury and non-injury data points overlap. Imbalanced and overlapping data classification represent a challenge for traditional machine learning models, which often fail to recognize patterns in such data (Shahee and Ananthakumar, 2021; Kiesow et al., 2021).

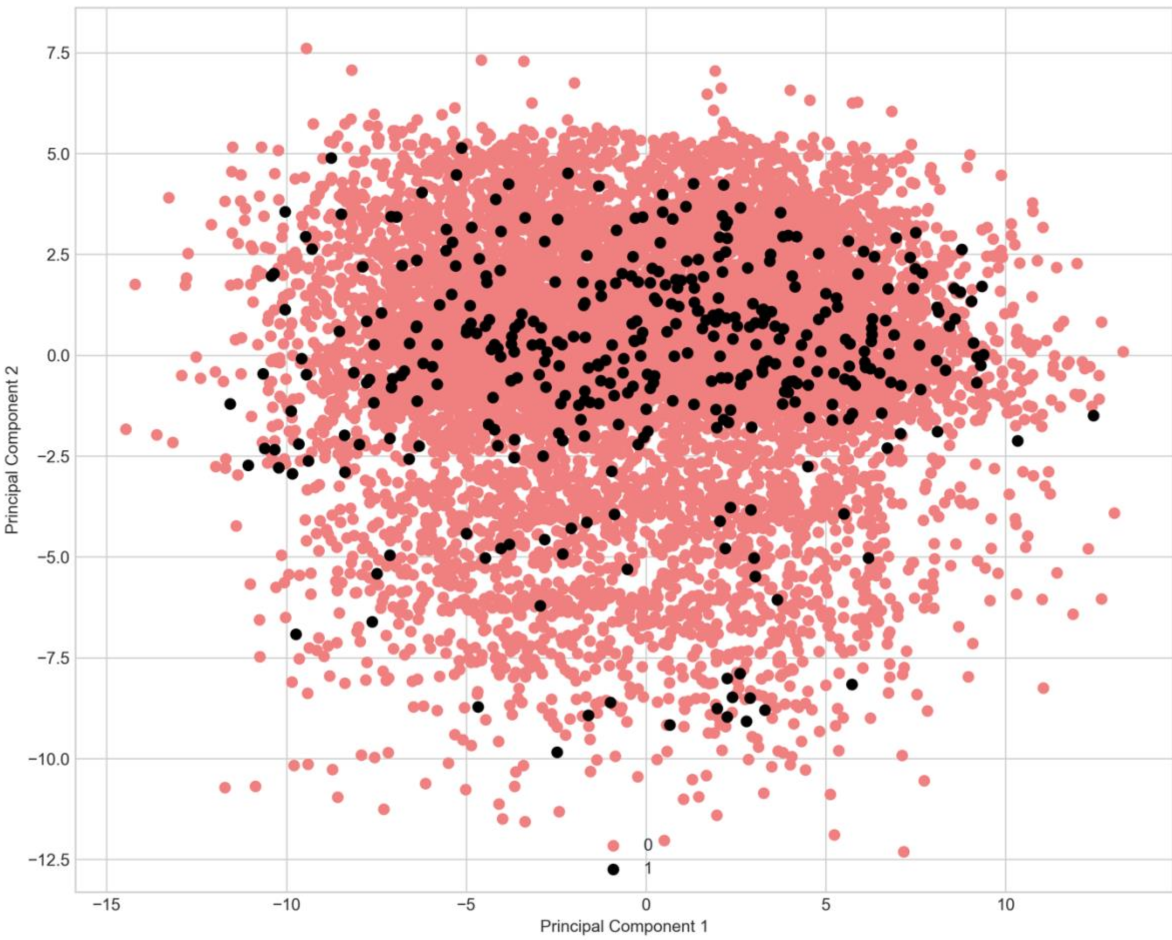
Figure 1 The Relationship Between Graphical Representation of Injury and Non-Injury Distribution in the Original and Seven-Day Injury Prediction Dataset using two training load variables.



Note. Top panel: Injury and non-injury distribution in the original dataset. Bottom panel: Injury and non-injury distribution in the seven-day injury prediction dataset. To present the injury and non-injury distribution in both the datasets, total duration and total Distance (m) were used.

In addition, for a better depiction of the classification problem and how our high-dimensional injury and non-injury datapoints appear in a two-dimensional plane we performed Principal Component Analysis. Below Figure 2 demonstrates that the injury and non-injury data points are overlapping (Tang et al., 2010; Sáez et al., 2019; Gupta and Gupta, 2018; Shahee and Ananthakumar, 2021; Kiesow et al., 2021). This is illustrated by many instances where similar training programs resulted in different outcomes, which is likely an indication that the features which would clearly separate the two classes are not being currently collected.

Figure 2 Principal Component Analysis of the Seven-Day Injury Prediction Dataset



Note: Principal component analysis on dataset D (the seven-day injury prediction) with 106 features. Red dots represent non-injury data points; black dots represent injury data points.

We should also note that, while calculating ACWR, MSWR and EWMA for each player, we used the training sessions which fell in the past seven days before each training session or match-day. The past seven days may be different from the past seven training sessions as the past seven training sessions might not fall into the past seven days.

3.2.3 Model construction

For model building, validation, and testing, we used the Python programming language. We used various machine learning algorithms—logistic regression, k-nearest neighbors, decision tree, and random forest resulted in poor model performance, failing to predict most of the actual injuries—with XGBoost (Chen and Guestrin, 2016) and Artificial Neural Network (ANN)

(Mehlig, 2019) providing the best results. In this paper, we thus focus from this point onwards on the use of and results from the XGBoost and ANN algorithms. We used various pre-processing techniques, such as oversampling the minority data points (i.e., the injury data), feature scaling (i.e., scaling each training load), and setting different hyperparameters.

We first split the entire dataset into two parts—the training data ($\mathrm{D_{Train}}$), containing the first four and half seasons' data, and the test data ($\mathrm{D_{Test}}$), containing the remaining half season. $\mathrm{D_{Train}}$ contained 9548 non-injury data points and 341 injury data points and $\mathrm{D_{Test}}$ contained 493 non-injury data points and 37 injury data points. The test set was further divided into three labelled months: Month 1 contained 161 non-injury data points and 14 injury data points; Month 2 contained 162 non-injury data points and 14 injury data points; and Month 3 contained 170 non-injury data points and 9 injury data points. Months 4 and 5 did not contain any injury data points.

We first trained XGBoost and ANN on $\mathrm{D_{Train}}$. During this model training we performed 10-fold cross-validation to check how well the model performed on different validation subsets of the data. Hyperparameter optimization techniques, including grid-search and Bayesian optimization, were implemented to refine the model's configuration. The overarching goal of hyperparameter tuning was to identify settings that would yield optimal outcomes when tested on the independent test dataset. To achieve this, the Bayesian optimization process yielded a set of hyperparameters that notably improved the prediction of instances associated with non-injuries. Complementary to this, grid-search contributed partially to the refinement of hyperparameters by predicting both injuries and non-injuries in a balanced way. These endeavors collectively provided preliminary estimates of hyperparameter values. It is noteworthy that the precise values obtained from the Bayesian optimization and grid-search hyperparameter optimization techniques were not adopted verbatim. Subsequent to the initial hyperparameter optimization, a further iterative phase ensued wherein the hyperparameters of both models were subject to adjustments. This iterative refinement process involved multiple iterations of cross-validation procedures to iteratively enhance the model configurations. We also performed feature selection techniques, such as Recursive Feature Elimination, Variance Threshold (i.e., removing low variance features) techniques to reduce the dimensionality of the feature space and risk of overfitting. The best results were obtained by simultaneously using all features (i.e., all the training load types).

Data imbalance in the training data was a concern, which, if left untreated, would heavily bias the outcomes towards non-injuries. To combat this data imbalance, while applying XGBoost, we (a) implemented the Synthetic Minority Oversampling Technique (i.e., SMOTE: Chawla et al., 2002) to create "synthetic" injury instances, and (b) set the weighting for injury at nine times higher than the non-injury weighting. On the other hand, while applying ANN, we (a) scaled the data, (b) implemented SMOTE, and (c) set the weighting for injury at 11 times higher than the non-injury weighting. The weight parameters were identified empirically, by meaning that we adjusted the weights for both the models by running them several times through cross-validation and noticed how they perform on the test data.

Following best practice, the test dataset was not included for any of the data balancing, training, and validation phases of the model development. Missing values in the test dataset were imputed by using the corresponding imputation values from the training data. Table 5 (in the 3.3 Results Section) provides a summary of the results, describing the machine learning algorithms employed, the pre-processing techniques for each employed algorithm, along with evaluation metrics. The two machine learning models were compared with two baseline models: Baseline 1 predicted the most frequent class (i.e., the non-injury datapoints); Baseline 2 randomly predicted the class (i.e., injury or non-injury) by respecting the distribution of the classes. Below Table 6 details the different hyperparameter settings and working architectures for both (XGBoost and ANN) algorithms.

Table 5 Training Algorithm Hyperparameter Settings and Architecture

Machine Learning Algorithm	Hyperparameter setting and architecture
XGBoost	Objective: binary (logistic) colsample_bytree: 0.9 learning rate: 0.09 maximum depth: 3 alpha: 5 gamma: 5 evaluation metric: error
Artificial Neural Network	Input layer: 106, Hidden layer 1: 200, Dropout: 0.5, Hidden layer 2: 100, Dropout: 0.5, Output layer: 1 Activation function for hidden layer 1, 2: Rectified Linear Unit (RELU) Activation function for output layer: Sigmoid Kernel initializer for input layer: Glorot Uniform Optimizer: ADAM Loss function: Binary crossentropy Learning rate: initial learning rate 0.0001 with an exponential decay rate 0.96 Epochs: 100 Batch size: 128

Note. Above, the hyperparameters that used in our chapter for each used algorithm is presented. In Section 2.3, we described how we came up with these specific hyperparameters for both the algorithms. These hyperparameters are not absolute and may vary according to data used in other studies.

3.3 Results

A model with XGBoost correctly predicted 13 of 14 injuries in Month 1, as well as 8 of 9 injuries in Month 3, but predicted just 5 of 14 injuries in Month 2. A model with ANN correctly predicted 11 of 14 injuries in Month 1, as well as 8 of 9 injuries in Month 3, but also predicted 9 of 14 injuries in Month 2. The latter model with ANN improved the precision and recall for injuries and non-injuries during cross-validation as across a combined value for Month 1, Month 2, and Month 3. The baseline models (i.e., Baseline 1 and 2) demonstrated AUC of 0.50, which demonstrates that they are in effect random models. The baseline models failed to predict injury. Thus, the results provided by both XGBoost and ANN represent a significant improvement when compared with the baseline models.

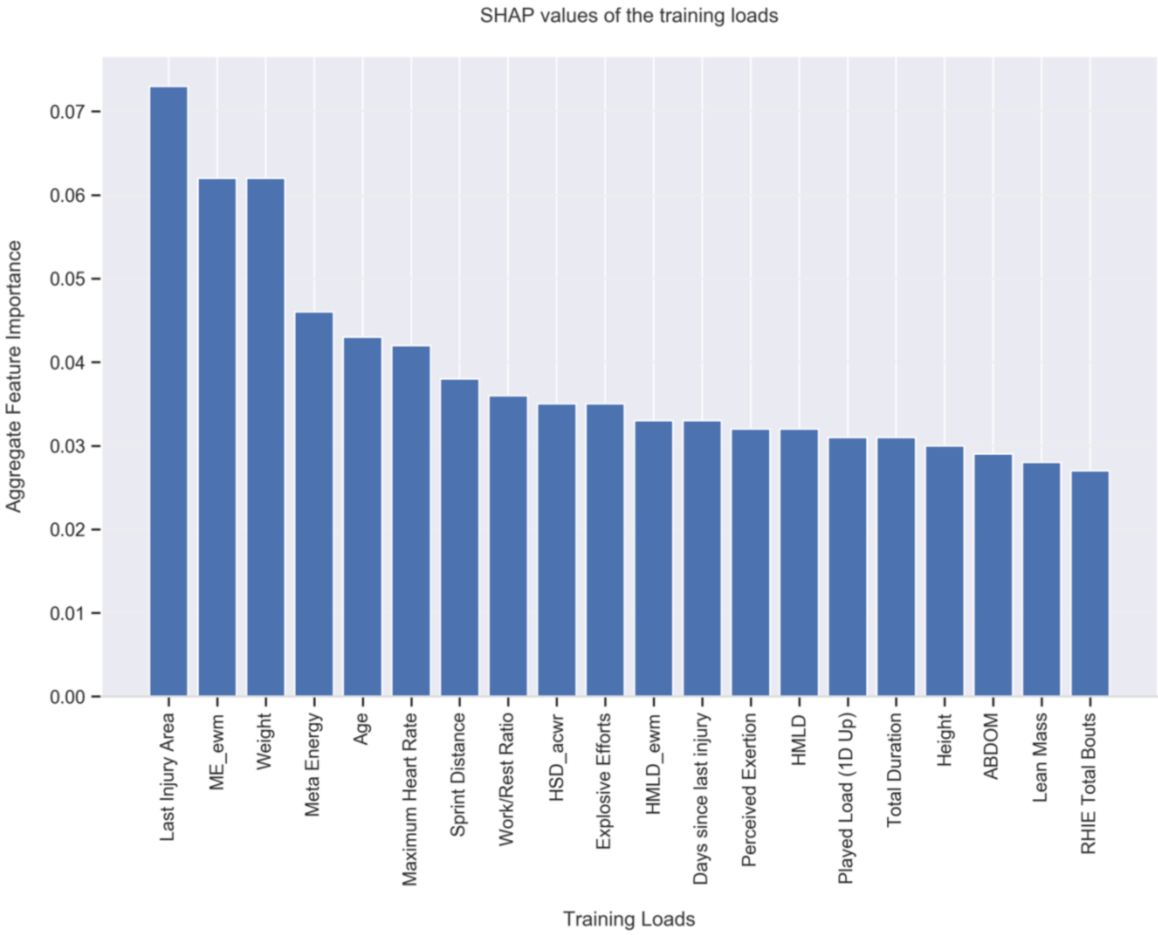
Table 6 Model Fit for the Best-Fitting Model from each Analysis

Machine learning algorithms, pre-processing technique(s)	Model evaluation	Non-injury and injury	Precision (%)	Recall (%)	AUC	Confusion matrix TN / FN	FP / TP
Algorithm 1:	Cross-validation (Training data)	Non-injury	0.99±0.00	0.72±0.02	0.74±0.04	N/A	
XGBoost		Injury	0.09±0.01	0.76±0.08			
Pre-processing:	Month 1	Non-injury	0.99	0.54	0.73	87	74
Oversample: SMOTE		Injury	0.15	0.93		1	13
	Month 2	Non-injury	0.92	0.61	0.48	99	63
Class weight:		Injury	0.07	0.36		9	5
{non injury: 1, injury: 9}	Month 3	Non-injury	0.99	0.55	0.72	93	77
		Injury	0.09	0.89		1	8
	Month 1 + Month 2 + Month 3	Non-injury	0.97	0.57	0.64	279	214
		Injury	0.10	0.73		11	26
Algorithm 2:	Cross-validation (Training data)	Non-injury	0.99±0.00	0.74±0.03	0.80±0.02	N/A	
Artificial Neural Network		Injury	0.10±0.01	0.86±0.04			
	Month 1	Non-injury	0.97	0.58	0.69	96	65
Pre-processing:		Injury	0.14	0.79		3	11
Feature scaling: Min max scaler with feature range (0.01, 0.99)	Month 2	Non-injury	0.95	0.60	0.62	98	64
		Injury	0.12	0.64		5	9
	Month 3	Non-injury	0.99	0.64	0.77	99	64
Oversample: SMOTE		Injury	0.12	0.89		1	8
Class weight:	Month 1 + Month 2 + Month 3	Non-injury	0.97	0.61	0.69	300	193
{non injury: 1, injury: 11}		Injury	0.13	0.77		9	28
Baseline 1 (most frequent) *	Cross-validation (Training data)	Non-injury	.97	1.00	0.50	N/A	
		Injury	0.00	0.00			
Baseline 2 (stratified) *	Cross-validation (Training data)	Non-injury	0.97	0.97	0.50	N/A	
		Injury	0.03	0.03			

Note. Each model was run 1000 times during cross-validation with stratified sampling to check model stability.
* We have not provided evaluation metrics for these two baseline models in month 1, 2, and 3, because they correctly predicted non-injuries only (i.e., they failed to predict any injuries).

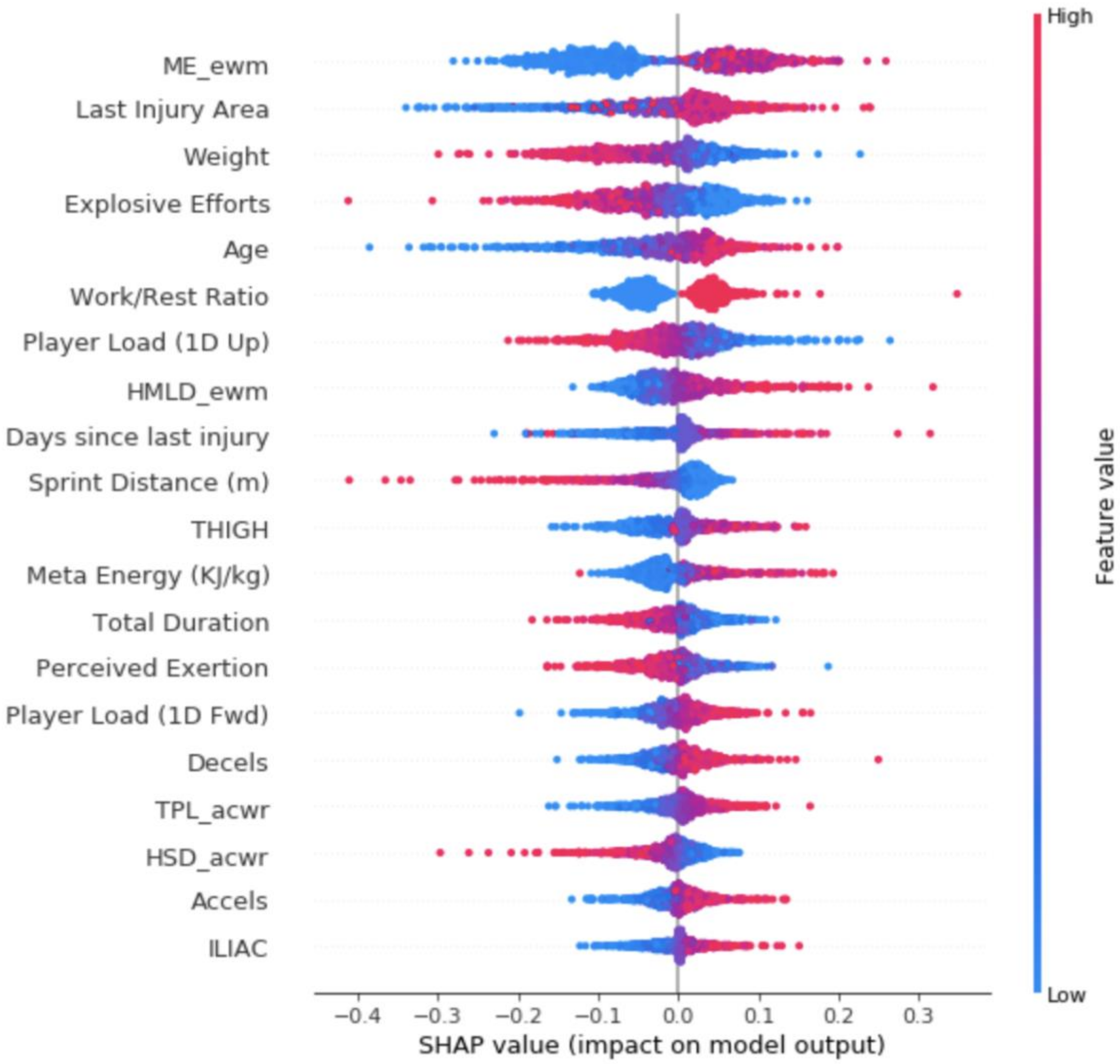
To interpret and visualize the output of each model and the contribution of each feature (i.e., training load) towards the model we used the Shaply Additive Explanations (SHAP) approach (Lundberg and Lee, 2017)—see Figure 3 and Figure 4. Higher SHAP values denote a higher contribution for that training load towards the model's prediction. Given the relatively improved model, when using ANN over XGBoost, the following SHAP explanations relate to the model with ANN. With this in mind, the five most important features for injury risk in the train and validation data appear to be as follows: last injury area; exponential weighted moving average of meta energy; weight; meta energy; and age. We also used SHAP to examine the key features for injury risk at Months 1, 2, and 3 predicted by our trained ANN model (see Figures 5-7). The two most important features that appeared in all models were last injury area and weight.

Figure 3 Top 20 Features According to SHAP Values in The Training and Validation Data



Note: The variables in the model are listed from relatively the most important (left) to the least (right) important by their average global impact on the model. Each bar shows the mean absolute SHAP value for each variable, the higher the value, the higher the importance on the classification model (i.e., a higher probability of a positive prediction which is injury). The same applies for the Fig. 4, 5 and 6 as well.

Figure 4 Distribution of SHAP values for top features in The Training and Validation Data



Note: The variables in the model are listed from relatively the most important (top) to the least (bottom) important by their average global impact on the model. Each dot represents the SHAP value of an individual sample in the dataset which is plotted horizontally next to the feature name. We get an estimation of the distribution of the SHAP values per variable, saying that the higher the absolute value the higher the importance on the model prediction also, positive SHAP values represent a higher probability of a positive prediction (i.e., Injury).

Figure 5 Top 20 Features According to SHAP Values for Month 1

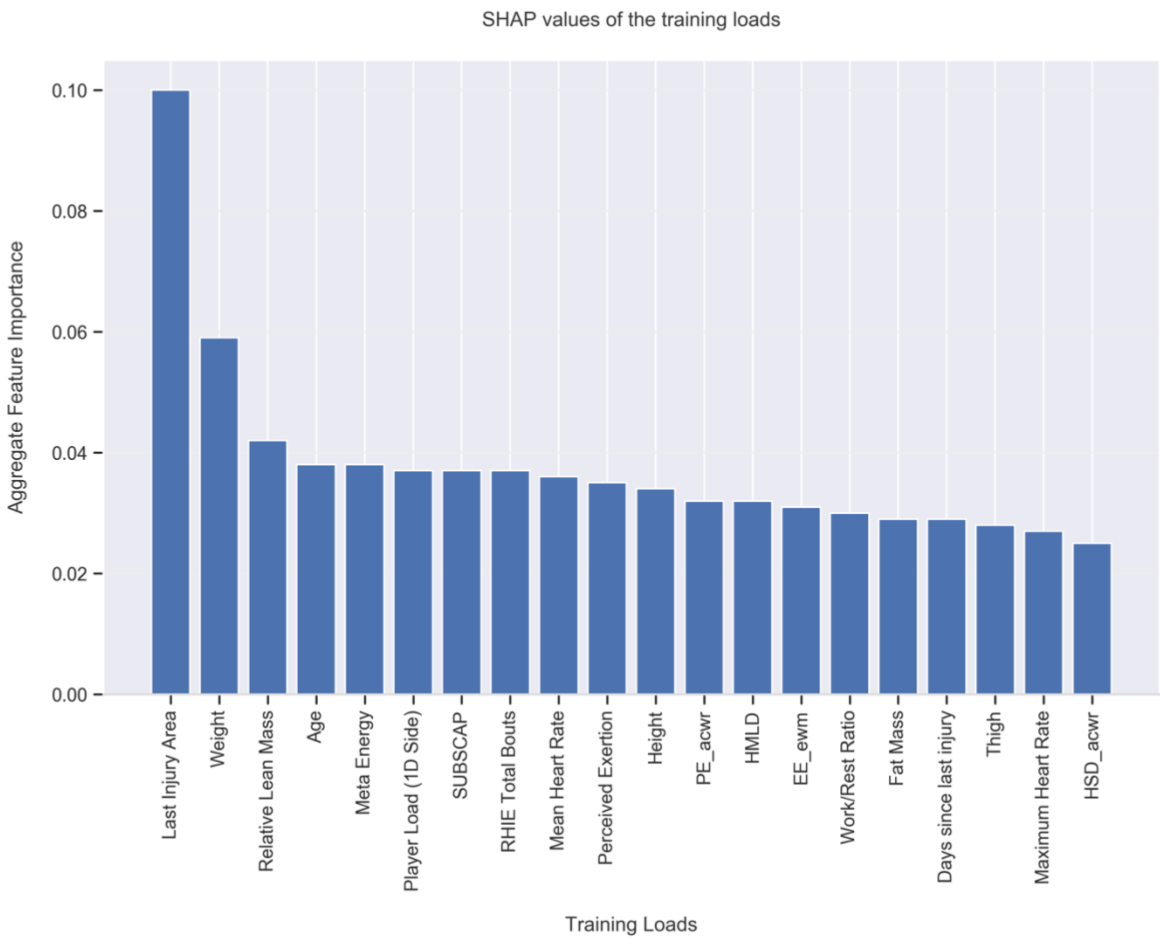


Figure 6 Top 20 Features According to SHAP Values for Month 2

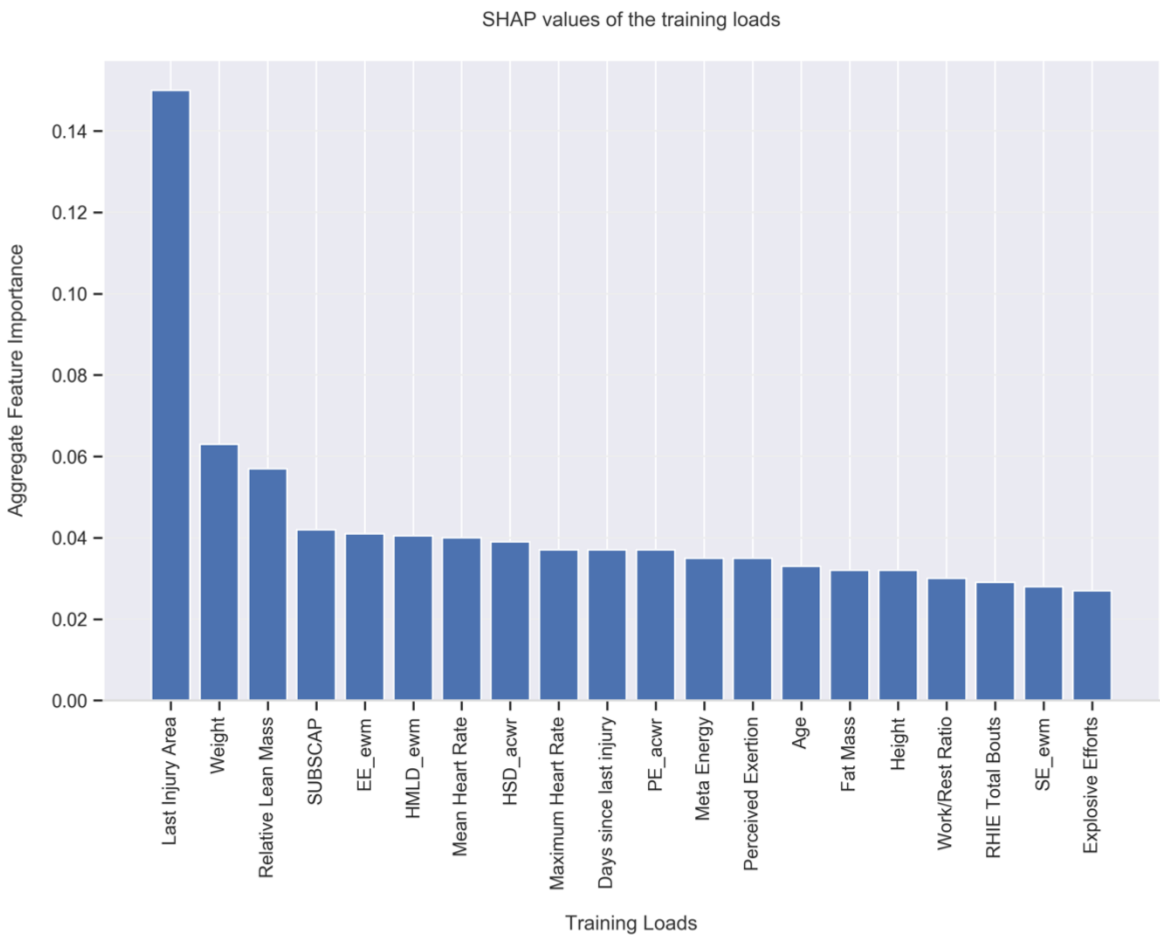
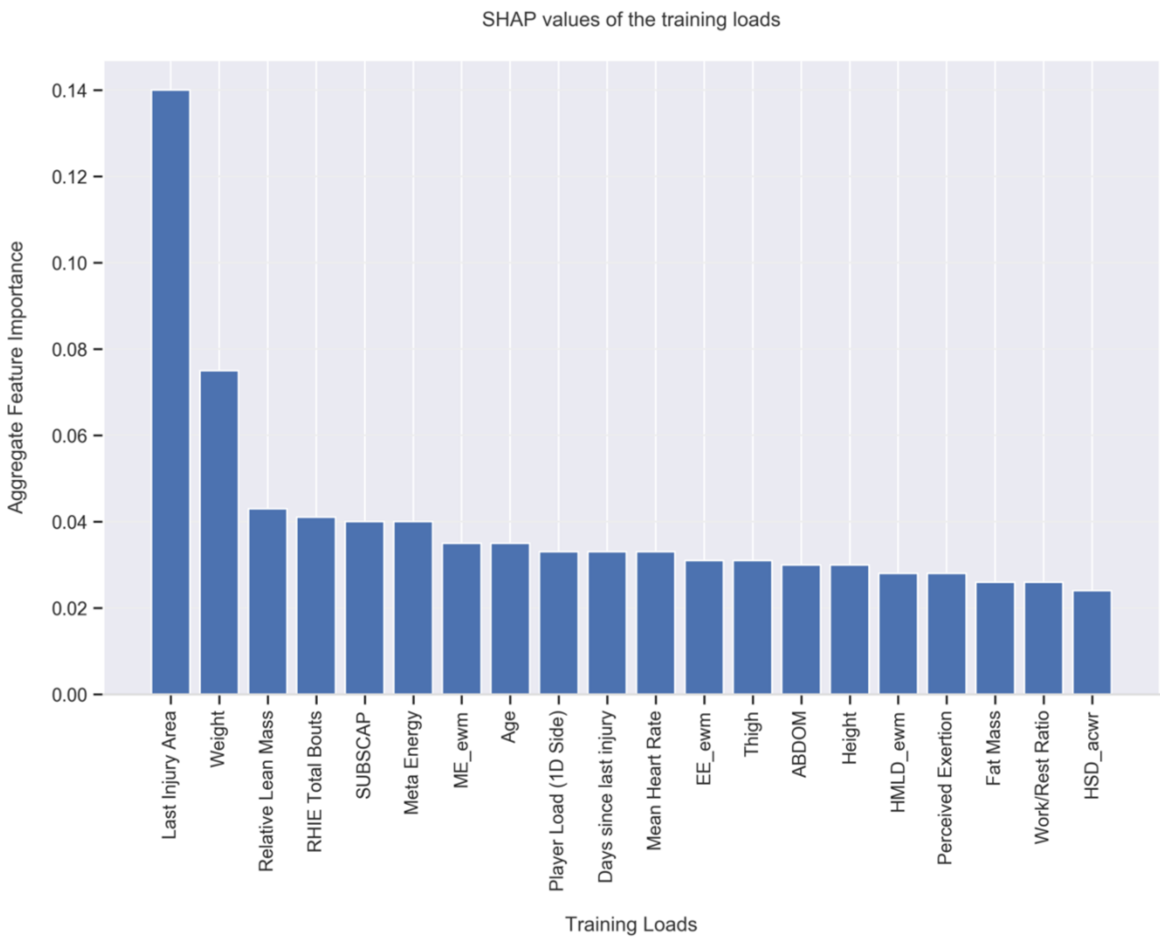


Figure 7 The Top 20 Features According to SHAP Values for Month 3



3.4 Discussion

The purpose of this chapter was to use machine learning to examine the relationship between training load and soccer injury with a multi-season dataset from one English Premier League club. Our results demonstrated that two algorithms (XGBoost and ANN) provided the best results. Correctly predicting 26 of 37 injuries, XGBoost produced a precision value of 10% and recall of 73%; correctly predicting 28 of 37 injuries, ANN produced a precision value of 13% and recall of 77%. For the latter relatively better model using ANN, the most important features contributing to injury were "last injury area" and "weight". Thus, although precision (i.e., the ratio of correctly predicted injuries to the total number of correctly and incorrectly predicted injuries) was relatively low (meaning that many of the model's predicted injuries were not in fact injuries), values for recall (the ratio of correctly predicted injuries to the total observed injuries) were relatively high, suggesting precision suffered at the expense of being able to accurately predict most of the actual injury cases. If this model were used in an applied setting, the "false alarms" (those non-injuries that were predicted as injuries) might lead to some players being unnecessarily rested from training; at the same time, however, the model's correctly predicted injuries would lead to most genuinely at-risk players rightfully being rested, and thereby saving players from injury and the club from losing players to injury, with the concomitant selection problems, rehabilitation time, and financial impact. Finally, the ANN model produced low false negatives, suggesting that if the model predicts that a player will not be injured, this is likely to be the case.

Injury prediction perspectives. Our chapter used a very high dimensional and highly imbalanced, overlapped dataset. Although ANN has been successfully employed to deal with such high dimensional, overlapped datasets in other fields of artificial intelligence (such as in object detection, image recognition, speech recognition, text processing, recommendation systems, and time series model building: Bohr and Memarzadeh, 2020; Emmert-Streib et al., 2020; Johnson and Khoshgoftaar, 2019), it has never been used for soccer injury prediction. In the present chapter, ANN out-performed “state-of-the-art” XGBoost, with better recall and precision values. In attempting to counter class imbalance in the present chapter’s dataset, data oversampling (i.e., Smote), in combination with setting the weights for injury at nine (for XGBoost) and eleven (for ANN) times higher than for non-injury (termed a cost-sensitive classification), we were able to maximize the accurate prediction of injuries.

Injury prediction is based on analysis of longitudinal data, with the goal of being able to accurately predict injuries in some pre-defined upcoming period of days. Thus, in order to ensure the independence of test and train data, in addition to the usual cross-validation, we also evaluated our models on unseen future (test) data. In terms of data pre-processing, differently from Lopez-Valenciano et al. (2018) and Ayala et al. (2019) who imputed missing values using the mean, we used different imputation techniques for different types of training loads. For example, some physical training load variables (such as weight and body fat percentage) are not measured on a daily basis, even though they naturally increase or decrease gradually over time. Imputing the missing values of these features by using the mean or the most frequent values may not reflect well the actual values over time. To combat this potential inaccuracy, we used interpolation for imputing the missing values of those time-dependent features. In a similar way, to better replicate the most practical and reasonable values with our GPS measures, ACWR, MSWR, and EWMA, we imputed missing values using k-nearest neighbor or weekly mean values.

Explainability. Compared with white-box machine learning models, “black-box” models, like those examined in the present research (i.e., XGBoost and ANN), can provide better predictive performance, but at the expense of being difficult to interpret and understand. With black-box models, then, additional post hoc methods are needed to interpret and understand results (Loyola-Gonzalez, 2019). Thus, in terms of the explainability of our model, we present (based on SHAP explanations) the important features (i.e., training loads) in Figures 3-6. Last injury area was a key feature in the ANN model—37% of injured players had a previous record of thigh injury, 30% had a previous record of knee injury, 16% had a previous record of lower leg injury, and 17% did not have any previous record of injury. Further, 84% of injuries occurred in those with body weights between 73kg and 85kg. It is worthy of note that, despite the power of SHAP explanations, the output from such global explanations can sometimes be misleading. For example, in our main dataset 122 of 133 injuries, and in our dataset D 334 out of 378 injuries, occurred when the exponential weighted moving average (EWMA) of Meta Energy exceeded a value of 6.14. On the contrary, with our test data (i.e., those data not included in the training and validation dataset), of 530 data points, there were no data points for which EWMA of Meta Energy exceeded a value of 6.14. Thus, although (see Figure 3) EWMA of Meta Energy was one of the top three features in the training and validation data, it failed to emerge as an important feature in Months 1-3 (as can be seen in Figures 4-6). What this means is that, despite its apparent importance during training and validation, EWMA of Meta Energy plays no major role in terms of explainability of the test data. Building from the above, if we had divided our entire dataset on a 10% train-test split basis (rather than using our process of testing on later data), we would have likely concluded that EWMA of Meta Energy plays a more prominent role in terms of explainability than it actually does in real life. Finally, it is also worth noting that in our data, values for the ACWR (the most well-researched model of injury monitoring in soccer) appeared to differ from those noted in the existing literature. Thus, in contrast with research demonstrating, for example, that values in excess of 2 (Bowen et al., 2019) or less than 1 (Rossi et al., 2018) might lead to greater injury risk, the majority of injuries in our data occurred when ACWR values were between 0.5 and 1.5.

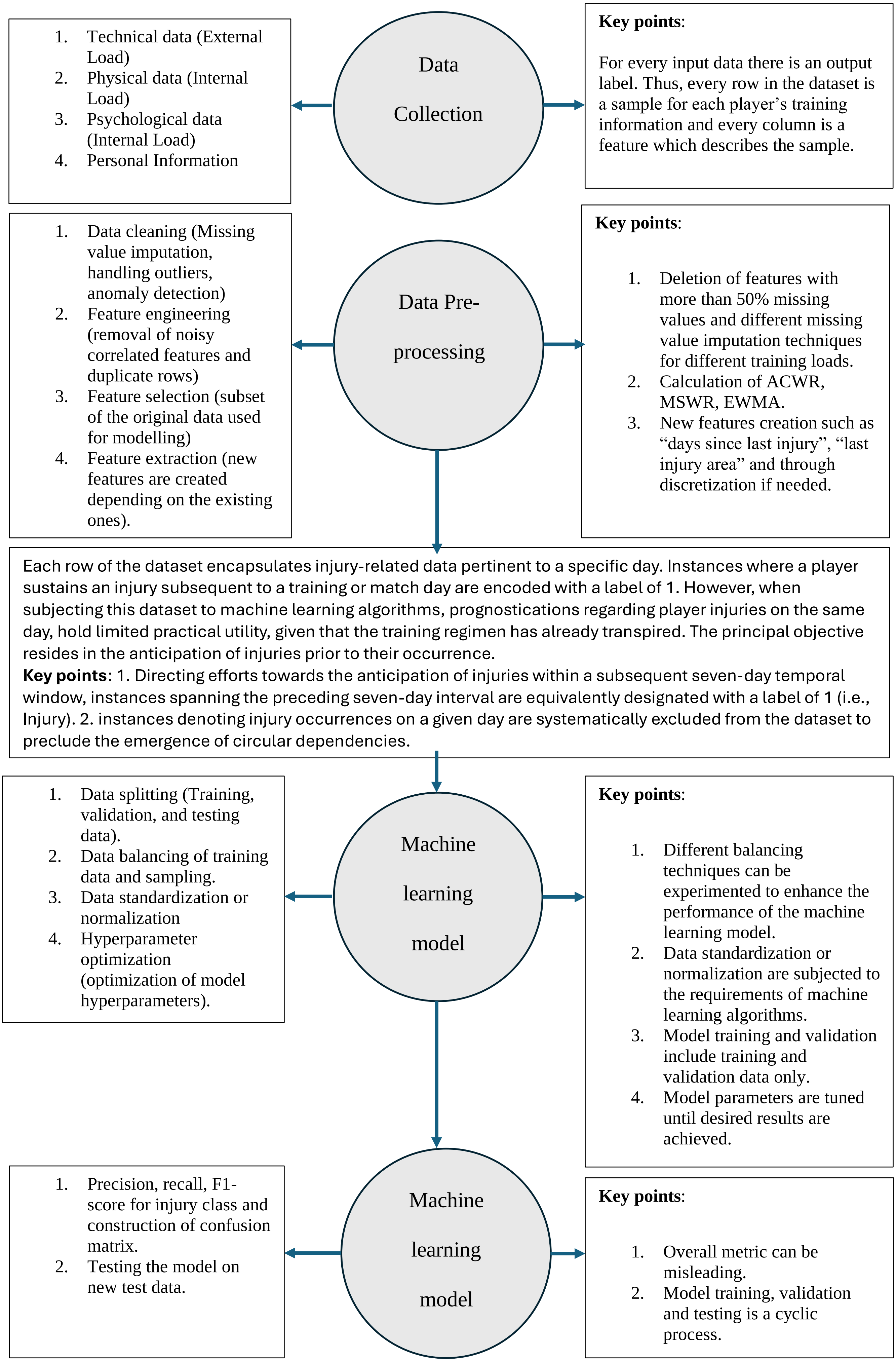
Practical applications. The models developed in this chapter could be used by clubs and practitioners to calculate the probability of a player getting injured in the next seven days. With the use of explainability (via SHAP), practitioners would also be well-positioned to have an essence of the cause of injuries predicted by the models. The results from the present chapter cannot be directly compared with other studies into soccer injury, because, unlike those studies, we used a multi-season dataset with a very high imbalance ratio. However, in seeking to make comparisons, we reproduced as closely as possible, with our data, the analysis strategy from two other well-regarded soccer injury studies—the work of Rossi et al. (2018) and Vallance et al. (2020). In attempting to predict injuries in the next day and in the next seven-day window, we used all the possible similar features (i.e., the training load variables) from Rossi et al. (2018) and Vallance et al. (2020) that were also available in our data, and followed their methods with regard to data pre-processing, feature selection, feature extraction, balancing techniques, model training, hyperparameter optimization, along with the model evaluation and validation techniques, where specified. In reproducing the work of Rossi et al. (2018), we used the Recursive Feature Elimination (Guyon et al., 2002) as the feature selection technique which yielded just one feature in our data, and the prediction based on that one feature was not as high as the results reported by Rossi at el. (2018). In reproducing the work of Vallance et al. (2020), we used the Bayesian optimization hyperparameter technique with our data, which predicted most of the non-injuries. Rossi et al. (2018) reported as their best algorithm Decision Tree, and Vallance et al. (2020) reported as their best algorithms k-nearest neighbors, random forest, decision tree, and XGBoost—conversely, only XGBoost performed well with our data. It is important to note that the differences we noted in our data are completely normal, and should be expected. All clubs have different philosophies and unique ways of handling their training load data. As a result, the number of training loads used and the training programs employed at clubs are frequently quite different. And thus, the choice of the best performing machine learning algorithm for each dataset is likely dependent on the context and the quality of those data.

Strengths and limitations. The present research had some notable strengths. First, to the best of our knowledge this is the first chapter that has considered multi-season data with elite male soccer players from the English Premier League. Second, we included many types of training loads, including GPS measures, physical and psychological loads, personal information, as well as ACWR, MSWR, and EWMA of certain training load variables (See Table 4). Third, we also created features—such as last injury area and days since the last injury—which appeared to enhance the predictive utility of our machine learning model and were among the most important injury predictors. Fourth, the proposed seven-day injury prediction window is unique to our chapter—and aligned well with the notion that English Premier League are generally played every seven days. Fifth, our use of ANN was a novel addition, which appeared marginally more effective than the state-of-the-art XGBoost in predicting injuries. All the above led us to conclude that the most important features in our chapter were "last injury area" and "weight", which are very general—these two features are monitored in almost every sporting organization to evaluate injury risk among players, and thus in practical terms the present research has genuinely real-world application. Against the backdrop of these many strengths, a major limitation for the process used in the present research (as is true for many machine learning processes) is that when new data are available, the model would have to be retrained, and thus the predictions may then vary. That said, given that we were able to demonstrate that machine learning models trained on a highly multi-dimensional and imbalanced dataset can indeed predict and explain injuries to address the needs of a professional soccer club, different clubs and organizations could use our approach with amendments to the feature set as required.

Future Research. As noted above, a limitation of the present research is the need to retrain the models when new data become available. Thus, a future research avenue could be to develop automation of the model training process with continuously incoming injury data, so that the models adapt to this new information. This would seem particularly important in soccer, wherein changes in training processes, team members, and injuries mean that the underlying distribution of the data does not remain constant across seasons. We believe that this limitation could be addressed by using adaptive streaming predictive methods (Yang, Manias and Shami, 2021), and we encourage future research to examine this further.

3.5 Conclusions

Using a highly imbalanced and high dimensional, overlapped, multi-season dataset from an English Premier League soccer club, we were able to predict soccer injuries with high recall. Our novel use of ANN in combination with explainable artificial intelligence also demonstrated its potential to unearth effective insights into the workload-injury relationship. Our data pre-processing techniques such as unique missing value imputation techniques, new features creation, handling of the high imbalance in non-injuries and injuries, train-validation process alongside testing of models on real-life in-coming data, and improving recall and precision techniques all have potential to lay the foundation for future research to employ machine learning in a more practical way to predict injuries.



Data Collection
1. Technical data (External Load)
2. Physical data (Internal Load)
3. Psychological data (Internal Load)
4. Personal Information
Key points:
For every input data there is an output label. Thus, every row in the dataset is a sample for each player's training information and every column is a feature which describes the sample.
Data Pre-processing
1. Data cleaning (Missing value imputation, handling outliers, anomaly detection)
2. Feature engineering (removal of noisy correlated features and duplicate rows)
3. Feature selection (subset of the original data used for modelling)
4. Feature extraction (new features are created depending on the existing ones).
Key points:
1. Deletion of features with more than 50% missing values and different missing value imputation techniques for different training loads.
2. Calculation of ACWR, MSWR, EWMA.
3. New features creation such as "days since last injury", "last injury area" and through discretization if needed.
Each row of the dataset encapsulates injury-related data pertinent to a specific day. Instances where a player sustains an injury subsequent to a training or match day are encoded with a label of 1. However, when subjecting this dataset to machine learning algorithms, prognostications regarding player injuries on the same day, hold limited practical utility, given that the training regimen has already transpired. The principal objective resides in the anticipation of injuries prior to their occurrence.
Key points: 1. Directing efforts towards the anticipation of injuries within a subsequent seven-day temporal window, instances spanning the preceding seven-day interval are equivalently designated with a label of 1 (i.e., Injury). 2. instances denoting injury occurrences on a given day are systematically excluded from the dataset to preclude the emergence of circular dependencies.
Machine learning model
1. Data splitting (Training, validation, and testing data).
2. Data balancing of training data and sampling.
3. Data standardization or normalization
4. Hyperparameter optimization (optimization of model hyperparameters).
Key points:
1. Different balancing techniques can be experimented to enhance the performance of the machine learning model.
2. Data standardization or normalization are subjected to the requirements of machine learning algorithms.
3. Model training and validation include training and validation data only.
4. Model parameters are tuned until desired results are achieved.
Machine learning model
1. Precision, recall, F1-score for injury class and construction of confusion matrix.
2. Testing the model on new test data.
Key points:
1. Overall metric can be misleading.
2. Model training, validation and testing is a cyclic process.

Chapter 4

4. Abstract

This chapter aimed to enhance soccer injury prediction by employing adaptive machine learning methodologies, including online continual learning and drift detection, applied to a multi-season dataset of Elite Premier League players. The dataset comprised 35 male professional soccer players (aged 25.79±3.75 years, range 18–37 years; height 1.80±0.07 m, range 1.63–1.95 m; weight 80.70±6.78 kg, range 66.03–93.70 kg) over five seasons (2014–2019). Data included 106 training load variables, spanning GPS metrics, personal information, physical and psychological data, and derived workload ratios (ACWR, MSWR, and EWMA), analysed in relation to 133 non-contact injuries (imbalance ratio: 0.013). Logistic Regression, AdaBoost, and Artificial Neural Networks were implemented across static, continual, and adaptive learning contexts. Techniques such as Synthetic Minority Oversampling Technique (SMOTE) addressed data imbalance. Performance was evaluated using precision, recall, ROC-AUC, Cohen's kappa, and confusion matrices. Adaptive learning models with drift detection consistently outperformed static and continual learning approaches, particularly in injury prediction for minority classes. Sliding window retraining, focusing on recent data, achieved the best overall performance, with high ROC-AUC and Cohen's kappa scores. Cumulative training, integrating multi-season data, further enhanced the models' adaptability and predictive accuracy across all classifiers. Neural networks demonstrated superior performance compared to traditional algorithms, especially in dynamic scenarios. This chapter highlights the superiority of adaptive machine learning methodologies, particularly drift detection and sliding window retraining, in addressing the dynamic and evolving nature of injury prediction. Cumulative training emerged as a critical factor in improving model performance. These findings offer practical applications for injury prevention and player well-being management

in professional soccer. Future research should focus on integrating real-time data from wearable devices and exploring advanced adaptive learning frameworks for enhanced predictive capabilities.

4.1 Introduction

The contemporary landscape of sports science highlights the critical importance of monitoring athletes' training and competition loads, a field that has garnered significant attention in recent years (Kalkhoven et al., 2021). Recognizing the pivotal role of load monitoring, professional sports organizations allocate substantial resources to optimize training adaptations, assess fatigue and recovery dynamics, and mitigate the risks of injury and illness (Kalkhoven et al., 2021; Halson, 2014). Soccer, as the world's most popular sport, involves an extensive player base of 200,000 professionals and 240 million amateurs (Rahnama, 2011; Owoeye et al., 2020; Jones et al., 2019). However, this widespread participation is accompanied by a notably high incidence of injuries compared to other sports (Rahnama, 2011; Owoeye et al., 2020; Jones et al., 2019). Consequently, soccer has become a focal point for research on load monitoring and injury prevention. Soccer-related injuries can result in prolonged player absences, negatively impacting team performance and incurring significant financial costs. For instance, injuries in the English Premier League (EPL) led to an estimated expenditure of approximately £45 million per season between 2012-2013 and 2016-2017 (Eliakim et al., 2020).

To better understand the intricate relationship between training loads and soccer injuries, recent research has increasingly adopted machine learning (ML) techniques (Majumdar et al., 2022). These methods provide a novel analytical framework, enriching the understanding of the complex interplay between training loads and injury risk. Historically, soccer injury analysis has heavily relied on the Acute: Chronic Workload Ratio (ACWR) (Majumdar et al., 2024), a widely used metric that has faced methodological critiques and inconsistent results. Machine learning approaches offer a more nuanced perspective by incorporating a broader range of explanatory variables, including external load metrics (e.g., GPS-derived features), physical and psychological factors, and personal player information (Hulin et al., 2013; Impellizzeri et al., 2020).

Recent ML studies have sought to overcome the limitations of traditional models by leveraging multiple explanatory variables and comprehensive datasets (Bowen et al., 2019; Rossi et al., 2018; Vallance et al., 2020; Naglah et al., 2018; López-Valenciano et al., 2018; Ayala et al., 2019; Rommers et al., 2020; Oliver et al., 2020; Venturelli et al., 2011; Hecksteden et al., 2022; Piłka et al., 2023). Although promising, existing ML research in soccer injury prediction often lacks clarity in evaluation metrics (e.g., per-class versus averaged metrics), effective pre-processing techniques, and longitudinal testing frameworks (Majumdar et al., 2022). These gaps have been addressed in prior research by Majumdar et al. (2024) but significant challenges remain. One such challenge is the high prevalence of false positives, where non-injuries are incorrectly classified as injuries, leading to unnecessary interruptions in training and unwarranted player rest periods (Majumdar et al., 2024).

The root of this issue lies in two key factors: (1) the pronounced class imbalance in injury prediction datasets and (2) the continuous evolution of the data's statistical properties over time, a phenomenon referred to as "concept drift" (Robles-Palazón et al., 2023). While data balancing techniques such as the Synthetic Minority Oversampling Technique (SMOTE) can address class imbalance, traditional ML models that rely on static historical data often struggle to adapt to dynamic shifts in data distribution. Consequently, these models fail to capture the evolving nature of soccer, characterized by changes in coaching strategies, player rosters, and training regimens, which can lead to week-to-week and season-to-season variability in data quality and distribution (Majumdar et al., 2022).

Concept drift, defined as the temporal alteration of the underlying data distribution, poses a critical challenge in the deployment of predictive models. It undermines the foundational assumption of static data, rendering pre-trained models obsolete or suboptimal as data evolves

(Janardan and Mehta, 2017; Lu et al., 2018; Webb et al., 2017). This phenomenon necessitates continuous model updates to maintain performance over time (Goel and Batra, 2021; Hussain et al., 2021; Wang et al., 2013; Disabato and Roveri, 2022; Zenisek et al., 2019). Addressing concept drift has led to the development of various strategies, including ensemble methods, online learning algorithms, adaptive modeling techniques, and advanced data pre-processing approaches (Janardan and Mehta, 2017; Lu et al., 2018; Webb et al., 2017; Goel and Batra, 2021; Hussain et al., 2021; Wang et al., 2013; Disabato and Roveri, 2022; Zenisek et al., 2019). Despite these advancements, a key limitation in current injury prediction systems is the need for frequent manual retraining. Automating model adaptation to incorporate new injury data remains an essential objective, particularly in soccer, where dynamic factors such as changing training practices and team compositions further exacerbate concept drift.

Continual learning, also known as lifelong or incremental learning, offers a promising solution to these challenges. This paradigm enables models to acquire and retain knowledge over time, adapting to evolving tasks and data distributions (Gomes et al., 2019; Lee and Lee, 2020; Gao and Lei, 2017; Sudharsan et al., 2021; Wang and Wang, 2023; Zenisek et al., 2019; Krawczyk et al., 2017). Unlike traditional machine learning, which typically addresses static, isolated tasks, continual learning is designed to operate in dynamic environments. Applications of this approach extend across various domains, including robotics, natural language processing, recommendation systems, and medical diagnostics (Lee and Lee, 2020). Continual learning frameworks emphasize adaptability, efficiency, and the ability to mitigate catastrophic forgetting, ensuring that new information is integrated without compromising previously learned knowledge. By incrementally updating models, continual learning reduces computational costs and improves resource efficiency, making it a vital component of modern AI systems.

In this chapter, we present two distinct paradigms for injury prediction: one utilizing traditional machine learning algorithms, such as Logistic Regression and AdaBoost, and the other leveraging neural network-based approaches. The traditional machine learning paradigm focuses on simplicity and interpretability, offering a reliable framework for identifying injury risks using structured data. In contrast, the neural network-based paradigm emphasizes adaptability and scalability, enabling the model to capture complex, nonlinear relationships within the data.

The first paradigm involves three distinct learning scenarios: Static learning, Continual learning, and Drift Retraining Adaptive learning. These scenarios are further evaluated through non-cumulative and cumulative perspectives to assess their effectiveness in handling evolving data. The second paradigm, based on neural networks, mirrors these methodologies but incorporates backpropagation for immediate updates, sliding window retraining, and adaptive mechanisms for drift detection.

By integrating these paradigms, this chapter aims to evaluate their comparative effectiveness in addressing the challenges posed by concept drift and dynamic injury data. We leverage a unique multi-season dataset encompassing five seasons of training and injury data from Elite Premier League players. The primary utility of these models lies in their ability to evaluate injury risk within a seven-day period, aligning with the regular match schedules of the EPL. To date, the application of online learning in soccer injury prediction remains underexplored. By addressing this gap, our chapter seeks to develop a comprehensive, multidimensional online continual predictive model that advances both academic research and practical applications in sports science.

4.2 Materials and Methods

4.2.1 Participants

This investigation included a cohort of 35 professional male soccer players with an average age of 25.79 years (±3.75), spanning a range from 18 to 37 years. On average, participants were 1.80 meters tall (±0.07) with a height range of 1.63 to 1.95 meters, and weighed 80.70 kilograms (±6.78), ranging from 66.03 to 93.70 kilograms. These athletes were part of a team competing in the English Premier League, and data collection occurred across five consecutive seasons, from 2014-2015 through 2018-2019. Players were categorized by their field positions, including 8 full-backs, 9 centre-backs, 7 central midfielders, 8 wing-forwards, and 3 strikers.

The dataset contained a total of 343 injury records, focusing specifically on 133 non-contact injuries. Injury types were distributed as follows: 43 injuries involved the thigh, 29 affected the knee, 24 targeted the hip, 19 were ankle-related, and 18 involved the lower leg. The frequency of injuries varied among players: 8 players experienced a single injury, 9 sustained two injuries, 4 encountered three injuries, 2 had four injuries, 4 endured five injuries, 2 faced six injuries, 4 incurred seven injuries, 1 experienced eleven injuries, and 1 player suffered sixteen injuries.

The seasonal distribution of injuries showed variability, with 11 injuries recorded during the 2014-2015 season, 6 injuries in the 2015-2016 season (the team's debut in the Premier League), 28 injuries in 2016-2017, 41 in 2017-2018, and 47 in 2018-2019.

4.2.2 Data collection and Feature creation

The dataset utilized in this chapter was obtained from the first-team sports science department of the club and was collected as part of their routine data monitoring procedures, with all necessary permissions secured. The focus of this research was on 133 non-contact injuries documented within the dataset. Notably, absences from training sessions that resulted in unavailable training load data were not classified as missing data.

The dataset comprised various categories of information, including Global Positioning System (GPS) metrics, physical measurements (e.g., skinfold thickness and body fat percentage), psychological parameters (e.g., Rating of Perceived Exertion, RPE), and demographic data. During feature selection, attributes with more than 60% missing values were excluded. Importantly, instances of training absences were not explicitly marked as missing data and were therefore not treated as such. To handle incomplete data, appropriate imputation techniques were employed. Additionally, two novel features were generated: "last injury area" and "days since last injury." Table 4 (In Chapter 3) outlines the training load variables included in the analysis, providing detailed descriptions, data sources, collection methods, collection frequencies (e.g., daily for GPS and psychological data, bi-weekly for physical data), and the imputation strategies used for addressing missing values.

4.2.3 Dataset construction

This chapter develops a multi-dimensional injury prediction framework to estimate the likelihood of injuries within a seven-day period, aligned with the English Premier League's weekly schedule. This timeframe is consistent with previous methodologies, such as those by Vallance et al. (2020). Players typically undergo three to four training sessions per week, with intensity peaking later in the week.

The master dataset consists of 106 variables (See Table 4), including 40 GPS-derived metrics, 6 personal details, 14 physical, 4 psychological, and 42 workload-related variables (e.g., ACWR, MSWR, EWMA). An injury label (1 for injury, 0 for non-injury) was assigned, forming a dataset of 10,653 data points. Of these, 10,520 represented non-injury cases, and 133 indicated injuries, creating a class imbalance (ratio: 0.013).

To address this, injury labels were redefined to mark data points from the seven days preceding an injury as at-risk (label 1), excluding the original injury points. This adjustment produced a refined dataset with 10,142 non-injury cases and 378 at-risk cases, reducing the imbalance ratio to 0.037.

Spanning five seasons, the dataset provides season-wise distributions: Season 1—1448 non-injury, 25 injury; Season 2—1980 non-injury, 17 injury; Season 3—2376 non-injury, 79 injury; Season 4—2242 non-injury, 123 injury; and Season 5—2096 non-injury, 134 injury. These distributions enable detailed multi-season analysis of injury prediction.

4.2.4 Data Drift Analysis

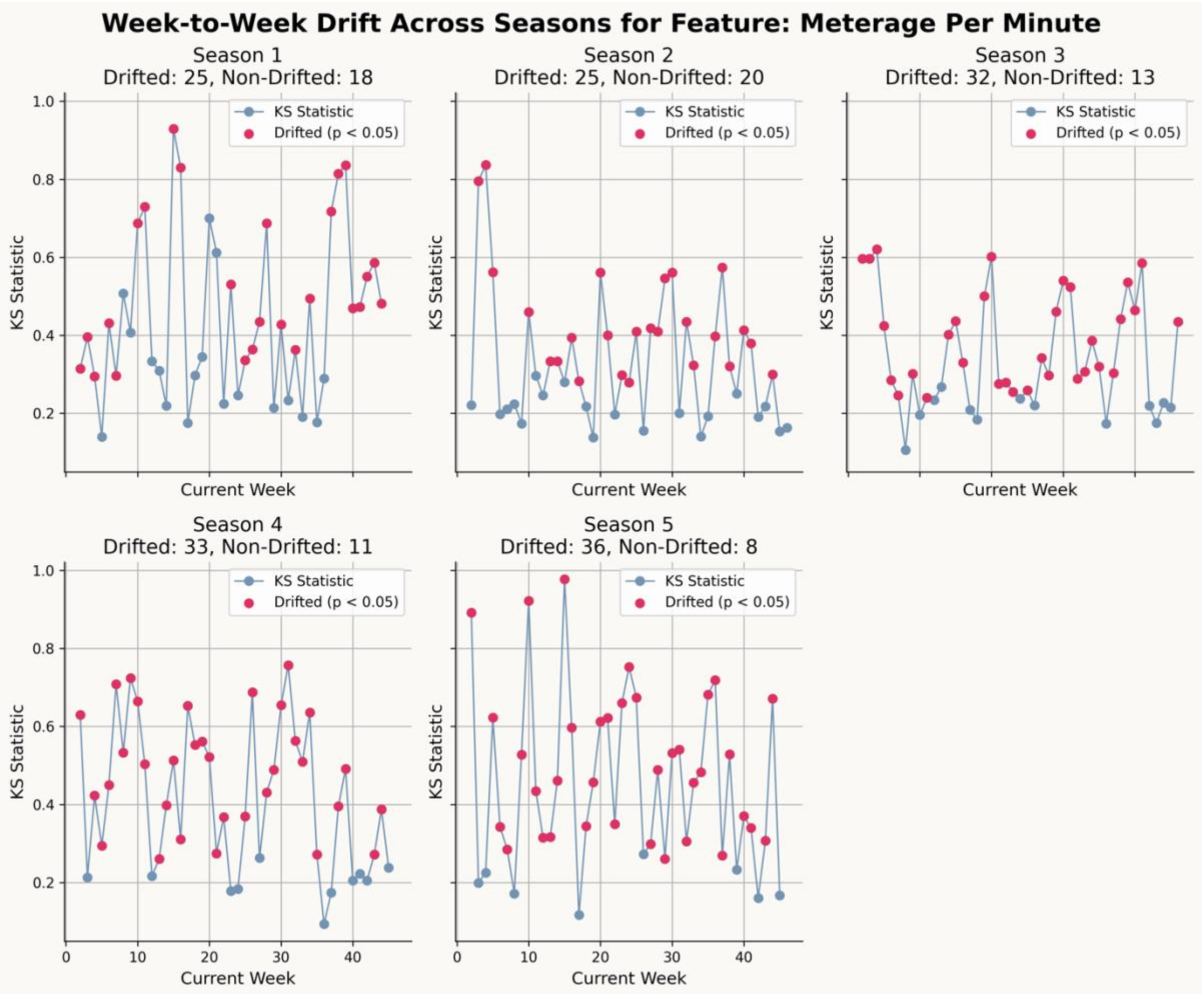
In data-driven systems, maintaining the consistency and reliability of the data distribution over time is crucial for achieving accurate and robust predictive outcomes. However, data drift—a change in the statistical properties of data over time—can compromise these objectives. In this chapter, we investigated data drift across five consecutive seasons using three complementary methodologies: the Kolmogorov-Smirnov (KS) statistic, Principal Component Analysis (PCA), and t-distributed Stochastic Neighbor Embedding (t-SNE). These methods allowed us to quantify, visualize, and interpret the occurrence of data drifts in training load metrics and their potential association with injuries.

Among the numerous features analyzed, we focused on Meterage Per Minute (MPM), which measures the distance covered per minute during activity. This metric is fundamental to training programs as it directly reflects an athlete's intensity and workload efficiency. From a sports science perspective, MPM serves as a critical indicator of an athlete's cardiovascular and muscular endurance. Its variations over time can reveal shifts in fitness levels, recovery states, and adaptation to training loads, making it a key parameter for injury prevention and performance optimization.

4.2.4.1 Week-to-Week Drift Analysis Using KS Statistic

The KS test is a non-parametric method that evaluates the maximum difference between two cumulative distribution functions, allowing us to compare the distributions of a given feature between consecutive weeks. For each feature and season, we calculated the KS statistic and its corresponding p-value to determine whether a significant drift occurred. A p-value below 0.05 was considered indicative of significant drift. Week-to-week KS statistics were plotted for each season, highlighting drifted weeks in red. Additionally, the number of drifted and non-drifted weeks was annotated on each plot (See Figure 8).

Figure 8 Week to Week Drift Visualisation

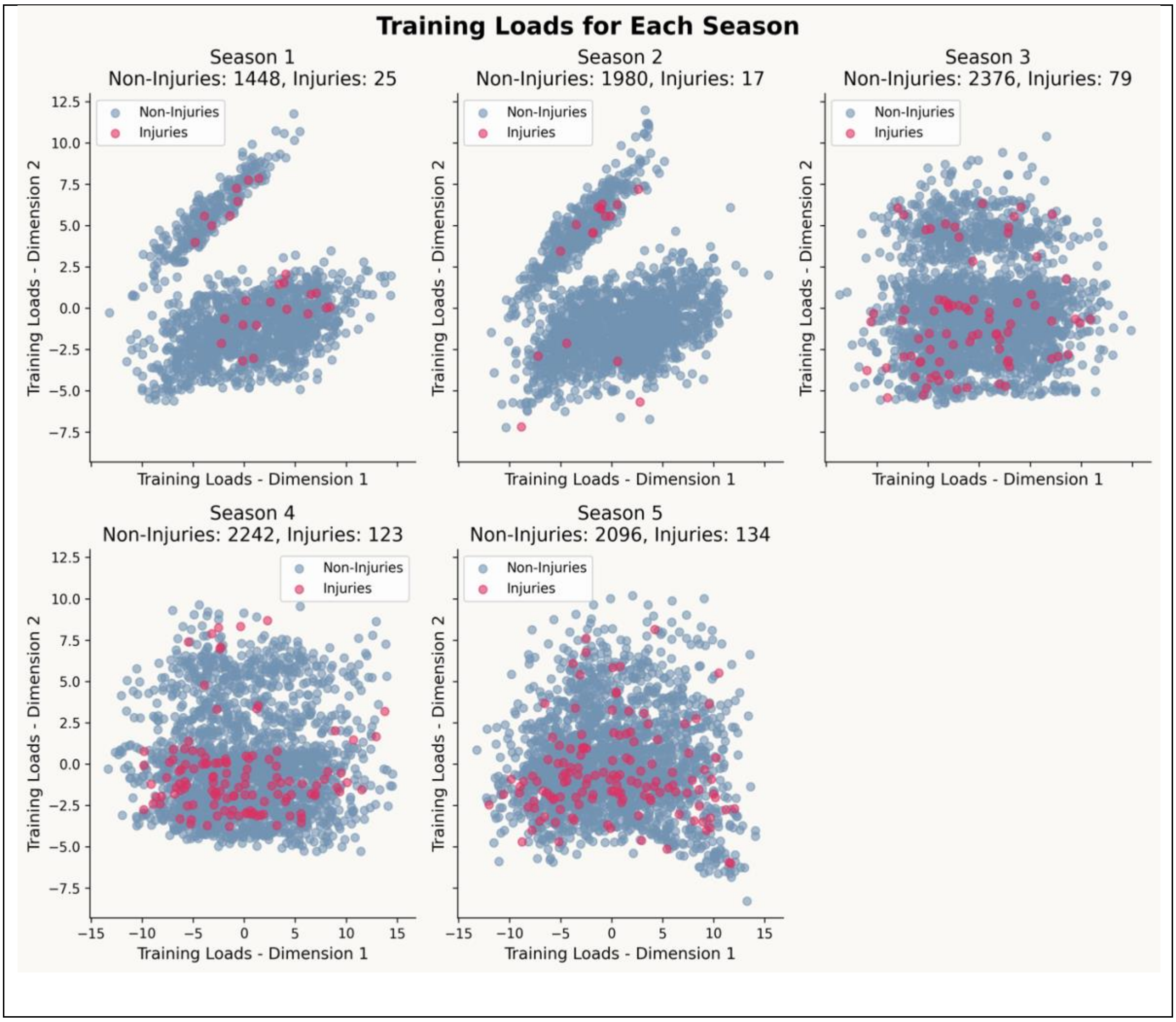


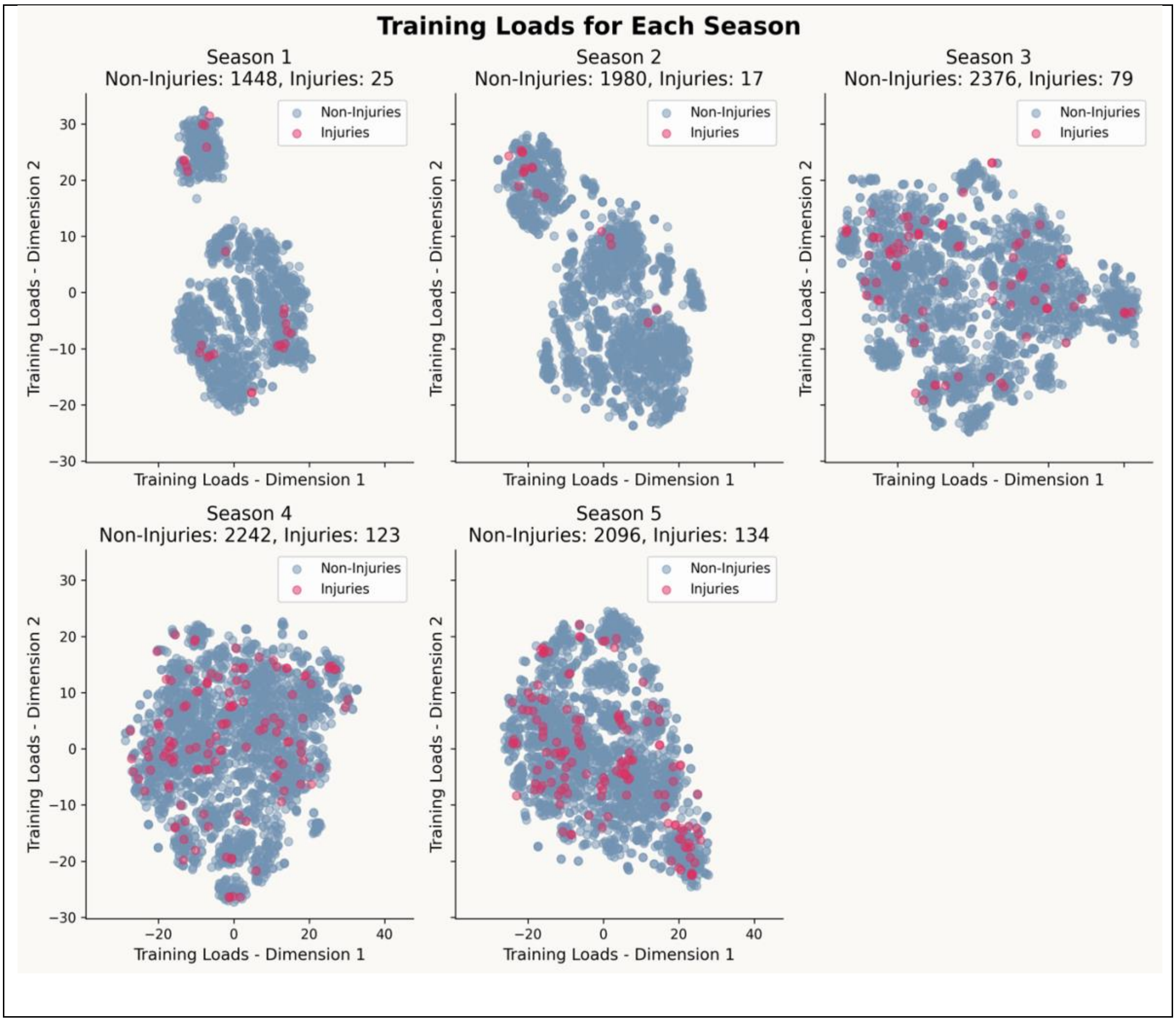
The Kolmogorov-Smirnov (KS) statistic plots across all five seasons indicated multiple instances of significant drift, with the number of drifted weeks varying each season. In Season

1, 25 weeks exhibited drift, while 18 weeks remained stable. Similarly, Season 2 had 25 drifted weeks and 20 non-drifted weeks. The frequency of drift increased in Season 3, with 32 drifted weeks and only 13 stable weeks. This trend continued in Season 4, where 33 weeks experienced drift, leaving just 11 weeks unaffected. The highest occurrence of drift was observed in Season 5, with 36 drifted weeks and only 8 weeks showing no drift. The drifts appeared more frequently in certain periods, potentially aligning with changes in training intensity or match schedules.

4.2.4.2 PCA and t-SNE for Multivariate Drift Visualisation

Figure 9 PCA and t-SNE for Multivariate Drift Visualisation





PCA scatter plots showed distinct clustering of data points within each season, with noticeable variations in the spread and orientation of clusters over time. These changes reflected shifts in the multivariate structure of training loads, suggesting the presence of systemic data drifts. (See Figure 9: Top Part)

The t-SNE visualizations provided a more nuanced view of the data, revealing localized clusters and overlaps between injury and non-injury data points. The distribution of injuries within these clusters varied across seasons, highlighting potential correlations between changes in training load patterns and injury occurrence. (See Figure 9: Bottom Part)

The combination of KS statistic, PCA, and t-SNE provided a comprehensive framework for detecting and visualizing data drift. The KS statistic allowed for precise identification of drifted weeks, while PCA and t-SNE offered insights into the multivariate and nonlinear patterns of drift, respectively. These findings underscore the importance of monitoring data drift in sports analytics, as it can influence model performance and injury prediction.

4.3 Model construction

4.3.1 Learning scenarios with Traditional Machine Learning Algorithms

This chapter investigates three distinct learning scenarios: Static learning, Continual learning, and Drift Retraining Adaptive learning (DDM-based non-static learning). Each scenario is further examined through two perspectives: non-cumulative and cumulative learning.

4.3.1.1 Learning Scenarios

Static Learning

In static learning, the machine learning model was trained on the designated training dataset and then used to predict outcomes on the entire test dataset without further updates or modifications.

Continual Learning

Continual learning involved iterative predictions on each data point in the test set. After each prediction, the predicted data point was appended to the training dataset, and the model was retrained before proceeding to the next prediction. This iterative approach allowed the model to gradually adapt as it processed new data points.

Drift Retraining Adaptive Learning (DDM)

Drift Retraining Adaptive learning followed a similar iterative prediction and retraining process as continual learning but incorporated the Drift Detection Method (DDM). DDM monitored the predictions to identify potential data drifts, triggering model retraining when significant drift was detected. This adaptation aimed to enhance model evaluation and performance by addressing evolving data distributions.

4.3.1.2 Learning Perspectives

Non-Cumulative Learning

Non-cumulative learning involved training the model on individual seasons and testing it on the subsequent season. For example, training on season 1 and testing on season 2, training on season 2 and testing on season 3, and so on, up to season 5.

Cumulative Learning

Cumulative learning expanded the training dataset by incorporating data from multiple seasons before testing on the next season. For example, training on seasons 1 and 2 and testing on season 3, training on seasons 1, 2, and 3 and testing on season 4, and continuing this pattern up to season 5.

4.3.1.3 Machine Learning Algorithms

To achieve the research objectives, two machine learning algorithms were utilized:

Logistic Regression:

This well-established algorithm is widely used in binary classification tasks due to its simplicity and interpretability. Logistic Regression served as the baseline model for comparison.

AdaBoost:

This ensemble learning algorithm combines multiple weak learners to create a stronger predictive model. AdaBoost was employed to evaluate the benefits of an ensemble-based approach in comparison to a single-model technique.

By employing these two algorithms, the chapter conducted a comparative analysis to assess the effectiveness of both conventional and ensemble-based methods across the different learning scenarios and perspectives. This comprehensive evaluation facilitated a deeper understanding of the strengths and limitations of each learning strategy in the context of injury prediction in professional soccer.

By combining the three distinct learning scenarios—Static Learning, Continual Learning, and Drift Retraining Adaptive Learning—with the two perspectives of non-cumulative and cumulative learning, we establish six comprehensive scenarios for evaluation with each machine learning algorithm. These scenarios are as follows:

Static Non-Cumulative: The model is trained on a specific season and evaluated on the subsequent season without incorporating additional data or updates.

Continual Non-Cumulative: The model is iteratively updated with each prediction by adding the predicted data point to the training set, retraining the model before the next prediction, while maintaining season-specific independence.

Drift Retraining Adaptive Non-Cumulative: Similar to Continual Non-Cumulative, this scenario incorporates Drift Detection Method (DDM) to trigger retraining only when significant data drift is identified, offering a more adaptive approach to evolving data distributions.

Static Cumulative: The model is trained on an aggregated dataset comprising multiple seasons and tested on the subsequent season without further updates.

Continual Cumulative: Extending the Continual Learning approach, this scenario iteratively retrains the model with predictions while using a cumulative dataset that expands to include data from prior seasons.

Drift Retraining Adaptive Cumulative: This scenario builds upon Drift Retraining Adaptive Learning by incorporating cumulative data, retraining the model when data drift is detected, ensuring the most adaptable approach to data evolution.

These six scenarios provide a robust framework for evaluating the performance of each algorithm across varying levels of static and adaptive learning, as well as data inclusion strategies. By systematically analyzing these scenarios, we aim to elucidate the nuances of different learning strategies and their efficacy in addressing the challenges of injury prediction in professional soccer.

4.3.2 Learning scenarios with Neural Network-Based Approaches

In addition to traditional machine learning methods, a neural network-based framework was explored. The neural network model implemented the same three learning methodologies:

Immediate Neural Updates

Similar to the basic online learning framework, the neural network's weights were updated after processing each data point using backpropagation. While highly adaptive, the absence of drift detection made this approach prone to instability in dynamic environments.

Drift Detection with Batch Retraining for Neural Networks

Neural network performance was monitored using a sliding window of recent loss values. When a performance drop was detected, the network was retrained on a combined dataset of previous and recent data points. This ensured adaptability while retaining historical patterns, though computational costs were higher.

Sliding Window Retraining for Neural Networks

Drift detection triggered retraining using only the most recent data points within a sliding window. This minimized computational overhead and improved responsiveness to current trends, making it suitable for environments with rapid data evolution.

By employing both traditional and neural network-based approaches, this chapter provides a robust evaluation of online learning methodologies for injury prediction. The comparative analysis highlights the trade-offs between flexibility, computational efficiency, and stability, offering valuable insights for real-world applications in professional soccer.

The three neural network-based approaches—Immediate Neural Updates, Drift Detection with Batch Retraining, and Sliding Window Retraining—were evaluated under both non-cumulative and cumulative learning scenarios, mirroring the framework used for traditional machine learning methods. This dual perspective allowed for a comprehensive assessment of the interplay between data inclusion strategies and adaptive methodologies. By systematically comparing these six scenarios, this chapter elucidates the trade-offs in performance, computational efficiency, and stability across traditional and neural network-based approaches for injury prediction in professional soccer.

4.4 Results

4.4.1 Results from Traditional Machine Learning Algorithms

Table 7 Model Fit Evaluation for the Optimal Models Derived from Each Analytical Approach with Logistic Regression

Learning scenarios	Data Points	Evaluation metrics							
Algorithms	Non-injury and Injury	Precision	Recall	Cohen's kappa	RMSE	Brier Score	ROC AUC Score	Confusion matrix TN FN	FP TP
Static non-cumulative	Non-injury	0.96	0.76	0.0045	0.5065	0.176	0.501	6637	2057
	Injury	0.04	0.25					264	89
Continual non-cumulative	Non-injury	0.98	0.98	0.541	0.1854	0.034	0.770	8539	155
	Injury	0.56	0.56					156	197
Drift Retraining Adaptive Non-Cumulative	Non-injury	0.98	0.98	0.571	0.18	0.032	0.781	8545	149
	Injury	0.58	0.59					144	209
Static Cumulative	Non-injury	0.96	0.70	0.0111	0.5614	0.198	0.523	6073	2621
	Injury	0.04	0.35					230	123
Continual Cumulative	Non-injury	0.98	0.98	0.5502	0.1839	0.034	0.779	8540	154
	Injury	0.57	0.57					152	201
Drift Retraining Adaptive Cumulative	Non-injury	0.98	0.98	0.5828	0.1775	0.032	0.793	8549	145
	Injury	0.59	0.60					140	213

Table 8 Model Fit Evaluation for the Optimal Models Derived from Each Analytical Approach with AdaBoost

Learning scenarios	Data Points	Evaluation metrics							
Algorithms	Non-injury and Injury	Precision	Recall	Cohen's kappa	RMSE	Brier Score	ROC AUC Score	Confusion matrix TN FN	FP TP
Static non-cumulative	Non-injury	0.97	0.67	0.0157	0.5814	0.338	0.536	5848	2846
	Injury	0.05	0.40					212	141
Continual non-cumulative	Non-injury	0.98	0.98	0.5952	0.1762	0.032	0.805	8545	149
	Injury	0.60	0.63					132	221
Drift Retraining Adaptive Non-Cumulative	Non-injury	0.99	0.98	0.6182	0.1711	0.03	0.816	8553	141
	Injury	0.62	0.65					124	229
Static Cumulative	Non-injury	0.96	0.73	-0.0033	0.54	0.225	0.494	6317	2377
	Injury	0.04	0.26					261	92
Continual Cumulative	Non-injury	0.99	0.98	0.6212	0.1708	0.03	0.819	8552	142
	Injury	0.62	0.65					122	231
Drift Retraining Adaptive Cumulative	Non-injury	0.99	0.98	0.6074	0.1728	0.029	0.808	8554	140
	Injury	0.61	0.63					130	223

Static Learning: Both Logistic Regression and AdaBoost struggled with injury predictions in static scenarios, as evidenced by low precision, recall, and Cohen's kappa values. The inability to adapt to evolving data distributions highlights the limitations of static learning.

Continual Learning: Iterative updates significantly improved injury-class performance for both algorithms. Logistic Regression achieved moderate gains, while AdaBoost demonstrated robust adaptability with higher recall (up to 0.63) and ROC AUC (up to 0.819), underscoring the benefits of incremental learning.

Drift Retraining Adaptive Learning: Drift detection further enhanced model performance, particularly in non-cumulative scenarios. Logistic Regression achieved its highest metrics with an ROC AUC of 0.793 and recall of 0.60 in cumulative settings. AdaBoost showed the best overall performance, achieving an ROC AUC of 0.816 and recall of 0.65, validating its strength in dynamic environments.

4.4.2 Results from the Learning scenario with Neural Network-Based Approaches

Table 9 Model Fit Evaluation for the Optimal Models Derived from Each Analytical Approach with Artificial Neural Network

Learning scenarios	Data Points	Evaluation metrics							
Algorithms	Non-injury and Injury	Precision	Recall	Cohen's kappa	RMSE	Brier Score	ROC AUC Score	Confusion matrix TN FN	 FP TP
Continual non-cumulative	Non-injury	0.97	0.99	0.353	0.1878	0.031	0.62	8635	59
	Injury	0.61	0.26					260	93
Continual Retraining Non-cumulative	Non-injury	0.98	0.99	0.473	0.1866	0.028	0.71	8580	114
	Injury	0.57	0.43					201	152
Continual Retraining Non-cumulative (Recent data)	Non-injury	0.98	0.99	0.519	0.1762	0.025	0.72	8604	90
	Injury	0.64	0.46					191	162
Continual cumulative	Non-injury	0.97	1.00	0.31	0.1875	0.031	0.61	8653	41
	Injury	0.65	0.22					277	76
Continual Retraining cumulative	Non-injury	0.98	0.99	0.51	0.1766	0.025	0.71	8610	84
	Injury	0.65	0.44					198	155
Continual Retraining cumulative (Recent data)	Non-injury	0.98	0.99	0.55	0.1721	0.024	0.74	8605	89
	Injury	0.66	0.49					179	174

The evaluation of neural network-based approaches under six scenarios revealed key differences in performance across Immediate Neural Updates, Drift Detection with Batch Retraining, and Sliding Window Retraining methods:

Immediate Neural Updates:

High precision (0.97–0.98) and recall (0.99) for non-injury predictions, but poor recall for injury predictions (0.22–0.26). Cohen's kappa values (0.31–0.353) and ROC AUC (0.61–0.62) were low, indicating poor overall balance and stability.

Drift Detection with Batch Retraining:

Non-cumulative scenarios showed improved performance, with injury recall increasing to 0.43 and ROC AUC reaching 0.71. Cumulative learning enhanced injury recall to 0.44, but the gains were marginal, indicating potential inefficiencies in incorporating extensive historical data.

Sliding Window Retraining:

Non-cumulative settings achieved higher injury recall (0.46) and improved Cohen's kappa (0.519), with an ROC AUC of 0.72. Cumulative sliding window retraining yielded the best results, with a Cohen's kappa of 0.55, injury recall of 0.49, and ROC AUC of 0.74, demonstrating its effectiveness in balancing adaptability, efficiency, and stability.

Across all scenarios, sliding window retraining consistently outperformed other approaches in terms of injury prediction, while drift detection proved essential for stability and adaptability. However, injury prediction remained a challenging task due to class imbalance.

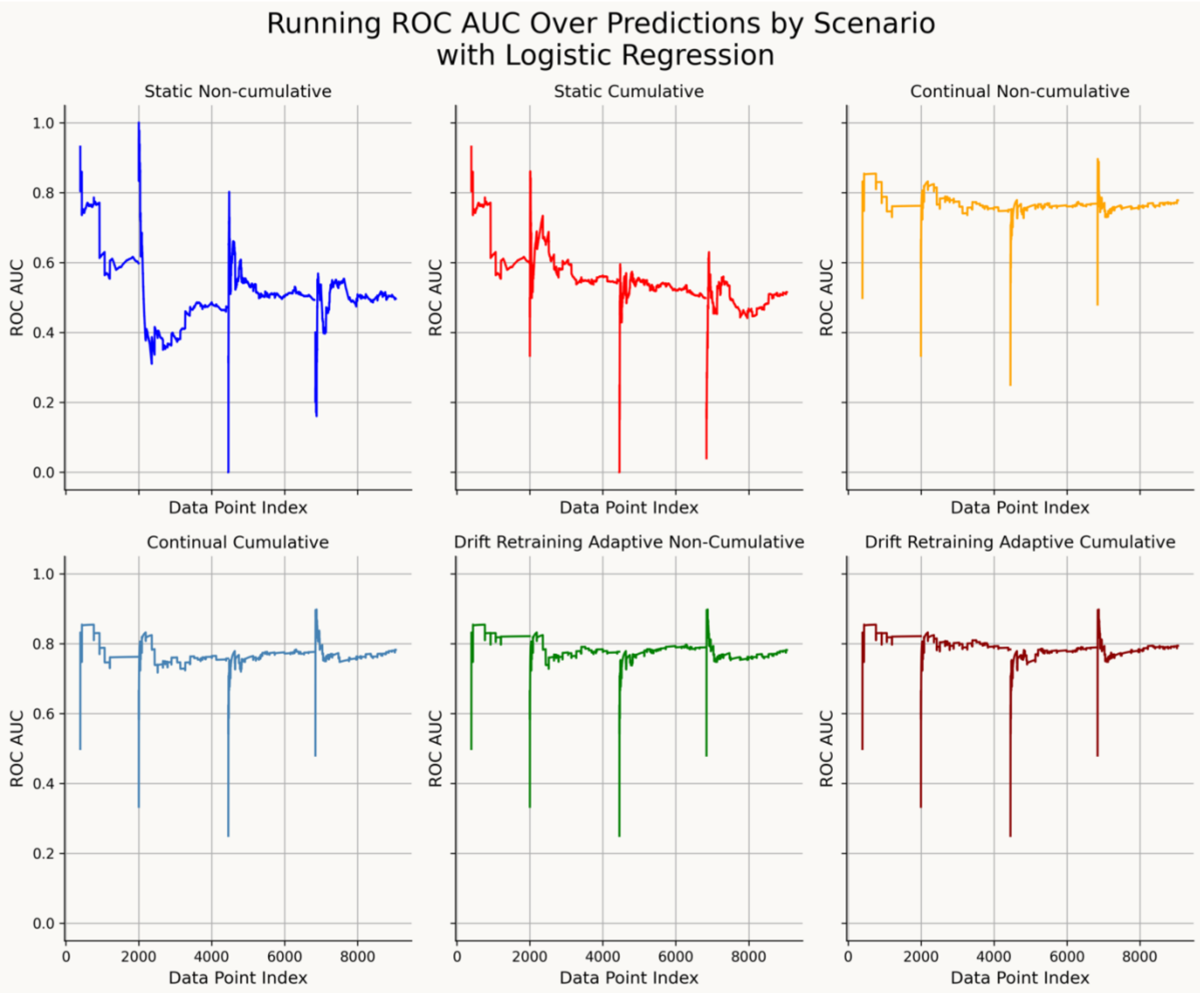
4.5 Discussion

This chapter evaluated the performance of traditional machine learning algorithms (Logistic Regression and AdaBoost) and neural network-based approaches across six distinct learning scenarios, combining static, continual, and adaptive methodologies with non-cumulative and cumulative perspectives. By systematically comparing these approaches, the research provides valuable insights into the trade-offs between adaptability, stability, and computational efficiency in the context of injury prediction in professional soccer. The findings highlight the critical role of adaptive strategies, particularly drift detection and sliding window retraining, in addressing the challenges posed by evolving data distributions and class imbalances.

Injury prediction perspectives.

Traditional machine learning approaches:

Figure 10 Running ROC AUC Over Predictions by Scenario with Logistic Regression



This figure illustrates the running ROC AUC across six learning scenarios (Static Non-Cumulative, Static Cumulative, Continual Non-Cumulative, Continual Cumulative, Drift Retraining Adaptive Non-Cumulative, and Drift Retraining Adaptive Cumulative) for the Logistic Regression model. It highlights the model's adaptability and performance in evolving data distributions.

Figure 11 Running ROC AUC Over Predictions by Scenario with AdaBoost

This figure presents the running ROC AUC across six learning scenarios for the AdaBoost model, demonstrating its ensemble learning capabilities and its ability to handle drift detection and cumulative retraining effectively.

The evaluation of traditional machine learning algorithms, Logistic Regression and AdaBoost, across six distinct learning scenarios provides important insights into their utility in predicting injuries in professional soccer. From a sports science perspective, the dynamic and high-intensity nature of football presents unique challenges for injury prediction, as player workloads, match conditions, and external factors such as training regimens and recovery cycles contribute to rapidly evolving data distributions. Static learning approaches, which lack adaptability, were unable to capture these dynamics effectively. Logistic Regression, in particular, exhibited poor performance, with injury recall as low as 0.25 and minimal agreement metrics in non-cumulative settings. AdaBoost demonstrated marginally better results in static scenarios, with an injury recall of 0.40, yet both algorithms struggled to address the temporal and contextual variability inherent in football injury data.

Continual learning offered a notable improvement by iteratively updating models with each new data point, aligning more closely with real-world football scenarios where injury risks evolve over time. This adaptive approach allowed for significant increases in injury recall, reaching 0.56 for Logistic Regression and 0.63 for AdaBoost in non-cumulative settings. These results highlight the potential of continual learning to integrate updated information about players' physiological states, training loads, and match exposures, which are critical in football injury prevention strategies. Cumulative learning, while adding historical context, produced diminishing returns, as excessive reliance on past data diluted the relevance of recent trends. Nonetheless, AdaBoost excelled in this scenario, achieving a recall of 0.65 and demonstrating its capability to balance historical and current injury predictors.

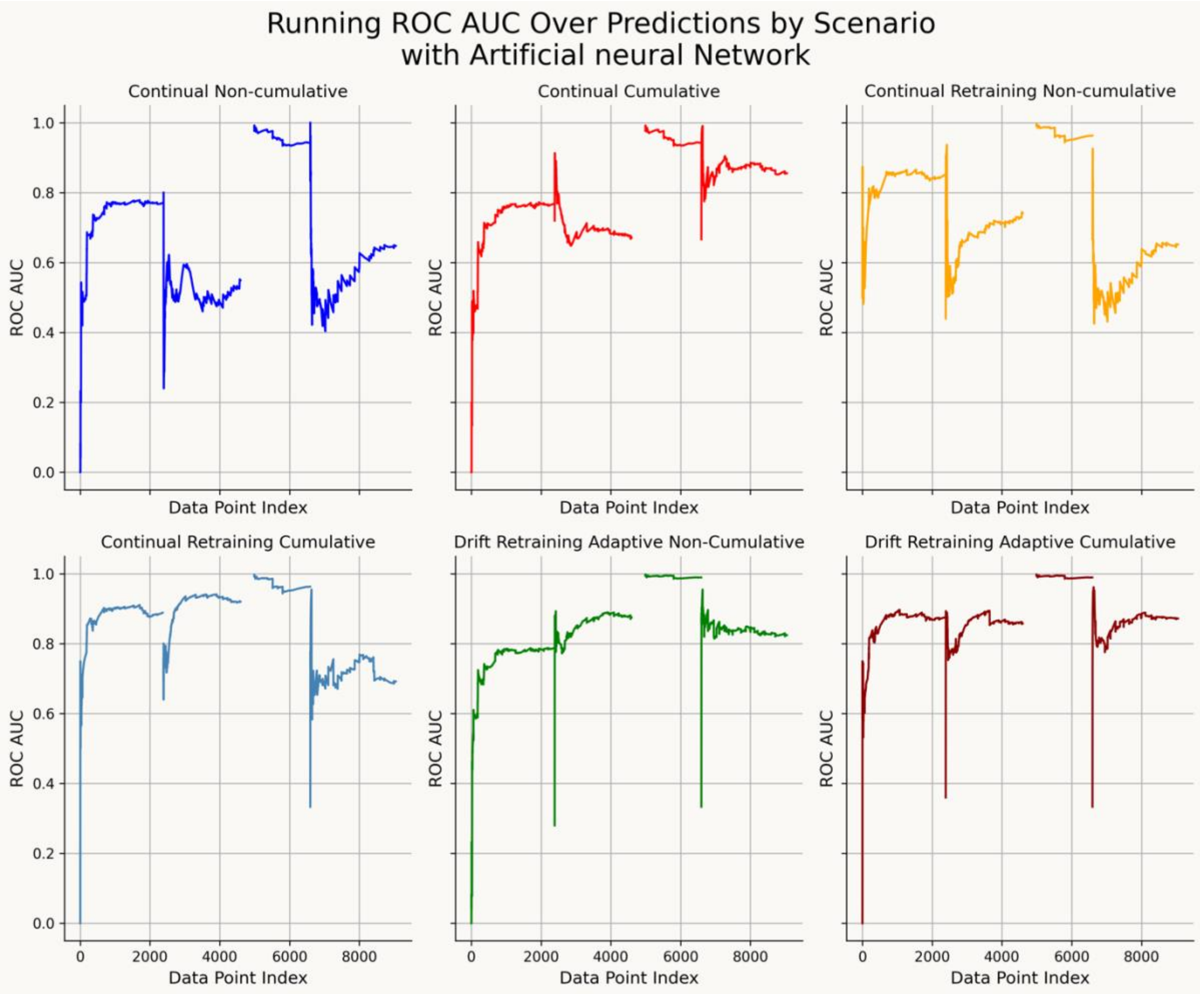
Drift Retraining Adaptive Learning emerged as the most effective methodology, particularly for managing the complex, non-linear nature of injury prediction in football. With the inclusion of drift detection, models were able to identify significant shifts in data distributions, such as those caused by mid-season workload spikes or changes in player form. This dynamic adjustment was especially beneficial for AdaBoost, which achieved an injury recall of 0.65 and a Cohen's kappa of 0.6182 in non-cumulative settings, while maintaining strong performance in cumulative learning with an ROC AUC of 0.808. Logistic Regression also saw its best performance in adaptive scenarios, with an injury recall of 0.60 and a kappa of 0.5828, underscoring the importance of retraining strategies in maintaining model relevance.

From a football sports science perspective, these findings highlight the necessity of incorporating adaptive learning strategies that align with the fluid nature of player health and performance data. Drift detection, in particular, mirrors the proactive injury prevention approaches employed by sports scientists, such as monitoring player fatigue or recovery metrics to adjust training loads in real time. Moreover, the superior performance of AdaBoost reflects its ability to capture the interplay of multiple risk factors, akin to how sports scientists consider a holistic view of a player's physical, tactical, and psychological state.

Overall, these results emphasize that static approaches are ill-suited for dynamic environments like football, where injury risks are constantly evolving. The integration of continual learning and drift detection offers a pathway to more accurate and context-aware injury predictions, enabling sports science teams to make data-driven decisions in load management, injury prevention, and player recovery. This synergy between machine learning methodologies and sports science practices could pave the way for enhanced injury prediction systems, ultimately contributing to better player welfare and team performance.

Neural Network based approaches:

Figure 12 Running ROC AUC Over Predictions by Scenario with Artificial Neural Network



This figure shows the running ROC AUC for six learning scenarios with Artificial Neural Networks, comparing the effectiveness of continual learning, retraining, and sliding window approaches. The trends underscore the neural network's dynamic adaptability to changing data conditions.

The evaluation of neural network-based approaches for injury prediction, tested across six distinct learning scenarios, reveals nuanced insights into their adaptability and performance. These scenarios included Continual Non-Cumulative, Continual Retraining Non-Cumulative, Continual Retraining Non-Cumulative with Recent Data, Continual Cumulative, Continual Retraining Cumulative, and Continual Retraining Cumulative with Recent Data. Neural networks leveraged their capability to iteratively adapt to new data while balancing computational efficiency and model stability, making them well-suited for dynamic injury prediction tasks.

In the Continual Non-Cumulative scenario, neural networks demonstrated strong non-injury prediction metrics with a precision of 0.97 and recall of 0.99. However, injury-class performance remained limited, with a recall of 0.26 and a Cohen's kappa of 0.353. This suggests that, while immediate updates can help the model adapt, the absence of drift detection or retraining hinders its ability to handle minority-class dynamics effectively. Similarly, the Continual Cumulative scenario, which incorporated historical data without explicit retraining mechanisms, showed a slight drop in injury-class recall to 0.22, indicating potential overfitting or an inability to focus on recent trends.

Incorporating retraining strategies significantly improved the model's performance. The Continual Retraining Non-Cumulative approach enhanced injury recall to 0.43 and improved Cohen's kappa to 0.473. These results highlight the benefits of retraining, which allows the model to recalibrate its parameters and respond to distributional shifts in the data. Notably, Continual Retraining Non-Cumulative with Recent Data further boosted performance, achieving an injury recall of 0.46, a Cohen's kappa of 0.519, and an ROC AUC of 0.72. By focusing retraining on recent data points within a sliding window, this approach optimized computational efficiency while maintaining adaptability to current trends.

Cumulative retraining approaches demonstrated similar trends, with Continual Retraining Cumulative improving injury recall to 0.44 and Cohen's kappa to 0.51. However, the inclusion of extensive historical data diluted the model's focus on recent patterns, particularly for the injury class. The Continual Retraining Cumulative with Recent Data approach addressed this limitation by leveraging a sliding window to prioritize recent information. This scenario achieved the highest overall performance, with an injury recall of 0.49, a Cohen's kappa of

0.55, and an ROC AUC of 0.74. These results underscore the effectiveness of sliding window retraining in balancing computational cost and adaptability to evolving data.

From a sports science perspective, the ability to dynamically adapt to changing player workloads, match conditions, and recovery profiles is critical in injury prediction. Neural networks equipped with retraining strategies, particularly sliding window retraining, align well with these requirements. The superior performance observed in these scenarios reflects their capacity to capture complex, non-linear relationships between injury risk factors, such as training load, player fatigue, and match intensity. Moreover, the emphasis on recent data mirrors the practices of sports scientists, who prioritize current player metrics when making load management decisions.

Overall, these results highlight the limitations of immediate updates and static cumulative approaches, emphasizing the need for adaptive strategies to address the challenges of evolving injury risks in football. Sliding window retraining emerged as the most effective methodology, offering a robust balance between stability, computational efficiency, and responsiveness to recent trends. These findings pave the way for more sophisticated injury prediction systems that integrate machine learning with sports science insights to enhance player welfare and performance optimization.

Practical applications.

The practical applications of our findings extend to multiple stakeholders in soccer injury prevention and management, providing actionable insights to enhance player well-being and optimize performance. Soccer clubs and medical staff can utilize the adaptive learning models developed in this chapter as a proactive tool to assess injury risks on a regular basis. By continuously monitoring players’ well-being and identifying potential vulnerabilities, these models enable timely interventions and the design of tailored training programs to mitigate injury risks effectively.

The cumulative training approach adds a valuable dimension to these applications. By integrating data across multiple seasons, teams can gain a more comprehensive understanding of injury patterns unique to their players. This longitudinal perspective allows for the creation of personalized injury prevention strategies that account for the individual risks, playing styles, and physiological characteristics of each player, ensuring a more targeted and effective approach to injury management.

Integrating adaptive learning models into injury prediction frameworks also aligns seamlessly with the broader movement toward data-driven decision-making in sports management. By incorporating predictive analytics into a holistic sports science approach, teams can combine physiological, performance, and contextual data to optimize player health and performance. Such frameworks not only enhance the precision of injury risk assessments but also support informed decision-making around player workload, recovery, and rotation strategies.

In a broader context, this research contributes to the growing intersection of machine learning, sports science, and medicine. The adaptive learning models presented in this chapter exemplify the potential of artificial intelligence to revolutionize injury prediction in dynamic and high-performance environments. As the field continues to evolve, the integration of these models offers immense promise for transforming injury management strategies and ensuring the long-term well-being of athletes in soccer and other sports. By bridging the gap between data science and practical sports applications, this work sets the stage for innovative approaches to athlete care and performance optimization.

Strengths and limitations.

Our chapter introduces several strengths that advance the field of soccer injury prediction while also acknowledging limitations that provide important considerations for future research.

Strengths.

Innovative Methodology

This research integrates online continual and adaptive learning methodologies into soccer injury prediction, offering a dynamic approach to capture evolving injury patterns. The use of multi-seasonal data sets our chapter apart from previous static or single-season analyses, providing novel insights into the interplay between training loads, player characteristics, and injury risks over time.

Comprehensive Multi-Season Dataset

Utilizing a rich dataset of Elite Premier League soccer players enhances the generalizability and applicability of our findings. This dataset, encompassing detailed information on training load, injury occurrences, and player performance metrics, enables a thorough examination of the multifactorial nature of soccer injuries.

Robust Performance Evaluation Metrics

Employing rigorous metrics such as ROC-AUC, Cohen's kappa, and Brier scores provides a robust quantitative framework to evaluate model performance across various scenarios. These metrics allow for detailed comparisons of different learning methodologies and configurations, ensuring a reliable foundation for model selection.

Adaptive Weekly Prediction Framework

The development of machine learning models tailored to a weekly prediction framework aligns with the practical scheduling of soccer matches. This real-time adaptability allows the models to reflect the dynamic nature of player performance and workload variations, ensuring timely and actionable injury predictions throughout the season.

Integration of Drift Detection

The incorporation of Drift Detection Method (DDM) enhances the models' ability to handle concept drifts in the data, such as changes in player workload or training regimens. This adaptive capability aligns with real-world scenarios where injury risks evolve dynamically.

Limitations.

Data Imbalance

Injury prediction inherently suffers from an imbalance between injury and non-injury data points. Although techniques such as Synthetic Minority Oversampling Technique (SMOTE) were applied, residual biases may persist, particularly in scenarios with limited injury instances. This imbalance can challenge the models' ability to generalize effectively for minority classes.

Feature Selection Complexity

While the inclusion of a wide range of features provides a holistic view of injury predictors, it introduces dimensionality and potential overfitting risks. Although methods like Recursive Feature Elimination and Variance Threshold were employed, further refinement of feature selection and interpretability remains an area for improvement.

Sensitivity of Drift Detection

The effectiveness of DDM is influenced by parameter tuning and the nature of the detected drift. Variations in drift patterns or poorly calibrated parameters may limit the models' adaptability in highly dynamic scenarios. Exploring alternative drift detection methods or parameter-agnostic techniques could enhance robustness.

External Validity

The chapter's findings, while highly relevant to Elite Premier League soccer, may not generalize directly to other leagues or sports. Variations in playing styles, training methods, and injury management strategies across different contexts necessitate cautious interpretation of the results when applied outside this setting.

Continuous Data Integration Challenges

The continuous evolution of soccer player data, influenced by matches and training sessions, poses challenges for real-time data integration and model adaptability. The absence of a streamlined software management system could hinder the model's ability to incorporate new data effectively, potentially impacting prediction accuracy. Developing automated pipelines for data preprocessing, integration, and evaluation would address this limitation.

Computational Overhead

Adaptive learning models, particularly those using cumulative and sliding window retraining, introduce significant computational demands. While effective, these approaches require careful resource management to ensure scalability for real-time applications.

Additional Considerations

Interpretability of Models

While neural network-based approaches demonstrated superior performance, their inherent complexity poses challenges for interpretability compared to traditional machine learning methods. Future work should explore methods to enhance explainability, enabling medical teams and coaching staff to better understand the drivers of injury predictions.

Integration with Wearable Technology

Incorporating real-time data from wearable devices (e.g., GPS trackers, heart rate monitors) could further improve the granularity and predictive power of the models by capturing nuanced workload and fatigue metrics.

This chapter's strengths lie in its innovative adaptive learning methodology, comprehensive dataset, rigorous evaluation metrics, and practical alignment with weekly soccer schedules. However, challenges such as data imbalance, feature selection complexity, drift detection sensitivity, external validity, and continuous data integration highlight areas for further exploration. Addressing these limitations will enhance the reliability and applicability of machine learning models in injury prediction, paving the way for more effective, data-driven player management strategies in soccer and beyond.

4.6 Future Research

While this chapter has established a strong foundation for adaptive machine learning approaches in soccer injury prediction, there are several avenues for future research that could further enhance the applicability, scalability, and effectiveness of these models. One key area involves addressing the persistent challenge of data imbalance, particularly between injury and non-injury data points. While techniques such as SMOTE were effective to some extent, future work could explore advanced methods like cost-sensitive learning or generative data augmentation to mitigate biases and improve model performance for minority classes.

Another critical direction involves improving the robustness of drift detection. Although the Drift Detection Method (DDM) was effective in handling concept drifts, exploring alternative techniques such as Adaptive Windowing (ADWIN), Kullback-Leibler Divergence, or ensemble-based drift detection methods could provide better adaptability to subtle and nonlinear shifts in player data. Such advancements could ensure that the models remain highly responsive to dynamic changes in injury risk factors.

The integration of wearable and tracking data is another promising avenue. Incorporating real-time data streams from GPS trackers, accelerometers, and heart rate monitors could enhance the granularity and timeliness of injury risk assessments. Future models should focus on seamless integration of these data sources to enable real-time predictions that accurately reflect player workload, fatigue, and recovery. This aligns with the growing emphasis in sports science on utilizing technology to monitor player well-being in real time.

Future research should also aim to improve the interpretability of neural network-based approaches. While these models demonstrated superior performance, their complexity poses challenges for practical implementation. Integrating explainable AI techniques, such as SHAP or LIME, could bridge this gap, providing actionable insights for medical teams and coaching staff to better understand the factors contributing to injury risks. This would help foster trust and usability of machine learning models in real-world sports environments.

The generalizability of these findings to other leagues, levels of competition, and sports domains is another area requiring further exploration. While this chapter focused on Elite Premier League players, extending the analysis to different settings could test the adaptability of the models and account for variations in playing styles, training regimens, and injury management strategies. Such efforts would broaden the impact of machine learning in sports science and provide valuable insights across diverse athletic contexts.

Additionally, future work should address the challenges of continuous data integration by developing automated pipelines for data preprocessing, integration, and evaluation. As player data evolves with each match and training session, robust systems are essential to ensure that models remain up-to-date and predictions remain accurate. Automating these processes would reduce manual intervention and enable the scalability of machine learning applications in injury prediction.

Moreover, exploring the fusion of multimodal data sources, such as biomechanical data, medical imaging, psychological metrics, and environmental factors, could provide a more holistic view of injury risks. This comprehensive approach could enhance model accuracy and offer deeper insights into the multifactorial nature of injuries. At the same time, future research should consider the ethical implications of deploying AI-based models in sports, including issues related to player data privacy, transparency, and equitable access to technology.

Lastly, future models could extend their scope beyond short-term injury prediction to include long-term health outcomes, such as chronic injury risk or career longevity. Customizing models to account for specific roles and positions on the field could also improve accuracy, as differing physical demands influence injury risks. These advancements would provide tailored recommendations for injury prevention and management, aligning with the ultimate goal of safeguarding player health and optimizing performance.

By addressing these research directions, future studies can build on the adaptive learning framework presented in this chapter, contributing to more effective, scalable, and ethically sound injury prediction systems. Such advancements will not only enhance player management strategies but also support the broader integration of machine learning into sports science, ensuring the long-term well-being of athletes across diverse contexts.

4.7 Conclusions

This chapter provides a comprehensive framework for soccer injury prediction by integrating adaptive machine learning approaches, including online continual learning, drift detection, and sliding window retraining, across both traditional algorithms and neural networks. By evaluating these methodologies under non-cumulative and cumulative scenarios, the findings highlight the critical importance of adaptability in managing the dynamic and evolving nature of injury risks in professional soccer.

The results underscore the limitations of static learning approaches, which fail to capture the temporal complexity of injury patterns, and emphasize the superior performance of adaptive strategies, particularly drift detection and sliding window retraining. These methods demonstrated their ability to balance stability, computational efficiency, and responsiveness to changing data distributions, making them practical tools for injury management in high-performance sports environments.

From a sports science perspective, the integration of these machine learning models into injury prediction frameworks has significant potential to enhance player health and performance. By leveraging multi-seasonal datasets and aligning predictive models with weekly match schedules, teams can proactively assess risks, design personalized training regimens, and optimize player workloads. Moreover, the application of robust evaluation metrics, such as ROC-AUC and Cohen's kappa, ensures the reliability of these predictions, supporting informed decision-making for medical teams and coaching staff.

While this chapter advances the state of the art in soccer injury prediction, it also highlights key challenges, including data imbalance, feature selection complexity, and the need for real-time integration of player data. Addressing these challenges through future research will further enhance the scalability, generalizability, and ethical deployment of machine learning models in sports. Ultimately, this work lays the foundation for more effective, data-driven injury management strategies, contributing to the long-term welfare of athletes and the broader adoption of artificial intelligence in sports science.

In conclusion, our research marks a significant step forward in the field of soccer injury prediction, showcasing the effectiveness of online continual and adaptive learning approaches.

As we navigate the intersection of sports science and machine learning, the foundations laid in this chapter pave the way for continued advancements, ultimately supporting the health, performance, and longevity of soccer players in elite competitions.

Chapter 5

5. Thesis Overview

This thesis explored the relationship between training load and injury prediction in professional soccer through machine learning. It presented a transition from traditional static models, which rely on fixed statistical relationships, to adaptive learning approaches that dynamically update based on evolving data trends. By integrating multi-season datasets and high-dimensional workload metrics, this research provided a more comprehensive and data-driven approach to understanding injury risk in elite-level soccer.

A key focus of this thesis was to address limitations in existing injury prediction methodologies by integrating machine learning insights with a structured, multi-season injury prediction framework. The research initially explored conventional injury prediction models, identifying key risk factors and assessing their predictive capabilities using machine learning techniques. Building on this foundation, the chapter advanced towards a more adaptive approach by incorporating continual learning and drift detection techniques, ensuring that predictive models remained relevant as training practices, match intensities, and player workload patterns evolved. Unlike previous studies that applied machine learning in a static manner, this research developed a real-time injury risk framework, capable of dynamically adjusting to changing data distributions and capturing subtle variations in injury risk factors over time.

Furthermore, this thesis examined the interplay between physiological, psychological, and biomechanical factors in injury prediction, offering a multi-faceted perspective on player health management. By leveraging advanced artificial intelligence models, including Artificial Neural Networks (ANNs) and ensemble-based approaches, the chapter demonstrated the superiority of adaptive learning over conventional statistical models in predicting injuries with higher accuracy and robustness.

This final chapter synthesizes the key findings from the research, evaluates their impact on both sports science and machine learning communities, and outlines future directions for further refining injury prediction systems. The insights gained from this work contribute to the broader goal of enhancing player safety, optimizing workload management, and improving long-term athletic performance through data-driven decision-making.

5.1 Summary of Key Findings

5.1.1 Literature Review and Identification of Gaps (Chapter 2)

Existing research on soccer injury prediction lacks a standardized and unified methodology, leading to inconsistencies in predictive accuracy and model reliability across different studies (Van Eetvelde et al., 2021; Rossi et al., 2022). While various machine learning and statistical models have been employed to assess injury risk, the absence of a cohesive framework limits the comparability and applicability of these findings across diverse sporting contexts. Furthermore, many prior studies have relied on single-season datasets, which restricts the generalizability of models when applied to subsequent seasons. The dynamic nature of soccer, influenced by evolving training regimens, player workloads, and match intensities, necessitates models that can adapt over time, yet most existing approaches fail to incorporate such adaptability (Bowen et al., 2019; López-Valenciano et al., 2018).

Another persistent issue in the field is the significant imbalance in injury data, where the number of injury instances is disproportionately small compared to non-injury cases. This data imbalance poses challenges for machine learning models, which often become biased towards predicting non-injury outcomes due to the overwhelming presence of negative cases in the dataset. Strategies such as oversampling and cost-sensitive learning have been proposed to address this issue, yet the problem persists, impacting the reliability of predictions (Rommers et al., 2020; Ayala et al., 2019).

A significant gap in existing research is the lack of consideration for concept drift—the phenomenon where injury patterns evolve over time due to factors such as changes in coaching strategies, medical interventions, and player conditioning. Most static models fail to adapt to these evolving trends, leading to reduced predictive performance as datasets become outdated (Venturelli et al., 2011; Naglah et al., 2018). This limitation highlights the need for adaptive learning methodologies capable of continuously updating predictive models, addressing data imbalance effectively, and incorporating drift detection mechanisms to maintain long-term model relevance in real-world applications.

5.1.2 Machine Learning for Multi-Season Injury Prediction (Chapter 3)

This chapter I applied Artificial Neural Networks (ANNs) to multi-season soccer injury prediction, marking a significant departure from traditional statistical and single-season machine learning models (Chen & Guestrin, 2016; Mehlig, 2019). The findings demonstrated that ANNs outperformed conventional machine learning algorithms, such as Logistic Regression and XGBoost, in capturing complex relationships between training load variables and injury occurrence (Rossi et al., 2018; Vallance et al., 2020). Through the integration of multi-season data, the model accounted for variations in training intensities, match demands, and evolving injury patterns, providing a more robust and generalizable framework for predicting injuries across different seasons.

The chapter identified key predictors influencing injury risk, including last injury area, player weight, and meta energy, reinforcing the importance of both historical injury records and physiological factors in determining susceptibility to future injuries (Oliver et al., 2020; Kampakis, 2016). Despite achieving a high recall rate of 77%—indicating the model's ability to correctly identify injury cases—the chapter observed lower precision due to an increased number of false positives, a challenge inherent in injury prediction due to the rarity of injury events compared to non-injury instances (Hastie et al., 2009; Ruddy et al., 2019). The imbalance between injury and non-injury cases often skews predictions toward the majority class, necessitating further advancements in data handling techniques and model optimization to enhance specificity.

One of the primary limitations of static models in injury prediction is their vulnerability to concept drift, where shifts in data distribution over time degrade model performance. Static models, which are trained on historical datasets, struggle to adapt to changes in player conditioning, tactical adjustments, and evolving risk factors that influence injury likelihood. This chapter addressed these challenges by advocating for adaptive learning approaches capable of recognizing and adjusting to data drift dynamically, thereby maintaining predictive accuracy over multiple seasons (Impellizzeri et al., 2020; Claudino et al., 2019). The research findings highlight the potential for ANN-based models to provide sports practitioners with a proactive tool for managing player workload and injury prevention strategies, reinforcing the necessity for further advancements in machine learning methodologies tailored to dynamic sports environments.

5.1.3 Continual and Adaptive Learning for Injury Prediction (Chapter 4)

This chapter introduced adaptive machine learning methodologies to address concept drift, a key challenge in sports injury prediction, ensuring that models remain relevant as player workload patterns and injury risk factors evolve over time (Janardan & Mehta, 2017; Webb et al., 2017). Traditional machine learning models often become outdated due to their reliance on static datasets, making them ineffective in the ever-changing environment of professional soccer. By incorporating continual learning techniques, this research allowed predictive models to update dynamically as new data arrives, thereby maintaining their predictive efficacy (Goel & Batra, 2021; Hussain et al., 2021).

A major advancement in this chapter was the implementation of drift detection mechanisms, which trigger retraining only when significant changes in data distribution occur, reducing unnecessary computational overhead while improving model adaptability (Wang et al., 2013; Disabato & Roveri, 2022). Unlike traditional retraining methods that require extensive historical datasets, this research employed sliding window retraining, which prioritizes recent data over outdated information. This approach enhances responsiveness to current injury risk trends while minimizing computational demands, ensuring models remain both efficient and effective (Zenisek et al., 2019; Krawczyk et al., 2017).

The findings of this chapter demonstrate that adaptive learning methodologies, particularly the combination of drift detection and sliding window retraining, yield superior predictive accuracy compared to static models. The use of Artificial Neural Networks (ANNs) further solidified this advantage, as they outperformed traditional machine learning approaches by exhibiting greater flexibility in identifying evolving injury patterns (Gomes et al., 2019; Lee & Lee, 2020). This research also validated the practicality of using a 7-day prediction window, aligning with the competitive schedules of the English Premier League and allowing coaching staff and medical professionals to make timely interventions in player workload management (Wang & Wang, 2023). These advancements position adaptive learning as a critical tool for enhancing injury prediction and prevention strategies in professional soccer, paving the way for future innovations in the intersection of sports science and machine learning.

5.2 Novel Contributions in Injury Risk Research

This thesis significantly advances the field of injury risk assessment by addressing limitations inherent in traditional models and offering a more nuanced, data-driven approach through machine learning. Traditional models for injury prediction have predominantly relied on predefined statistical relationships that fail to capture the dynamic and evolving nature of injury risk in professional soccer. These models often suffer from rigidity, assuming a static relationship between workload metrics and injury occurrence without accounting for contextual factors such as variations in training regimens, player conditioning levels, and tactical strategies.

By leveraging multi-season datasets, this thesis provides a broader and more generalizable framework for injury prediction, overcoming the constraints of single-season studies that lack long-term applicability. The integration of high-dimensional workload metrics, encompassing GPS-derived data, physiological markers, and psychological indicators, ensures a holistic understanding of injury risk. This multi-faceted approach recognizes that injuries are rarely the result of a single factor but rather the cumulative outcome of multiple interacting variables, a limitation often overlooked in previous research.

A key innovation in this thesis is the application of adaptive learning techniques, particularly continual and drift-aware machine learning models, which address the challenge of concept drift—where the statistical properties of input data change over time, leading to reduced predictive performance in static models. By implementing drift detection and sliding window retraining, this research ensures that predictive models remain relevant and effective even as training methodologies and match demands evolve. Furthermore, the integration of artificial neural networks (ANNs) enhances the capacity to model complex, nonlinear relationships within the dataset, outperforming traditional machine learning approaches in both accuracy and robustness.

The practical implications of this research extend beyond theoretical advancements, offering actionable insights for sports scientists, coaches, and medical staff in professional soccer. The development of a 7-day injury prediction framework aligns with the competitive schedules of elite teams, providing timely risk assessments that facilitate informed decision-making regarding player workload management and recovery strategies. This predictive model empowers teams to implement preventative measures, potentially reducing injury incidence and improving overall team performance.

By bridging the gap between sports science and artificial intelligence, this thesis establishes a new paradigm for injury risk modeling, demonstrating the efficacy of adaptive machine learning in a real-world sports setting. The findings underscore the necessity of continuously evolving predictive frameworks to maintain their applicability in dynamic environments, setting the stage for future innovations in injury prevention and athlete performance optimization.

5.2.1 Addressing Limitations in Traditional Injury Risk Models

Historically, injury risk research in sports science has relied on heuristic-based models such as the Acute Chronic Workload Ratio (ACWR), Monotony and Strain Models, and Physiological Thresholds (Hulin et al., 2013; Impellizzeri et al., 2020). While these models have provided valuable insights into training load management, they suffer from several critical limitations that reduce their effectiveness in practical applications. One major drawback is their reliance on static and rigid assumptions, where fixed threshold values determine workload-related injury risks without considering individual adaptation responses. This oversimplification disregards the dynamic nature of player conditioning and recovery, leading to inaccurate assessments of injury susceptibility (Windt & Gabbett, 2017; Drew et al., 2016). Furthermore, these traditional models often employ binary injury classification, categorizing athletes as either at risk or not at risk without accounting for the continuum of risk levels. This dichotomous approach overlooks the progressive accumulation of stress and fatigue, which can predispose players to injury over time rather than at a singular threshold (Soligard et al., 2016; Gabbett, 2016).

Another key limitation of traditional injury risk models is their failure to incorporate the complex interdependencies between multiple workload variables. These models typically analyse one or two factors in isolation, such as training volume or intensity, without considering the intricate relationships between biomechanical, physiological, and psychological stressors. By failing to integrate a broader range of contributing variables, these models do not provide a holistic assessment of injury risk (Bourdon et al., 2017; Bowen et al., 2019). Additionally, the lack of adaptability in traditional models means they cannot account for evolving injury patterns, changing training methodologies, or fluctuations in match demands over multiple seasons. As a result, their predictive utility diminishes over time, necessitating the development of more sophisticated, data-driven approaches that can dynamically adjust to changing conditions. Addressing these shortcomings, this chapter presents an advanced machine learning framework that incorporates multi-season data, adaptive learning techniques, and real-time workload monitoring to offer a more accurate, individualized, and comprehensive approach to injury risk prediction in professional soccer.

5.2.2 Advancing Injury Risk Prediction through Machine Learning

This research introduces a novel machine learning-driven framework that overcomes the limitations of traditional injury risk models by shifting from static, rule-based approaches to adaptive, data-driven systems (Chen & Guestrin, 2016; Mehlig, 2019). Traditional models rely on fixed relationships between workload metrics and injury likelihood, often failing to account for the evolving nature of player conditioning, match intensity, and external stressors. This chapter builds upon these challenges by leveraging multi-season data to capture the temporal evolution of injury risk factors, a key step in enhancing predictive accuracy and generalizability across different competitive environments (Rossi et al., 2018; Vallance et al., 2020). By incorporating a high-dimensional dataset that integrates GPS-tracked workload metrics, physiological responses, and psychological markers, the proposed framework ensures a more holistic understanding of the factors contributing to injury risk (Oliver et al., 2020; Kampakis, 2016).

A central innovation in this thesis is the introduction of a 7-day forecasting window, which allows for short-term injury prediction that aligns with real-world match schedules and workload management strategies (Janardan & Mehta, 2017; Webb et al., 2017). This approach provides actionable insights that enable coaching and medical staff to proactively adjust training intensities, implement recovery protocols, and optimize player rotation strategies to mitigate injury risks. Unlike static models that become obsolete as injury patterns shift, the machine learning-based system adapts dynamically, ensuring that predictions remain relevant as new data becomes available.

Furthermore, by integrating artificial neural networks (ANNs) and advanced machine learning techniques, this thesis demonstrates significant improvements in predictive robustness and precision. The use of drift detection and continual learning mechanisms ensures that the model remains responsive to changes in player workload patterns and emerging risk factors, a key advantage over traditional methodologies. These advancements position the current chapter at the forefront of modern injury risk modeling, bridging the gap between sports science and artificial intelligence while ensuring higher predictive accuracy, robustness, and real-world applicability in elite football settings. The findings underscore the transformative potential of machine learning in enhancing injury prevention strategies, ultimately contributing to improved athlete performance, reduced injury incidence, and more efficient player management.

5.3 Future Research Directions

Future research should focus on expanding the application of machine learning methodologies in injury risk prediction by integrating real-time data from wearable technology, addressing data imbalance issues, and generalizing predictive models across different leagues and sports. The integration of wearable technology, such as GPS trackers, accelerometers, and heart rate monitors, offers the potential to enhance the precision of injury prediction models. These devices continuously capture an athlete's physiological and biomechanical data, providing real-time insights into workload fluctuations and fatigue levels (Halson, 2014; Gabbett, 2016). By incorporating these dynamic data sources, future models can refine their predictive accuracy, allowing for more personalized and immediate interventions to prevent injuries.

Handling data imbalance remains a persistent challenge in injury prediction research. Injuries are relatively rare events compared to the number of non-injury instances, leading to biased models that favor the majority class. Existing solutions such as oversampling, undersampling, and cost-sensitive learning techniques have demonstrated moderate success, but further advancements are needed to improve model performance without distorting the underlying data distribution (Krawczyk, 2016; Leevy et al., 2018). Future studies should explore novel synthetic data generation techniques, such as generative adversarial networks (GANs), which could create more realistic injury cases to balance datasets. Additionally, transfer learning approaches that leverage knowledge from related datasets may enhance injury prediction in environments where limited injury data are available.

Another critical area for future research is the generalization of predictive models to different leagues and sports. While the current chapter focuses on professional soccer, injury risk factors vary significantly across different sports due to variations in gameplay intensity, biomechanics, and training methodologies. To enhance the applicability of machine learning models, future research should assess their performance across multiple leagues, age groups, and competitive levels (Kamiri & Mariga, 2021; Gibert et al., 2016). Cross-sport validation will provide insights into whether injury risk predictors remain consistent or if sport-specific adaptations are necessary. Expanding model applicability beyond soccer will contribute to the broader field of sports science, improving injury prevention strategies for athletes across various disciplines.

Ultimately, these research directions will ensure the continued advancement of machine learning applications in sports science, fostering innovations that lead to more effective injury prevention, better athlete performance management, and optimized training regimens tailored to individual needs.

5.4 Final Conclusion

This thesis marks a significant step forward in the application of machine learning to sports science, particularly in the domain of soccer injury prediction. By introducing adaptive learning methodologies, it effectively addresses longstanding challenges in prior research, such as concept drift and data imbalance, which have historically limited the reliability and generalizability of injury prediction models (Ribeiro et al., 2016; Goldstein et al., 2014). The integration of continual learning and drift detection techniques ensures that predictive models remain dynamic and adaptable, allowing them to respond to evolving player workload patterns, tactical variations, and medical advancements in injury prevention strategies. These innovations collectively contribute to a more robust and practical framework for monitoring injury risk in elite-level athletes.

The findings of this thesis hold substantial practical implications for stakeholders in professional soccer, including clubs, coaches, and medical staff. By leveraging data-driven insights, teams can make informed decisions regarding training loads, recovery protocols, and player rotation strategies, ultimately reducing the likelihood of injuries and optimizing player performance (Friedman, 2001; Hastie et al., 2009). Moreover, the ability to predict injuries within a seven-day timeframe aligns with real-world match scheduling, providing an actionable tool for medical professionals to implement targeted interventions before injuries occur.

Beyond its immediate applications in soccer, this research sets a precedent for the broader integration of machine learning in sports science. The methodologies developed in this chapter can be extended to other high-intensity sports where injury prevention and workload optimization are critical concerns. By bridging the gap between artificial intelligence and applied sports science, this work paves the way for future advancements aimed at protecting athlete health, enhancing performance longevity, and refining training methodologies in professional sports. As machine learning continues to evolve, the insights gained from this research will contribute to the development of increasingly sophisticated models capable of revolutionizing injury prevention strategies across multiple sporting disciplines.

References

Anderson, C. and Sally, D. (2014) *The numbers game: why everything you know about football is wrong*. London: Penguin.

Akenhead, R. and Nassis, G.P. (2016) 'Training load and player monitoring in high-level football: current practice and perceptions', *International Journal of Sports Physiology and Performance*, 11(5), pp. 587–593.

Akenhead, R., Hayes, P., Thompson, K. and French, D. (2013) 'Diminutions of acceleration and deceleration output during professional football play', *Journal of Science and Medicine in Sport*, 16(6), pp. 556–561.

Anderson, L., Di Michele, R., Close, G., et al. (2016) 'Quantification of seasonal-long physical load in soccer players with different starting status from the English Premier League: implications for maintaining physical fitness', *International Journal of Sports Physiology and Performance*, 11(8), pp. 1038–1046.

Anderson, L., Triplett-McBride, T., Foster, C., et al. (2003) 'Impact of training patterns on incidence of illness and injury during a women’s collegiate basketball season', *Journal of Strength and Conditioning Research*, 17(4), pp. 734–738.

Ayala, F., López-Valenciano, A., Gámez Martín, J.A., et al. (2019) 'A preventive model for hamstring injuries in professional soccer: learning algorithms', *International Journal of Sports Medicine*, 40(5), pp. 344–353.

Bourdon, P.C., Cardinale, M., Murray, A., et al. (2017) 'Monitoring athlete training loads: consensus statement', *International Journal of Sports Physiology and Performance*, 12(s2), pp. S2-161–S2-170.

Bahr, R. (2009) 'No injuries, but plenty of pain? On the methodology for recording overuse symptoms in sports', *British Journal of Sports Medicine*, 43, pp. 966–972.

Bahr, R. and Holme, I. (2003) 'Risk factors for sports injuries – a methodological approach', *British Journal of Sports Medicine*, 37, pp. 384–392.

Buist, I., Bredeweg, S.W., Lemmink, K.A.P.M., et al. (2007) 'The GronORun chapter: is a graded training program for novice runners effective in preventing running-related injuries? Design of a randomized controlled trial', *BMC Musculoskeletal Disorders*, 8(1), p. 24.

Brink, M.S., Visscher, C., Arends, S., et al. (2010) 'Monitoring stress and recovery: new insights for the prevention of injuries and illnesses in elite youth soccer players', *British Journal of Sports Medicine*, 44(11), pp. 809–815.

Buchheit, M., Racinais, S., Bilsborough, J.C., et al. (2013) 'Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players', *Journal of Science and Medicine in Sport*, 16(6), pp. 550–555.

Belle, V. and Papantonis, I. (2020) 'Principles and practice of explainable machine learning', *arXiv*, arXiv:2009.11698 [cs, stat].

Bourdon, P.C., Cardinale, M., Murray, A., et al. (2017) 'Monitoring athlete training loads: consensus statement', *International Journal of Sports Physiology and Performance*, 12(S2), pp. S2-161–S2-170.

Bowen, L., Gross, A.S., Gimpel, M., et al. (2019) 'Spikes in acute:chronic workload ratio (ACWR) associated with a 5–7 times greater injury rate in English Premier League soccer players: a comprehensive 3-year chapter', *British Journal of Sports Medicine*.

Bohr, A. and Memarzadeh, K. (2020) 'The rise of artificial intelligence in healthcare applications', *Artificial Intelligence in Healthcare*, pp. 25–60.

Claudino, J.G., Capanema, D.O., de Souza, T.V., et al. (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).

Carey, D.L., Blanch, P., Ong, K.L., et al. (2017) 'Training loads and injury risk in Australian football: differing acute:chronic workload ratios influence match injury risk', *British Journal of Sports Medicine*, 51(16), pp. 1215–1220.

Calvert, T.W., Banister, E.W., Savage, M.V. and Bach, T. (1976) 'A systems model of training for athletic performance', *Australian Journal of Sports Medicine*, 8, pp. 57–61.

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) 'SMOTE: synthetic minority over-sampling technique', *Journal of Artificial Intelligence Research*, 16, pp. 321–357.

Chen, T. and Guestrin, C. (2016) 'XGBoost: a scalable tree boosting system', *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pp. 785–794.

De Silva, V., Caine, M. and Skinner, S. (2018) 'Player tracking data analytics as a tool for physical performance management in football: a case chapter from Chelsea football club academy', *Sports*, 6(4), p. 130.

Drew, M.K., Cook, J. and Finch, C.F. (2016) 'Sports-related workload and injury risk: simply knowing the risks will not prevent injuries: narrative review', *British Journal of Sports Medicine*, 50(21), pp. 1306–1308.

Disabato, S. and Roveri, M. (2022) 'Tiny machine learning for concept drift', *IEEE Transactions on Neural Networks and Learning Systems*, 34(1), pp. 1–12.

Ekstrand, J., Lundqvist, D., Davison, M., et al. (2019) 'Communication quality between the medical team and the head coach/manager is associated with injury burden and player availability in elite football clubs', *British Journal of Sports Medicine*, 53, pp. 304–308.

Ekstrand, J., Walden, M. and Hagglund, M. (2016) 'Hamstring injuries have increased by 4% annually in men's professional football since 2001: a 13-year longitudinal analysis of the

English Football League (2018) 'Charter for academy players and parents: EFL youth development players' and parents' guide season 2018/19'.

Emmert-Streib, F., Yang, Z., Feng, H., et al. (2020) 'An introductory review of deep learning for prediction models with big data', *Frontiers in Artificial Intelligence*, 3.

Eliakim, E., Morgulev, E., Lidor, R. and Meckel, Y. (2020) 'Estimation of injury costs: financial damage of English Premier League teams' underachievement due to injuries', *BMJ Open Sport & Exercise Medicine*, 6(1), p. e000675.

Foster, C. (1998) 'Monitoring training in athletes with reference to overtraining syndrome', *Medicine and Science in Sports and Exercise*, 30(7), pp. 1164–1168.

Gabbett, H., Windt, J. and Gabbett, T. (2016) 'Cost-benefit analysis underlies training decisions in elite sport', *British Journal of Sports Medicine*, 50, pp. 1291–1292.

Gabbett, T. (2004) 'Influence of training and match intensity on injuries in rugby league', *Journal of Sports Sciences*, 22(5), pp. 409–417.

Gabbett, T. (2005) 'Changes in physiological and anthropometric characteristics of rugby league players during a competitive season', *Journal of Strength and Conditioning Research*, 19(2), pp. 400–408.

Gabbett, T. (2010) 'The development and application of an injury prediction model for non-contact, soft-tissue injuries in elite collision sport athletes', *Journal of Strength and Conditioning Research*, 24(10), pp. 2593–2603.

Gabbett, T. (2016) 'The training-injury prevention paradox: should athletes be training smarter and harder?', *British Journal of Sports Medicine*, 50(5), pp. 273–280.

Gabbett, T. (2018) 'Debunking the myths about training load, injury and performance: empirical evidence, hot topics and recommendations for practitioners', *British Journal of Sports Medicine*.

Gabbett, T. and Blanch, P. (2019) 'Research, urban myths and the never-ending story', *British Journal of Sports Medicine*, 53, pp. 592–593.

Gabbett, T. and Domrow, N. (2007) 'Relationships between training load, injury and fitness in sub-elite collision sport athletes', *Journal of Sports Sciences*, 25(13), pp. 1507–1519.

Gabbett, T. and Mulvey, M. (2008) 'Time-motion analysis of small-sided training games and competition in elite women soccer players', *Journal of Strength and Conditioning Research*, 22(2), pp. 543–552.

Gabbett, T. and Ullah, S. (2012) 'Relationship between running loads and soft-tissue injury in elite team sport athletes', *Journal of Strength and Conditioning Research*, 26(4), pp. 953–960.

Gabbett, T. and Whiteley, R. (2017) 'Two training-load paradoxes: can we work harder and smarter, can physical preparation and medical be teammates?', *International Journal of Sports Physiology and Performance*, 12(S2), pp. S2-50–S2-54.

Gabbett, T., Jenkins, D. and Abernethy, B. (2012) 'Physical demands of professional rugby league training and competition using microtechnology', *Journal of Science and Medicine in Sport*, 15(1), pp. 80–86.

Gabbett, T., Nielsen, R., Bertelsen, M., et al. (2019) 'In pursuit of the "unbreakable" athlete: what is the role of moderating factors and circular causation?', *British Journal of Sports Medicine*, 53, pp. 394–395.

Gabbett, T., Whyte, D., Hartwig, T., et al. (2014) 'The relationship between workloads, physical performance, injury and illness in adolescent male football players', *Sports Medicine*, 44(7), pp. 989–1003.

Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002) 'Gene selection for cancer classification using support vector machines', *Machine Learning*, 46, pp. 389–422.

Gabbett, T.J. (2016) 'The training-injury prevention paradox: should athletes be training smarter and harder?', *British Journal of Sports Medicine*, 50(5), pp. 273–280.

Goel, K. and Batra, S. (2021) 'Adaptive online learning for classification under concept drift', *International Journal of Computational Science and Engineering*, 24(2), pp. 128–141.

Gupta, S. and Gupta, A. (2018) 'Handling class overlapping to detect noisy instances in classification', *The Knowledge Engineering Review*, 33.

Halson, S.L. (2014) 'Monitoring training load to understand fatigue in athletes', *Sports Medicine*, 44(S2), pp. 139–147.

Hulin, B.T., Gabbett, T.J., Blanch, P., et al. (2014) 'Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers', *British Journal of Sports Medicine*, 48(8), pp. 708–712.

Hussain, S.S., Hashmani, M., Uddin, V., et al. (2021) 'A novel approach to detect concept drift using machine learning', *2021 International Conference on Computer & Information Sciences (ICCOINS)*, pp. 1–6.

Ibrahimović, M., Mustafović, E., Causevic, D., et al. (2021) 'Injury rate in professional football: a systematic review', *International Journal of Physical Education, Fitness and Sports*, 10(1), pp. 52–63.

Impellizzeri, F.M., Woodcock, S., Coutts, A.J., et al. (2021) 'What role do chronic workloads play in the acute to chronic workload ratio? Time to dismiss ACWR and its underlying theory', *Sports Medicine*.

Johnson, J.M. and Khoshgoftaar, T.M. (2019) 'Survey on deep learning with class imbalance', *Journal of Big Data*, 6(1).

Jones, A., Jones, G., Greig, N., et al. (2019) 'Epidemiology of injury in English professional football players: A cohort chapter', *Physical Therapy in Sport*, 35, pp. 18–22.

Kalkhoven, J.T., Watsford, M.L., Coutts, A.J., et al. (2021) 'Training load and injury: causal pathways and future directions', *Sports Medicine*, 51(6), pp. 1137–1150.

Kampakis, S. (2016) 'Predictive modeling of football injuries'.

Kiesow, H., Spreng, R.N., Holmes, A.J., et al. (2021) 'Deep learning identifies partially overlapping subnetworks in the human social brain', *Communications Biology*, 4(1), pp. 1–14.

Krawczyk, B., Minku, L.L., Gama, J., et al. (2017) 'Ensemble learning for data stream analysis: a survey', *Information Fusion*, 37, pp. 132–156.

López-Valenciano, A., Ayala, F., Puerta, J.M., et al. (2018) 'A preventive model for muscle injuries', *Medicine & Science in Sports & Exercise*, 50(5), pp. 915–927.

Loyola-Gonzalez, O. (2019) 'Black-box vs. white-box: understanding their advantages and weaknesses from a practical point of view', *IEEE Access*, 7, pp. 154096–154113.

Lundberg, S. and Lee, S.-I. (2017) 'A unified approach to interpreting model predictions', *arXiv*, arXiv:1705.07874 [cs, stat].

Murray, N.B., Gabbett, T.J. and Townshend, A.D. (2017) 'Calculating acute:chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages', *British Journal of Sports Medicine*, 51(9), pp. 749–754.

Majumdar, A., Bakirov, R., Hodges, D., et al. (2022) 'Machine learning for understanding and predicting injuries in football', *Sports Medicine - Open*, 8(1).

Majumdar, A., Bakirov, R., Hodges, D., McCullagh, S. and Rees, T. (2024) ‘A multi-season machine learning approach to examine the training load and injury relationship in professional soccer’, *Journal of Sports Analytics*, 1 January, pp. 47–65.

Mehlig, B. (2019) 'Artificial neural networks', *arXiv*, arXiv:1901.05639 [cond-mat, stat].

Naglah, A., Khalifa, F., Mahmoud, A., et al. (2018) 'Athlete-customized injury prediction using training load statistical records and machine learning', *IEEE Xplore*.

Owoeye, O.B., VanderWey, M.J. and Pike, I. (2020) 'Reducing injuries in soccer (football): an umbrella review of best evidence across the epidemiological framework for prevention', *Sports Medicine - Open*, 6(1).

Oliver, J.L., Ayala, F., De Ste Croix, M.B.A., et al. (2020) 'Using machine learning to improve our understanding of injury risk and prediction in elite male youth football players', *Journal of Science and Medicine in Sport*.

Rein, R. and Memmert, D. (2016) 'Big data and tactical analysis in elite football: future challenges and opportunities for sports science', *SpringerPlus*, 5(1).

Racinais, S., Buchheit, M., Bilsborough, J., et al. (2014) 'Physiological and performance responses to a training camp in the heat in professional Australian football players', *International Journal of Sports Physiology and Performance*, 9(4), pp. 598–603.

Rahnama, N. (2011) 'Prevention of football injuries', *International Journal of Preventive Medicine*, 2(1), pp. 38–40.

Soligard, T., Schwellnus, M., Alonso, J.M., et al. (2016) 'How much is too much? (Part 1). International Olympic Committee consensus statement on load in sport and risk of injury', *British Journal of Sports Medicine,* 50(17), pp. 1030–1041.

Rommers, N., Rössler, R., Verhagen, E., et al. (2020) 'A machine learning approach to assess injury risk in elite youth football players', *Medicine & Science in Sports & Exercise*, 52(8), pp. 1745–1751.

Rossi, A., Pappalardo, L. and Cintia, P. (2021) 'A narrative review for a machine learning application in sports: an example based on injury forecasting in soccer', *Sports*, 10(1), p. 5.

Rossi, A., Pappalardo, L., Cintia, P., et al. (2018) 'Effective injury forecasting in soccer with GPS training data and machine learning', *PLOS ONE*, 13(7), p. e0201264.

Sáez, J.A., Galar, M. and Krawczyk, B. (2019) 'Addressing the overlapping data problem in classification using the one-vs-one decomposition strategy', *IEEE Access*, 7, pp. 83396–83411.

Shahee, S.A. and Ananthakumar, U. (2021) 'An overlap sensitive neural network for class imbalanced data', *Data Mining and Knowledge Discovery*.

Tang, W., Mao, K.Z., Mak, L.O. and Ng, G.W. (2010) 'Classification for overlapping classes using optimized overlapping region detection and soft decision', *IEEE Xplore*.

UEFA Elite Club Injury Chapter', *British Journal of Sports Medicine*, 50(12), pp. 731–737.

Vallance, E., Sutton-Charani, N., Imoussaten, A., et al. (2020) 'Combining internal- and external-training-loads to predict non-contact injuries in soccer', *Applied Sciences*, 10(15), p. 5261.

Van Eetvelde, H., Mendonça, L.D., Ley, C., et al. (2021) 'Machine learning methods in sport injury prediction and prevention: a systematic review', *Journal of Experimental Orthopaedics*, 8(1).

Venturelli, M., Schena, F., Zanolla, L. and Bishop, D. (2011) 'Injury risk factors in young soccer players detected by a multivariate survival model', *Journal of Science and Medicine in Sport*, 14(4), pp. 293–298.

Windt, J. and Gabbett, T.J. (2017) 'How do training and competition workloads relate to injury? The workload-injury aetiology model', *British Journal of Sports Medicine*, 51(5), pp. 428–435.

Wang, S., Minku, L.L., Ghezzi, D., et al. (2013) 'Concept drift detection for online class imbalance learning', *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8.

Yang, L., Manias, D.M. and Shami, A. (2021) 'PWPAE: an ensemble framework for concept drift adaptation in IoT data streams', *2021 IEEE Global Communications Conference (GLOBECOM)*.

Zenisek, J., Holzinger, F. and Affenzeller, M. (2019) 'Machine learning based concept drift detection for predictive maintenance', *Computers & Industrial Engineering*, 137, p. 106031.