MGI

Master Degree Program in

Information Management

The Use of AI in Sports to Prevent Injuries

João Nuno Leong Chang

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

The Use of AI in Sports to Prevent Injuries

by

João Nuno Leong Chang

Master Thesis presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Business Intelligence.

Supervised by

Vítor Duarte dos Santos, PhD, NOVA Information Management School

April 2025

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism, any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

[Lisbon, April 07, 2025]

João Chang

ACKNOWLEDGEMENTS

I would like to start by thanking Professor Vítor for his guidance, availability, and time. His support was crucial to the development of this master thesis. His positivity and organization inspired me, keeping me motivated and helping me enjoy every day of research. I wish him all the best, both professionally and personally.

I would like to thank the participants for their availability and curiosity in providing feedback for this product development. Their knowledge and insights were crucial to this study. Special thanks to Ana Moura, Marco Jorge, and Rui Mendes for their contributions.

Thank you to all my friends who supported me, especially those I met during this master's degree. It was truly a pleasure to share classes with you and becoming great friends. To my family, thank you for your support and for providing me with the opportunity to embark on this academic journey.

Finally, thank you to the NOVA IMS institution. From starting my undergraduate program in 2020 to (hopefully) completing this master's degree, these five years have been filled with knowledge and enriching experiences that I will carry with me for the future.

ABSTRACT

In the fast-evolving context of sports, where performing at the highest level is key, the impact of athlete injuries has become major, affecting not only the athletes themselves but also everyone involved in the sports ecosystem. These injuries pose significant challenges to athlete safety and performance, leading to costly consequences for clubs and organizations, thereby highlighting the urgent need for effective prevention strategies. This research addresses the critical need for innovative solutions by developing a comprehensive framework that integrates Artificial Intelligence (AI) and Internet of Things (IoT) technologies to enhance injury prevention across various sports. The framework guides data scientists through four key steps: Context Identification, Anatomical Location, Data Requirements, and Final Solution, ensuring that predictive models are tailored to the unique characteristics of each sport. By leveraging player-specific data and real-time metrics collected through IoT devices, the framework aims to improve predictive accuracy and facilitate timely interventions. Furthermore, it has been validated by specialists, confirming its usefulness in real-world scenarios. This study highlights the transformative potential of AI and IoT technologies in revolutionizing injury prevention strategies, ultimately promoting athlete safety and optimizing performance.

KEYWORDS

Sports; Health; Injury Prevention; Artificial Intelligence; Internet of Things

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
ANN	Adaptive Neural Network
BMI	Body Mass Index
CNN	Convolutional Neural Network
DSR	Design Science Research
GPS	Global Positioning System
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
LEMS	Lower extremity muscle strain
LSTM	Long Short-Term Memory
LTL	Linear Temporal Logic
ML	Machine Learning
NFL	National Football League
NHL	National Hockey League
OSIICS	Orchard Sports Injury and Illness Classification System
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SEMG	Surface Electromyography
SLRQ	Systematic Literature Review Question
SMDCS	Sport Medicine Diagnostic Coding System
STL	Single Task Learning
SVM	Support Vector Machine
VAR	Video Assistant Referee
XGBoost	Extreme Gradient Boosting

1. INTRODUCTION

In today’s fast-paced world of emerging technologies, executives and organizations are shifting to new digital technologies requiring them to carefully analyse the way they operate, and in many cases, forcing them to implement a digital transformation program reshaping the current business model. Essentially, the adoption of digital technologies may open doors for new possibilities for efficiency gains, customer intimacy, and innovation (Volberda et al., 2021).

The advancement of technology has brought significant changes in the world of sports, from smart clothing, sports management software, and innovative data analytics capabilities, with perceived impacts on athletes' performance, injuries, operations, and fan experience (Qi et al., 2024).

Injuries always pose a liability in sports, affecting athletes’ performance or even ending their athletic career. This can lead to physical and mental injuries, requiring lengthy rehabilitation. This has prompted sports entities to take a keen interest in injury prevention (Yang et al., 2022), leveraging technologies to optimize training programs, minimize the risk of injury, and improve athletes’ safety and performance (Rebelo et al., 2023).

Participating in sports activities is widely recognized for its health benefits, it can help reduce weight, improve blood pressure and cardiorespiratory fitness, among many other things. However, there is evidence of health risks associated with sports, highlighting the importance of effective injury prevention measures (Oja et al., 2024). Among these risks, some injuries can be fatal, with sudden cardiac death affecting 1-3 out of 100,000 competitive athletes aged under 35 years each year, making it a common goal for many sports institutes to ensure safety and prevent fatal events during sports practice (Robles et al., 2022).

With the use of advanced Artificial Intelligence (AI)-powered solutions, player-specific data can be used to identify potential injury risks by examining patterns and anomalies that might indicate high fatigue levels or muscle stress, providing fundamental insights for enhanced decision-making (Van Eetvelde et al., 2021), thereby contributing to player longevity in the sport (Li & Huang, 2023).

Thus, it’s important to study the available solutions and understand how to apply them properly to each specific situation, with a focus on injury prevention, which is a major concern for athletes. This situation led to the formulation of the following research question: What are the most suitable AI-powered solutions that can be implemented in different sports according to their unique characteristics, focusing on injury prevention?

The goal of this research is to develop a comprehensive framework indicating the most suitable AI-powered solutions that can be implemented in different sports according to their unique characteristics, focusing on injury prevention.

To achieve the main goal, the following intermediate objectives were defined:

- Perform a study on health, sports, injuries in sports, and AI;
- Analyse relevant health factors and metrics associated with injury;
- Make a comprehensive literature review on AI and injuries prevention;
- Understand which AI-powered solutions best serve the research proposal;
- Propose a framework that combines the AI solutions for different types of sports;
- Evaluate the framework.

The use of AI to prevent injuries in sports is still in the early stages. One major challenge is the lack of a standardized dataset on sports injuries across various studies, making it difficult to test and compare new modeling approaches. This issue arises because the data is often owned by private companies and is not publicly available due to privacy concerns (Kumar et al., 2024). Numerous studies on AI applications in sports indicate that most successful implementations have been in team sports, such as American football and basketball. However, there is still room for improvement in data collection methods and the application of AI technologies in these sports (Hammes et al., 2022).

This study aims to extend the research on AI applicability to both team and individual sports, focusing on the high degree of injury risk and enhancing safety across sports. It is expected that valuable information will be available for sports entities to act on viable solutions to prevent injuries based on data-driven approaches, leading to improved performance, reduced downtime due to injuries, and overall better health outcomes for athletes (Rebelo et al., 2023).

To achieve this, in this dissertation, a comprehensive framework is developed to assess the best AI solutions suited to the characteristics of the different types of sports and associated injury risks— Contact Sports and Non- Contact Sports (Prieto-González et al., 2021). By investing in technologies with preventive injury insights, sports entities can ensure long-term financial benefits, not only by enhancing performance for longer periods, but also by reducing rehabilitation costs. With this, it is hoped that this study will contribute to the state of the art, adding value for further scientific investigations on related topics.

2. LITERATURE REVIEW

The primary objective of this review is to develop a comprehensive overview of key concepts, identify trends, and uncover research gaps that require further investigation. For this, an exploratory and systematic literature review are conducted to gather information from various sources relevant to the area of study of this dissertation: AI, sports, and injury prevention. This phase is important to establish a knowledge base on topics for further development and analysis.

2.1. Sports

The term “sport” has its origins in the Old French word *deporter*, a word used to describe a pleasant activity, an enjoyment, or to seek amusement on an active exercise or taking part in some game (Sport | Etymology of Sport by Etymonline (2023)). Nowadays, the concept of sport has extended its roots not only being a leisure activity but also having competitive characteristics (SPORT | English Meaning - Cambridge Dictionary (n.d.)). Sports encompass a wide range of activities, and it has been subject to numerous studies, with no consensus on its definition. It is widely known for being positively associated with health benefits and having a major impact on social and psychological impacts of engaging in sports.

Sports can be defined as a rule- and skill-based competitive activity that can involve diverse factors according to each specificity. It is aimed at achieving a common goal, usually referred to as winning or not losing. Often involves physical, strategic, and teamwork efforts, providing both participants and spectators with enjoyment and entertainment (Borge, 2020).

Sports are played widely by people around the world with distinct types of commitment, some play recreationally, and others are engaged competitively practicing on a regular basis and participating in official sports competitions at any level. Within the competitive group, there is a subgroup of elite and professional athletes who achieve athletic excellence, often competing internationally and making it their full-time job (Pressler & Niebauer, 2020).

Sports offer numerous health benefits, including promoting weight loss, improving cardiovascular fitness, and regulating heart and lung function (Oja et al., 2024). From a psychological point of view, engaging in sports can help reduce symptoms of depression and anxiety, contributing to improved mental health and well-being (Eather et al., 2023).

Sports are multifaceted, contributing to life satisfaction and happiness, significantly impacting both younger and older individuals, as well as those in poor health. The specific sports people engage in are often influenced by their cultural and social environments. For instance, certain sports may be more popular in specific regions due to historical, cultural, or social factors. This cultural influence can determine the accessibility and popularity of different sports, thereby affecting overall participation rates and the associated benefits (Eather et al., 2023).

2.1.1. Types of Sports

There are various types of sports, each with their own characteristics and uniqueness, making people fall in love with the sport. It can be practiced indoors or outdoors, playing with a racket or racing on foot or under water. There is also a seasonal variable commonly distinguishing winter sports from summer sports.

Multiple approaches exist for classifying sports based on different perspectives. Depending on the objective of evaluation, one type of classification might be more appropriate than another, so each terminology should be considered beforehand. However, it’s important to note that there is no universal definition for categorizing sports, as no method considers all the characteristics of sports. This remains a topic of ongoing study (Pressler & Niebauer, 2020). For the purpose of this study, the two most relevant classifications are presented.

Team Sports and Individual Sports

Team sports are a popular form of physical activity that involves collaboration and promoting teamwork among players to achieve a common goal, usually competing against another team. The essence of team sports lies in the collective effort, where each player’s role contributes to the overall performance and outcome. Examples include Football, Basketball, Rugby, and Hockey. These sports not only offer physical benefits but also play a significant role in social development, particularly for younger individuals. They help create a sense of community and camaraderie, which is essential for building interpersonal relationships and enhancing overall well-being. Participating in team sports is also associated with long-term mental health benefits, boosting confidence and self-esteem when practicing on a regular basis (Eather et al., 2023).

On the opposite side, individual sports emphasize self-reliance and personal responsibility, as athletes depend solely on themselves for success. Individual sport athletes often participate for goal-oriented reasons requiring high levels of discipline and dedication, fostering mental toughness and resilience. On the other hand, compared to team sports, it involves less social interaction, which can potentially affect overall mental health and psychological well-being (Eather et al., 2023). Training focuses on mastering specific skills, and athletes may benefit from flexible schedules tailored to their needs. Examples of individual sports are Tennis, Swimming, Boxing, and Gymnastics.

Contact Sports and Non-Contact Sports

Another way to classify sports is by the level of physical contact involved. Contact Sports and Non-Contact Sports define this distinction.

In Contact Sports, players engage in activities that involve physical interactions with one another. The nature of these sports involves frequent physical confrontations during gameplay, which can play a fundamental role in reaching strategic goals. This can include tackles, blockings, and collisions, as seen in football, rugby, hockey, or in combat sports where athletes strike each other to score points. Usually, the nature of these sports lead to higher injury rates caused by common physical interactions (Hind et al., 2020).

On the other spectrum, Non-Contact Sports do not involve physical contact with other players during gameplay. Sports like volleyball, Badminton, and gymnastics are classified within this category. Usually these sports rely on skills, precision, and strategy without physical collisions (John, 2024.). Although these sports do not involve physical contact with others, injuries can still occur, though of a different nature. Typically, the injury rate in this category is lower than in contact sports (Prieto-González et al., 2021).

Both categories offer unique perspectives on classification and often intersect, highlighting their distinct importance. Team sports aren't limited to contact sports, nor are individual sports exclusively non-contact. This emphasizes the need to choose the appropriate classification method for the relevance of research.

2.1.2. The Role of Technologies in Sports

Information technologies in sports have evolved from a basic utility function to be an indispensable element that enhances performance and ensures health and safety. Its use has also led to faster and more accurate organizational processes, including the management of sports-related data, which enhances overall productivity and efficiency. This evolution reflects a reciprocal relationship between sports and technologies, as the human capabilities limits are reached, future advancements in sports will increasingly rely on technology (Frevel et al., 2022). Based on the findings from Qi et al. (2024), the prominent technological applications can be described as follows:

Enhancement of Athlete Performance

Technological advancements are significantly transforming the landscape of sports, enhancing athletes' performance in various ways and helping them achieve their goals. Improved athletic gear, such as rackets, swimsuits, clothing, and footwear, produced with better quality materials, allows players to feel more comfortable and confident while playing, giving them an extra boost in performance. The integration of analytics and artificial intelligence platforms helps coaches analyse performance data, leading to more informed training decisions and allowing for tailored training regimens that can promote overall athletic performance (Habibi et al., 2023). The continuous evolution of these technologies promises to further improve how sports are played and to break records once thought impossible for humans to achieve.

Injury Prevention and Recovery

Technologies are now a major asset in managing and preventing injuries in sports. They can also boost recovery by tailoring rehabilitation plans, analysing real-time data, and identifying potential biomechanical issues during specific movements. Technological integrations in wearable devices and gear, like smartwatches or smart helmets, are also applications to support this cause (Van Eetvelde et al., 2021). For example, the National Football League (NFL) created the software Digital Athlete that incorporates artificial intelligence and Machine Learning (ML), that is being used by all league players, to enhance safety and prevent injuries. This software can conduct real-time risk analysis to identify potential risk injuries as they happen, allowing immediate intervention to protect players (Langton, 2024). This shift is enhancing the safety of sports, focusing on athletes' health.

Fan Engagement

Digital technologies have transformed how fans interact with sports. Organizations are leveraging fan engagement platforms and tools to help understand audience preferences, leading to improved marketing strategies and fan engagement. Examples of these technologies are social media, podcasts and live streaming. With this, fans can be up to date, feel connected with their favorite athletes and teams through digital channels, and watch them from anywhere in the world (Romero-Jara et al., 2023).

In-Stadium technologies like mobile ordering and contactless payments help reduce waiting time enhancing convenience when ordering food and beverages during an event. This reflects a shift towards a more interactive and engaging fan experience adopting digital solutions.

Truth in Sports

Intelligent tools are significantly impacting this field by enhancing fairness and transparency. They assist referees in making more precise temporal and spatial decisions during sports activities, thereby maintaining the integrity of the sport. Examples include the use of Video Assistant Referee (VAR) in football, Hawk-Eye in tennis and badminton, and Photo Finish technology to determine the fastest human on the planet. These technologies can effectively reduce human error in critical situations and contribute to preserving the authenticity of sports, ensuring that the outcomes are a true reflection of athletes’ abilities and efforts (Spitz et al., 2020).

In essence, yearly technological innovations in sports significantly impact various industries, creating an economic ‘life cycle’ that spans from research and manufacturing to optimized operations, cost savings, enhanced athletic performance, merchandizing sales, ultimately boosting revenue generation. Despite the potential economic benefits, challenges such as financial constraints, skill gaps, and ethical risks associated may arise highlighting the need to carefully address these issues before integration.

2.2. Injuries in Sports

Sports injuries refer to physical harm that athletes sustain during their participation in sports activities, caused by quick or repetitive transfer of kinetic energy (Bahr et al., 2020). They often occur due to overuse, improper technique, or accidents during play.

These injuries can lead to a range of consequences, including mental health issues, reduced performance levels for athletes, financial losses for both professional teams and the athletes themselves, and lifelong physical pain or reduced mobility. After sustaining an injury, it is important to integrate a rehabilitation program and receive psychological support, if necessary. Effective rehabilitation can help athletes manage their cognitive, emotional, and behavioral responses to injury, facilitating a smoother return to sport (Tranaeus et al., 2024).

Thus, it's of major importance to follow preventive measures like stretching and flexibility training to enhance muscle resilience and joint stability (Behm et al., 2021). Athletes should also follow a balanced training plan, avoid muscle overuse, and taking rest days to prevent fatigue-related injuries (Wang et al., 2024).

2.2.1. Types of Injuries

Injuries are a broad topic and can have different degrees of complexity. According to Wang et al. (2024) the most common injuries in sports occur to the musculoskeletal system, which is crucial for movement. However, injuries can also impact on other organ systems, such as the neurological (e.g., concussions and peripheral nerve injuries) and cardiovascular systems (e.g., arrhythmias), for example.

To address this field, based on the article from Bahr et al. (2020), the most common categorization types when evaluating an injury are mechanism, anatomical location, and the tissue type. It is relevant to note that this classification is followed by the Orchard Sports Injury and Illness Classification System (OSIICS) and Sport Medicine Diagnostic Coding System (SMDCS).

Mechanism

Before treating an injury, it is important to understand the context and mechanisms that led to it. This includes the two main classifications – Overuse/Repetitive and Acute injury. Subclassifications also arise such as the specific movement that led to the injury or if it involved any contact with another person, object, or no contact at all.

Overuse injuries are defined as injuries that result from repetitive micro-trauma in specific body parts, such as muscles, tendons, and bones. These injuries develop gradually rather than suddenly, and the accumulation of stress over time tends to lead to chronic conditions, meaning they can persist over extended periods, often with phases of improvement and worsening (Martin et al., 2024).

Opposed to overuse injuries, acute injuries occur suddenly and are often the result of a specific incident or trauma during sports activities. This can include events like falls, collisions, or sudden twists that lead to immediate physical damage. These incidents are significantly influenced by psychological factors, particularly stress responses affecting athlete's decision-making abilities, making them more prone to errors, collisions, and compromised motor control, which ultimately increases the risk of acute injuries (Tranaeus et al., 2024).

Anatomical Location

Injuries are often grouped by anatomical location helping in summarizing data for injury surveillance and accurately identifying where the injury has occurred. The categories are classified as follows:

- Head and Neck (Head and Neck);
- Upper Limb (Shoulders, Upper Arm, Elbow, Forearm, Wrist, and Hand);
- Trunk (Chest, Thoracic Spine, Lumbosacral, and Abdomen);
- Lower Limb (Hip, Groin, Thigh, Knee, Lower Leg, Ankle, and Foot);
- Unspecified Area;
- Multiple Regions (Single Injury Crossing Two or More Regions).

Tissue Type

The type of tissue injured, and its respective pathology is essential to evaluate the complexity of the damage. This classification is crucial for understanding the nature of the injury and its implications for treatment and recovery. The tissue types are:

- Muscle/Tendon (Muscle Injury, Muscle Contusion, Tendon Rupture, etc.);
- Nervous (Brain/Spinal Cord Injury, Peripheral Nerve Injury);
- Bone (Fracture, Bone Stress, Bone Contusion, etc.);
- Cartilage/Synovium/Bursa (Cartilage Injury, Arthritis, etc.);
- Ligament/Joint Capsule (Joint Sprain, Chronic Instability);
- Superficial Tissues/Skin (Contusion, Laceration, Abrasion);
- Vessels (Vascular Trauma);
- Stump (Stump Injury, in Amputees);
- Internal Organs (Organ Trauma);
- Non- Specific (Injury Without Tissue Type Specified).

In summary, these three categorization types should be considered in injury surveillance and sports medicine research. This structured approach allows for better data collection and comparison across studies.

2.2.2. Health Factors and Associated Metrics

There is no consensus on an official list of variables regarding sports injury risk. Therefore, the most appropriate variables should be chosen for each research case. Following the research conducted by Zadeh et al. (2021), several broad categories of specific variables related to health factors in injuries have been identified. These categories are Characteristic, Physiological, and Psychological variables. By analysing these three fields, a comprehensive data collection can be performed for injury analysis.

Characteristic Variables

To study the individual proneness to injury, its relevant to understand their biological metrics such as, its age, sex, height, weight, and Body Mass Index (BMI). These metrics provide essential baseline information that can influence injury risk. For instance, age can affect the body's ability to recover from injuries, while sex can influence the types of injuries more commonly experienced. Height and weight are crucial for understanding the mechanical stresses on the body, and BMI can indicate a comprehensive view of a person's health, alerting for higher or lower injury risk. By analysing these biological metrics, researchers can identify patterns and correlations that may predict injury risk, allowing for more targeted prevention strategies (Zadeh et al., 2021).

Physiological Variables

As discussed above, biological factors are highly important. Following this, physiological variables such as different types of heart rate, breathing rate, and caloric burn are frequently mentioned, as seen in the article from Zadeh et al. (2021).

Inside the broad field of physiology, physical factors are also relevant to studying injuries. The current fitness level assesses important strength metrics. This can be studied by different tests and activities, for example, peak force, jump height, and explosiveness (Zadeh et al., 2021). Having good flexibility is also highly correlated with lower injuries rates in sports, as strengthens the muscles while improving the overall range of motion, Behm et al. (2021) says.

Additionally, physical factors are commonly subject to extrinsic factors. It is relevant to know the type of sport the athlete is involved in, if the training load and intensity is adequate, and how loads are distributed across different body parts during physical activities. The equipment used such as clothing, footwear, and protective gear are also elements that can affect the athletes' performance and risk injury. The playing field can also be influential, regarding type of surface, maintenance, weather conditions, and lightning (Zadeh et al., 2021).

Psychological Variables

The mind can also influence sport outcomes and potential injury risks. Mental health issues such as anxiety, depression, and stress are critical for the overall well-being of athletes. These psychological states can impact performance and directly increase the likelihood of sustaining injuries during sports activities. Additionally, poor sleep quality can significantly affect an athlete's performance and increase the risk of injury by causing slower reaction times and reduced cognitive function (Clemente et al., 2021). Finally, psychological factors can negatively impact rehabilitation outcomes, with athletes often experiencing concerns about their competence and ability to handle the demands of competitive sports post-injury (Tranaeus et al., 2024).

2.3. Challenges and Opportunities

Sports participation offers numerous benefits, but injuries pose significant short- and long-term negative consequences across various domains, necessitating effective injury prevention strategies. The challenge begins with establishing clear injury prevention frameworks and defining key terms, which need refinement to enhance their applicability and effectiveness in real-world contexts. This highlights the need for a comprehensive understanding of injury mechanisms and risk factors. A multidisciplinary and holistic approach, involving fields such as sports medicine, psychology, and data science, is essential for addressing the complexities of sports injuries and developing effective strategies to reduce their occurrence and reoccurrence (Edouard & Ford, 2020).

Moreover, leveraging advancements in technology can offer innovative solutions and tailored approaches, further enhancing injury prevention efforts. In addition, collaboration between researchers, practitioners, and stakeholders is crucial to implement these strategies effectively and promote safer sports environments (Rebelo et al., 2023).

2.4. Related Work - AI and Prevention of Injuries

In this section of the literature review, a systematic analysis is conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology. Meticulously chosen articles related to the research topic are reviewed to provide a comprehensive understanding of the subject matter and an overview of the existing work.

2.4.1. PRISMA Methodology

The PRISMA methodology is a systematic scientific approach designed as a set of methods for identifying, selecting, appraising, and synthesizing studies, determining their inclusion or exclusion based on search criteria and relevance to the study. This comprehensive framework was designed to enhance the quality and transparency of systematic reviews (Page et al., 2021).

The PRISMA 2020 statement is composed of seven sections providing a high-level overview of the systematic review process, with a detailed 27 item steps guide for reporting. It also provides a four-phase flow diagram template that researchers can use and adapt to suit their needs for developing a systematic review (Page et al., 2021) . The four phases diagram can be described as follows:

1. **Identification –** Gather all potential studies from common scientific databases based on a certain search criterion. Remove duplicate and ineligible studies.
2. **Screening –** Exclude studies that do not meet the inclusion criteria.
3. **Eligibility –** Evaluate the studies for eligibility and exclude those that do not meet the predefined set of reasons.
4. **Included –** Final selection of studies that meet all eligibility criteria for further analysis.

2.4.2. PRISMA Execution

From the previous comprehensive study carried out on sports (in Section 2.1), injuries in sports (in Section 2.2), and their respective subtopics, a knowledge base was developed to better understand key concepts for this research. The objective is to review related work on previous years on the use of AI to predict sports injuries. To guide this review, Systematic Literature Review Questions (SLRQs) were formulated in Table 1, and are further analysed in Section 2.4.3.

Table 1 – Systematic Literature Review Questions

SLRQ1	What is the current status of research in this area?
SLRQ2	What are the major issues of AI in Sports Injuries?
SLRQ3	What kind of AI techniques are currently useful in this area?
SLRQ4	What are the most suitable AI-powered solutions that can be implemented in different sports according to their unique characteristics, focusing on injury prevention?
SLRQ5	What are the advantages and disadvantages of applying AI techniques in this field?

To answer these questions using the PRISMA methodology, a set of keywords for each main topic was meticulously chosen based on prior theoretical study. This was then followed by creating a search string to accurately filter and select relevant studies to support this research.

It is important to note that the keywords are in English, and only articles in English were considered. The selected keywords are defined in Table 2.

Table 2 – Systematic Review's Keywords

	Sports Injuries	**Artificial Intelligence**
Keywords	Sports Injuries	Artificial Intelligence
	Acute Injuries	Deep Learning
	Overuse Injuries	Machine learning
	Contact Sports	Artificial Neural Networks
	Non-contact Sports	Predictive Analysis

After defining the keywords, a specific search string was built composed by the combination of the above keywords with the objective of finding them either on titles, abstracts or keywords of articles and papers available in the resource databases. The orchestration of keywords ensures the retrieval of relevant articles encompassing the two main topics, Sports Injuries and AI. The search string used was:

("Sports Injuries" OR "Acute Injuries" OR "Overuse Injuries" OR "Contact Sports" OR "Non-contact Sports") AND ("Artificial Intelligence" OR "Deep Learning" OR "Machine Learning" OR " Artificial Neural Networks" OR "Predictive Analysis").

The search was conducted in October 2024 on the scientific information resource databases in Table 3.

Table 3 – Systematic Review's Resource Databases

Resource Database	**Resource URL**
Scopus	https://www.scopus.com/home.uri
Web of Science	https://www.webofknowledge.com/
Institute of Electrical and Electronics Engineers (IEEE)	https://www.ieee.org/

Following the PRISMA methodology, the next step was to define the inclusion and exclusion criteria, presented in Table 4, for the articles identified in the search. This crucial step ensures that the study remains within the desired scope and adheres to specific criteria and format during the screening and eligibility phases.

Table 4 – Systematic Review's Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Any scientific article showing evidence of AI utilization in Sports Injuries	Papers focusing on Sports Injuries but without focusing on AI techniques utilization
Paper must be a peered reviewed conference or journal paper written in English	Articles not in English and duplicate papers
Paper is published between 2020 and 2024	Articles published before 2020
	Non-academic or non-scientific papers (e.g., websites, magazines reports, newspapers, consulting articles, books, citations)
	Papers with titles outside the scope of this work

Following the PRISMA workflow's Identification phase, a total of 490 articles were retrieved from three resource databases using the search string. After removing 91 duplicates, 399 records remained.

During the Screening phase, inclusion and exclusion criteria were applied to ensure the articles were from 2020 to 2024, written in English, published as scientific articles or conference papers, and relevant to the scope of the research. This step eliminated 160 records, leaving 239 for further review.

In the Eligibility phase, the first step involved assessing the titles of the articles and removing those unrelated to the study topic, resulting in 192 removals and leaving 47 articles. The second step involved evaluating the abstracts and main texts of the remaining articles, excluding those that did not contribute to the use of AI in preventing sports injuries or addressing research questions. Inaccessible and retracted articles were also excluded at this stage. This phase excluded 32 records, concluding the workflow process with a list of 15 articles included in the study. The described PRISMA workflow is represented in Figure 1.

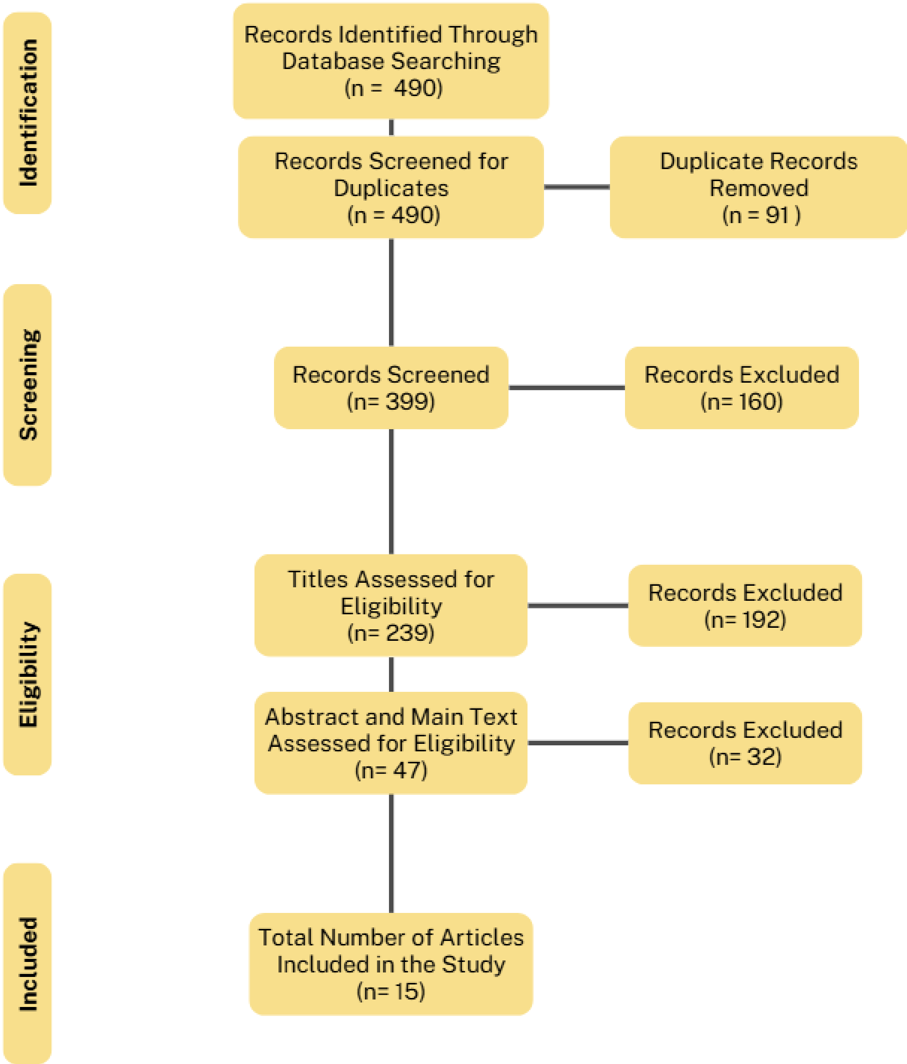


Figure 1 – PRISMA Execution

In Table 5 a list of 15 articles from the PRISMA workflow is presented, including 13 journal articles and two conference papers. Each entry includes a brief description of the article's contribution, conclusions, gaps, and future work suggestions.

Table 5 – PRISMA Results Table - Included Articles

Authors	Article	Contribution	Publication Type
(Castellanos et al., 2021)	*Predicting Risk of Sport-Related Concussion in Collegiate Athletes and Military Cadets: A Machine Learning Approach Using Baseline Data from the CARE Consortium Study*	The article studies a ML predictive model using the Support Vector Machine (SVM) model to assess/predict the risk of an athlete sustaining a concussion based on baseline data across various sports. Sports-specific models were considered but performed similarly to the Single Task Learning (STL) model approach that was assumed to have the same risk and protective factors. Concussion risk factors are discussed.	Journal Article
(He, 2021)	*Prediction Model of Juvenile Football Players' Sports Injury*	This article explores the nature and causes of sports injuries and uses a text classification model based on machine	Journal Article

Authors	Article	Contribution	Publication Type
	Based on Text Classification Technology of Machine Learning	learning to create an injury prediction in football. Data is collected through surveys and integrated equipment to measure exercise volume and load. This helps identify the causes and impacts of injuries for future preventive effects.	
(Hecksteden et al., 2023)	*Forecasting football injuries by combining screening, monitoring and machine learning*	This article explores how screening and monitoring features, within the concept of the "Web of Determinants", contribute to predicting injury risk in professional football. Utilizing a Gradient Boosting model, the study focuses on acute and non-contact injuries. It underscores the added value of combining screening and daily monitoring data, rather than relying solely on monitoring data. Additionally, the study emphasizes the importance of continuous model refinement and validation.	Journal Article
(Kolodziej et al., 2023)	*Predictive modeling of lower extremity injury risk in male elite youth soccer players using least absolute shrinkage and selection operator regression*	Through Lasso regression, the three most significant predictors of injury are identified, and using Lasso model for a classification task, injury outcomes of non-contact lower body extremity injuries in football are predicted. Laboratory-based data collection, including 3D motion analysis, postural control testing, and strength testing, is conducted before the season begins. The study highlights the potential of collecting daily internal and external workload data using Global Positioning System (GPS) tracking technology to produce real-time insights. The most common injuries observed were ankle sprains and muscle strains. The authors suggest that future evaluations of the model should involve larger sample sizes.	Journal Article
(Liu et al., 2023)	*Sports Injury Prediction in Professional Tennis*	This paper addresses two different approaches in predicting injuries in Tennis. It uses ML Decision trees approach whereas the best model was CatBoost and assesses the probability of a player incurring an injury on the following month. It also finds the Pattern Mining model based on Linear Temporal Logic (LTL) learning approach efficient and interesting alternative method, finding temporal patterns in players' injury history data. Three types of data were collected, player features, injury history, and historical match data.	Conference Paper
(Lu et al., 2022)	*Machine Learning for Predicting Lower Extremity Muscle Strain in National Basketball Association Athletes*	This article creates an injury predictive model on Lower Extremity Muscle Strain (LEMS) in basketball. The best performing model was the Extreme Gradient Boosting (XGBoost) that used 20 years of data from publicly available online platforms. It was found that a recent back injury significantly increased the risk of LEMS. The final predictive model was integrated into a web-based application designed for research and educational purposes that generated predictions for probabilities of LEMS.	Journal Article

Authors	Article	Contribution	Publication Type
(Luu et al., 2020)	*Machine Learning Outperforms Logistic Regression Analysis to Predict Next-Season NHL Player Injury: An Analysis of 2322 Players From 2007 to 2017*	This article evaluates various models for the probability of a hockey athlete incurring an injury in the upcoming season on the National Hockey League (NHL). XGBoost emerged as the top performer for both goalies and position players. The data, spanning 10 years, was sourced from two databases. All injury types were considered, with prior injury count identified as the strongest predictor.	Journal Article
(Lyubovsky et al., 2022)	*A pain free nociceptor: Predicting football injuries with machine learning*	This article explores various methods for predicting in-game injuries in American football. Data was gathered from player statistics, training loads, surveys, and various Internet of Things (IoT) wearable devices. Each type of data was analysed using different models. The survey data demonstrated the highest predictive power, and combining player statistics with survey data yielded the best results when using logistic regression. The Long Short-Term Memory (LSTM) model, utilizing only player statistics and survey data, also performed well. The GPS data, however, may not have been as effective due to the nature of the sport and the quality of the data.	Journal Article
(Martínez-Gramage et al., 2020)	*A random forest machine learning framework to reduce running injuries in young triathletes*	Utilize a Random Forest (RF) classification model to analyse triathletes, with a focus on running, to identify the key features influencing the likelihood of injury. These significant contributors to injury risk provide insights into which biomechanical and kinematic patterns can be modified to prevent injuries through a seven-month post-retraining program. The data, sourced from various origins, includes biomechanical data from Surface Electromyography (SEMG) sensors and video recordings.	Journal Article
(Martins et al., 2022)	*Predictive Modeling of Injury Risk Based on Body Composition and Selected Physical Fitness Tests for Elite Football Players*	The study aimed to develop a predictive regression model to estimate the number of potential football injuries per season. Multiple-input single-output models were employed, with the Ridge model emerging as the best, highlighting the most relevant injury risk factors. The data used focused on body composition and physical fitness, and only continuous variables were considered. Additionally, several IoT devices were used to collect data.	Journal Article
(Ren et al., 2024)	*Real-time sports injury monitoring system based on the deep learning algorithm*	This article proposes a sports injury monitoring system that tracks physiological parameters during exercise to assess injury risk. If abnormal conditions are detected, the system sends an alert with preventive measures to avoid injury. The system uses video detection, human key points detection, and polynomial fitting analysis to capture human movements. Among various models tested, the SVM model performed the best. The system was tested in running, aerobic activities, and table tennis.	Journal Article

Authors	Article	Contribution	Publication Type
(She, 2024)	*Application of Big Data Analysis in Model Construction to Prevent Athlete Injury in Training*	In this article, a predictive model is developed to forecast injuries in basketball. Initially, key factors influencing athlete injury risk are identified using univariate analysis. Subsequently, a predictive model is constructed, with the combination of Adaboost and Random Forest algorithms showcasing the best performance.	Journal Article
(Tsilimigkras et al., 2024)	*Enhancing Sports Injury Risk Assessment in Soccer Through Machine Learning and Training Load Analysis*	This study aims to assess the risk of non-contact muscle injuries in football, excluding goalkeepers from the analysis. It utilized physiological and mechanical load variables, recording players' physical activity with wearable GPS devices and heart rate monitors. The seven most relevant features were selected for the final classification model using an SVM classifier. The study highlights a connection between intense running activities and muscle injuries, and the importance of physiological data such as heart rate for the analysis. The Polar Team Pro System was the IoT used in the study to collect data.	Journal Article
(Xu et al., 2023)	*Soccer Sports Injury Risk Analysis and Prediction by Edge Wearable Devices and Machine Learning*	This paper addresses a binary classification problem to predict injuries in football. The best-performing model was an SVM. A wearable sensor was used, tracking real-time movement of athletes during sports activities. Only data from non-contact injuries was used to train the model.	Conference Paper
(G. Yang & Xu, 2022)	*Sequence Video and Artificial Intelligence Assisted Basketball Injury Risk Early Warning Method*	This article explores the use of sequential video and AI to detect early warning signs of injury risk in basketball. The videos are converted into images, and noise is filtered out. A Convolutional Neural Network (CNN) model is employed to identify injury-prone areas. Subsequently, an Adaptive Neural Network (ANN) model evaluates these areas, performs feature extraction, and generates an early warning value.	Journal Article

2.4.3. PRISMA Results Analysis

Upon completion of the PRISMA methodology, performed in the previous section, relevant information has been extracted and is analysed in this section. This information is important to answer systematic literature research questions.

Regarding SLRQ1 - Status of Research in This Area

It is widely acknowledged that the application of AI techniques is still in its early stages. Evaluations of the state of the art in this research topic reveal that many models require further research and development to enhance their results. Nevertheless, promising outcomes have been achieved, offering hope for the future of AI in this field.

It was found that the primary focus is on developing programming models rather than on applying AI technologies in specific contexts. The gathered and analysed studies were tested across seven different sports, with over a third focusing on football. While this concentration is beneficial for the development of AI in football, as new studies can build upon previous ones, it also introduces a potential bias towards the characteristics of football, potentially affecting the applicability of findings to other sports.

Various approaches and hypotheses are being explored, such as assessing injury risk before a season starts (Castellanos et al., 2021), during a season (Hecksteden et al., 2023), or in real-time during activities (Ren et al., 2024). For instance, researchers have tested forecasting potential injuries for a season (Martins et al., 2022), identifying key features influencing injury likelihood (Martínez-Gramage et al., 2020), and using video movement detection to identify risky player movements (Ren et al., 2024), among other methods. Additionally, some studies focus on specific injuries and body parts, including concussion risk (Castellanos et al., 2021), lower body injuries (Kolodziej et al., 2023), while others take a holistic approach, considering all body parts (Luu et al., 2020). This diverse range of research highlights the multifaceted nature of injury prevention and the importance of tailored strategies for different sports and activities.

In the development of these studies, various types of data were collected. The most common type was characteristic data, including standard metrics such as age, weight, height, and BMI. Additionally, some studies incorporated physiological data (Tsilimigkras et al., 2024), physical test data (Xu et al., 2023), questionnaire responses (He, 2021), sensor data (Lyubovsky et al., 2022), and video data (G. Yang & Xu, 2022) for their research. Various articles, for example from Luu et al. (2020), have identified prior injury count as the strongest predictor in the predictive model. These diverse data types provide a comprehensive view of the subjects being studied, allowing for more nuanced analysis and insights.

Regarding SLRQ2 – Major Issues Related to AI in Sports Injuries

There were several common limitations and difficulties encountered during these studies. The most common was regarding data collection, this occurred due to several reasons. The article from Martínez-Gramage et al. (2020) mentions the difficulty in obtaining a big sample of elite athletes, where only 19 athletes were included in the study, resulting in a small sample size. This scarce data problem was connected to the data provider, since various cases did not collect the data themselves, requesting from third parties (G. Yang & Xu, 2022). This scarce data issue was mentioned in various articles. Another challenge in data collection is the limited availability of wearable technologies or sensors, limiting the number of athletes who could use the sensors. This was primarily due to their high costs, Lyubovsky et al. (2022) notes. These limitations may be a factor in some poor performances of some model results affecting model training.

With the rise of innovative technologies, new methods of data collection are emerging. Various studies have utilized sensors and GPS wearables to capture data, aiming to improve predictive abilities for injury occurrence. However, new things do not always mean better. The diversity of data types is vast, and the challenge remains to find the optimal combination of data to achieve the best results, highlights Lyubovsky et al. (2022).

It was commonly identified that achieving optimal results is challenging, as predicting injuries involves various factors. For instance, the evaluation of physical performance related to biomechanical and neuromuscular factors can change or evolve for multiple reasons. This underscores the importance of regular daily data tracking rather than relying on one-time data collection, Kolodziej et al. (2023) points out. Studies, such as those by Hecksteden et al. (2023), concluded the necessity of continuous model refinement and validation whether for testing larger datasets or different types of data.

Regarding SLRQ3 – AI Techniques Currently Useful in This Area

The application of AI techniques to contribute to injury prevention has several approaches for application. It is worth noting that in most articles, several models were tested, and the recommended algorithm(s) for each article was used for further analysis during research.

Among the 15 retrieved articles, it was noticed that 15 of them proceeded with a binary classification model, predicting or assessing whether an athlete would be or was in risk of injury or not. The most common and effective algorithm was the SVM algorithm which outperformed the other models in four different studies (Castellanos et al., 2021; Ren et al., 2024; Xu et al., 2023; Tsilimigkras et al., 2024). Other prevalent algorithms in classification were Gradient boosting (Hecksteden et al., 2023), Catboost (Liu et al., 2023), XGBoost (Lu et al., 2022; Luu et al., 2020), Logistic regression (Lyubovsky et al., 2022), Random Forest (Martínez-Gramage et al., 2020), Lasso (Kolodziej et al., 2023), Adaboost-Random Forest (She, 2024), and CNN and ANN in by G. Yang & Xu (2022). The study from He (2021) leveraged text classification algorithm structured around a vector space model. This approach is designed to analyse various data points related to athletes' training and physical conditions to predict potential injuries accurately.

Interestingly, Martínez-Gramage et al. (2020) focus shifts towards the practical applicability of the modeling results. After testing several models and selecting Random Forest for the study, the variables that best discriminated between injured and non-injured individuals were identified as pelvic kinematics, knee flexion, ankle dorsiflexion at initial contact, and gluteus medius. Following this, a seven-month gait retraining program was conducted, showing positive results with improved running mechanics and a reduction in injuries among triathletes.

It can be observed that the natural way to address injury prevention is through a classification problem. Nonetheless, Martins et al. (2022) utilizes a regression procedure to estimate the number of potential injuries to occur during a football season. In this research, several multiple-input single-output regression-type models were tested, and the Ridge model performed the best. Additionally, Kolodziej et al. (2023) use a regression model as an intermediary step to identify prevalent predictors. In this case, Lasso regression model confidently highlights that concentric knee extensor peak torque, hip transversal plane moment in single leg drop landing task and center of pressure sway in single leg stance test as the three most important predictors for injury.

A unique approach was presented by Liu et al. (2023) that not only leveraged modern machine learning techniques to address a classification problem but also incorporated a pattern mining method using LTL learning. The author addresses the complexity and black-box nature of contemporary ML models, which often make it difficult for players and coaches to interpret the results. To overcome this, a pattern mining approach using LTL learning was proposed. This alternative model automatically extracts player-specific temporal patterns related to tournament schedules over the past three months and their correlation with injuries. This method has proven insightful in understanding the primary causes of injuries, tailored to individual players and different player types.

Regarding SLRQ4 – The Most Suitable AI-Powered Solutions That Can Be Implemented in Different Sports According to Their Unique Characteristics, Focusing on Injury Prevention

The studies that leveraged the SVM model (Castellanos et al., 2021; Ren et al., 2024; Tsilimigkras et al., 2024; Xu et al., 2023) for binary classification, proved to be a reliable solution when paired with advanced data collection tools such as sensors, video data, and wearable GPS trackers (Ren et al., 2024; Tsilimigkras et al., 2024; Xu et al., 2023). The researchers Castellanos et al. (2021) and Ren et al. (2024) effectively assess sports injuries, demonstrating a generalizable approach adaptable to various contexts. Tsilimigkras et al. (2024) and Xu et al. (2023) shift their focus on football, highlighting the model's compatibility with the sport's characteristics and its strong adaptability for predicting injuries in other sports. This adaptability is further enhanced by the model's ability to integrate diverse data sources, making it a versatile tool for sports analytics and injury prevention.

The use of video technology to extract and analyse data has proven to be an effective approach for assessing biomechanical and kinematic patterns in players' movements (Martínez-Gramage et al., 2020; Ren et al., 2024; G. Yang & Xu, 2022). G. Yang & Xu (2022) leveraged two different variants of neural networks, while Martínez-Gramage et al. (2020) and Ren et al. (2024) applied a Random Forest and SVM model, respectively. The sports analysed in these articles varied, including studies on triathletes (focused on running) (Martínez-Gramage et al., 2020), basketball players (G. Yang & Xu, 2022), and sports including table tennis and aerobic exercises (Ren et al., 2024). This indicates the potential of video technology to support AI models in analysing players' movements across different types of sports.

The article from Ren et al. (2024) highlights a promising innovation in real-time monitoring for training and competitive games. This system uses a portable mobile terminal to evaluate the human body during movement, while the main monitoring system collects real-time data metrics during exercise. If an injury risk is detected, the system issues a warning along with preventive measures to reduce or avoid injury. Tested across various sports, this preventive technique shows great potential for widespread implementation.

The pattern mining approach described by Liu et al. (2023) is particularly suited to individual sports, as it leverages a player's physical and mental capabilities to effectively manage tournament scheduling and optimize planning based on performance outcomes. This technique could be studied in sports like track and field and gymnastics, for example.

Regarding SLRQ5 - Advantages and Disadvantages of Applying AI Techniques in This Field

Although some models may perform reasonably well, AI techniques not always are accurate, and false positives or false negatives can occur, Lyubovsky et al. (2022) shows. This variability can lead to misplaced trust or skepticism from athletes and coaches due to the black-box nature of these AI predictors, Liu et al. (2023) remarks.

The data collection process can be exhaustive and raise concerns. Data privacy and ethics, especially when collecting personal and physiological data from athletes, must be handled with care. Following the code of ethics and obtaining informed consent from participants is essential, Martins et al. (2022) indicates. Acknowledging the difficulty in predicting acute injuries due to their unpredictable nature (Kolodziej et al., 2023), some studies, for example Xu et al. (2023) opted to only collect and include non-contact injuries to aim for a more accurate model. This approach, however, limits the models to these types of injuries, which can be a disadvantage since acute injuries constitute a significant percentage of total sports injuries and are unavoidable, as Lyubovsky et al. (2022) mentions. Nonetheless, the article from Hecksteden et al. (2023) focuses on predicting both acute and non-contact injuries in their solution.

More than half of the studies analysed leveraged the use of technologies to collect data, such as sensors, GPS wearable devices, or video. Ren et al. (2024) points out that reliance on these tools creates a dependency on AI and technology, potentially diminishing the use of human intuition, which can sometimes detect injury symptoms and indicators that machines cannot. Additionally, the acquisition, implementation, and maintenance of such technologies can be expensive and inaccessible for small teams and organizations (Lyubovsky et al., 2022), forcing them to rely on traditional methods. Castellanos et al. (2021) states that technical issues, such as malfunctions or data inaccuracies, can also negatively impact the effectiveness of injury prevention.

On the positive side, AI techniques offer numerous benefits in sports. By analysing vast amounts of data, AI can identify patterns and predict potential injuries before they occur (Liu et al., 2023). This allows for early intervention and the creation of personalized training programs. For example, as seen in the research by Martínez-Gramage et al. (2020), key injury indicators were identified, and preventive measures were successfully applied, leading to improvements in athletes' performance and safety. Additionally, as seen in the article by Ren et al. (2024), leveraging wearable technology and advanced sensors enables real-time monitoring, providing immediate feedback during training sessions or competitive events.

Preventing injuries is similar to safeguarding the future. Data-driven decisions help prevent unfortunate incidents while enhancing athletes' safety (She, 2024). Although investing in these AI systems may cause some controversy within the organization, they can significantly reduce medical and treatment costs that would arise without preventive measures (He, 2021). Moreover, they help maintain higher performance levels and reduce the need for athletes to take days off due to severe injuries (Martínez-Gramage et al., 2020).

3. RESEARCH METHODOLOGY

The primary objective of the Design Science Research (DSR) methodology is to create artefacts. This dissertation aims to develop a comprehensive framework that identifies the most appropriate AI-powered solutions for various sports, focusing on injury prevention. Since a framework is considered an artefact, the DSR methodology is the most suitable approach.

3.1. Design Science Research

Design Science Research (DSR) is a problem-solving methodology aimed at enhancing human knowledge through the creation of artefacts. This approach fosters innovation and supports the sustainable transformation of society. The DSR methodology consists of six main phases (vom Brocke et al., 2020):

1. Problem Identification and Motivation.
2. Definition of Objectives.
3. Design and Development of Artefacts.
4. Demonstration.
5. Evaluation.
6. Communication.

Each step is designed to guide the researcher through their work in a step-by-step process, promoting an organized thought process and ensuring the topic's relevance. Figure 2 illustrates the framework and its steps (Peffers et al., 2007):

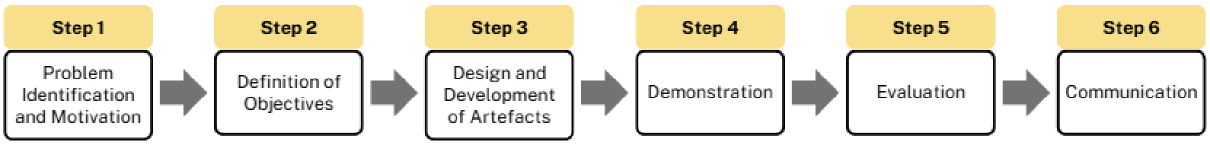


Figure 2 – Design Science Research Phases

1. **Problem Identification and Motivation -** Define the specific research problem and explain the value of a solution to the context of study. It’s required to have knowledge of the state of the problem and the importance of its solution.
2. **Definition of Objectives -** Set the objectives for a solution based on the problem scenario and an understanding of what is possible and feasible. (Quantitative or Qualitative objectives).
3. **Design and Development of Artefacts -** Design and develop the artefact, this includes determining its desired functionality, its architecture, and creating the actual artefact. For this step, it’s required to have knowledge of theory that can be applied to propose the solution.

4. **Demonstration -** Demonstrate how the artefact can solve one or more parts of the problem. This step requires to have effective knowledge of how to use the artefact and apply it to the problem context.
5. **Evaluation -** Evaluate how effectively the artefact addresses the problem by comparing the solution's objectives with the observed results from the artefact demonstration. This step requires specific metrics for evaluation. At this point, the researcher can either return to step three to improve the artefact's effectiveness or proceed to the next step.
6. **Communication –** Present the problem and its significance, share the findings, discuss the artefact's impact, and outline any limitations to relevant stakeholders. This step requires an understanding of the disciplinary culture.

While the process is typically structured in a sequential order, researchers don't always need to follow it from step one to step six. For a problem-centered approach, starting at step one is expected. An objective-centered solution might begin at step two, while a design- and development-centered approach could start at step three. Additionally, a client- or context-focused solution may start at step four, resulting in a reverse process. The order can vary depending on the research focus, context, and goals (Peffers et al., 2007).

3.2. Research Strategy

The research strategy employed for the development of this research followed the DSR methodology and was developed as follows.

Starting from step one, the development of this dissertation began by conducting a literature review of the main topics and studying how AI technology is currently helping to prevent injuries in the sports world. This topic is relatively recent, with limited research available. There are also many factors related to sports injuries, creating some challenges for the development of this topic. Therefore, it was of great importance to further analyse this topic and find measures to promote a safer environment for athletes across various sports.

In phase two, the major events to be accomplished were defined. This phase started with studying and understanding the current state of the art and analysing relevant health factors and metrics. This provided a clear vision for addressing the research gap, with the main objective being to propose a framework to achieve this goal.

For phase three, the design and development of the artefact began by defining core assumptions regarding algorithms and IoT devices related to sports injuries, based on literature. This was followed by developing the artefact's architecture and outlining its functionalities for various types of sports, guiding each decision for future applications. Considering the variety of factors to be accounted for in building the solution, two approaches were considered: a tree-based solution where each sport and injury follows a path to finding the optimal approach, or a unique solution capable of preventing injuries across all sports. Ultimately, the first option was chosen.

In the demonstration phase, a use case focused on the research context was conducted to showcase how the framework works and can be applied in a real-case scenario to prevent injuries. The next step was to evaluate how the framework addressed the problem and met the defined objectives, by conducting interviews with specialists from different professional areas. The answers were analysed and used for further improvements.

To conclude, in the last phase, the relevant findings and main takeaways from the research were shared and explained, including the limitations encountered during its development and considerations for future research and development.

4. EMPIRICAL STUDY

This chapter focuses on the artefact, from its development to validation. Assumptions are constructed based on the literature review, followed by an explanation of the framework, including a fictitious use case to demonstrate its applicability. An evaluation step through interviews is then performed, and the results are analyzed and discussed.

4.1. Assumptions

In Chapter 2, a meticulous literature review was conducted, providing essential information to support the framework's assumptions. At this stage, it was concluded that there is no obvious solution to all cases. As previously observed, many different approaches can be used for the same sport. Therefore, each one's characteristics and goals must be considered when building a comprehensive framework.

Section 2.1.2 identified that sports can be classified as Team Sports, Individual Sports, Contact Sports, or Non-Contact Sports. These classifications were applied to each sport in this section. Additionally, Section 2.2.2 examined injury classifications, revealing various levels of granularity. Given the specificity of the articles reviewed, classifying injuries by their anatomical location was considered the most appropriate approach for the forthcoming structure.

Predictive Techniques

In Table 6, following the analysis of the literature review articles, a summary was created to identify which predictive techniques were used for each type of sport and injury location. This provides an overview of the characteristics of the previously conducted studies and highlights the most used predictive techniques.

IoT Devices

Similarly, the same table structure was used in Table 7 to identify the types of IoT devices utilized across all 15 articles. In some cases, no IoT devices were used, while in others, multiple devices were employed. These devices varied in type, with their primary function being data collection to enhance the predictive model with additional features. In some cases, the use of IoT devices with predictive models incorporated were used to directly evaluate the risk of injury.

Table 6 – Algorithms Recommendations

Predictive Techniques		**Team Sports**		**Individual Sports**		**Not Specific/ Various**
		Contact Sports	Non-Contact Sports	Contact Sports	Non-Contact Sports	
Anatomical Location	Head and Neck	Gradient Boosting, XGBoost, Logistic Regression, Ridge, Adaboost-Random Forest, SVM, CNN, ANN			SVM, CatBoost, Pattern Mining	SVM
	Upper Limb	Gradient Boosting, XGBoost, Logistic Regression, Ridge, SVM, Adaboost-Random Forest, CNN, ANN			SVM, CatBoost, Pattern Mining	SVM
	Trunk	Gradient Boosting, XGBoost, Logistic Regression, Adaboost-Random Forest, SVM, CNN, ANN			Random Forest, SVM, CatBoost, Pattern Mining	SVM
	Lower Limb	Gradient Boosting, Lasso Regression, XGBoost, Logistic Regression, Ridge, SVM, Text Classification, Adaboost-Random Forest, CNN, ANN			Random Forest, SVM, CatBoost, Pattern Mining	SVM
	Unspecified Area;					
	Multiple Regions					
	All Body Parts	Gradient Boosting, XGBoost, Logistic Regression, Adaboost-Random Forest, SVM, CNN, ANN			CatBoost, Pattern Mining, SVM	SVM

Table 7 – IoTs Recommendations

<table>
<tr><th colspan="2" rowspan="2">IoTs</th><th colspan="2">Team Sports</th><th colspan="2">Individual Sports</th><th rowspan="2">Not Specific/ Various</th></tr>
<tr><th>Contact Sports</th><th>Non-Contact Sports</th><th>Contact Sports</th><th>Non-Contact Sports</th></tr>
<tr><td rowspan="6">Anatomical Location</td><td rowspan="2">Head and Neck</td><td>2D Video Recordings, InBody 770, Jamar Plus+, Optojump Next System, Witty-Gate Photocells, Polar Team Pro System, Catapult - Playertek</td><td rowspan="2"></td><td rowspan="2"></td><td>2D Video Recordings</td><td>2D Video Recordings</td></tr>
<tr><td>IoT with Predictive Model: Omegawave Sensor</td><td>IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System</td><td>IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System</td></tr>
<tr><td rowspan="2">Upper Limb</td><td>Wearable Sensor, 2D Video Recordings, InBody 770, Jamar Plus+, Optojump Next System, Witty-Gate Photocells, Polar Team Pro System, Catapult - Playertek</td><td rowspan="2"></td><td rowspan="2"></td><td>2D Video Recordings</td><td>2D Video Recordings</td></tr>
<tr><td>IoT with Predictive Model: Omegawave Sensor</td><td>IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System</td><td>IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System</td></tr>
<tr><td rowspan="2">Trunk</td><td>2D Video Recordings, Polar Team Pro System, Catapult - Playertek</td><td rowspan="2"></td><td rowspan="2"></td><td>SEMG Sensors, 2D Video Recordings</td><td>2D Video Recordings</td></tr>
<tr><td>IoT with Predictive Model: Omegawave Sensor</td><td>IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System</td><td>IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System</td></tr>
</table>

	Lower Limb	Wearable Sensor, Qualisys - 3D motion capture system, Pegasus 3-D system, IsoMed 2000 Isokinetic Dynamometer, 2D Video Recordings, InBody 770, Jamar Plus+, Optojump Next System, Witty-Gate Photocells, Polar Team Pro System, Catapult - Playertek			SEMG Sensors, 2D	2D Video Recordings
		IoT with Predictive Model: Omegawave Sensor			IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System	IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System
	Unspecified Area;					
	Multiple Regions					
	All Body Parts	Catapult – Playertek, 2D Video Recordings, Polar Team Pro System			2D Video Recordings	2D Video Recordings
		IoT with Predictive Model: Omegawave Sensor			IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System	IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System

In the analysed articles, 15 IoT devices were identified. Table 8 classifies them by their data extraction characteristics and purposes: GPS/ Motion, Video Data, and Physiological Data. Additionally, the column "IoT with Predictive Model" is marked when an IoT has predictive capabilities for assessing injury risk.

It is worth noting that the IoT "Wearable Sensor" and "2D Video Recordings" are generic devices as the articles in the literature did not specify the brand and model. The "Proposed idea: Portable + Main Monitoring System" is a proposed system idea not available on the market and subject to continuous research.

Table 8 – IoTs Characteristics

Devices (IoT)	GPS/ Motion	Video Data	Physiological Data	IoT with Predictive Model
Wearable Sensor (GPS, Accelerometer)	X			
Qualisys - 3D Motion Capture System (Consisting of 12 Infrared Cameras 120 Hz)	X	X		
Pegasus 3-D System (Strength Testing)			X	
IsoMed 2000 Isokinetic Dynamometer (Strength Testing)			X	
Omegawave Sensor (Brain, Heart, Nervous and Cardiac System)			X	X
SEMG Sensors (Bioelectrical Signals from the Body)			X	
2D Video Recordings (Camera)		X		
InBody 770 (Body Composition)			X	
Jamar Plus+ (Hand Dynamometer)			X	
Optojump Next System (Optical Measurement System)			X	
Witty-Gate Photocells (Timing System)			X	
Proposed Idea: Optical Wearable Sensor (Motion, Physiological Data, Framework of Sensors)	X		X	
Polar Team Pro System (GPS, Accelerometer, Heart Rate)	X		X	
Catapult - Playertek (GPS, Accelerometer, Heart Rate)	X		X	

4.2. Framework

Upon collecting the necessary information in the literature review followed by a structured and organised set of ideas in the assumptions sector, conditions are met to develop the framework. A high-level structure is presented to provide a logical structure and serve as a guideline for data professionals, aiming to support them in generating insightful information with evidence-based support.

The final solution features a tree-based guided process, where each sport and injury follow a path to find the optimal approach tailored to the requirements and scope, aiming to develop the best possible product for injury prevention.

It is important to note that implementing this framework requires technical knowledge in data science to complete its various steps. Additionally, external research on specific IoTs is necessary, as this dissertation does not fully cover them. Furthermore, the required information to successfully complete the workflow should be transparently shared between entities. Finally, the prices of IoT devices, infrastructure, etc., are not considered in this study as they can become outdated in a short period of time and depend on several factors.

The framework is composed of four main steps (Figure 3), each focusing on different aspects that are further unfolded in Figure 4. The steps are as follows:

1. Step 1 - Context Identification.
2. Step 2 - Anatomical Location.
3. Step 3 - Data Requirements.
4. Step 4 - Final Solution.

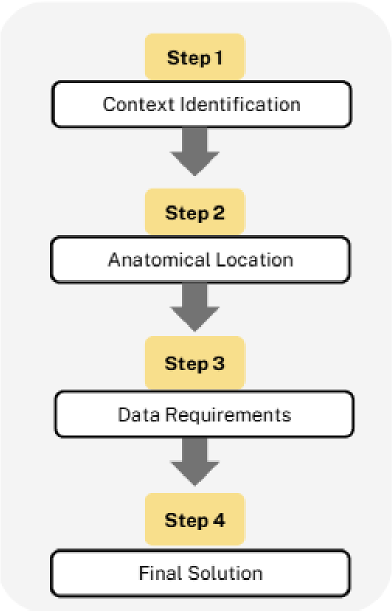


Figure 3 – Framework Step

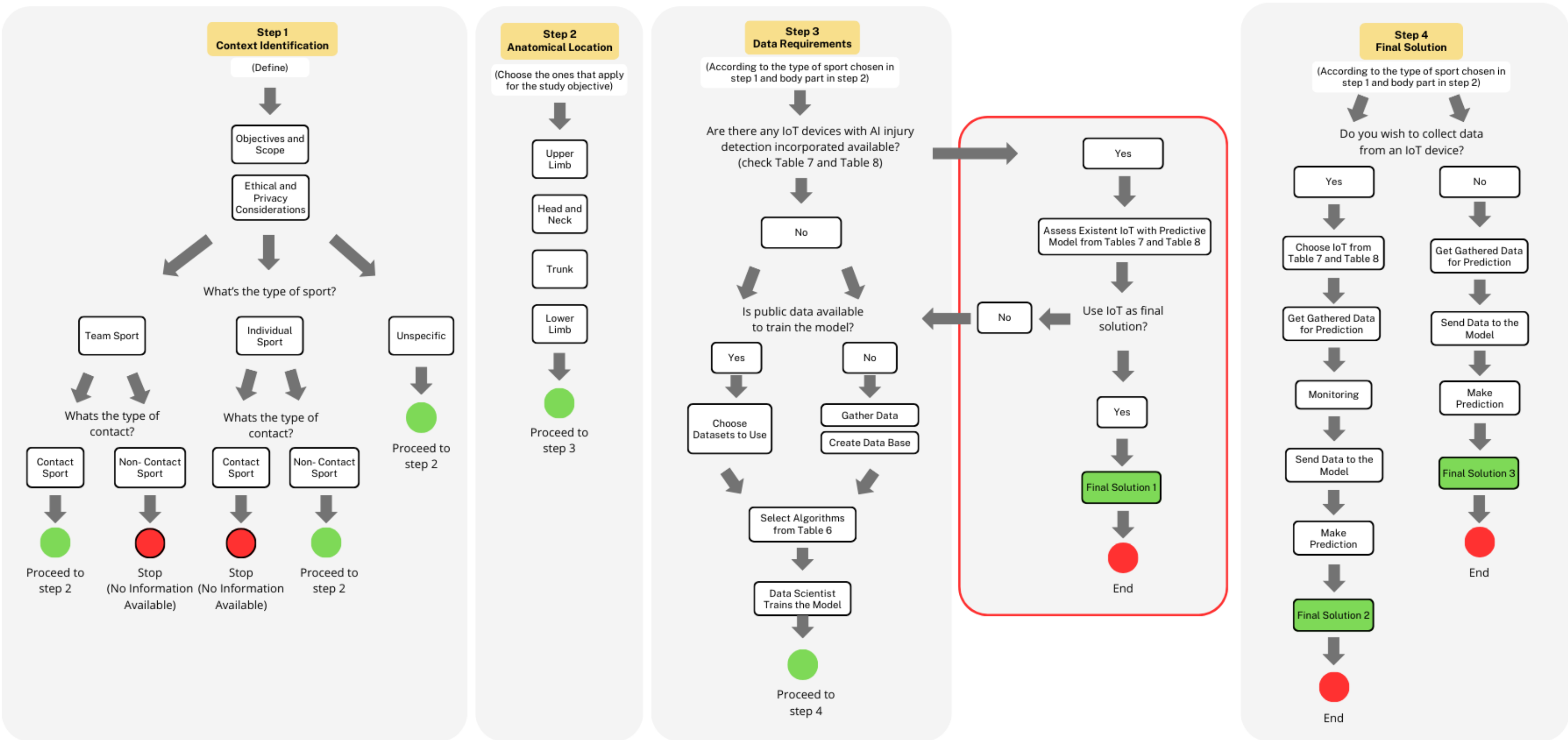


Figure 4 – Framework Steps Detailed Version

4.2.1. Step 1 - Context Identification

The first step of the workflow involves identifying the context problem. The user should outline clear objectives, whether the goal is to reduce injury risk rates, improve recovery times, or enhance overall athlete performance.

Having qualified personnel, such as data scientists, sports scientists, and medical professionals, adds significant value during the development process. It’s also important to ensure the necessary conditions, including adequate infrastructure, to support the solution’s development. Since athletes’ data will be collected, it's crucial to inform them about how it will be used and to obtain their written consent.

The next step is to classify the type of sport under study. It starts by determining whether it is a team sport, an individual sport, or if the sport is not specified. If it is unspecified, the user proceeds directly to step three. For team or individual sports, the next question to address is the type of contact involved, distinguishing between contact and non-contact sports.

If it is a team sport and a contact sport, or an individual sport and a non-contact sport, proceed to step three. However, if it is a team sport and a non-contact sport, or an individual sport and a contact sport, the process is terminated due to the lack of information regarding the applicability of AI to these types of sports.

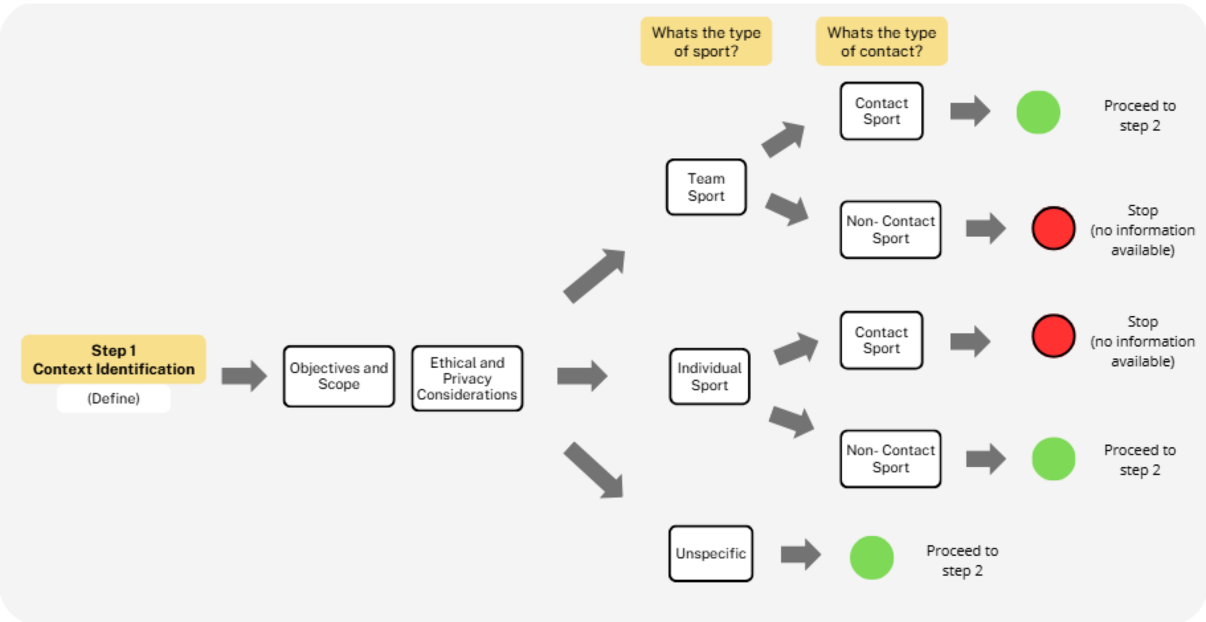


Figure 5 – Framework Step 1 - Context Identification

4.2.2. Step 2 - Anatomical Location

In step two, the anatomical location is chosen based on the study objective. Researchers may focus on a specific body part or cover the entire body. It's important to have this defined and clear. This simple step is then followed by step three.

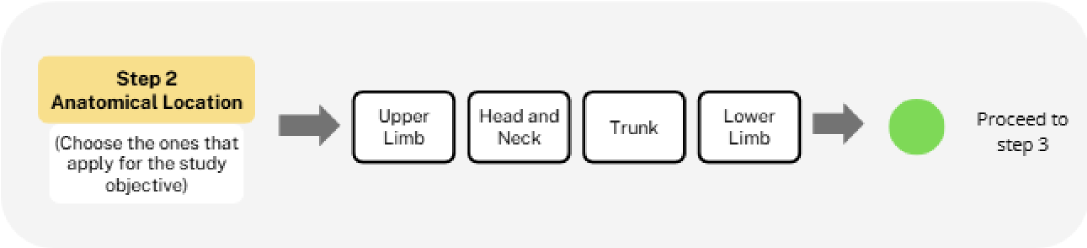


Figure 6 – Framework Step 2 - Anatomical Location

4.2.3. Step 3 - Data Requirements

Step three begins by branching into two paths, depending on whether IoT devices with built-in predictive models for assessing injury risk are available. If such devices are available, the user exits the workflow and can evaluate the IoT devices with AI capabilities in Table 7 and Table 8. The user then decides whether to adopt the device as the final solution, if he chooses to use those devices, the process concludes with Final Solution 1. Contrarily, he can opt by returning to the workflow, potentially using the IoT device with AI as a complementary tool, choosing it further in step four, or not using it at all.

To ensure a smooth and effective process, it's essential to define the type of data and key metrics to guarantee quality for the subsequent steps. This data may include characteristics, psychological information, physiological metrics, and other types. The user assesses if public data is available and if so, chooses a dataset containing the metrics previously defined. If there is no public data available, the user should gather data and create a database that meets the necessary data requirements allowing him to progress on the workflow. This step may require subsequent steps including creating a metadata, requirement analysis, among others.

After having disposable data, according to the type of sport and body part under study, Table 6 serves as a support indicating a set of recommended algorithms to test. After some testing the best performing model is chosen and used to train the model by the data scientist. It is important to note that essential technical steps are required during model training, which are not specified in this framework.

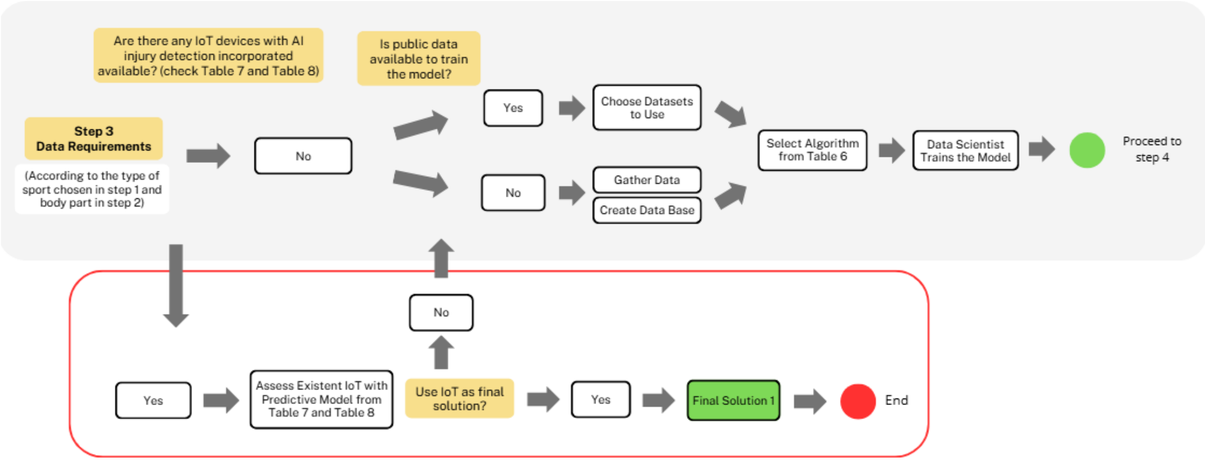


Figure 7 – Framework Step 3 - Data Requirements

4.2.4. Step 4 - Final Solution

The last step starts by evaluating whether the user wants or needs to collect data from IoT devices. If yes, Table 7 and Table 8 are recommended to evaluate the best device for each case, providing suggestions based on the type of sport, body location, and data to be extracted.

The following activities encompass gathering data for prediction, perform monitorization (with IoTs) for additional information, and sending all data to the model to perform the prediction. This ends with the Final Solution 2.

Similarly, if the user wishes not to use any IoT device for additional information, the following activities are gathering the data from the athletes, sending it to the model, and concluding with the prediction. This ends the with the Final Solution 3.

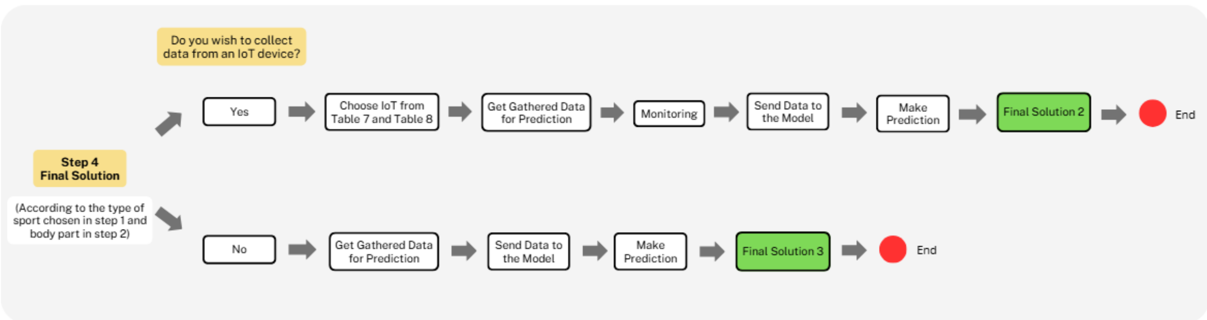


Figure 8 – Framework Step 4 - Final Solution

4.3. Use Case

The purpose of this subsection is to provide a detailed understanding of the framework's functionality and its potential impact on injury prevention. To achieve this, a fictional use case is presented, illustrating the step-by-step process, highlighting the benefits of implementing such a system, and clarifying any doubts while obtaining valuable feedback.

The NOVA Football Team wants to reduce the rate of injury of their players to minimize days off due to injury, therefore improving overall performance and improving the chances of winning the league. The team decided to consult Sports AI to find a solution to make this happen. Sports AI decided to implement an injury risk prediction based on an artificial intelligence solution, following the framework presented in this dissertation. The most suitable solution is assessed to produce insightful information for the team.

Before the implementation of the framework, some acknowledgments should be addressed to the NOVA football team:

- The customer must be available for regular contact with Sports AI;
- The results won't lead to a 100% chance of accuracy as there is a margin for error;
- Results are not instant as they may take some time to see significant results;
- The final solution may and can be subject to modifications.

Starting the implementation of the framework and to ensure the best possible outcome, the NOVA Football Team must have full transparency with the information asked, considering the complexity of the workflow. To conduct this analysis, the framework presented in Figure 4 was consulted. Additionally, Table 6, Table 7, and Table 8 are subject to consultation as support information.

Table 9 – Use Case Step 1 - Context Identification

Step 1 - Context Identification	
Objectives and Scope	Identify athletes at risk of injury in the lower extremity of the body by leveraging predictive capabilities and open to use IoT devices to monitor their physical activity and key health metrics. The results will offer actionable insights and recommendations to athletes, coaches, and medical staff to adjust training regimens and implement preventive measures.
Ethical and Privacy Considerations	Full consent from the athletes for Sports AI to collect data throughout the ongoing project
Type of Sport?	Team Sport
Type of Contact?	Contact Sport

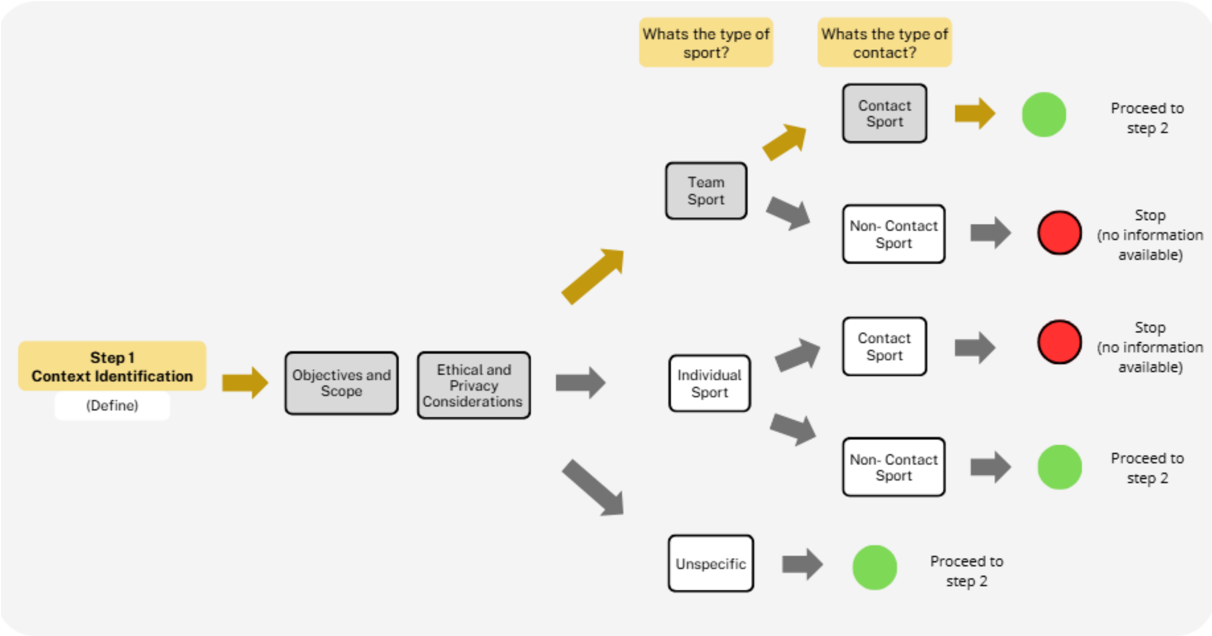


Figure 9 – Use Case Step 1 - Roadmap

In step one the objectives and scope were outlined, ethical and privacy matters were ensured, and the type of sports was defined as Team Sports, and Contact Sport. The process proceeds to step two.

Table 10 – Use Case Step 2 - Anatomical Location

Step 2 - Anatomical Location	
Anatomical Location	Lower Limb

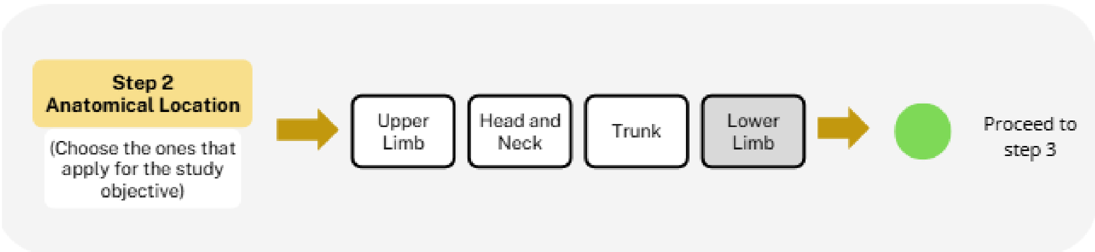


Figure 10 – Use Case Step 2 - Roadmap

In step two the anatomical area chosen was solely the Lower Limb region and the process proceeds to step three.

Table 11 – Use Case Step 3 - Data Requirements

Step 3 - Data Requirements	
Are IoT Devices with AI Injury Detection Available?	Yes
Use IoT as Final Solution?	No
Is Public Data Available?	Yes
Select Algorithms from Table 6	Gradient Boosting, Lasso Regression, XGBoost, Logistic Regression, Ridge, SVM, Text Classification, Adaboost-Random Forest, CNN, ANN SVM performs the best and it's the chosen one

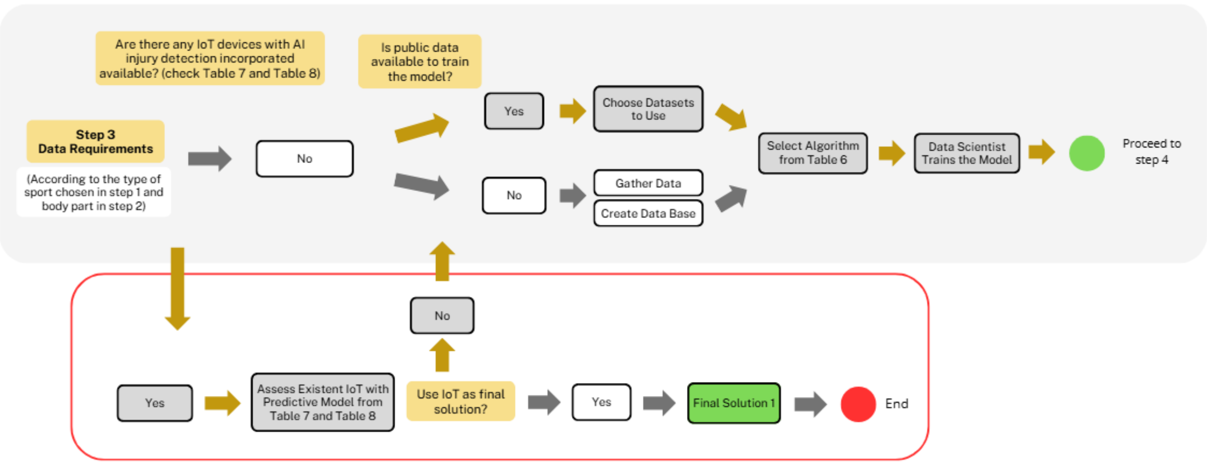


Figure 11 – Use Case Step 3 - Roadmap

In the third step, it is noted that there are viable IoT options with AI available for implementation. However, after evaluating these options with the client, the user decides to go back to the workflow for a more personalized approach. As there is available data to train the model, the user selects a publicly available dataset with the desired metrics for the next steps. Following the guidelines in Table 6, the most prevalent algorithms are carefully assessed, and the SVM model, being the best performer, is chosen to proceed to step four.

Table 12 – Use Case Step 4 - Final Solution

Step 4 - Final Solution	
Do You Wish to Collect Data from an IoT Device?	Yes
Choose IoT from Table 7 and Table 8	Wearable Sensor, Qualisys - 3D motion capture system, Pegasus 3-D system, IsoMed 2000 Isokinetic Dynamometer, 2D Video Recordings, InBody 770, Jamar Plus+, Optojump Next System, Witty-Gate Photocells, Polar Team Pro System, Catapult – Playertek, Omegawave Sensor <u>Polar Team Pro System and Omegawave sensor are the chosen IoT devices</u>

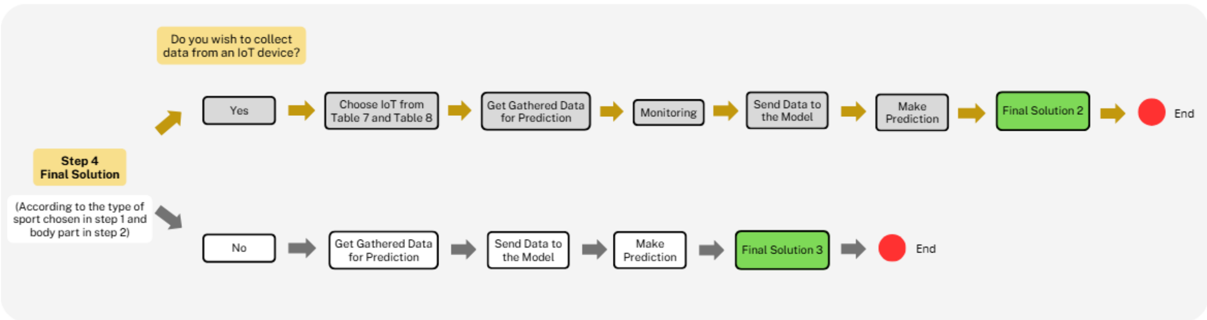


Figure 12 – Use Case Step 4 - Roadmap

In the final step, the user opts to use IoT devices to assist with data collection. Following the guidelines in Table 7, the recommended IoT devices are identified. Using additional inputs from Table 8, the Polar Team Pro System and Omegawave sensor are selected meeting the metrics requirements.

The next steps involve gathering the necessary data and conducting the monitoring phase with the IoT devices. Finally, the collected data is sent to the model to perform the prediction, completing the workflow process.

Having successfully terminated every step of the workflow, the final solution suggests gathering data from public sources with the desired metrics and choose the algorithms that works the best following the Table 6 recommendation. The training phase was performed with the SVM model performing the best.

As the client wants to use an IoT device in the study, following the recommendations from Tables 7 and Table 8, the Polar Team Pro System and the Omegawave Sensor were identified as a good fit for the sport's context and requirements, contributing to the study as comprehensive IoT solutions providing important metrics for analysis. After monitorization, the model is able to make the prediction providing insightful information for the team, concluding the workflow.

4.4. Evaluation

Following the DSR phases presented by Peffers et al. (2007) in Section 3.1, the evaluation phase, which assesses the usability of the proposed framework, involved interviewing three participants. This step was invaluable for obtaining feedback from professionals with diverse specializations, who shared their thoughts and improvement suggestions on the current model. A brief description of the three experts' backgrounds and domains is presented in Table 13.

Table 13 – Interviewees Description

Interviewee	Background	Domain
Data Scientist	Works as Data Scientist for more than 5 years	Data Scientist/ Machine Learning Engineer at Redlight Software
Coach	Worked with all indoor sports and is currently working with the Women's Football U19 team as Physical Trainer	Physical Trainer at Sport Lisboa e Benfica
Athlete	Competed at the Beijing Olympic Games 2008 in Badminton and is now a full-time Medical Doctor	Former Olympic Athlete in Badminton and Medical Doctor

The interviews began with an introductory presentation outlining the research objectives and providing context. The framework was then presented, explaining its application in real-world scenarios. This was followed by the focus questions listed in Table 14. The slides of the presentation are available in Appendix D.

Table 14 – Interview Questions

Question 1	Do you consider the proposed framework useful? Why?
Question 2	Would you consider implementing the proposed framework? Why?
Question 3	Do you have any recommendation or suggestions for further improvements of the proposed framework?

The interviews were conducted via Zoom, with immediate transcription to capture the essence of each response, and each participant consented to the recording and documentation of their answers for research purposes. The complete responses are available in the Appendixes section (Appendix A, Appendix B and Appendix C), while this section includes only highlighted parts from the answers to the questions.

Regarding Question 1 - The Answers Were the Following

Data Scientist:

“I think that in terms of usefulness it makes perfect sense. It is a framework that is solving a problem that exists and will continue to exist in sport. It's flexible to the point of being able to have different approaches, which is also natural taking into account, if you want to make a framework that is adaptable to any sport, you also have to be able to handle different types of inputs and outputs.

I think it's particularly interesting out there, both the connection with IoTs, which all high-level sportsmen usually use in one way or another (…) I think it gives a very good and impactful notion of what the sporting life of people whose coaches or whose structure uses this framework can be."

Coach:

“I consider it useful in the sense that you identify a problem and then from the various areas you have, you can deepen. (…)

(…) You can get to the "best" models more easily. Because there are several algorithms, but one can work for one type of sport and not work for another, and if we try to go directly to the solution, which is to find the best model, maybe we are not applying the right one."

Athlete:

"Yes, I think it has a very great potential to be used, even on a day-to-day basis. (…) you have more data from team sports, but I think that in the future there will also be more and more individual sports, and this can even be very useful in trying to predict what the moment of risk is for each athlete and try to find solutions for that. (…)”

Regarding Question 2 - The Answers Were the Following

Data Scientist:

“Yes, I feel that although there is a lot of scientific knowledge about injury prevention and about the human body and anatomy, the truth is that each case is different and has a lot to do with the specific context of the athlete at a given time (…)

A framework that helps those who plan training to understand the patterns of the athlete and to be able to pick up on patterns of athletes who came before and who went through the same type of training I think is very beneficial because statistics in this case will help a lot.

Something that can take and generalize a little bit in the predictive way of the issue I think is great, I think any athlete would benefit and well used, it would have benefits both for the athletes themselves and for the clubs that obviously manage to have the athletes perform well without injuries, without stops that are obviously costly."

Coach:

"I would implement it because there is a huge variety of injuries in football, and there are so many anatomical structures that are in contact and that an injury can occur, that we, being one or two people doing this work, is a lot of work for one person. (…)

(…) I think it would be useful in this sense to have models to help and facilitate our work, and to make us more effective in other areas.

I think that in the world of professional sport, because I am also in contact with professional sport in football, it is very much done "old-fashioned". We look at the data and start to see patterns, but it's very naked-eye and it takes a lot of manpower. It's not so certain, nor can we look at things so globally.

(…) Every type of collection we do is for descriptive analysis. After that, we can, for future cases, try to act differently, but we don't have predictive models."

Athlete:

"Yes, for the same reason. I think it can be very useful in trying to identify the moments when each athlete, whether team sports or individual sports, can be in a moment of risk. Whether due to training overload, tournament overload, or others... (...)"

Regarding Question 3 - The Answers Were the Following

Data Scientist:

"(…) one thing you can do is organize the algorithms by groups, by clusters of algorithm types and simplify the table a little bit, because pointing out to each sport what type of algorithms, can make it a little easier for those who are going to apply it (…)

You could also add in step three the option to use any of the models that are already available, that are already trained, and say what type of input, data, is compatible with that same model. As a result, the "Train the model" step would be optional."

Coach:

“After step two, I think I would put an intermediate step, to differentiate the type of injuries. (…)

(…) For example, instead of moving on to the model, identify first the structure, and then the type of injury, whether muscular, joint, etc., within that same structure."

Athlete:

“It is important to detail the periods of greatest stress for each sport. That is, physical stress, which can condition the athlete's greater risk of developing an injury. (…)

(…) Individual factors are important even for team sports. (…) For example, genetic factors. Your body as an athlete is different from anyone else. And you may have a greater or lesser predisposition. At the moment, not having access to the person's genetic data, it is not yet possible to implement it.

But it may be that in the future, 10 years at most, we will be able to have this technology more accessible and add to other data (…) I think this can be a very important step. The genetic part counts a lot. (…)”

4.5. Discussion

After evaluation of the artefact, this section presents a careful analysis of the feedback received. The discussion focuses on the utility of the artefact, the viability of its implementation, and possible improvements based on interview responses. Additionally, some general observations from the participants are included.

Regarding the utility of the proposed framework, all three participants found it useful and highlighted its potential for implementation. Its importance in addressing the existing problem of injuries in sports was emphasized, noting its flexibility and adaptability across various sports and contexts. The incorporation of IoTs in the framework was seen as a significant step, as high-level athletes generally integrate these technologies into their training processes.

The benefit of the step-by-step process was also highlighted, starting from a macro perspective and gradually delving deeper to address the problem. This approach guides users to avoid missteps, thereby building confidence in finding the best possible solution.

When asked if they would consider implementing this framework, all interviewees unanimously agreed that they would. It was noted that each athlete is unique in terms of injury occurrences. Leveraging this framework to plan training sessions and understand patterns based on previous injuries will help identify when an athlete may be at risk, thanks to the simplicity of the statistics. Furthermore, any athlete would benefit from it, leading to better performance and fewer costly injuries for the clubs.

It was highlighted that even in professional sports, injury assessment methods are still outdated, relying on descriptive data judged by the naked eye and previous human knowledge, without predictive measures. This can lead to inaccurate evaluations. By implementing this framework, efficiency levels are promoted, injury assessments will improve, and professionals can focus on other areas, enhancing the overall productivity of the team.

For further improvement, it was suggested to provide a clearer way to recommend algorithms. For example, the algorithms in Table 6 could be organized into groups by type, making the table easier to interpret. Additionally, it was recommended to add the option to implement available pre-trained models in step three and specify the type of input data necessary, without making it mandatory for the data scientist to train the model.

From another perspective, it was suggested to add an intermediary step between steps two and three to specify the tissue type within the injured location. Since most injuries are muscular, this method would allow for segregation by tissue within the framework, adding a layer of specificity after selecting the anatomical region, rather than maintaining a global view of injuries. Although this issue could provide valuable insights, it was carefully considered but not pursued, as explained in Section 4.1.

The importance of considering stress variables in the model that can trigger an injury was also highlighted. For example, in football, when a team has to play games every two days, there is a small window for recovery, which should be taken into account. Additionally, the importance of specific individual variables that are “the future” of injury prevention was emphasized. Genetic factors were used as an example, noting that each body may have a higher tendency to incur an injury. Although this technology is not yet available, studies and research are being conducted. Polymorphisms and genetic markers can provide risk predictions, and in the future, this technology may be accessible and serve as input for valuable predictions.

5. CONCLUSIONS

This chapter concludes the development work of this dissertation by reflecting on the key conclusions, discussing the limitations encountered during its development, and outlining potential areas for future research and development. The evaluation stage and collected feedback confirm that the initial objectives have been fulfilled.

5.1. Synthesis of Developed Work

The development of this work began with an investigation into various subjects, including sports, injuries, and technologies within this context. The knowledge gained from the literature review allowed the creation and development of the artefact, which was subsequently validated by both technical and non-technical specialists to verify its utility. As a result, it can be concluded that the research objectives were achieved, answering the research question: What are the most suitable AI-powered solutions that can be implemented in different sports according to their unique characteristics, focusing on injury prevention? This was accomplished through the creation of an artefact that guides the user to the best solution.

5.2. Limitations

This research encountered several limitations during its development. The literature review revealed significantly more articles on team sports than individual sports, creating an unbalanced sample bias that was reflected in the assumptions' tables. Additionally, during the construction of the artefact, setting a standard for the granularity level for both technical and non-technical steps as a high-level framework was challenging to keep a balanced level of detail.

Technological advancements induce volatile price ranges in the short term, and various factors can cause these values to fluctuate. Therefore, prices were not considered in this work. Furthermore, with emerging technologies, those addressed in this work may become obsolete in the future, constituting a significant limitation.

In the use case, due to the complexity of the practical application of the framework and the limited time available for this dissertation, a fictitious use case was conducted. This approach may have limited the discovery of significant insights that could have been revealed through a real-world use case.

5.3. Future Work

For future work, the validation process could be enhanced by gathering insights from a larger and more diverse population, ensuring greater confidence in the artefact. Additionally, conducting a real use case and exposing it to different sports and contexts, ideally in collaboration with a club or sports entity, would be beneficial for research and academic purposes.

The insightful suggestions from the interviewees should be considered for implementation in future research and development. Moreover, given the rapid advancements in technology, it is recommended to continuously update the algorithms and IoTs within the artefact. Incorporating new knowledge from emerging studies into the framework will also be essential.

Lastly, making this research available in academic forums or scientific platforms would increase its accessibility, potentially sparking public interest and promoting further investigations on the topic.

BIBLIOGRAPHICAL REFERENCES

Bahr, R., Clarsen, B., Derman, W., Dvorak, J., Emery, C. A., Finch, C. F., Hägglund, M., Junge, A., Kemp, S., Khan, K. M., Marshall, S. W., Meeuwisse, W., Mountjoy, M., Orchard, J. W., Pluim, B., Quarrie, K. L., Reider, B., Schwellnus, M., Soligard, T., Chamari, K. (2020). International Olympic Committee consensus statement: methods for recording and reporting of epidemiological data on injury and illness in sport 2020 (including STROBE Extension for Sport Injury and Illness Surveillance (STROBE-SIIS)). *British Journal of Sports Medicine*, *54*(7), 372–389. https://doi.org/10.1136/BJSPORTS-2019-101969

Behm, D. G., Kay, A. D., Trajano, G. S., Alizadeh, S., & Blazevich, A. J. (2021). Effects of Stretching on Injury Risk Reduction and Balance. *Journal of Clinical Exercise Physiology*, *10*(3), 106–116. https://doi.org/10.31189/2165-6193-10.3.106

Borge, S. (2020). What Is Sport? *Sport, Ethics and Philosophy*, *15*(3), 308–330. https://doi.org/10.1080/17511321.2020.1760922

Castellanos, J., Phoo, C. P., Eckner, J. T., Franco, L., Broglio, S. P., McCrea, M., McAllister, T., Wiens, J., Hoy, A., Kelly, L., Jackson, J., Kelly, T., Buckley, T., Clugston, J. R., Ortega, J., Kontos, A., Giza, C. C., Mihalik, J., & Rowson, S. (2021). Predicting Risk of Sport-Related Concussion in Collegiate Athletes and Military Cadets: A Machine Learning Approach Using Baseline Data from the CARE Consortium Study. *Sports Medicine*, *51*(3), 567–579. https://doi.org/10.1007/s40279-020-01390-w

Clemente, F. M., Afonso, J., Costa, J., Oliveira, R., Pino-Ortega, J., & Rico-González, M. (2021). Relationships between Sleep, Athletic and Match Performance, Training Load, and Injuries: A Systematic Review of Soccer Players. *Healthcare 2021, Vol. 9, Page 808*, *9*(7), 808. https://doi.org/10.3390/HEALTHCARE9070808

Eather, N., Wade, L., Pankowiak, A., & Eime, R. (2023). The impact of sports participation on mental health and social outcomes in adults: a systematic review and the 'Mental Health through Sport' conceptual model. In *Systematic Reviews* (Vol. 12, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13643-023-02264-8

Edouard, P., & Ford, K. R. (2020). Great Challenges Toward Sports Injury Prevention and Rehabilitation. *Frontiers in Sports and Active Living*, *2*. https://doi.org/10.3389/fspor.2020.00080

Frevel, N., Beiderbeck, D., & Schmidt, S. L. (2022). The impact of technology on sports – A prospective study. *Technological Forecasting and Social Change*, *182*. https://doi.org/10.1016/j.techfore.2022.121838

Habibi, F., Khairandish, M. O. (2023). Evolution of technology in sports: Impact on performance, management, and fan experience. *International Journal of Science and Research Archive*, *10*(1), 995–1000. https://doi.org/10.30574/ijsra.2023.10.1.0831

Hammes, F., Hagg, A., Asteroth, A., & Link, D. (2022). Artificial Intelligence in Elite Sports—A Narrative Review of Success Stories and Challenges. *Frontiers in Sports and Active Living*, *4*. https://doi.org/10.3389/fspor.2022.861466

He, K. (2021). Prediction Model of Juvenile Football Players' Sports Injury Based on Text Classification Technology of Machine Learning. *Mobile Information Systems*, *2021*, 2955215. https://doi.org/10.1155/2021/2955215

Hecksteden, A., Schmartz, G. P., Egyptien, Y., Aus der Fünten, K., Keller, A., & Meyer, T. (2023). Forecasting football injuries by combining screening, monitoring and machine learning. *Science and Medicine in Football*, *7*(3), 214–228. https://doi.org/10.1080/24733938.2022.2095006

Hind, K., Konerth, N., Entwistle, I., Theadom, A., Lewis, G., King, D., Chazot, P., & Hume, P. (2020). Cumulative Sport-Related Injuries and Longer Term Impact in Retired Male Elite- and Amateur-Level Rugby Code Athletes and Non-contact Athletes: A Retrospective Study. *Sports Medicine*, *50*(11), 2051–2061. https://doi.org/10.1007/s40279-020-01310-y

John, A. (2024). *What are the Differences between Contact And Non Contact Sports? Unveiling the Distinctive Elements - The Sports Ground*. Retrieved October 16, 2024, from https://thesportsground.com/what-are-the-differences-between-contact-and-non-contact-sports/

Kolodziej, M., Groll, A., Nolte, K., Willwacher, S., Alt, T., Schmidt, M., & Jaitner, T. (2023). Predictive modeling of lower extremity injury risk in male elite youth soccer players using least absolute shrinkage and selection operator regression. *Scandinavian Journal of Medicine and Science in Sports*, *33*(6), 1021–1033. https://doi.org/10.1111/sms.14322

Kumar, G. S., Kumar, M. D., Reddy, S. V. R., Kumari, B. V. S., & Reddy, C. R. (2024). Injury Prediction in Sports using Artificial Intelligence Applications: A Brief Review. *Volume 5, Issue 1, Pages 16 - 26*, *5*(1), 16–26. https://doi.org/10.18196/jrc.v5i1.20814

Langton, J. (2024). *Building a digital athlete: Using AI to rewrite the playbook on NFL player safety*. https://www.nfl.com/playerhealthandsafety/equipment-and-innovation/aws-partnership/building-a-digital-athlete-using-ai-to-rewrite-the-playbook-on-nfl-player-safety

Li, A., & Huang, W. (2023). A comprehensive survey of artificial intelligence and cloud computing applications in the sports industry. *Wireless Networks*. https://doi.org/10.1007/s11276-023-03567-3

Liu, Z., Jiang, K., & Dong, J. S. (2023). Sports Injury Prediction in Professional Tennis. *Proceedings of IEEE Pacific Rim International Symposium on Dependable Computing, PRDC*, 304–308. https://doi.org/10.1109/PRDC59308.2023.00048

Lu, Y., Pareek, A., Lavoie-Gagne, O. Z., Forlenza, E. M., Patel, B. H., Reinholz, A. K., Forsythe, B., & Camp, C. L. (2022). Machine Learning for Predicting Lower Extremity Muscle Strain in National Basketball Association Athletes. *Orthopaedic Journal of Sports Medicine*, *10*(7). https://doi.org/10.1177/23259671221111742

Luu, B. C., Wright, A. L., Haeberle, H. S., Karnuta, J. M., Schickendantz, M. S., Makhni, E. C., Nwachukwu, B. U., Williams, R. J., & Ramkumar, P. N. (2020). Machine Learning Outperforms Logistic Regression Analysis to Predict Next-Season NHL Player Injury: An Analysis of 2322 Players From 2007 to 2017. *Orthopaedic Journal of Sports Medicine*, *8*(9). https://doi.org/10.1177/2325967120953404

Lyubovsky, A., Liu, Z., Watson, A., Kuehn, S., Korem, E., & Zhou, G. (2022). A pain free nociceptor: Predicting football injuries with machine learning. *Smart Health*, *24*, 100262. https://doi.org/10.1016/j.smhl.2021.100262

Martin, S., Edouard, P., Sanchez, X., & Ivarsson, A. (2024). Overuse injury and affects in competitive sport: A prospective longitudinal study. *Scandinavian Journal of Medicine and Science in Sports*, *34*(5), e14644. https://doi.org/10.1111/sms.14644

Martínez-Gramage, J., Albiach, J. P., Moltó, I. N., Amer-Cuenca, J. J., Moreno, V. H., & Segura-Ortí, E. (2020). A random forest machine learning framework to reduce running injuries in young triathletes. *Sensors (Switzerland)*, *20*(21), 1–12. https://doi.org/10.3390/s20216388

Martins, F., Przednowek, K., França, C., Lopes, H., de Maio Nascimento, M., Sarmento, H., Marques, A., Ihle, A., Henriques, R., & Gouveia, É. R. (2022). Predictive Modeling of Injury Risk Based on Body Composition and Selected Physical Fitness Tests for Elite Football Players. *Journal of Clinical Medicine*, *11*(16), 4923. https://doi.org/10.3390/jcm11164923

Oja, P., Memon, A. R., Titze, S., Jurakic, D., Chen, S. T., Shrestha, N., Em, S., Matolic, T., Vasankari, T., Heinonen, A., Grgic, J., Koski, P., Kokko, S., Kelly, P., Foster, C., Podnar, H., & Pedisic, Z. (2024). Health Benefits of Different Sports: a Systematic Review and Meta-Analysis of Longitudinal and Intervention Studies Including 2.6 Million Adult Participants. In *Sports Medicine - Open* (Vol. 10, Issue 1). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1186/s40798-024-00692-x

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906. https://doi.org/10.1016/J.IJSU.2021.105906

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pressler, A., & Niebauer, J. (2020). Textbook of Sports and Exercise Cardiology. *Textbook of Sports and Exercise Cardiology*. https://doi.org/10.1007/978-3-030-35374-2

Prieto-González, P., Martínez-Castillo, J. L., Fernández-Galván, L. M., Casado, A., Soporki, S., & Sánchez-Infante, J. (2021). Epidemiology of sports-related injuries and associated risk factors in adolescent athletes: An injury surveillance. *International Journal of Environmental Research and Public Health*, *18*(9). https://doi.org/10.3390/ijerph18094857

Qi, Y., Sajadi, S. M., Baghaei, S., Rezaei, R., & Li, W. (2024). Digital technologies in sports: Opportunities, challenges, and strategies for safeguarding athlete wellbeing and competitive integrity in the digital era. *Technology in Society*, *77*, 102496. https://doi.org/10.1016/J.TECHSOC.2024.102496

Rebelo, A., Martinho, D. V., Valente-dos-Santos, J., Coelho-e-Silva, M. J., & Teixeira, D. S. (2023). From data to action: a scoping review of wearable technologies and biomechanical assessments informing injury prevention strategies in sport. *BMC Sports Science, Medicine and Rehabilitation*, *15*(1). https://doi.org/10.1186/s13102-023-00783-4

Ren, L., Wang, Y., & Li, K. (2024). Real-time sports injury monitoring system based on the deep learning algorithm. *BMC Medical Imaging*, *24*(1), 122. https://doi.org/10.1186/s12880-024-01304-6

Robles, A. G., Palamà, Z., Nesti, M., Tunzi, R. M., Delise, P., Cavarretta, E., Penco, M., Romano, S., & Sciarra, L. (2022). Sport Related Sudden Death: The Importance of Primary and Secondary Prevention. In *Journal of Clinical Medicine* (Vol. 11, Issue 16). MDPI. https://doi.org/10.3390/jcm11164683

Romero-Jara, E., Solanellas, F., Muñoz, J., & López-Carril, S. (2023). Connecting with fans in the digital age: an exploratory and comparative analysis of social media management in top football clubs. *Humanities and Social Sciences Communications*, *10*(1). https://doi.org/10.1057/s41599-023-02357-8

She, H. (2024). Application of Big Data Analysis in Model Construction to Prevent Athlete Injury in Training. *Applied Mathematics and Nonlinear Sciences*, *9*(1), 20241723. https://doi.org/10.2478/amns-2024-1723

Spitz, J., Wagemans, J., Memmert, D., Williams, A. M., & Helsen, W. F. (2020). Video assistant referees (VAR): The impact of technology on decision making in association football referees. *Journal of Sports Sciences*, *39*(2), 147–153. https://doi.org/10.1080/02640414.2020.1809163

SPORT | English meaning - Cambridge Dictionary. (n.d.). Cambridge Dictionary. Retrieved October 16, 2024, from https://dictionary.cambridge.org/dictionary/english/sport

sport | Etymology of sport by etymonline. (2023). Online Etymology Dictionary. https://www.etymonline.com/word/sport

Tranaeus, U., Gledhill, A., Johnson, U., Podlog, L., Wadey, R., Wiese Bjornstal, D., & Ivarsson, A. (2024). 50 Years of Research on the Psychology of Sport Injury: A Consensus Statement. *Sports Medicine (Auckland, N.Z.)*, *54*(7), 1733–1748. https://doi.org/10.1007/S40279-024-02045-W

Tsilimigkras, T., Kakkos, I., Matsopoulos, G. K., & Bogdanis, G. C. (2024). Enhancing Sports Injury Risk Assessment in Soccer Through Machine Learning and Training Load Analysis. *Journal of Sports Science and Medicine*, *23*(3), 537–547. https://doi.org/10.52082/jssm.2024.537

Van Eetvelde, H., Mendonça, L. D., Ley, C., Seil, R., & Tischer, T. (2021). Machine learning methods in sport injury prediction and prevention: a systematic review. *Journal of Experimental Orthopaedics*, *8*(1), 1–15. https://doi.org/10.1186/S40634-021-00346-X/TABLES/5

Volberda, H. W., Khanagha, S., Baden-Fuller, C., Mihalache, O. R., & Birkinshaw, J. (2021). Strategizing in a digital world: Overcoming cognitive barriers, reconfiguring routines and introducing new organizational forms. *Long Range Planning*, *54*(5). https://doi.org/10.1016/j.lrp.2021.102110

vom Brocke, J., Hevner, A., & Maedche, A. (2020). *Introduction to Design Science Research* (pp. 1–13). https://doi.org/10.1007/978-3-030-46781-4_1

Wang, C., Stovitz, S. D., Kaufman, J. S., Steele, R. J., & Shrier, I. (2024). Principles of musculoskeletal sport injuries for epidemiologists: a review. *Injury Epidemiology 2024 11:1*, *11*(1), 1–13. https://doi.org/10.1186/S40621-024-00507-3

Xu, S., Zhang, X., & Jin, N. (2023). Soccer Sports Injury Risk Analysis and Prediction by Edge Wearable Devices and Machine Learning. *Proceedings - 2023 International Conference on Artificial Intelligence of Things and Systems, AIoTSys 2023*, 38–43. https://doi.org/10.1109/AIoTSys58602.2023.00027

Yang, G., & Xu, X. (2022). Sequence Video and Artificial Intelligence Assisted Basketball Injury Risk Early Warning Method. *Mobile Information Systems*, *2022*, 4778245. https://doi.org/10.1155/2022/4778245

Yang, S. X., Cheng, S., & Su, D. L. (2022). Sports injury and stressor-related disorder in competitive athletes: A systematic review and a new framework. In *Burns and Trauma* (Vol. 10). Oxford University Press. https://doi.org/10.1093/burnst/tkac017

Zadeh, A., Taylor, D., Bertsos, M., Tillman, T., Nosoudi, N., & Bruce, S. (2021). Predicting Sports Injuries with Wearable Technology and Data Analysis. *Information Systems Frontiers*, *23*(4), 1023–1037. https://doi.org/10.1007/S10796-020-10018-3/TABLES/10

APPENDIXES

Appendix A

Appendix A presents the transcription of the interview conducted with the Data Scientist, originally in Portuguese and later translated to English, which was included in the evaluation of the framework solution in Section 4.4. The original version of the three answers is available below, following a brief introduction to the topic and an explanation of the framework and its development.

Answer to Question 1

“Acho que a nível da utilidade faz todo o sentido. É uma framework que está a resolver um problema que existe e que vai continuar a existir no desporto. É flexível ao ponto de conseguir ter diferentes abordagens, que também é natural tendo em conta, se queres fazer uma framework que seja adaptável a qualquer desporto, também tens de ser capaz de lidar com diferentes tipos de inputs e outputs.

Acho que é particularmente interessante por aí, tanto a conexão com os IoTs, que todos os desportistas de alto nível geralmente usam de uma maneira ou de outra, mas também porque permite guardar os dados e ir iterando e melhorando e permitindo ter uma noção muito própria, ou seja, quer para clube, quer mais mais high level por desporto. Acho que dá uma noção muito boa e impactante naquilo que pode ser a vida desportiva de pessoas cujos treinadores ou cuja estrutura usam esta framework.”

Answer to Question 2

“Sim, eu sinto que apesar de haver muito conhecimento científico sobre prevenção de lesão e sobre o corpo e a anatomia humana, a verdade é que cada caso é um caso e tem muito que ver com o contexto específico do atleta em determinado momento e esse contexto é definido por que competições é que ele está a fazer, qual é a sua carga de treinos, que tipo de treinos é que está a fazer, entre outros, e é muito fácil o mesmo atleta não reagir da mesma maneira a diferentes tipos de treinos e uma framework que ajude quem planeia os treinos a entender os padrões do atleta e a conseguir pegar em padrões de atletas que vieram antes e que passaram pelo mesmo tipo de treinos acho que é muito benéfico porque a estatística neste caso vai ajudar bastante.

Algo que consiga pegar e generalizar um bocadinho da maneira predictiva da questão acho que é ótimo, acho que qualquer atleta beneficiaria e bem usada, teria benefícios tanto para os atletas em si como para os clubes que obviamente conseguem ter os atletas a ter boa performance sem lesões, sem paragens que obviamente são custosas.”

Answer to Question 3

“Acho que a nível de algoritmia se calhar ficou um pouco complexo porque ao nos limitarmos a expor aquilo que já foi estudado e com os algoritmos que foram estudados em determinado contexto, em determinado paper, podemos estar a reduzir o leque de opções para o utilizador. Posso dar o exemplo que, claramente nos individual sports, nos papers que foram dados, foi usado, por exemplo, o CatBoost mas em vez do CatBoost usar o Gradient Boosting ou usar o XGBoost, os resultados não vão diferir assim tanto e portanto uma coisa que se pode fazer é organizar os algoritmos por grupos, por clusters de tipos de algoritmos e simplificar um bocadinho a tabela, porque apontar para cada desporto qual o tipo de algoritmos, pode facilitar um bocadinho a quem for aplicar porque pondo-me no papel de um user que não conhece os algoritmos é um pouco overwhelming, acaba por ser um bocadinho assustador a quantidade de nomes e de algoritmos estranhos obviamente para alguém que não é da área.

Poderias também adicionar no step three a opção de usar algum dos modelos que já estão disponíveis, que já estão treinados, e dizer qual o tipo de input, de dados, compatível com esse mesmo modelo. Deste modo, o passo “Train the model” seria opcional.”

Appendix B

Appendix A presents the transcription of the interview conducted with the Coach, originally in Portuguese and later translated to English, which was included in the evaluation of the framework solution in Section 4.4. The original version of the three answers is available below, following a brief introduction to the topic and an explanation of the framework and its development.

Answer to Question 1

“Eu considero útil no sentido em que tu identificas um problema e depois a partir das várias áreas que tens, consegues aprofundar. A partir do macro vais caminhando para o problema em si e só fazendo esse exercício de qual é o caminho que eu quero seguir, tu consegues identificar melhor o problema do que se for diretamente a ele.

É útil, essencialmente, para isso porque no fundo tu queres ver que modelo aplicar para a problemática, e tu tendo este tipo de framework, consegues chegar aos “melhores” modelos mais facilmente. Porque há vários algoritmos, mas um pode funcionar para um tipo de desporto e não funcionar para outro, e se nós tentarmos ir diretamente à solução, que é encontrar o melhor modelo, se calhar não estamos a aplicar um modelo certo.”

Answer to Question 2

“Eu implementaria porque existe uma enorme variedade de lesões no futebol, e são tantas as estruturas anatômicas que estão em contacto e que pode ocorrer uma lesão, que nós, sendo uma ou duas pessoas a fazer esse trabalho, é muito trabalho para uma pessoa. Especialmente no meu caso, que eu tenho várias áreas de intervenção, sendo esta a parte da prevenção de lesões uma delas, eu acho que ia ser útil neste sentido de ter modelos a ajudar e facilitar o nosso trabalho, e a fazer com que sejamos mais eficazes noutras áreas. (...)

Acho que no mundo do desporto profissional, porque também estou em contacto com o desporto profissional no futebol, é muito feito meio “à antiga”. Olhamos para os dados e começamos a ver padrões mas é muito a olho nu e é preciso muita mão de obra. Não é tão certo, nem conseguimos olhar de forma tão global para as coisas. (...)

E acho que esse trabalho, através deste tipo de frameworks, podia ser facilitador nesse sentido. Todo tipo de recolha que nós fazemos é para análise descritiva. Depois a partir disso nós podemos, para casos futuros, tentar agir de forma diferente, mas não temos modelos preditivos.”

Answer to Question 3

“Depois do passo dois, eu acho que punha um passo intermédio, para diferenciar o tipo de lesões. Podia-se especificar mais por exemplo, analisar só lesões musculares. Porque eu acho que dentro desse tipo de lesões são as lesões tem melhor aplicabilidade para este tipo de frameworks, são mais comuns. Pela minha experiência, são as lesões mais associadas à sobrecarga. Não acontece uma lesão, por exemplo, articular, ou uma rotura de ligamentos, pode acontecer, mas eu acho que está mais ligada às lesões musculares.

A recomendação é focar num tipo de lesão mais concreto, em vez de pegares em todas as lesões musculares. Por exemplo, Em vez de passar logo para o modelo, identificar primeiro a estrutura, e depois o tipo de lesão, seja muscular, articular, etc., dentro dessa mesma estrutura."

Appendix C

Appendix A presents the transcription of the interview conducted with the Athlete, originally in Portuguese and later translated to English, which was included in the evaluation of the framework solution in Section 4.4. The original version of the three answers is available below, following a brief introduction to the topic and an explanation of the framework and its development.

Answer to Question 1

“Sim, acho que tem um potencial muito grande de ser utilizada, mesmo no dia-a-dia. Eu acho que tanto para desportos de equipa, pronto, lá está o que me mostraste, tens mais dados de desportos coletivos, mas eu acho que no futuro também vai cada vez haver mais desportos individuais e isto pode mesmo ser bastante útil a tentar prever até qual é o momento de risco para cada atleta e tentar arranjar soluções para isso. Acho que sim, acho que pode ser bastante útil.”

Answer to Question 2

“Sim, pela mesma razão. Eu acho que pode ser bastante útil em tentarmos identificar quais são os momentos em que cada atleta, quer desportos coletivos, quer desportos individuais, possa estar num momento de risco. Seja por supercarga de treino, seja por supercarga de torneios, seja por outros... Tu não especificaste exatamente quais são, dentro de cada IoT, quais são os pontos que eles monitorizam, mas eu imagino que no futuro cada vez vai haver mais capacidade de ir até buscar fatores emocionais, sono, hidratação... Tudo, coisas que podem pôr em risco o aparecimento de uma lesão, maior ou menor gravidade. Eu consideraria, sem dúvida.”

Answer to Question 3

“Eu acho que esta parte ética e da privacidade é extremamente importante. Considerar desportos de contacto, não contacto, de equipa, individuais... É importante detalhar para cada desporto os períodos de maior stress. Ou seja, de stress físico, que podem condicionar o maior risco do atleta de desenvolver uma lesão.

Por exemplo, pensando no caso do futebol. Quando eles têm jogos para a Liga dos Campeões e depois têm os jogos do Campeonato Nacional e quase que jogam dois em dois dias, naqueles períodos de maior stress, ter todos esses dados.

Mas também é importante, mesmo para desportos de equipa, fatores individuais. E no futuro isso vai ser tudo... Eu acho que nós vamos ter acesso a toda essa informação. Por exemplo, fatores genéticos. O teu corpo, o teu organismo como atleta é diferente de outra pessoa qualquer. E podes ter uma predisposição maior ou menor. Portanto, eu imagino que neste momento, não tendo acesso a dados genéticos da pessoa, não se consegue ainda colocar.

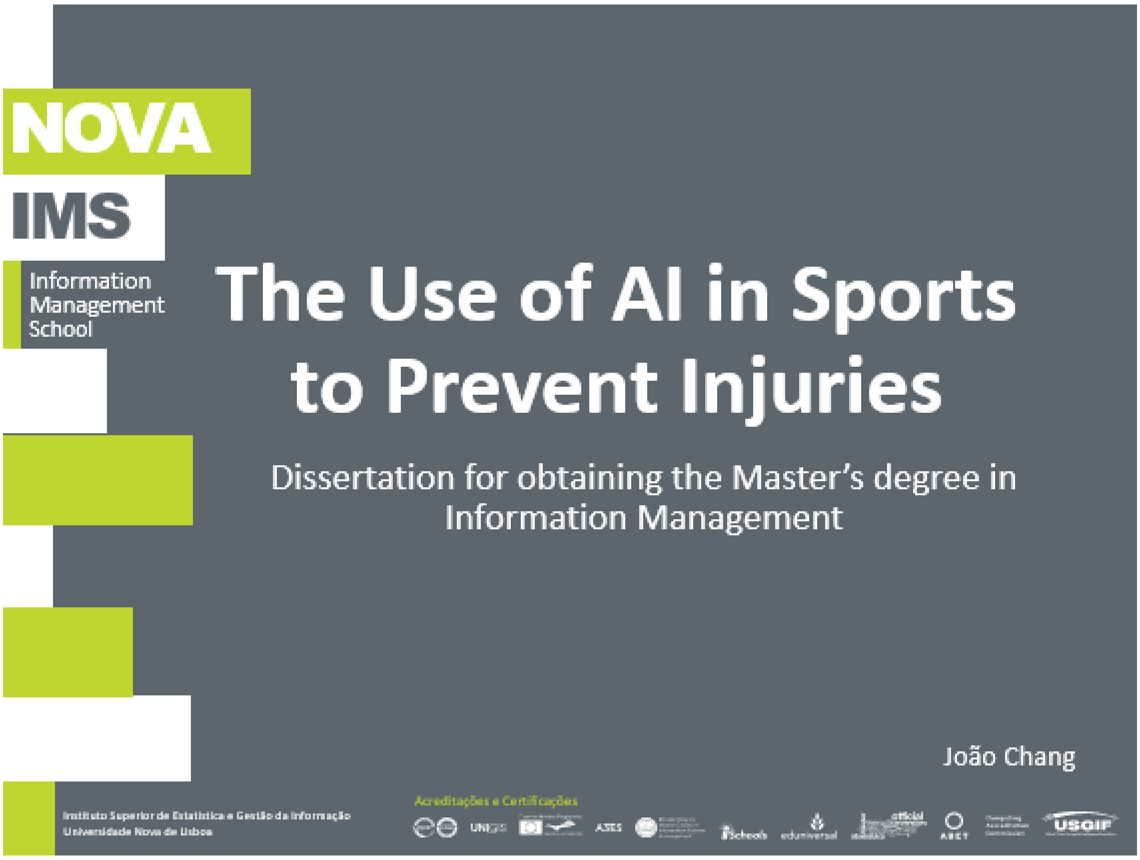
Mas pode ser que no futuro, dez anos no máximo, nós já vamos conseguir ter essa tecnologia mais acessível e juntar a outros dados, como por exemplo a idade, coisas que nós já utilizamos corriqueiramente no dia-a-dia, como os relógios de frequência cardíaca, de VO2 máximo, de carga de treino, de fatores físicos. Eu acho que isto pode ser um passo bastante importante. A parte genética conta bastante.

Por exemplo, a Carolina Marin (spanish badminton athlete), se tu comparares o corpo de uma europeia com uma chinesa, geneticamente os chineses têm mais facilidade em aguentar cargas. Tu vês que a Marin já teve três roturas de ligamentos. Eu acho que isso também vai ser bastante importante.

Mas ainda estamos um bocadinho longe. Não é por não se pensar nisto nisso, é porque não há ainda tecnologia para isso. Há já polimorfismos e marcadores genéticos que nos dão uma previsão de risco. Nós já sabemos muita coisa mas ainda há muito para descobrir...”

Appendix D

Appendix D displays the PowerPoint presentation used as support to conduct the interviews.

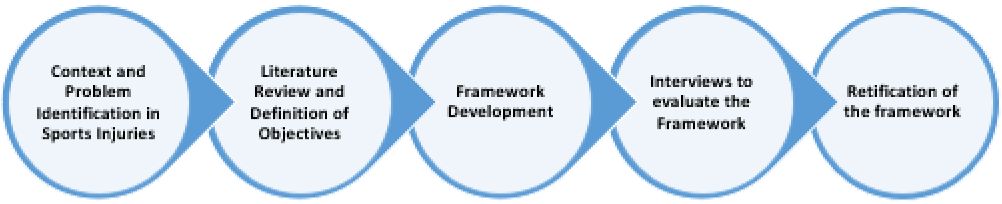


Problem Statement

1 Sports organizations face challenges in effectively implementing digital transformation programs to enhance operational efficiency, customer intimacy, and innovation.

2 Sports entities need to leverage advanced technologies to optimize training programs, minimize injury risks, and improve athlete safety and performance.

Research Strategy



"What are the most suitable AI-powered devices and tools that can be implemented in different sports according to their unique characteristics, focusing on injury prevention?"

Framework

01

Identification of scope

- Create a framework to serve as guidance for a data scientist to implement a predictive model regarding injury prevention
- Types of predictive models: Alert for prevention (training loads, historical, physical tests), risk assessment of occurrence, risk based on incorrect movements...

02

Definition of requirements

- Assess the most popular algorithms and IoTs for each type of sport in previous studies;
- Describe each of IoTs by their functionality.
- Note: Prices of IoTs, infrastructure, staff, etc. are not considered for this study.

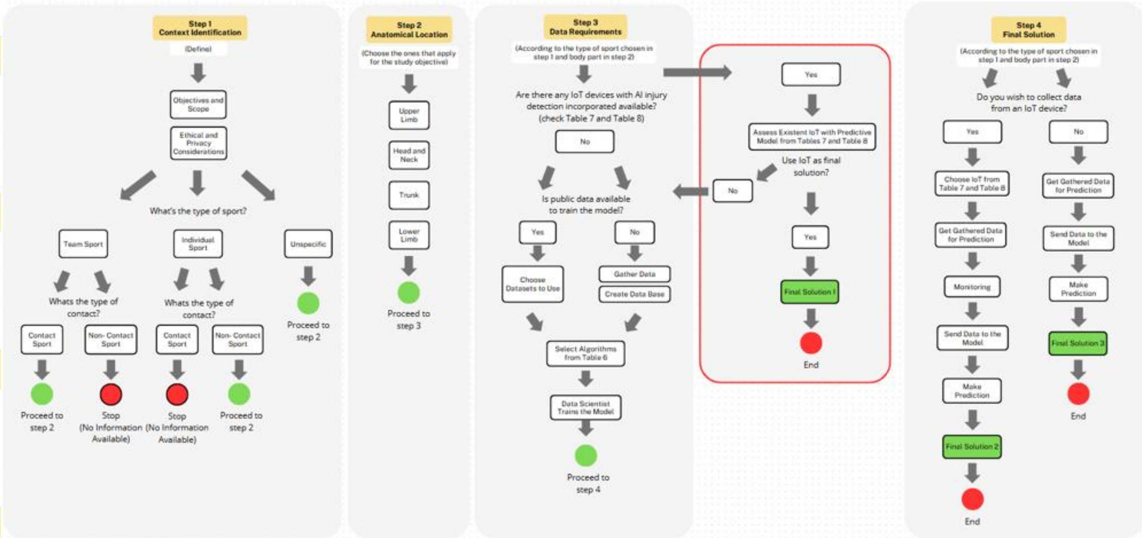
03

Final Solution

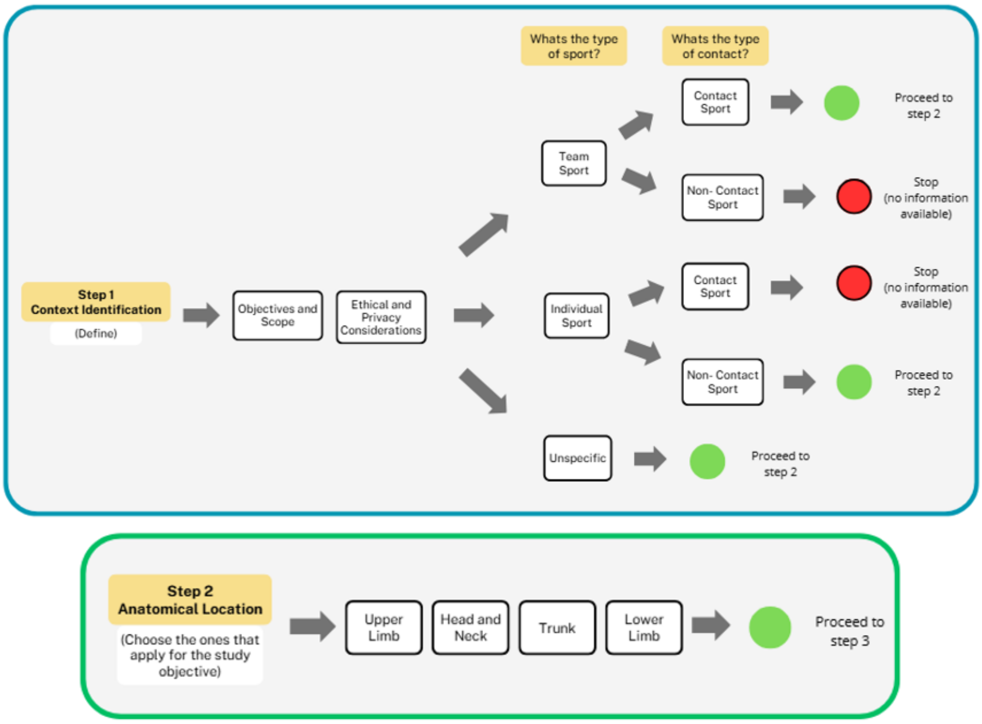
- Guided steps for different approaches to be implemented by a Data team to create a predictive model focusing on injury prevention according to the sport.

Framework

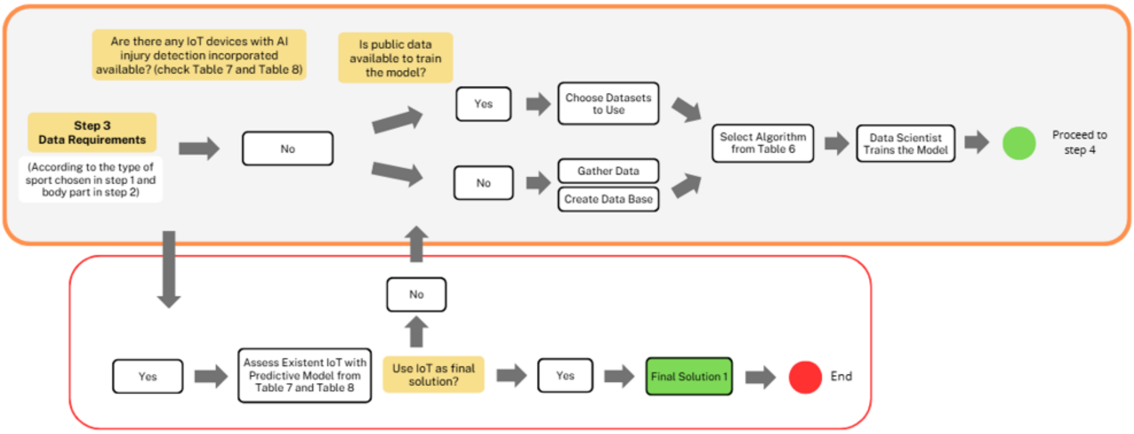
Framework



Framework - Step 1 and 2



Framework – Step 3



Framework – Step 4

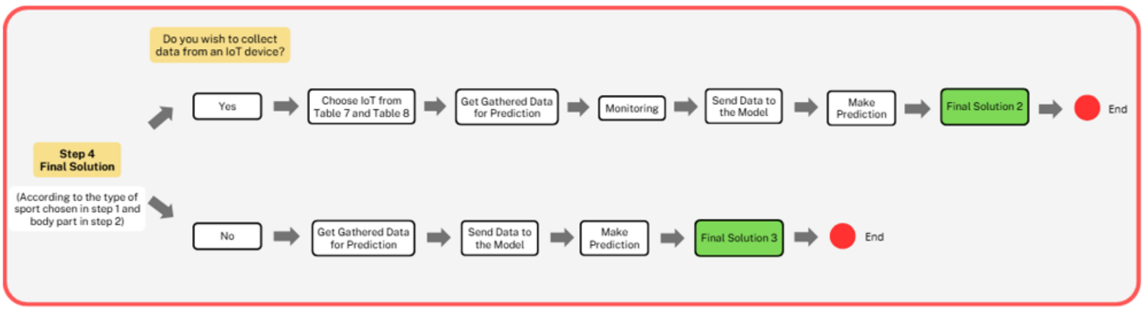


Table Algorithms

Predictive Techniques		Team Sports		Individual Sports		Not Specific/ Various
		Contact Sports	Non-Contact Sports	Contact Sports	Non-Contact Sports	
Anatomical Location	Head and Neck	Gradient Boosting, XGBoost, Logistic Regression, Ridge, HVARPNN, Adaboost-Random Forest, SVM, CNN, ANN			SVM, CatBoost, Pattern Mining	SVM
	Upper Limb	Gradient Boosting, XGBoost, Logistic Regression, Ridge, SVM, HVARPNN, Adaboost-Random Forest, CNN, ANN			SVM, CatBoost, Pattern Mining	SVM
	Trunk	Gradient Boosting, XGBoost, Logistic Regression, HVARPNN, Adaboost-Random Forest, SVM, CNN, ANN			Random Forest, SVM, CatBoost, Pattern Mining	SVM
	Lower Limb	Gradient Boosting, Lasso Regression, XGBoost, Logistic Regression, Ridge, SVM, Text Classification, HVARPNN, Adaboost-Random Forest, CNN, ANN			Random Forest, SVM, CatBoost, Pattern Mining	SVM
	Unspecified Area;					
	Multiple Regions					
	All Body Parts	Gradient Boosting, XGBoost, Logistic Regression, HVARPNN, Adaboost-Random Forest, SVM, CNN, ANN			CatBoost, Pattern Mining, SVM	SVM

Table IoTs

IoTs		Team Sports		Individual Sports		Not Specific/ Various
		Contact Sports	Non-Contact Sports	Contact Sports	Non-Contact Sports	
Anatomical Location	Head and Neck	2D Video Recordings, InBody 770, Jamar Plus+, Optojump Next System, Witty-Gate Photocells, Proposed Idea: Optical Wearable Sensor, Polar Team Pro System, Catapult - Playertek			2D Video Recordings	2D Video Recordings
		IoT with Predictive Model: Omegawave Sensor			IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System	IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System
	Upper Limb	Wearable Sensor, 2D Video Recordings, InBody 770, Jamar Plus+, Optojump Next System, Witty-Gate Photocells, Proposed Idea: Optical Wearable Sensor, Polar Team Pro System, Catapult - Playertek			2D Video Recordings	2D Video Recordings
		IoT with Predictive Model: Omegawave Sensor			IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System	IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System
	Trunk	2D Video Recordings, Proposed Idea: Optical Wearable Sensor, Polar Team Pro System, Catapult - Playertek			SEMG Sensors, 2D Video Recordings	2D Video Recordings
		IoT with Predictive Model: Omegawave Sensor			IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System	IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System
	Lower Limb	Wearable Sensor, Qualisys - 3D motion capture system, Pegasus 3-D system, IsoMed 2000 Isokinetic Dynamometer, 2D Video Recordings, InBody 770, Jamar Plus+, Optojump Next System, Witty-Gate Photocells, Proposed Idea: Optical Wearable Sensor, Polar Team Pro System, Catapult - Playertek			SEMG Sensors, 2D	2D Video Recordings
		IoT with Predictive Model: Omegawave Sensor			IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System	IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System
	Unspecified Area;					
	Multiple Regions					
	All Body Parts	Catapult – Playertek, 2D Video Recordings, Proposed Idea: Optical Wearable Sensor, Polar Team Pro System			2D Video Recordings	2D Video Recordings
		IoT with Predictive Model: Omegawave Sensor			IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System	IoT with Predictive Model: Proposed idea: Portable + Main Monitoring System

Table IoTs Characteristics

Devices (IoT)	GPS/ Motion	Video Data	Physiological Data	IoT with Predictive Model
Wearable Sensor (GPS, Accelerometer)	x			
Qualisys - 3D Motion Capture System (Consisting of 12 Infrared Cameras 120 Hz)	x	x		
Pegasus 3-D System (Strength Testing)			x	
IsoMed 2000 Isokinetic Dynamometer (Strength Testing)			x	
Omegawave Sensor (Brain, Heart, Nervous and Cardiac System)			x	x
SEMG Sensors (Bioelectrical Signals from the Body)			x	
2D Video Recordings (Camera)		x		
InBody 770 (Body Composition)			x	
Jamar Plus+ (Hand Dynamometer)			x	
Optojump Next System (Optical Measurement System)			x	
Witty-Gate Photocells (Timing System)			x	
Proposed Idea: Optical Wearable Sensor (Motion, Physiological Data, Framework of Sensors)	x		x	
Proposed idea: Portable + Main Monitoring System (Motion, Physiological Data)	x		x	x
Polar Team Pro System (GPS, Accelerometer, Heart Rate)	x		x	
Catapult - Playertek (GPS, Accelerometer, Heart Rate)	x		x	

Interview Questions

1 - Do you consider the proposed framework useful? Why?

Interview Questions

2 - Would you consider implementing the proposed framework? Why?

Interview Questions

3 - Do you have any recommendation or suggestions for further improvements of the proposed framework?


Thank you for your time and expertise!
Address: Campus de Campolide, 1070-312 Lisboa, Portugal
Phone: +351 213 828 610
Fax: +351 213 828 611
Acreditações e Certificações
Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa

Appendix E

Appendix E contains the Ethics Committee's approval report.

NOVA IMS Information Management School

This is to certify that

Project No.: **DSCI2025-3-63359**

Project Title: **The Use of AI in Sports to Prevent Injuries**

Principal Researcher: **João Chang**

according to the regulations of the Ethics Committee of NOVA IMS and MagIC Research Center this project was considered to meet the requirements of the NOVA IMS Internal Review Board, being considered **APPROVED** on 3/6/2025.

It is the Principal Researcher's responsibility to ensure that all researchers and stakeholders associated with this project are aware of the conditions of approval and which documents have been approved.

The Principal Researcher is required to notify the Ethics Committee, via amendment or progress report, of

- Any significant change to the project and the reason for that change;
- Any unforeseen events or unexpected developments that merit notification;
- The inability of the Principal Researcher to continue in that role or any other change in research personnel involved in the project.

Lisbon, 3/6/2025

NOVA IMS Ethics Committee

ethicscommittee@novaims.unl.pt

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa