UNIVERSITY OF TURKU
Turku School of Economics

Data-driven evaluation of on-field player performance in football using sensor and video technologies

Information Systems Science
Master's thesis

Author:
Atte Arjovuo

Supervisor:
Jukka Heikkilä

5.4.2023
Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Bachelor's thesis / **Master's thesis** / Licentiate thesis / Doctoral thesis

Subject: Information Systems Science
Author: Atte Arjovuo
Title: Data-driven evaluation of on-field player performance in football using sensor and video technologies
Supervisor: Jukka Heikkilä
Number of pages: 85 pages + appendices 5 pages
Date: 3.4.2023

Data has become increasingly relevant and used in football over the years. Technological development has made it possible to gather data from various aspects of the game. However, despite the growing popularity of sports analytics, relatively little research, especially qualitative, has been done on the topic. The purpose of this thesis is to create understanding and practices for taking advance of data for evaluation of on-field player performance in football using sensor and video technologies. This is done by identifying and combining technological possibilities with sports knowledge and suggesting an approach for data-driven evaluation of the on-field player performance. Review of previous literature and semi-structured theme interviews have been used as a method to achieve the purpose of the thesis. The findings of the thesis show that data can be used in the evaluation of on-field player performance in football by assessing players' physical, technical, tactical, and mental attributes. These attributes have several different metrics, the value of which depends on several factors such as the team's objectives. Furthermore, an approach is presented in the thesis which suggests that the selection of team-specific attributes and metrics guides the user to consider which data is needed to be able to evaluate the desired metrics, which then can be linked to certain technologies and analytical solutions presented in the thesis.

Key words: 1. Sports Analytics 2. Sensor technology, 3. Data, 4. Decision-making, 5. Performance

Table of contents

List of figures

List of tables

1 Introduction

1.1 Background and motivation

Data has become increasingly relevant in football over the years. Teams are using various data analysis tools to gain insights into players, tactics, and strategies. Furthermore, data supports in decision-making by providing objectivity and helping coaches and analysts making informed decisions. (Muniz & Flamand 2022.)

Traditional measures such as experience and intuition have not disappeared. However, data has given teams an extra tool to gain a competitive edge among the rivalry. In the face of limited resources, teams that utilize data enjoy the advantage of having objective information that can be used to evaluate players, assess performance, identify areas of improvement, and acquire suitable new players. (de la Torre et al. 2022.)

Technological development has made it possible to gather data from various aspects. Today's capabilities for efficient data gathering and processing has undoubtedly created an opportunity to disrupt the sports industry. There is already real-world evidence of teams successfully realizing benefits of data, which has led to the formation of a concept "sports analytics". (Morgulev et al. 2018.)

Sports analytics has drawn large attention since the beginning of the 21st century. Technological advancements, increased research on the topic, and many real examples of team's realizing benefits and gaining competitive advantage have all contributed making sports analytics a hot concept. The market value of sports analytics has increased from 125 million dollars reported in 2014, to 885 million dollars reported in 2020 with expected annual growth rate of 21.3% until 2028 (PR Newswire 2015, Grand View Research 2021).

In competitive football, favourable rule changes have made it possible to gather and process data during official games using real-time tracking technologies (Rein & Memmert 2016). This has allowed teams to start using data to monitor various metrics and recognize tactical patterns by gathering data-driven insights from various aspects. However, data is not useful if there are no people who know how to interpret it.

For example, Danish mid-level league team FC Midtjylland surprised by winning the Danish Super League for the first time in the 2014/2015 season. Before the season, the team had surprisingly invested in a tracking technology and acquired technical know-how. Since then, the Danish team has won the league championship three times and the Danish cup twice. The team has also been actively involved in the qualification phases of UEFA Europa League and even the Champions League. The team management has told that data has brought the team added value by optimized player performance and improved decision-making. (KMD 2018.)

1.2 Research Gap

Due to its topicality, sports analytics is an interesting and relevant research area. However, despite the growing popularity of sports analytics, relatively little research has been done in terms of football analytics. While analytics have already been used in sports, there is a large gap between research community and sports professionals. Referring to Mondello & Kamke (2014) the information sharing between sports organizations and academics is still limited.

From an academic point of view, sports analytics generally lacks mutually accepted definitions and framework. Goes et al. (2021) executed a systematic literature search for their research paper in football analytics resulting in 2338 identified studies and 73 papers on the topic. However, most of the research are still scattered and mostly written from the perspective of mathematics, statistics, and marketing (Caya & Bourdon 2016).

This makes sense, as the large amount of data available in sports creates a great opportunity for both the sports teams and academics to be used for statistical analysis and quantitative research. For example, professional team FC Bayern München captures over 60 million datapoints from each of its games to be further processed with various statistical models (Tan et al. 2017).

Nonetheless, the fundamental objective of data analytics and the quantitative work is to improve decision-making. For this reason, qualitative research is much needed on the discipline. The objective of this thesis is to fill this research gap and go beyond the statistics.

1.3 Purpose and Structure

The purpose of this thesis is to create understanding and practices for taking advance of data for evaluation of player performance in football. This is done by identifying and combining technological possibilities with sports knowledge and providing an approach for data-driven evaluation of player performance. Previous research and expert interviews have been used to achieve the purpose of the thesis.

The main research question is the following:

- How can data be used in evaluating the performance of football players?

To answer the research question, the thesis is carried out as a qualitative study which consists of a theoretical and empirical section. The theoretical section is carried out as a literature review, presenting previous research made from the subject. This sets a basis for the empirical part to be able to supplement the theory and answer the research question. The empirical part is done by expert interviews bringing theory and practice together, providing concreteness to the research.

Combining technological possibilities with sports knowledge is crucial for finding the possible and most optimal means for data-driven evaluation of player performance. The main research question will be addressed through support research questions. These support research questions are split to four categories, which also creates the structure of the literature review. The support research questions are the following:

- Technology: How can data be collected in football?
- Data: What kind of data can be collected in football?
- Analytics: How can data be used in football?
- Decision-making: What factors affect the evaluation of players' performance?

The structure of the theoretical section is organized in categories to answer these support research questions. The categories and the chapters are technology, data, analytics, and decision-making.

Technology chapter uses previous research to discuss the most used technologies in the modern football. In the chapter, various technologies are presented, as well as their strengths and weaknesses. The chapter finally answers the first support research question.

Once the data sources (technology) are known, the next chapter discusses the possible data types used in football using findings from previous research. The chapter presents possible data types as well as their potential use, answering the second support research question.

Next, the analytics chapter uses previous research to create an overview and understanding of how data can be used in football. Various analytical areas are discussed as well as their potential for football. Analytics chapter provides an answer to the third support research question.

Lastly, chapter on decision-making is the least technical, as its purpose is to ground the empirical section by creating an understanding and theoretical view of the factors influencing the evaluation of player performance. The chapter acts as the last chapter of the literature review by answering the final support research question and setting up a theoretical base priming the empirical section.

The empiric section of the thesis consists of methodology, results and discussion, and conclusion. Methodology presents the methods used in the thesis. As a method, the data is collected through semi-structured theme interviews. Results and discussion present the findings of the interviews and makes preliminary conclusions by connecting these findings with previous research. Finally, conclusion chapter presents and summarizes the final research conclusions by answering the main and support research questions. Lastly, ideas for future research are presented.

2 Technology

2.1 Manual to automated

Technology and information processing have advanced dramatically. Technological developments continue to drive change in different industries and in our everyday lives. Technological innovations shape the way how we live and perform different tasks. These innovations have brought us to times, when using technology in sports is plausible. (Torres-Ronda et al. 2017.)

At its best, technology enables manual tasks to be performed automatically, significantly improving the depth and accuracy of information (Windt et al. 2020). The factors mostly driving the evolution to be able to use technology with an increasing extent are the shrinking size of devices, the acceleration of data processing speed, and the decline in costs (Torres-Ronda et al. 2017.) Continuous technological development enables an opportunity of exploitation into new application areas.

Due to these reasons, during footballs World Cup 2014 various new technologies were first time deployed and seen on official games. The World Cup was in many ways a turning point for the future of football. The German national team acted as a pioneer and set an example for others on how technology can be used to improve decision-making. By doing so, the Germans ended up winning the tournament. (Bojanova 2014.)

2.2 Technology in sports

Due to the technological developments, taking advance of technology in sports has highly increased during the last decade. The use of technology in sports has created new opportunities and made various areas more efficient. Especially the development of tracking technologies, such as local positioning systems (LPS) and electronic performance and tracking systems (EPTS) has enabled various new opportunities, especially in team sports such as football (Pino-Ortega et al. 2021).

The time for taking advance of technology in football is favourable. The International Football Association (FIFA) allowed the use of wireless sensor technologies during official games in 2015. Since then, some football teams have begun to utilize technology more widely. (Rein & Memmert 2016.)

Technology has also been brought to be part of the sport itself, reflecting a change to a more technology accepting direction in general. VAR (Video Assistant Referee) system was introduced in 2018, which has significantly changed the nature of the game. VAR is a system based on a video and virtual offside line technology, which can reliably identify factors such as offsides, goals, and fouls. (Spitz et al. 2021; FIFA VAR 2018.) Before VAR, if these factors were not seen by the referee, it was a natural aspect of the game.

In a dynamic sport as football, technology enables capturing a comprehensive view of an athlete or a team considering many relevant factors. These factors can be combined together with a help of data. This enables better understanding of how one factor might affect another. Cross-disciplinary knowledge can be used by mixing for example medical knowledge with physiology, physical performance with technical skill, and injuries with training load. (Windt et al. 2020.)

The impact of teams using of tracking technologies has been noticeable in a short time. It will be interesting to see what direction technological development takes sports in the future. Currently, the most used technologies in football are Global Positioning System, video-based systems, radar and microwave technology, and the wearable technology (Memmert & Raabe 2018). All of these has their own strengths and weaknesses, which will be discussed next in the following subsections.

2.2.1 Global Positioning System

Global Positioning System (GPS) is a satellite-based wireless navigation and timing technology invented and taken to more practical use by several scientist and the U.S. Department of Defence in the 1970's. GPS provides positional data of people and objects and is by far the most used tracking technology in the world. (Bajaj et al. 2002.)

The functioning of GPS is based on a network of satellites in the orbit sending radio signals in frequencies to receiver devices on earth. This process enables the location to be measured by triangulation, calculating the distance between the satellite and the receiver device. As the positional data provided by the GPS is easily accessible, it has been widely used in many commercial and not-commercial applications. Due to its usability, GPS technology is being used constantly in new application areas such as in sports. The full benefits of GPS utilization are still to be realized in the future. (Yi et al. 2013.)

In 2022, GPS technology is the most used tracking technology in team sports and football. GPS provides relatively easy, flexible, and cheap way to gather positional data of the players. Many teams and leagues make use of the technology, and there are various alternative commercial or self-made solutions based on GPS technology on the market to collect data on player performance. (Rago et al. 2020.)

GPS technology enables tracking the movement of several players at the same time. In addition, data processing and analysis can be done in real-time, during games and training sessions (Aughey & Falloon 2010). Yet, GPS technology has its own strengths and weaknesses in football (Table 1).

The true benefits of GPS lays in the cost-efficiency and simplicity of the technology. The usability of GPS in sports is inherently good since there is no need for expensive investments in hardware installation. Data is being collected by small receiver devices or transponders which are placed on the player's equipment and accessed through various software solutions. The transponders are usually small and light, making them suitable for sports. (Malone et al. 2020.)

Another major advantage of GPS technology is its flexibility. The transponders can be integrated to gather various meaningful information. Besides physical load monitoring, other features such as fitness data, for example data of heart rate and respiratory rate, can be integrated to the same transponder. In addition, GPS technology is not locally bound, and it can be flexibly used on a different sites and events such as in away games and training camps. (Malone et al. 2020.)

However, using GPS technology in football still comes with its weaknesses. These weaknesses are related to usability and quality. As the technology depends on the satellites, the quality is strongly dependent on the location of use. In practice, one of the major weaknesses of GPS technology is that the technology is not usable indoors in a sport that has naturally quick changes of direction which requires extreme precision. Furthermore, even outdoors the functioning and quality is dependent on the number of satellites available and physical obstacles. In addition, atmospheric conditions can affect the strength of the signal. (Schulze et al. 2021.)

Finally, the leagues are still using their own game balls, which means that there is not yet possibility to use teams own GPS technology to gather data on the ball movements, as the ball does not have a transponder set in. In addition, GPS technology requires manual maintenance, such as charging the battery of the transponders. This can be time-consuming for a football team which might practice on a daily level. (Pons et al. 2019.)

Table 1 Strengths and weaknesses of GPS in football

Strengths	**Weaknesses**
+ Cost efficiency	- Quality
+ No installation required	- Not usable indoors
+ Integrable	- Ball tracking
+ Not locally bound	- Maintenance
+ Real-time	- Signal strength

2.2.2 Video technology

Video technology has been used in football since 1980s to support in decision-making. Video recordings have been used to better understand qualitative aspects in sports such as players performance, tactical and technical aspects. (Ying et al. 2011.)

Video technology has been an efficient, cost-efficient, and relatively easy way of gaining more understanding of the players performance and events of the games. However, the processes related to using video technology have been mostly manual. Manual tasks, such as watching video recordings has been time-consuming and inflexible. Additionally, it has completely lacked the possibility to gain real-time insights of performance. (Ying et al. 2011.)

During the last decade, technological development has enabled automation of video technology and its processes. Various algorithms, computer vision, artificial intelligence and other applications have been introduced to automatically track the game. Defined as video-based target tracking, the video image can be used to automatically identify desired objects and events. This has received a lot of interest amongst researchers in the field of computer vision. (Hui 2019.)

Automated tracking system based on video technology requires video cameras and a software to be able to record the game and process the information. There are various systems for different objectives, which might use different kind of logic. For example, systems based on multiple object tracking (MOT) is able to automatically identify desired factors from the video image and present the location of the players and the ball as well as identify desired game events. (Xing et al. 2010.)

The location can be processed for example in a X and Y coordinate format and be listed against relative time stamp. For identifying relevant events, in general the technology is able to learn typical features of certain event and classify the events accordingly. These events could be for example goals, fouls, or shots. At its simplest, when an event takes place, it is possible to add this to the team and individual player's statistics automatically, which enables teams and players to be compared with each other. (Yu & Farin 2005.) Hereby, target tracking combined with event detection enables possibilities for an in-depth analysis of player performance.

Video technology also has its own strengths and weaknesses (Table 2). The major strength of this technology is that it can be operated without a need of any additional transponders or receiver devices on the players' equipment. This enables the usage of this technology regardless the league and game. For its usability, most of the top leagues use video technology for the official games. (Linke et al. 2020.)

Furthermore, video technology has been found to be especially good for evaluating tactical aspects of the game. In addition of gathering knowledge of one's own team, the technology theoretically gives the possibility to track and gather data of the opponent team as well as their positioning on the field in relation to own locations. (Meng et al. 2020.)

However, video technology still has a lot of room for improvement. The need for multiple cameras from different angles might need a fixed installation decreasing the flexibility and usability of the technology. This means that it might be hard to use the technology outside of the home stadium, depending on the system. Portable versions have been developed, but they are still rare and often expensive. (Li et al. 2022.)

Also, the algorithms are still being further developed. Special events such as celebrating a goal, lost sight of the ball, and weather conditions can still cause problems and weaken the reliability of the data and the automated functionality of the technology. (Cuevas et al. 2020). In addition, adjusting the algorithms requires technical know-how, and might still be expensive for the teams.

Table 2 Strengths and weaknesses of video technology in football

Strengths	Weaknesses
+ No receiver devices required	- Reliability: algorithm-based
+ Able to track both teams & ball	- Might require installation & configuration
+ Easy to use	- Cost
+ Able to identify events	- Often locally bound (portable solutions?)
+ Real-time	- Adjustability

2.2.3 Radar and microwave technology

Radar- and microwave-based systems have gained a lot of attention in the sports world for its ability to gather reliable positional data. Systems based on radio frequencies tend to have similar elements compared to the GPS system, but generally with more reliable and accurate results. As with GPS technology, the radar and microwave-based systems can collect positional data of the players and the ball and provide the data as location coordinates in relation to time. (Botwicz et al. 2017.)

Ultra-wideband technology (UWB) is a popular tracking technology operating with radio frequencies. The technology has gathered popularity and is seen to be promising and well-suited radar- and microwave-based system to be exploited in sports. The functioning of UWB is based on multiple anchors installed across and above of the object to be measured, and small transponder units or tags, similar to the transponders used in GPS technology. In football the anchors are usually installed to the roof of the stadium, covering the entire surface area of the field. (Waqar et al. 2021.)

The location data can be gathered directly in the stadium or training ground through anchors continuously transmitting radio frequency across the field. With the tags on the players equipment, the accurate position of objects can be derived using the angle of arrival (AOA), received signal strength (RSS), or time of arrival (TOA) depending on the positioning algorithm. For example, algorithms based on time of arrival methodology, the location is calculated through anchors sending signal and this signal meeting different tags at differing points in time depending on their location. (Waqar et al. 2021.)

Another radio frequency -based suitable solution for player tracking in sports is Bluetooth-based tracking systems. These systems function similarly to UWB technology with an anchor-like locator devices installed on the field, and wearable tags worn by the objects of measurement. (Figueira et al. 2018.) Bluetooth-based solutions offer alternative, cost-efficient way of gathering data of players (Zafari et al. 2019).

Radio and microwave technologies have many strengths for player tracking in sports. Among these strengths, data quality and reliability can be considered one of the most important. For example, UWB-technology is able to capture accurate real-time location data generally with an accuracy of 1-10 centimetres, depending on the technology algorithm used. (Waqar et al. 2021.)

Another major strength in relation to alternative solutions such as GPS, is the indoor usability of most radio and microwave technologies. Since the systems are installed directly at the measurement location, the signal strength is not affected by objectives or other factors. (Wu et al. 2021.) This feature is a driving factor for the usability and popularity of radio and microwave technologies in sports also played indoors.

However, radio and microwave technologies still have their weaknesses and development areas, which affects the technology's flexibility and usability. These weaknesses include that the systems are usually based on fixed anchors requiring installation. This can make the system to be expensive and require high effort as the installation needs precise calibration. (Pino-Ortega et al. 2021.)

In terms of flexibility and usability, this also means that the system can be used, and data collected only from the field where the technology is installed. Portable options have started to be developed, but they are still very much in testing and development phase. (Pino-Ortega et al. 2021.)

Table 3 Strengths and weaknesses of UWB technology in football

Strengths	**Weaknesses**
+ Accuracy	- Often locally bound (portable solutions?)
+ Usability (indoors)	- Installation & configuration
+ Signal strength	- Cost
+ Real-time	- Ball tracking

2.2.4 Wearable technology

Wearable technology or wearables refers to a category of wireless electronic devices able to measure its user. The devices are typically small in size and can be worn or attached directly to the object of measurement, for example equipment of athlete. Wearables have found to be convenient and cost-effective way of capturing different data, and thus it has been widely used for many purposes and industries from leisure to work, and healthcare to professional sports. (Godfrey et al. 2018.)

Wearables have gained a lot of attention during the last decade. The market for these devices has been hot for a while with a more than 400 million devices sold in 2020 from which sports related devices accounted for 60%. (Arnold & Sade 2017.)

Wearable's functioning is based on various electronic sensors able to capture different data depending on measurement needs. Sensors takes advantage of electrical signal to detect physiological responses. Commonly, these signals are different based on the identified action, allowing built in algorithms interpret the signal and create specific outcomes. These outcomes can include data of activity, sleep, stress, and movement few to mention. (Godfrey et al. 2018.)

Depending on the data needs, sensors can include inertial measurement units (IMUs) and microelectromechanical sensors (MEMS). In general, these include a combination of accelerometers, gyroscopes, magnetometers, and flex sensors able to measure kinetic and kinematic parameters of the user. (Adesida et al. 2019.)

According to Godfrey et al. (2018) wearables can in general be subdivided in two categories: primary and secondary. Primary wearables include devices which are operating independently and acting as a central connector for other devices. Wearables of this category can for example be different fitness trackers. Secondary wearables include devices monitoring and measuring specific actions sending it to a primary wearable for further processing and analysis. This category can include devices such as heart rate monitor or fitness belt. (Godfrey et al. 2018.)

Wearables are commonly used in sports to due to its ability to capture data flexibly and cost-efficiently with a possibility of real-time monitoring and analysis. Wearables are especially used in football to capture health and fitness related data. (Aroganam et al. 2019.)

Wearables come with many strengths. Most of these advantages in football are related to usability, cost, and flexibility. Compared to many other solutions, wearables are cost-efficient way of gathering versatile data, even indoors, and many teams can afford it. Also, the technology fits sports well due to the small size of the devices. In addition, open-source applications enable the purpose of the use to be adjusted depending on the need. (Steijlen et al. 2021.)

However, the technology is still under development and is expected to bring its full potential in the future. The weaknesses are related to reliability and effort. The cost-efficiency might come with quality and durability issues as the devices in sports are still under development. The data quality and accuracy still vary a lot with different solutions. Additionally, battery replacement or recharging requires manual effort which can be a chore for a usage of a football team. (Chambers et al. 2015.)

Table 4 Strengths and weaknesses of wearable technology in football

Strengths	Weaknesses
+ Cost efficiency	- Reliability
+ Usability (adjustable & easy to use)	- Quality
+ Not locally bound	- Durability
+ Real-time	- Maintenance

2.3 Choice of technology

The technologies presented in the chapter are all popular and widely used solutions for capturing data in football. As presented in the subchapters, all these technologies possess their own strengths and weaknesses. Choosing the right system is influenced by many different factors.

These factors may include resources, objectives, strategies, and capabilities. For a team with a lot of resources, a combination of several systems might be the optimal and most comprehensive solution. On the other hand, smaller team could be satisfied with a GPS system to collect indicative data to support in planning the content of the training sessions.

Windt et al. (2020) suggests a critical decision-making framework for choosing a suitable technology for a team in sports. The framework arises four relevant questions to critically consider as team is deciding on integrating technology. (Figure 1) The questions are:

- Would the promised information be helpful?
- Can you trust the information you'll be getting?
- Can you integrate, manage, and analyze the data effectively?
- Can you implement the technology in your practice?

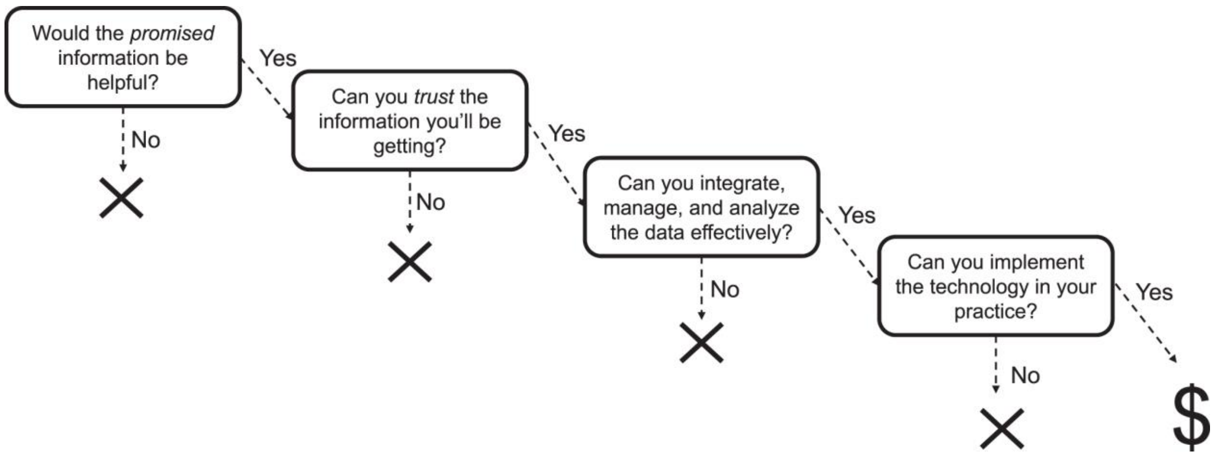

Figure 1: A critical decision-making framework for integrating technology in sport. (Windt et al. 2020)

Each of the questions have an important follow-up question, sources of evidence, and take-home messages which are presented in the following table (Table 5).

Table 5: A critical decision-making framework for integrating technology in sports. (Windt et al. 2020)

	Would the promised information be helpful?	Can you trust the information you will be getting?	Can you integrate, manage, and analyze the data effectively?	Can you implement the technology in your practice?
Follow-up questions	What question will you answer or what decision will you inform? Has a need already been identified for the promised information?	How much validity-related evidence is available regarding the new technology? Are you confident enough with the limitations of the technology to inform practice?	In what format and by what means is information from the technology delivered, and how much cleaning needs to be done to integrate it with other measurements? Do you have the analytical resources to handle and analyze the data?	What burden is placed on athletes and practitioners to collect the data? Does your culture allow for technology to be implemented and data to be collected, and will the technology affect the culture? Does your context allow for data to inform and alter practice?
Sources of evidence	Understand the challenges of your own context. Consult with other researchers and practitioners who have faced similar questions and challenges.	Scientific literature surrounding the validity-related evidence for the technology Internal validation and reliability Professional network	Data samples from the company, short-term trials Internal discussions or methodologic and statistical consultancy Professional network	Qualitative scientific evidence Internal communication (formal and informal) and education
Take-home messages	Start with the end in mind. Evaluate the existing environment and infrastructure to see whether you need new technology to get the information.	Evaluate continually. Consider the consequences. Pilot where possible. Partner where appropriate.	Plan ahead. Educate practitioners involved in collection on proper formatting and process. Automate processes where possible. Audit data and proactively set up quality controls.	Understand the implementation context. Look for invisible monitoring opportunities. Build technological implementation into existing routines.

Furthermore, it is also important to consider that in technologies that requires transponders or tags, such as GPS and UWB, tracking the location of the ball would require an installation of the receiver device inside the ball. This is not yet possible in all the official games. Often leagues use their own balls for official games. However, monitoring the position of the ball would give teams important data that can be used both in evaluating performance and tactical side of the game. For this, video technology has its strengths as it does not require transponders.

To answer the first support research question, table 6 below presents the most used technologies in football and summarizes their strengths and weaknesses.

Table 6 Strengths and weaknesses of popular technologies used in football.

	GPS	Video	UWB	Wearables
Strengths	+ Cost efficiency + No installation + Integrable + Not locally bound + Real-time	+ No receivers + Teams + ball + Easy to use + Identify events + Real-time	+ Accuracy + Usability (indoors) + Signal strength + Real-time	+ Cost efficiency + Easy to use + Not locally bound + Real-time
Weaknesses	- Quality - Not usable indoors - Ball tracking - Maintenance - Reliability	- Reliability - Installation - Cost - Locally bound - Adjustability	- Locally bound - Installation - Ball tracking - Cost	- Reliability - Quality - Durability - Maintenance

3 Data

3.1 Action to Data

The amount of existing data has increased enormously in the last decade. There were more than 64 billion gigabytes of data created, captured, copied, and consumed in the beginning of 2021 with a speed of approximately 2,5 billion gigabytes of data generated daily. (Statista 2022; Dobre & Xhafa 2014).

Data has become important factor in terms of efficiency and decision-making. According to Chen et al. (2014) data can be seen as mutually important factor as traditional factors such as material assets and human resources. This applies to the sports industry, where the introduction of different technologies has enabled the collection of large amount and various types of data.

With increasing amount and complexity of data, this chapter first defines and discusses big data. Thereafter possible data sources and types used in football are presented as well as their potential use cases.

3.2 Big data

The enormous growth in the amount of data has led to the inadequacy of traditional data tools in data processing. Concept of big data has arisen, which generally refers to a complex and large set of data, which processing requires more advanced tools, so that the data can be analysed, managed, and processed quickly enough. (Chen et al. 2014.)

Big data has been a hot trend for a while now, however despite the public attention, the concept has been problematic for precise definitions in the scientific literature due to its complex nature. Its definitions vary in different contexts and among different stakeholders, for example between organizations and academic research. (Ward & Barker 2013.)

According to Awan et al. (2022) big data can be defined as high-volume, composite, and real-time volumes of data, which requires advanced tools and management to be able to take advantage of the data (Awan et al. 2022). Big data has also been defined as “new paradigm of knowledge assets” (Hagstrom 2012), and as the “the next frontier for innovation, competition, and productivity” (Manyika et al. 2011).

Different models have been established to present the typical elements of big data. Five Vs of big data is one of the established models in the literature. The model identifies five elements specific to the nature of big data. These elements are volume, velocity, variety, value, and veracity. (Thudumu et al. 2020.)

Volume refers to the large size of the big data. The growing amount of data forces re-evaluation of the available tools. Analytical tools have to be able to handle large and ever-growing sets of data efficiently to be able to improve decision-making by producing insights (Ghasemaghaei & Calic 2019). According to Gandomi & Haider (2015), the size categories of individual big data sets can typically be described in terabytes and petabytes, but it is not meaningful to define an exact limit value for the size that qualifies as a big data set.

Velocity of big data refers to the time it takes to create the data. This feature of the five Vs evaluates how quickly data is generated and in what time the data should be processed. Due to digitalization, there has been large increase in the number of digital devices which has led to an unprecedented velocity in the generation of data. This speed causes a growing need for screen-based planning and real-time analysis of data, which on the other hand, if successful, enables efficiency to improve. (Gandomi & Haider 2015.)

Variety refers to the structural non-uniformity of the data. Meaning that data is available from several different sources. Data collected from different sources can be structured, partially structured or unstructured. There is least of structured data because it is already processed and sorted data. (Gandomi & Haider 2015.) Unstructured data refers to a random and inconsistent data which is inherently difficult to analyze. It is estimated that around 95% of the existing data is unstructured. (Cukier 2010.) Partially structured data contains characteristics of both above (Gandomi & Haider 2015).

Introduced by SAS, big data’s element of value refers to the complex nature of data, which requires special attention in analyzing the data. Value can be achieved by successfully analyzing large data sets, thus increasing the value density of the data set. Additionally, data does not bring value if there are no people who can interpret the results. (Gandomi & Haider 2015.)

IBM has introduced the element of veracity. Veracity refers to the data accuracy of various sources. Other sources may be in nature more reliable than others. This has to be taken into account as using analytical tools to analyze data from various sources. It is not necessarily possible to directly compare or combine data from several sources. (Gandomi & Haider 2015.)

The characteristics of big data presented above make the processing of data challenging (Johnson 2012). The large amount of data additionally arises challenges related to information security (Yaqoob et al. 2016). If sensitive personal data ends up leaked, it might significantly decrease reputation and weaken the trust of people towards the responsible entity (Tankard 2012).

3.3 Data sources and types in football

The collection of data in sports has increased and become easier during the last decades due to the technological development. The development of various technologies presented in the previous chapter has enabled progression in the data collection in football as well. (Castellano et al. 2014.) Football teams have started to collect data from their players not only during training but also during official games, after the national football association (FIFA) approved the use of wireless sensor technology during competitive sports events in 2018 (Rein & Memmert 2016; Di Salvo & Modonutti 2009; Bush et al. 2015).

Technological development has made it possible to improve the reliability and quality of data collected in sports. In addition, it has enabled the development of new data sources to produce increasingly relevant information (Rein & Memmert 2016). However, Mondello and Kamke (2014) point out that: "Although the availability of data in sports has grown considerably, the information shared between different organizations, academics and professionals is still often limited and anecdotal."

Sports teams use both internal and external sources to collect data. The teams' internal data sources include for example video data captured by video cameras, biometric data collected by wearables, and location data collected by tracking technologies such as GPS and UWB technology (Morgulev et al. 2018).

External data sources refer to third-party operators, often companies, that collect data from competitive games. For example, Opta (Optasports 2023), a company specialized in collecting and providing sports data, is a well-known third-party operator that collects game data and provides it for teams and the most enthusiastic fans.

In addition, sports leagues collect data to improve the customer experience and offer it to teams and fans. Third parties also provide private data to teams, which enables the use of data which is not available for the other teams. Additionally, if the teams do not have required technical capabilities, the third parties also offer data analysis and other services for the teams. (Davenport 2014a.)

See below illustration of data sources in football (Figure 2). It is good to note, that teams do not have to decide between internal and external data source but can use both.

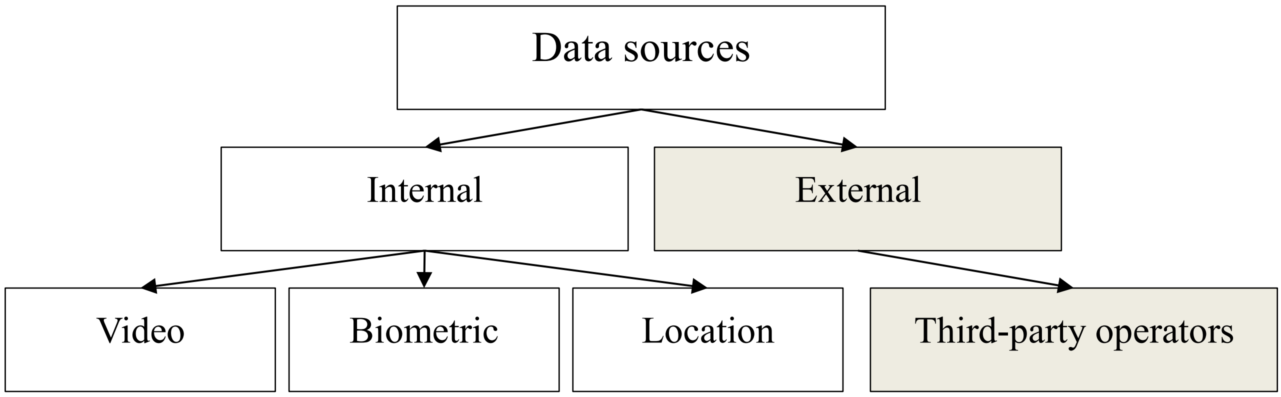

Figure 2 Data sources in football

In football, it is possible to capture versatile data, and monitor various metrics from individual players, teams, and games (Haghighat et al. 2013). The data can contain several different metrics depending on the objective of the data usage. According to Schumaker et al. (2010) data already plays a big role in supporting teams' decision-making.

The following subsections present in more detail the most used data types in football. These include location data collected by the tracking technologies, biometric data collected by the wearables, and video data collected with the video technology.

3.3.1 Location data

Location data is one of the most used data types in sports. Location data can come in many formats, for example as a X and Y coordinates related to time. True to its name, location data provides information of where the players and ball are in the field in differing points of time. This information can be used in many ways to calculate various other metrics.

Location data can provide teams with insights into the physical performance of the players, the physical requirements of the games, as well as the health of the players. This information can be used many ways depending on the objective, for example for monitoring players workloads and planning the content of team's exercises. (Aughey & Falloon 2010; Ehrmann et al. 2016.)

Location data can be used to derive metrics for physical performance such as players' endurance, speed, and agility. With existing technologies, accurate real-time physical data can be used to detect fatigue in games. This is done by identifying the most intense periods of the game and can help in decision-making related to making substitutes during a game. (Aughey & Falloon 2010)

Location data can also be used to gather tactical information. Knowledge on players movement and physical attributes can support in decision-making related to creating alternative strategies during games based on players performance (Aughey & Falloon 2010). Location data combined with other types of data enables monitoring the performance of the players in different game positions or locations of the field (Aughey 2011).

The strength of current days technologies providing location data, is the possibility to analyze players real-time. This enables analysis to be made during games or, for example, at halftime of a game. Location data provides teams and scouts with meaningful information of players' performance and physical attributes. (Sacha et al. 2014.)

3.3.2 Biometric data

The technological developments together with the allowed use of wireless sensor technology in games have enabled an optimal environment for the collection of biometric data in football. Today, biometric data can be collected from all individual players during games with the wearable technology presented in the previous chapter. (Carling et al. 2008; Di Salvo & Modonutti 2009.)

Biometric data refers to information of various aspects of human body. Biometric data is used to measure players' physical performance, fitness, and health. Biometric data collected by sensors provide coaches and players versatile information which can be used to recognize players current state, identify development points, and determine the direction for future planning. (Kos & Kramberger 2017; Bates 2020.)

Biometric data can be used to monitor players physical and physiological characteristics and performance. Monitored physical metrics can include intensity, activity, motor skills, and body and muscle functioning. Knowledge on physical performance can be used to optimize workloads and trajectories, prevent excessive loads, and identify areas of development and potential risks. (Sikka et al. 2019.)

Furthermore, biometric data can be used to monitor players fitness. This can include determining players power output, muscle strength and oxygen uptake. This gives further knowledge to better understand the underlying variables affecting other metrics such as physical performance. In addition, these support recognizing players physiological strengths and weaknesses to formulate personal training programs. (Elliot et al. 2019.)

In addition, biometric data is increasingly used for health monitoring. Collected health data includes metrics such as body temperature, pulse, and overall physical condition. Analyzing these can provide knowledge on players' health and even mental state. (Carling et al. 2008; Kos & Kramberger 2017.) For its usability for health monitoring, biometric data has begun to be used in the medical research, for example in the research of head injuries (Kos & Kramberger 2017).

3.3.3 Video data

Video analysis in sports has developed a lot over the recent years. Its suitability for use in sports has drawn attention for both sports teams and research community (D'Orazio & Leo 2010). Video analysis consists of processing of video data which can include data points on positions, tactics, and events. (Wang & Parameswaran 2004).

Furthermore, processing of video data in sports has found many different applications to the delight of both teams and fans. Slow-motion playback, automatic identification of players, analysis of repeated patterns and events, and statistics of game events are examples of what video data enables. In addition, video data enables teams to find problem areas, interpret tactics, and identify recurring situations and similarities in games. (Wang & Parameswaran 2004.)

Football video analysis includes three main application areas, which are video summarization, provision of additional information, and high-level analysis. The most important task of the video summarization is to automatically find the most important events and performances of the games, such as goals and fouls. After this, the highlights of the game can be separated into their own entity and the user can analyze only the events that interest him. (D'Orazio & Leo 2010.)

The automatic feature of video analysis makes it possible to collect and produce statistical data of games without manual action in real-time (Shah et al. 2015; Xing et al. 2010). In the future, this will create opportunities for more extensive real-time analysis of statistics.

Provision of additional information refers to the interpretation of video data, the goal of which is to produce improved content for the user. This interpretation includes, for example, the analysis of the individual players participating in the various events of the game and the identification of the nature of the events. (D'Orazio & Leo 2010.)

The objective of provision of additional information is to enable automatic recognition of different players and different game situations. It can therefore be used to identify and distinguish individual players and the ball. In addition, video cameras enable the identification and separation of different areas of the playing field, such as the goalkeeper's area, the centre circle, and the corners. (Wang & Parameswaran 2004.) This enables to automatically gather statistical data of where different events usually take place.

High-level analysis aims to detect characteristics such as player skills or team strategies. This requires a more complex assessment to be successful. The high-level analysis uses recognition of the players' movements and position, as well as multidimensional analysis of the ball's trajectory. (D'Orazio & Leo 2010.)

3.4 Data discussion

The data types presented in this chapter are all useful in their own way. They can be used and combined for different purposes depending on the team's needs and objectives. At its best, the team is able to utilize different data types from different sources, forming useful cross-disciplinary knowledge. (Windt et al. 2020.) Table below summarizes the data types used in football answering to the second support research question (Table 7). Table illustrates the different use cases of each data types as well as their general metrics.

Table 7 Data types used in football

	Location data	Biometric data	Video data
Measure:	Physical performance Tactical performance	Physiological performance Health and fitness	Strategy & Tactics Tactical performance
Common metrics:	Endurance Speed Aglity	Intensity Activity Strength	Game statistics Events Ball tracking
Enables:	Physical and tactical monitoring and planning	Physiological monitoring Health promotion & Injury prevention	Performance monitoring Tactical planning

4 Analytics

4.1 Data to Insights

Having a large amount of data available, analytical tools are needed to process the data. These days, many football teams have internal analytics teams responsible of the analytics and data processing. For example, Liverpool FC's analytics team have used analytical tools for a long-time, transforming data to insights (Lichtenthaler 2022).

In this chapter, relevant concepts such as data analytics, big data analytics and sports analytics are discussed and defined. Furthermore, different methods and analytical areas are presented. Finally, the third support research question is answered by presenting performance and health analytics, used in football.

4.2 Big data analytics

Data analytics refers to a tools and methods of processing data to improve efficiency and support in decision-making. Data can be processed by various statistical and quantitative tools and methods, for example by creating descriptive analysis and predictive models. (Davenport & Harris 2007.) Wilder and Ozgur (2015) define data analytics as the application of techniques and processes aimed at improving decision-making by turning unstructured data into insight.

Insights refers to a deep and intuitive understanding of various phenomena, so that people can make use of them (Sharma et al. 2014). In this thesis, insights refer to findings made from data with the help of analytics, which sports teams can use to improve their decision-making. Data-driven insight refers to utilizing current and previously collected data with the help of analytics (Ghasemaghaei & Calic 2019).

The objective of analytics is to generate insights which can be used in decision-making of the using entity. The generation of insights can be seen as an active engagement process between different stakeholders, such as analysts and decision-makers, using different analytical processes and techniques. The benefit from data in decision-making requires that several different actors from different areas of the organization cooperate. (Sharma et al. 2014.)

Due to the increasing amount of data, the possibilities of analytical tools have had to be re-evaluated. Big data analytics refers to analytical tools of a new era, which enables the complex and large data sets to be processed (Najafabadi et al. 2015). Many organizations have tried to respond to the increasing amount of data by investing in advanced analytical tools to improve the quality of their decision-making (Müller et al. 2016).

Big data analytics enables the discovery of hidden patterns and relationships in large data sets and thus helps generating insights (Kwon et al. 2014). Big data analytics have been used widely in the business but also within healthcare, e-government, security, and education few to mention (Frizzo-Barker et al. 2016). Big data analytics is also used in sports due to the large amount of data and various data sources (Rein & Memmert 2016).

The utilization of modern technology, big data and advanced analytical tools has created both new opportunities and challenges (Lam et al. 2017). In particular, the processing and analysis of big data have not only brought opportunities but also several challenges. For example, generating insights from big data (big data mining) is significantly more difficult compared to data which does not contain characteristics such as large size, speed, and diversity. (Chen & Zhang 2014.)

In addition, the fact that big data comes from several independent and disparate sources makes its processing and analysis more complicated (Wu et al. 2013). According to Wang et al. (2016), companies also have the challenge of finding competent employees who, in addition to managing technical issues, can interpret and utilize insights obtained from big data.

4.3 Sports analytics

Literature and research in the context of sports analytics is still relatively new and limited, but its amount is growing. The amount of research on sports analytics in the peer-reviewed scientific journals of Information Systems Science is also low. So far, most research on sports analytics has been written from the perspective of mathematics, statistics, and marketing. (Caya & Bourdon 2016.)

However, sports analytics research goes back 50 years (Wright 2009). Despite, only in the last two decades has sports analytics received more public attention. Interest in the subject was sparked at the beginning of the millennium. Around the same time, Michael

Lewis' published a popular book on the subject, Moneyball: The Art of Winning an Unfair Game (Shah et al. 2015).

Today, most top-level football teams often have either an analytics department or an analytics specialist (Steinberg 2015). The English Premier League teams are internationally known for using analytics to support in decision-making. Sports analytics has also established its own conference, MIT Sloan Sports Analytics, whose number of participants has grown from 175 to 3,500 participants since its foundation in 2007 (MIT Sloan Sports Analytics Conference 2019).

Sports analytics simply refers to the use of data analytics in the field of sports (Davenport 2014b). More precisely defined, sports analytics means extensive utilization and management of historical and real-time data to generate insights. Its objective is to inform the managers of sports clubs with the help of various information systems and thus improve decision-making to gain a competitive advantage. (Morgulev et al. 2018.)

The definition of sports analytics identifies three basic components which are information gathering, data management, and data analysis. Together, these three basic components form a framework of sports analytics (Figure 3). The combination of the basic components enables the data to be collected, managed, and processed with an objective of improving the decision-making process of desired stakeholders, such as coaches, athletes, or other decision-makers. (Morgulev et al. 2018.)

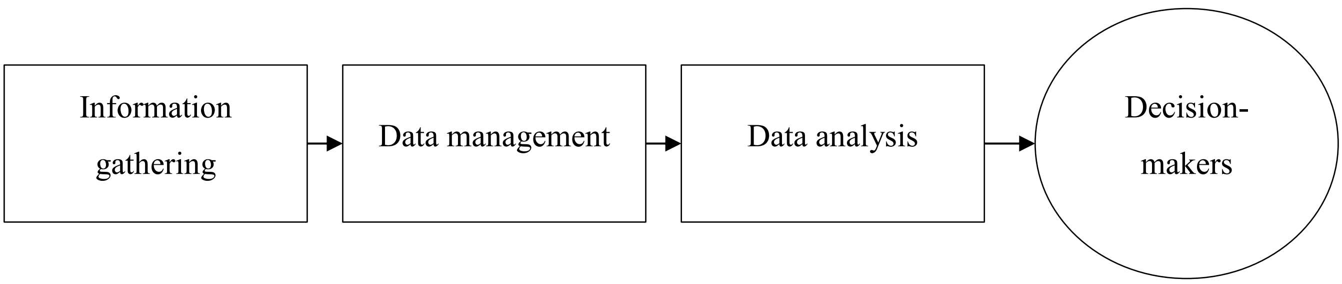

Figure 3 A sports analytics framework (Morgulev et al. 2018).

4.4 Methods

In general, analytics can be divided according to its purpose into three different areas: descriptive, predictive, and prescriptive analytics (Mortenson et al. 2015). Besides performance requirements, teams recognize the strong entertainment purpose of professional sports, and competition for fan interest and popularity occurs both on and off the field of play (Caya & Bourdon 2016).

Tools from different areas of analytics are used in sports clubs for different purposes depending on the objective, but the boundaries between these areas are theoretical and blurred. Often, analytics applications require tools from each of the following areas to produce insights. (Evans & Lindner 2012.)

4.4.1 Descriptive analytics

Descriptive analytics uses data to understand both past and current performance of the organization. The purpose of descriptive tools is to classify, describe, and combine data to be able to understand and analyze the certain phenomena. The objective is to convert data into descriptive insights that can be used to make informed decisions. (Evans & Lindner 2012.) Descriptive insights refer to the information produced by analytics about the current events of the organization (Ghasemaghaei & Hassanein 2016).

In football, descriptive analytics can be used to understand the events of games and create statistics that can be used by coaches, players, and supporters alike. For example, the English team West Ham United F.C. takes 15 minutes after every game, to go through the collected game data and statistics together with the team's players and analysts. This way, the coaches and players receive relevant information about the events of the game, which can be used to improve the performance of future games. (Davenport 2014a.)

Game statistics of individual players and events are also used to improve the spectator experience by providing real-time data to be seen during the games. For example, different diagrams presented in the games, social media, or websites have been found to increase the interest of the viewers. (Caya & Bourdon 2016.)

Furthermore, descriptive analytics enables to rank individual players based on certain metrics (e.g., number of goals). This enables to objectively compare players which can be used to formulate game strategies, in player selection, and for the player transfer market. (Davenport 2014a.)

4.4.2 Predictive analytics

Predictive analytics uses predictive models utilizing for example machine learning (Kibria et al. 2018). Its objective is to process historical data to be able to predict the future. Predictive models can be taught to identify hidden and recurring patterns and relationships within the data. This information can be then used for example for planning and predicting risk. (Evans & Lindner 2012.) Hereby, predictive insights refer to knowledge related to the future (Ghasemaghaei & Hassanein 2016).

In recent years, there has been a growing interest among researchers in creating different predictive models for predicting the end results of football games. For example, Liti et al. (2017) built an algorithm based on machine learning in their research. The algorithm used game data collected over 25 years of English Premier League's end results. With historical data of past end results, they were able to create model predicting the future game results somehow accurately. (Liti et al. 2017.)

Sports betting brokers are using predictive tools that try to predict the outcome of games to be able to accurately set the bet odds. When predicting results, it must be considered that although analytics can be used to obtain useful indicative information about the possible outcomes of football games, due to the complex nature of sports and football, it is not possible to use technology to create a model that would predict the results perfectly. Many bettors who are analytically gifted or play with strong experience and intuition try to take advantage of the inaccuracies found in betting companies' predictions and use it to their financial advantage. (Dixon & Coles 1997.)

4.4.3 Prescriptive analytics

The goal of prescriptive analytics is to find the best alternatives and select them. For this purpose, prescriptive analytics can be used to optimize, minimize or maximize the desired goal. The tools of prescriptive analytics can be used in organizations for example to determine the best possible pricing strategy to maximize revenues. (Evans & Lindner 2012.) Prescriptive information refers to insights that is used to determine alternatives to reach the optimal goal (Ghasemaghaei & Hassanein 2016).

Research published as early as 1983 shows that optimization has been used in sports, for example, in determining the ideal trajectories of athletes' movements (Hatze 1983). In addition, football teams are using prescriptive analytics to improve game performance, for example in planning of game strategy, tactics, and training. Optimization is also used in a financial sense, for example in the pricing of season tickets, and other fan products, and defining the salaries of players and staff. (Caya & Bourdon 2016.)

Furthermore, prescriptive analytics are used in organizing sport events (Ghoniem et al. 2017). For example, the International Football Association FIFA uses optimization in the organization of world championships. FIFA has used optimization to plan resources and in games scheduling. (Ghoniem et al. 2017.)

4.5 Sports analytics sub-areas

Davenport (2014b) identifies three strongly growing areas of sports analytics, which are player performance and game analytics, sports business analytics, and health promotion and injury prevention analytics. Following sub-chapter will present the player performance and game analytics and the analytics of health promotion and injury prevention.

4.5.1 Performance analytics

The processing of data to analyze the performance of sports teams and players has improved considerably over the last decade (Travassos et al. 2013). Performance analytics is widely used in football, and it generally refers to the area of sports analytics, which evaluates the players and teams’ performance for example from technical, tactical, and physical point of view (Carling 2019). Davenport (2014b) especially elevates the European football teams globally in the use of the most advanced sports analytics to assess performance.

Performance analytics is the most referred area of sports analytics in the literature. This sub-area includes for example evaluating teams and players performance and choosing the right game strategy and tactics with the help of an analysis made of one's own and the opponent team. (Davenport 2014a.) However, due to the lack of generally agreed frameworks, the literature does not always refer to this sports analytics area under the same title.

The amount of data collected in the English Premier League is so large that team analysts must carefully choose what and how they use it. Performance analytics mostly uses location and video data to assess the teams and players performance and evaluate strategies. To gather this data, teams are using tracking technologies such as GPS and UWB, and video technologies. (Davenport 2014b.)

The insights created by performance analytics enables analysts to evaluate players and team performance and communicate the insights to enable players to make the right decisions in different situations of the game (Davenport 2014a). However, this is not simple, as many factors have to be considered when evaluating player performance. These factors include internal factors such as objectives, player attributes, and positioning. Additionally, external factors contribute to the player performance such as related to the opponent, whether the game is played at home or away, and even the weather conditions. (Rein & Memmert 2016.)

In football, descriptive analytics is still mainly used to produce analysis of players performance. Descriptive data includes statistics, such as players possession, activity, passes, key passes and so on. This information is relatively easy to process and understand, and it allows the players to be compared with each other. (Davenport 2014b.)

Additionally, prescriptive analytics are being used to evaluate player performance. Prescriptive data can be used to identify different kind of strategies and tactics, such as defensive or attacking playing styles of a team. (Ruan et al. 2022.)

Furthermore, predictive analytics are being used increasingly to evaluate players performance. However, the utilization of predictive models in football to evaluate the performance of players is still relatively limited. Predictive analytics is currently mostly used for analysing players physical attributes, and in health analytics presented in the following sub-chapter. (Martins et al. 2022.)

In general, players performance is often assessed through players attributes and contribution to certain events. In this analysis, for example, players' activity in games and the usefulness of participation in different game situations can be determined. Sacha et al. (2014) divides the analyzed attributes of individual players into three groups, which are individual attributes, game characteristics and the player's game events.

Individual attributes include attributes such as the player's position and speed in different game situations. Additionally, players attributes can be evaluated through statistics such as passes, shots, interceptions, goals, mistakes and so on. When evaluating the players' game characteristics, one can focus on the player's movement. The player's position is studied by comparing his distance to other players and to the ball in different situations. The insights made from individual players can be used in analyzing the team and game, and in comparing different players. Analyzing the game events examines what kind of events take place on the field and why. (Sacha et al. 2014.)

In the performance analytics, external data is also used extensively. For example, the company Opta, which specializes in gathering sports data, offers game data. In addition to the teams, the teams' most passionate supporters also make use of this data by making their own interpretations of the available data. (Optasports 2023.)

In addition, many football teams openly share the data they collect with the public. For example, Manchester City F.C., who plays in the English Premier League, openly share game data with its supporters for open-source analysis. Manchester City openly says that it has used the open-source analysis made by its supporters in its operations. (Davenport 2014a.)

Furthermore, teams’ strategies and tactics can be assessed through performance analytics. The aim of the center of gravity analysis is to create an optimal tactic based on the players positioning and behaviour of the geometric center. The geometric center refers to the center of the team's position, i.e. the center of the different positions of all the players. The team's behaviour on the field is studied by analyzing the movement of the geometric center during different game situations. The research results have shown a strong connection between the changes in the distances between the center of the team and different game situations. Recognizing and analyzing the connection helps the team's coaches to choose the best possible tactics to play against teams with different strategies. (Frencken et al. 2011; Rein & Memmert 2016.)

In the table below, I have compiled examples of methods used in the performance analytics presented in the chapter and divided them into general methods of analytics and advanced methods (Table 8). The general methods of analytics are already widely used in some form in most teams that uses analytics. With these, teams can create value and improve their decision-making. Advanced methods are used by teams that act as pioneers in the use of sports analytics. (Davenport 2014a.)

Table 8 Methods of performance analytics in football

General methods	Advanced methods
External data sources	Video analytics
Descriptive analytics	Open-source analysis
Player’s performance analysis	
Tactical analysis	

4.5.2 Health analytics

Health promotion and injury prevention analytics is a very important and natural area of sports analytics, as player health and well-being can be seen as one of the main drivers of sports team performance (Davenport 2014a). Analytics for health promotion and injury prevention refers to the collection, management, and analysis of relevant information about the health of players to maintain performance and predict and prevent injuries (Hanisch & Hanisch 2007; Davenport 2014a).

Health and injury monitoring dates to the 1950s. In this case, the coaches manually wrote down descriptions and possible causes of injuries that occurred during training or games. (Hanisch & Hanisch 2007.) In the scientific literature, the analytics of health promotion and injury prevention is, however, a relatively new concept. Back in 1996, Twellaar et al.’s (1996) research on athletes' injuries identified that it is almost impossible to collect appropriate and reliable data on sports injuries. Based on the research, the only sure way to prevent injuries was to stop playing sports.

Due to the technological development during the last decade, the huge amount of data available, together with advanced information systems and technologies, has made it possible to use analytics also in monitoring and promoting the health of players. Analytics for health promotion and injury prevention enable accurate data on players' health to be collected automatically. At best, observations made from data create insights and improve the safety of players in game situations. This happens by making predictive analytical models to prevent injuries by identifying potential risk factors. (Wilkerson et al. 2018.)

Thanks to video, location and biometric data, relevant information about the health of the players is constantly collected, which is used to make more efficient and better decisions. Based on several scientific studies, it appears that biometric data is the main source of data for health promotion and injury prevention analytics. Most of the analytics tools for health promotion and injury prevention produce descriptive data. Predictive analytics are also increasingly being used to prevent player injuries. In addition, for example video data can be utilized by studying game situations that are particularly risky for injuries. Video data therefore enables the identification of risk factors. (Davenport 2014a.)

The biometric data collected by the wearable technology and sensors on the players is constantly monitored to study the current state of health of the players and to identify and prevent possible risk factors (Wilkerson et al. 2018). The use of location data and biometric data together enable the monitoring and optimization of the players' workload, which can be used in planning exercises according to the players' stress level (Ehrmann et al. 2016).

Most professional athletes have been injured at least once during their playing career, which may have caused an increasing risk of re-injury (Stanton et al. 2008, according to Wilkerson et al. 2018). According to Wilkerson et al. (2018), the most common injuries of athletes are sprains, stress injuries and concussions. Injuries prevent players from participating in team training and season games. The absence of a player often has a negative impact on the team's performance and finances. Injuries also affect the players' own mental and physical well-being and cause disappointment among the team's supporters. (Wilkerson et al. 2018.)

Today, sports clubs have moved from monitoring health and injuries to comprehensive management of athletes' long-term health and well-being and injuries. Injury management refers to the anticipation and prevention of injuries through the continuous collection and monitoring of players' health-related biometric data. Such information includes, for example, body temperature, pulse, physical condition, and sleep quality. Key player health metrics are constantly monitored, and injured players are consistently rehabilitated. Comprehensive management of injuries may at best even prevent team players from getting injured and thus create a strategic competitive advantage for the team. (Hanisch & Hanisch 2007; Kos & Kramberger 2017.)

Various analytical solutions have been introduced to monitor and analyze related to players health. With IBM’s predictive analysis - fatigue, intensity, and collisions can be monitored to produce accurate predictions and decrease the injury risk. (Millington &

Millington 2015.) As another example, SAP has introduced Injury Risk Monitor technology to monitor players health and forecast risks (Kim & Park 2015).

The team AC Milan, which plays at the highest league level in Italy, founded its own player health research and analytics center in 2002 and acts as a pioneer and trendsetter in the analytics of health promotion and injury prevention. The center, called MilanLab, utilizes a database consisting of biometric data collected from the team's players, which is used to create predictive analytical models to promote the players' health and prevent injuries. (Davenport 2014a.)

AC Milan collects mental, biochemical, and musculoskeletal data on its players. Data is collected using various scientific devices developed for sports, which are automated to alert the players when some of the measured values, for example pulse, exceed or fall below a certain set limit value. The team also uses video data, and for example, from a single jump in a game, more than 200 data points can be collected with the help of technology for more detailed analysis. (Davenport 2014a.)

In the very first year of the research center's operation, the team experienced a dramatic reduction of more than 90% in injuries compared to the previous five years. Thanks to MilanLab, the team's level of injuries has remained low since the establishment of the research center. (Davenport 2014a.) Many other professional teams have also subsequently established their own research centers that utilize analytics for health promotion and injury prevention (AC Milan 2018).

The table below shows the general and advanced methods of health promotion and injury prevention analytics presented in the chapter (Table 9). General methods of analytics, such as the use of descriptive analytics, are in common use in teams that utilize analytics. Advanced analytics methods refer to analytics tools that are used only by teams that are pioneers in the use of sports analytics. (Davenport 2014a.)

Table 9 Methods of health analytics in football

General methods	**Advanced methods**
Internal data sources	Predictive analytics
Descriptive analytics	Video data for predictive analysis
Biometric data	

5 Data-driven decision-making

5.1 Insights to decisions

As analytics have become more common in football, more and more teams are considering adopting it. Although teams continue to invest in the new technologies and analytical software (Davis 2013), the real value created by sports analytics for decision-making has been studied little in the scientific literature so far. Especially, using data to support in evaluating player performance lacks general frameworks in the literature.

How exactly does data leverage teams in decision-making and particularly in the evaluation of players performance? In this chapter, first data-driven decision-making is discussed. Thereafter, to understand how analytics can be used to evaluate player performance, player performance is considered through discussing the most relevant attributes of the players.

5.2 Data-driven decision-making

Decisions are made daily. Some decisions happen unconsciously without us even realizing the process. Other decisions might be results of careful chain of thoughts or work. This is usually the case with decisions with a large impact, for example in business context. Decisions always have an impact, small or large depending on the decision and context. Referring to Williams & Ford (2013), decision-making can be defined as the ability to use information of the current situation and select and execute actions depending on the goal.

In football, decisions are traditionally often made based on experience, sports knowledge, and intuitive feeling (Adesida et al. 2019). Hereby, the aspects of data-driven decision-making in sports have gained attention in the research, particularly in economics and psychology (Balafoutas et al. 2019).

There are many factors influencing decision-making. These can be, for example uncertainty, time pressure, emotions, and consequences. According to Sjöberg (2003) analytical decision-making is seen unfavourable compared to intuitive decision-making in non-professional roles. However, intuitive decisions are often impacted by emotions and feeling. Decision made based on statistical analysis rely on data and thus has the element of neutrality and objectivity, keeping the feelings out of the decision-making. (Sjöberg 2003.)

In a very controversial industry as sports is, coaches, players, and fans often have their own subjective views and opinions on how the sport should be played. However, as seen in the last decades for businesses, analytics can greatly improve the decision-making and bring objectivity to assess the performance. Football teams are no different, with an increasing opportunity to use technology, data, and analytics as presented in the previous chapters.

According to Ghasemaghaei et al. (2018) analytics can play a great role in improved decision-making. However, the benefits of analytics are rarely fully realized. The impact of analytics in improved decision-making is based on five dimensions: data quality, bigness of data, analytical skills, domain knowledge, and sophisticated tools (Figure 4).

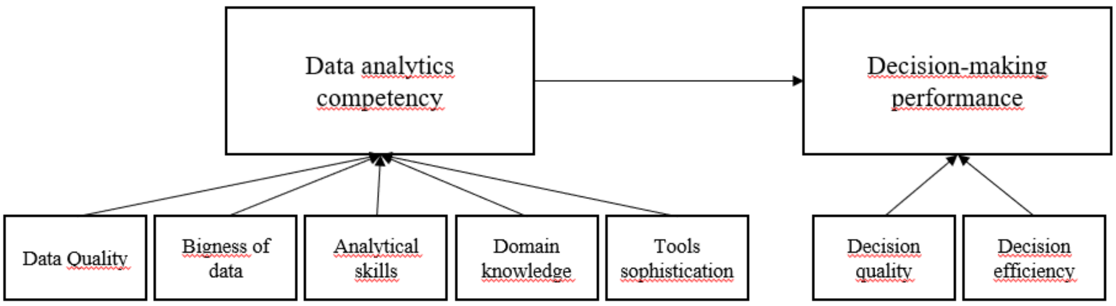

Figure 4 Analytics dimensions for improved decision-making (Ghasemaghaei et al. 2018).

All these competency dimensions are essential, as they all contribute into the overall analytics competence of an organization. In addition, all the dimensions have an impact on the decision efficiency and quality. To improve data-driven decision-making performance, organization should focus on developing each of the dimensions. (Ghasemaghaei et al. 2018.)

According to Patel et al. (2020) research on how big data affects the decision-making in sports, bigness of data, analytical skills, and sophisticated tools can indeed enhance the decision-making vastly, as the decisions can be done with a help of data instead of counting solely on intuition. Data provides objective information and thus enriches evaluation, which adds value to decision-making. Thereafter, data analytics competencies of a team provide a significant opportunity to achieve a competitive edge amongst the competitors.

However, it is important to note that as the technologies used in sports are still relatively new and constantly evolving, not all teams have the resources to implement expensive technologies. Decision-makers in sports need to recognize that adopting of sports technology to improve decision-making is time-consuming and requires a lot of resources. (Ratten 2019.)

To benefit from analytics for decision-making, the target state must be defined. To be able to answer the main research question, we must first answer the fourth support research question. The next subsection discusses team and player performance to identify factors which affect the players performance.

5.3 Team and player performance

Football is a multidimensional sport where the teams and player’s characteristics consist of many different factors. This makes comparing teams and players to each other complicated and not always the easiest task. Depending on the team’s objectives, strategy, and tactics, different attributes bring more added value to the team than others.

Football team consists of 11 players operating simultaneously on the field. High performance requires collaboration between the players to execute the chosen game strategy and tactics. According to Trequattrini et al. (2015) it is easier to evaluate individual player’s performance than team performance as it requires understanding of relationships and their quality. In their study, network analysis was conducted to assess team performance through the underlying relationships. As a results, they identified that team structure and dynamics can be assessed through network analysis evaluating indicators consisting of four main categories – indicators of efficiency, vulnerability, cohesion, and activity and leadership. (Trequattrini et al. 2015.)

Traditionally, other way of evaluating team performance has been analyzing historical data and past game performance. Game data has been also used to recognize different strategies such as passing strategies in different games. (Cintia et al. 2015.) Moreover, teams’ strategy and tactical decisions, such as chosen formation, affects the teams and players performance (McLean et al. 2018). The players’ compatibility on the game strategy and contribution towards team dynamics affect the team's performance. High team performance requires ability to do quick decisions and anticipate the game and the actions of team members and opponents around.

As individual players capability to contribute for the benefit of the team depends largely on the combination of individual attributes (Duarte et al. 2013), is player performance relevant to examine. To be able to evaluate players performance, has the most relevant player attributes be considered. In general, individual player attributes can be assessed and divided various ways. According to Williams et al. (2000) players' skills can be assessed already from early-stage through physical, physiological, psychological, and sociological factors. However, the prioritization of player core attributes and team-specific needs change coming to the professional level, where for example the sociological factor no longer has much or any importance.

Individual player attributes can in more general be categorized into four factors: physical, technical, tactical, and mental attributes. (Yang et al. 2018; Ryoo et al. 2018.) These attributes form a whole of player attributes, which is used in this thesis as a definition of the player's overall skills.

Physical attributes of a player consist of all the elements and measures indicating and contributing to the players physical performance and physics. These attributes include metrics such as speed, agility, strength, endurance, motor skills, coordination, jumping skills and so on. All these attributes are tangible as they are easy to identify, monitor, develop, and compare on different players. (Kelly & Williams 2020.)

Technical attributes of a player consist of metrics forming the overall technical performance level of a player. These metrics include variables for example of dribbling, passing, shooting, tackling, and heading. The level of these variables on a player can be identified through game-related statistics. However, there is no unequivocal way to evaluate the technical skills of players, but it depends on the goals and perspectives of the evaluator. (Kubayi & Larkin 2020.)

Tactical attributes of a player encompass the ability to understand the game and make tactical decisions. Tactical attributes consist of for example player's decision-making, positioning, and tactical understanding, and behaviour. These are in general more abstract variables compared to physical and technical attributes and their monitoring is not completely unambiguous. (Young et al. 2020.)

Finally, players mental attributes form a whole which often is as relevant, despite the most abstract and difficult to quantify. These attributes consist of variables such of players motivation, confidence, concentration, anxiety control and so on. These attributes highly effect on the players performance, but their interpretation is not unambiguous and always completely reliable. (Abdullah et al. 2016.)

All these player specific attributes contribute to the overall player performance, and can be assessed in several different ways, some of which have been presented in previous chapters. Table below summarizes the internal factors, player attributes, contributing to the player performance (Table 10). In addition, external attributes, such as team strategy, formation, structure, and dynamics can have an impact on players performance.

Table 10 Internal factors of player performance: Player attributes

	Internal factors			
Attribute:	Physical	Technical	Tactical	Mental
Description:	Physical performance	Technical performance	Game understanding	Mental capabilities
Example metrics:	Speed Endurance Strength	Dribbling Passing Shooting	Decision-making Positioning Tactical awareness	Motivation Confidence Concentration

6 Methodology

This chapter sets the empiric decisions made in the thesis presenting the chosen research approach and methodological choices. In addition, the chapter includes description of data collection and data analysis. Lastly, reliability and validity are measured using four different criteria: credibility, transferability, dependability, and confirmability (Lincoln & Guba 1985).

6.1 Research design

The empiric study of the thesis is carried out as a qualitative study. The research question is the following: How can data be used in evaluating the performance of football players? To be able to answer the research question, semi-structured theme interviews are conducted using the previous literature as a base.

Qualitative research approach is appropriate research method for the study as answering the research question requires deep understanding of a complex phenomenon. According to Eriksson & Kovalainen (2008), qualitative approach is especially suitable for this purpose. The objective of the semi-structured theme interviews is to conclude versatile and deep understanding of subject with no right or single answers. Subjectivity is a strong part of football-related opinions, because of this, interviews have been organized with people with different expertise. These people are professionals in different areas, consisting of coaches, physical trainers, players, managers, and scouts.

Furthermore, narrative analysis has been used to conduct answers to the research question. Narrative analysis is based on stories of experiences presented in the form of data (Merriam 2014). Narrative research is typically used so that the interviewees are asked to discuss the phenomena in a form of life stories (Flick 2009). In the interviews conducted for this study, the interviewees have been asked to express their feelings and thoughts of the themes and questions based on their previous experiences.

This method is specifically usable for the semi-structured theme interviews with a subjective area of discussion and interviewees with variable experience from different areas. The interview data gathered can be considered as narrative, as the interviewees are asked to answer through stories, feeling and experiences.

The interviews for this study are carried out as an individual face-to-face interviews. In a theme interview, the discussed topic is divided to different themes (Paavilainen-Mäntymäki 2017). These themes are Team performance, Player performance, and Data. All these themes are covered in full with all the interviewees. Often conducting theme interviews would not require drafting interview questions beforehand. However, for the objective of this study, theme interviews have been combined with the logic of semi-structured interviews so that under each theme, some of the questions have been pre-drafted beforehand the interviews.

See below an operationalization chart (Table 7) used as a base to for the empirical research of this thesis. The chart presents the purpose of the research together with the ideology how research question has been turned into interview questions. The support research questions have not been included in the chart, as they have been answered in the theoretical section of the thesis.

Table 11 Operationalization chart

Purpose: The purpose of this thesis is to create understanding and practices for taking advance of data for evaluation of player performance in football.		
Research Question	**Theme**	**Sample questions**
How can data be used in evaluating the performance of football players?	Team performance	What do you think are the most important elements and qualities of a good team and why? How can you evaluate these elements of a good team? How can you evaluate if a player fits to the team? (Team dynamics) How can you evaluate team performance?
	Player performance	What do you think are the most important elements and qualities of a good player and why? How can you evaluate these elements of a good player? What are the main attributes of a football player? How can you evaluate these attributes? What are the most relevant metrics? How can you measure these metrics?
	Data	How is data being used in football currently? Please, describe by using experiences. How is data being used to evaluate player performance? How do you see the use of data in football in the future?

6.2 Data collection

The data for this study is collected through semi-structured theme interviews. For getting the most delightful insights, the interviewees consist of different parties and expertise, such as coaches, physical coaches, players, experts and so on. The pre-drafted interview questions have been divided into different themes and formulated, as shown in the operationalization chart, in a way that there is room for interviewee's own opinions and discussion. This way subjective phenomena can be discussed to receive versatile data for being able to answer the research question and conclude reliable results.

The interviewees are selected on the base of their different roles and experience. The main theme related to all the sub-themes is player attributes. Therefore, the selected interviewees are mainly parties with experience in football coaching, player attributes, and team and game strategy. When choosing who to interview also interest and experience in utilizing technology and data was also considered.

Total of six persons were interviewed for the study. The duration of interviews were 60 minutes, and all the interviews were conducted within total of three weeks. All the interviews were recorded, based on which the data has been collected. See below a table (Table 8) listing all the interviewees and their roles.

Table 12 Interviewees and their roles

Interviewee	**Role**	**Interview duration (min)**
Interviewee 1	Physical coach	60
Interviewee 2	Sports analytics entrepreneur	60
Interviewee 3	Main coach	60
Interviewee 4	Player scout	60
Interviewee 5	Professional player	60
Interviewee 6	Sports analyst	60

6.3 Data analysis

Data analysis in research can be defined as the concept of utilizing research data to find the most relevant findings related to the research questions (Saaranen-Kauppinen & Puusniekka 2006). There are multiple ways of doing data analysis in qualitative research with a different focus and strengths depending on the research question and collected data.

In this study, the data gathered in the semi-structured theme interviews are analyzed with a method of deductive content analysis. Deductive content analysis is found to be efficient method for analyzing data in a form of text with a systematic and objective way. The deductiveness in the analysis means that the logic of reasoning used in the research is applied with a principle of enriching information from general to delimited. The analysis method includes three phases which are the reduction of data, categorization of data, and abstraction of the results. The abstraction of the results is mirrored against the theoretical background. (Tuomi & Sarajärvi 2018.)

Deductive content analysis is appropriate analysis method in the study as the complex phenomena requires a systematic way of finding the most relevant insights based on a collected data through experiences and stories. In addition, the three phases of the content analysis fit perfectly in the research as the interviewee experiences requires reduction of data, the themes used in the interviews allows for useful categorization of the data, and lastly the theoretical background is applicable and well suited to be mirrored against the abstraction of the results.

In this study, the reduction of data is carried out in a following manner: the collected data is first transcribed and reviewed with different perspectives to get a general view of the data. Once the data is reviewed and found to be sufficient, the reduction happens by identifying the most important findings related to the main research question and extracting this data from the total data set. As some of the interviews are in Finnish language, in this phase the data is also translated to English.

Next, the findings are categorized in different categories. This is made naturally, as the theme interviews allow easy categorization and is natural choice for theme-based interviews (Saaranen-Kauppinen & Puusniekka 2006). Hence, the categories in this study follow the themes of the interviews, and the findings are categorized to Team performance, Player performance, and Data.

Lastly, to conduct the last part of the deductive content analysis, an abstraction of the results is presented by mirroring the interview findings with the theoretical background. The structure of the findings and discussion chapter is formulated to follow the structure of the interviews comparing the findings with the theoretical part of the thesis. Finally, conclusions chapter summarize the abstraction of the results in a logical way by providing clear answers to the main research question and each of the support research questions.

6.4 Reliability

Referring to Lincoln & Guba (1985), the reliability of a qualitative study can be evaluated through four criteria: credibility, transferability, dependability, and confirmability. These criteria are used to evaluate the reliability of this thesis.

Credibility critically assess that how sufficiently the study fits the researched phenomena in terms of researchers familiarity with the topic and conclusions drawn from the data. For credibility, it is suitable to consider whether other research could use material of the study and come up with relatively close assumptions. (Eriksson & Kovalainen 2008). As the research area of this study is sports analytics, which is relatively new and subjective topic, has the credibility been taken into account by reading through a lot of previous studies made from the subject, and selecting the most common findings as the scientific base of the thesis. Additionally, the interviewees have been selected with diverse roles and experiences to enable as versatile views from the phenomena as possible. However, from these versatile views, clear trends in answers have been selected to the results of the thesis to emphasize credibility.

Transferability refers to the connection between previous research and this study. Transferability assesses whether there are similarities found in other research done from the phenomena. (Eriksson & Kovalainen 2008). As the research area of this study is sports analytics, and in more detail analytics in football which is relatively new research area - have all the findings been based on the interview results reflected with the findings from the previous research made from the subject. In terms of transferability, this study has its roots and deep connections within the previous research. The study has been built on existing research with an objective of enriching the research phenomena by combining findings from previous research with results of the interviews.

Dependability refers to that the research is done in logical and traceable manner. Dependability ensures that the reader is able to understand and trace the process of the research done. (Eriksson & Kovalainen 2008). In terms of dependability, the literature review of this thesis on which the interviews are based on, is built in a logical and traceable manner. All the previous literature used in this thesis has been referenced in a ethical way, and the order of presenting the literature is based on logic coming from the real world: technology, data, analytics, decision-making. Additionally, all the interview themes and questions have been formulated so that they could be asked in a similar way from any interviewee despite their role or experience.

Finally, Conformability refers to the degree of objectivity of the research. Objectivity should be able to be confirmed to ensure that the results have not been concluded biasedly. Thereafter, the findings should be linked to the data in a easily understandable way. (Eriksson & Kovalainen 2008). As this thesis has its roots in the previous literature, has objectivity been ensured by reflecting all the conclusions on combining the previous literature with the interview answers. Moreover, the interview questions have been formulated based on the previous research emphasizing objectivity.

7 Results and discussion

This chapter presents the results of the empirical research and discusses their connection to previous literature. The results are presented under the three themes: Team performance, Player performance, and Data.

7.1 Team performance

First theme of the interview was team performance. This was logical starting point for the interviews in terms of the research question, as evaluating team performance eventually leads us to the evaluation of performance of individual players. For some of the interviewers, the questions related to team were more challenging because of their different roles - for example, the answers of sports analyst and the team's head coach were very different.

The interviewees saw a football team as a larger whole, which includes the team's players, coaches, and support. However, as the study aims to evaluate player performance, when talking about a team in this study, it refers to its players. Common theme in the interview was that team performance ultimately stems from its players, and that the high-level purpose is to play the best possible combination of players on the field at the same time to achieve the highest possible team performance. This is in line with the previous findings of Duarte et al. (2013), who concluded in their research that team performance stems from individual players performance combined with the interaction and synergy with others.

“A good team consists of optimal formation with best possible players for each position complementing each other’s strengths and weaknesses.”

When discussing what elements generally makes a good team, several factors were highlighted. It was pointed out that a good team generally consists of high-performance players. Other factors repeated the most were team's structure, team dynamics, and strategy. The concept of a good team was approached in the interview by ignoring the limited resources, which in practice largely guides the decision-making, especially in Finnish football as was highlighted in the interviews.

For a team’s structure point of view, formation and playing position-related thinking was underlined. Following common observations were made in the interviews: a good team has a certain playing formation for a certain reason which is guided by the team's strategy. A good team is also able to reorganize itself according to situation and their opponent. In a good team every playing position includes multiple player options for different games and situations. Preferably so that if there is a weaker player somewhere, the surrounding players cover these shortcomings with their good qualities. This is in line with the findings of previous research, where means of performance analytics, such as center of gravity analysis had been recognized to support with strategy and tactics planning (Frencken et al. 2011; Rein & Memmert 2016).

Additionally, common theme in the interview was that formation was seen to have impact on the team's performance by affecting the individual players performance which is in line with previous research of McLean et al. (2018). Players might have personal strengths and qualities, which suits well a certain formation. This is relevant to notice, as formation is mainly decided by the team's coach which should understand the qualities, strengths, and weaknesses of its players. These qualities will be further discussed under the theme "Player performance".

As discussing team dynamics, in general the interview answers highlighted that a good team consisted of a suitable group of individuals who know how to work together and are fit for working together. This consists of many different personality elements, which in a good team are presented in a right ratio. Furthermore, it was highlighted that a good team's players clearly understand the roles and responsibilities of the team and has certain personal qualities for those roles. Required qualities vary depending on the role, but the sum of different qualities should generally include features such as leadership, trust, togetherness, teamwork, and a winning mentality. Other relevant qualities for a good team discussed were team cohesion, communication, tactical awareness, adaptability, discipline, and dedication.

However, quantifying these characteristics was perceived as challenging. Team dynamics were seen more as a characteristic that comes as a given. The dynamics can be influenced by creating a certain type of team culture but identifying and evaluating the team's dynamics requires an in-depth knowledge of the team and its players which requires time and experience. This means that, for example, when acquiring a new player, it can be challenging to a team to assess the player's impact and suitability for the team. This is in line with the previous research of Trequattrini et al. (2015) who highlighted the challenge of measuring mental attributes reliably.

"I think this is (team dynamics) much more difficult concept than you might imagine. At worst, you have to put together a team based on a few training sessions or video clips. It is generally too short time to be able to get to know the players in a deep way."

Team strategy, tactics and playing style was seen as a one of the most important factors for the team performance. It was highlighted that strategy generally guides the behaviour of the players: what they do on the field, and how they do it. Selected strategy of a good team was seen to be strongly connected to the characteristics of its players. Thereafter, it was pointed out that the strategy should support the strengths of the players so that they can be utilized as much as possible.

On the other hand, team strategy and tactics were seen as a changing variable. A good team's playing style was seen to be able to change depending on the situation and the opponent. This emphasises the team dynamics in terms of understanding the team roles and responsibilities which might be changing. Strategical adaptability to be able to adapt according to changing game situations were seen as a highly important factor contributing to a team performance.

"It is highly relevant that the players are able to adapt to different situations, strategies, and roles even the circumstances can change quite quickly. It requires strategical understanding and adaptability to be able to maintain one's level in varying conditions or situations."

Previous research had identified that formulation of team strategy could be supported by monitoring the performance data of team's players or opponent. In terms of teams own strategy, performance data could be used to identify teams' strengths and use this knowledge as a base for formulating playing strategy. Performance under playing with different strategies could also be compared when data has been collected over a longer period. For these objectives, location data, video data, and in overall descriptive information such as statistics were identified as suitable. (Aughey & Falloon 2010; D'Orazio & Leo 2010; Davenport 2014a.)

On the other hand, by identifying trends in opponents historical game data, it would be possible to gain knowledge of opponents playing strategy. Furthermore, it would be then possible to formulate own strategy to be the most fit for playing against this certain playing style. (Davenport 2014a.) However, collecting or obtaining data from opponents is still at a low level. In addition, as it becomes more common, teams would adopt it widely, which in part reduces the benefit it gives as everyone would use it.

Evaluating the team's performance, many things were highlighted, but the most important was the team's game performance in a long-term. Game’s performance in general consists of many already discussed factors, but in the end, it practically always culminates to team’s wins and losses. Wins and losses were unanimously seen as the most important and reflective measure of the team performance. A winning team consists of individual players who play well together. Next, we will discuss the second theme of the interviews, the player performance.

“At the end of the day, the best team is the winning team.”

7.2 Player performance

The discussion of player performance aimed to create an understanding of the most important attributes of players and their metrics. At first, the discussion took place on a general level, without taking a position on the current available data on the players' performance. This was aimed at providing an unbiased view of the players' actual most important attributes, which could best be used to assess the players' performance.

During the interview, it became clear that the interviewees were mixing up the attributes and metrics of the players such as physical attributes and speed. For the sake of clarity, the thesis uses a division where the player's attributes mean upper-level attributes such as physical and mental attributes. Primary metrics, on the other hand, are subcategories of these attributes, for example, speed is a primary metric of physical attribute. Finally, secondary metrics tell about the way metrics are observed, for example a secondary metric of speed could be the maximum speed or meters per second value.

In the interviews, the objective was approached first by letting the interviewees define a good player. The definitions included that a good player meets and exceeds the criteria of the team’s position requirements and eligible player profile in terms of the player qualities. It was strongly emphasized that the players' attributes are strongly dependent on the team's strategy, goals, and playing position.

When interviewees were portraying the attributes of a good player, the most relevant player attributes included player's physical, technical, tactical, and mental attributes, which is in line with the previous research (Yang et al. 2018; Ryoo et al. 2018). These categories were seen to cover all the elements necessary for a good football player. The categories were also seen to be strongly interconnected, however the prioritization varying depending on the interviewee's role.

7.2.1 Physical evaluation

For player's physical attributes a common theme in the answers was that these are the most tangible attributes of an athlete and easily identified, measured, and compared. Physical attributes are a lot about metrics of absolute values, and the interviews revealed that a relatively large amount of evaluation of players' physical performance was already done with the help of data.

According to the interviewees, the most relevant primary metrics of player's physical attributes were agility, strength, speed, endurance. The importance of these in relation to each other varied a lot in the answers. However, there was a common trend that in general agility was seen as more important than speed in football, depending on the playing position. In addition, interviewees with an experience in physical coaching emphasized strength as the most relevant attribute among all the attributes as it contributes to the other attributes as well.

Common theme was that evaluation of these attributes were in general done by different physical tests, data collection, and comparison of the players to desired player and position specific requirements and amongst the players on same position. In addition, the answers revealed that evaluation of physical attributes were already done partly with the help of data, such as location and heart rate data. Common technology for tracking these measures were either GPS technology or wearables. This reinforces the findings from the previous research where these technologies were generally considered to be among the most used and typical technologies in football (Rago et al. 2020; Aroganam et al. 2019).

Common theme in the interviews was that observations from the players physical attributes mainly guides the tracking of physical performance, planning of training, and management of weekly loads and intensity. This in turn was said to assist in identifying and preventing players injuries. These observations were in line with Kos & Kramberger's study (2017) of utilizing data in players health monitoring.

For agility, the secondary metrics of measurement were player's starting speed, acceleration, turning speed, reaction times and coordination. These were mainly measured by different agility and running courses and watching game performance through video material. However, the answers highlighted that obtaining reliable quantified absolute values for agility is more challenging than, for example, for speed and endurance.

"A lot depends on the playing position, but in general agility might be the most relevant attribute because it's rare to run more than a hundred meters at full speed. It is valuable if you can run fast a shorter distance from standing legs, recover from it quickly, and do it again."

Strength was seen as an important primary metric as it helps to win different situations in the game, contributes to other physical primary metrics, and for example assists in recovery. The answers had a common theme that strength in football contributes to both player's muscular power and endurance. The most relevant means of measurement was seen to be power output and muscular endurance.

"Muscular strength and endurance do not only help on the field but helps the player to recover outside of the field."

Speed was perceived as an important primary metric, very specific to the playing position. Secondary metrics of speed included monitoring of top, average, and starting speed. Other interesting answers included aspects of ability to maintain speed throughout the game which is highly correlated to other primary metrics such as endurance.

"It is not always the speed itself, but the knowledge when to use the speed, all players does not have this knowledge."

Endurance was seen as a base metric for the player to be able to execute all the other attributes, and so that the game performance and weekly loads are possible. Main means of measurement were simple running and stress tests executed with different paces and times. These tests were used to monitor metrics such as distance covered, heart rate, and oxygen uptake.

“Endurance is the engine – without endurance, the players are not able to perform the other attributes…”

7.2.2 Technical evaluation

Discussion on player’s technical attributes highlighted the importance of the category when evaluating player’s performance. As an average of the interview responses, the players' technical attributes were seen as the single most important category related to player performance evaluation. Technical attributes were also seen to be highly linked with the different requirements of different playing positions.

In general, the most repeated answers for technical attributes primary metrics highlighted in all the interviews were passing, ball control, shooting, finishing, and dribbling. General trend was that all these metrics were important depending on the team’s playing style and player’s position. The technical metrics were seen to enable the players to execute the teams’ strategies by performing specific technical tasks required for their position.

As a general criterion for evaluating player’s technical metrics were their ability and consistency in executing different tasks in game, in particular their technical ability to make and execute good decisions in different game situations, especially under pressure. Most repeated answers for secondary metrics included possession, opportunities created, scoring, passing and shooting technique and accuracy, and ability to play under pressure.

These metrics were said to be mainly monitored and evaluated by game statistics, watching video material, and by monitoring player’s performance live. However, statistics were mentioned to be not always available and video material was considered as time-consuming manual task. In addition, evaluation of players technical attributes was seen to be somewhat subjective matter, as other coaches preferred different kind of players and skills.

7.2.3 Tactical evaluation

The tactical attribute of players was seen as a highly important, difference-making quality. Players' tactical abilities determine the players' ability to understand the game and implement the team's strategy and tactics. Tactical attributes and tactical maturity were seen as an important player aspect required for evaluating a player's performance, yet a much more abstract category than physical and technical attributes.

Common theme in the interviews was that the most relevant tactical primary metrics included game reading skills, decision-making skills, positioning, and adaptability. General opinion was that players often either has quality on these tactical metrics or then not. In addition, quantifying these skills for measurement was perceived as challenging, a common theme of the responses was that a player's tactical qualities are often discerned from the game and compared amongst players with experience and feeling.

"A good player observes and sees the field and goes to the right place at the right time. In theory, these skills can also be learned and obtained, but at the professional level the small differences come precisely from whether these skills and understanding are found naturally or not."

7.2.4 Mental evaluation

When discussing players mental attributes there was a clear polarization among the interviewees. Some of the interviewees considered mental attributes as the least relevant category for evaluating player performance. However, some of the interviewees had completely different view as they considered the mental attributes to be the most significant factor distinguishing similar players from each other.

Regardless of the difference in opinion, mental attributes were seen as the most complex category to provide objective insights for evaluating player performance. Mental attributes were seen as even more abstract category than the tactical attributes. This is because while there are clear indicators available for physical and technical metrics, at best in the form of a number, equivalents are not as unambiguously available for mental metrics. This is in line with the previous research of Abdullah et al. (2016) who identified that mental metrics are often not as unambiguous and clear.

During the interviews, clearly the most relevant mental primary metrics were motivation, self-confidence, concentration, mental toughness, resilience, and commitment. Some of these were seen to be related to each other, and the combination of these was generally referred to as self-management. Self-management or "athleticism" involved commitment and motivation in an athlete's life: taking care of oneself and one's body, self-confidence and concentration on the field, and tolerating discomfort.

"Mental attributes make the small difference –mentally strong athlete is able to commit to the activity and is there on time, concentrates on the training, and takes care of one's body before and after the training"

As mentioned above, measuring, or evaluating the mental metrics were found difficult. Ways to measure mental metrics included using optimized sports psychological questionnaires. These are series of questions presented to the player facilitated by a professional with an objective of measuring players mental state, confidence, and motivation.

7.3 Data

The final theme of the interview was data. In particularly using data to evaluate player performance in football. The general objective was to find out real experiences of using data in football and by this reflect the answers to the findings of previous research.

Common theme in the interviews was that interviewees acknowledged that data could be measured from the players and that it would benefit a team vastly. However, the reality was that using data in football generally is still on a low level in Finland, mostly due to resources in football. Despite, the interviewees showed a general enthusiasm and interest towards the use of data and felt that in general there would be a need and demand for it in Finnish football.

Examples of using data to evaluate players attributes are concluded by combining previous research and the interview answers. The interview answers gave information for what are the relevant metrics to track, and previous research gave information of the possible technology and data source in sports. The data utilization use cases provided are simply examples, as the criterion of evaluating certain player attributes are subjective and team specific. Additionally, due to subjectivity, secondary metrics that came up in the interviews may be discussed, but they are left out of the final conclusions.

Currently, the experiences from using data in football came mostly from the physical data. Most of the interviewees had experience that physical data has been gathered through either GPS technology or various wearables technologies. This data had been used to evaluate players physical condition and to plan the content of the trainings. However, interviewees had a common view that monitoring of physical metrics could be used more widely to guide training from an individual and team perspective. Currently it seemed that data was merely used in silos, usually seen as an outside area for example for the main coach.

According to the interviewees, the most relevant primary metrics of player's physical attributes were agility, strength, speed, endurance. From these attributes, the monitoring and evaluating was mainly done with more traditional methods such as various tests, running courses, and watching video material. In terms of endurance and speed, GPS technology was used to evaluate players in terms of total distance ran during a game or practice, speed was monitored but not completely trusted on, and heart rate was monitored through wearables.

This is in line with the previous research (Aughey & Falloon 2010; Ehrmann et al. 2016.), location data can be used to derive different physical metrics such as endurance and speed with using players position and time. On a secondary metrics level, player's speed and distance covered could be presented as an absolute value in different formats, on real time, for example during a game. Additionally, player speed and endurance such as average intensity and top speed could be tracked off-game, and then compare player-specific game metrics to the known top value to give information about effort and different game situations. Also, heart rate and oxygen up-take can be monitored through wearables technology to give insights of most intense periods of a game or general knowledge of players endurance (Carling et al. 2008; Elliot et al. 2019).

For monitoring agility reliably, data accuracy would be needed. As identified in the previous research (Schulze et al. 2021), for example UWB would give more accurate data of player's positions on the field over the GPS technology. This data could be then used to calculate starting speed, turning speed, and reaction times using player locations related to time. As identified in the literature (Elliot et al. 2019), additionally power output can be monitored through wearable technologies. Information about muscles ability to produce power could be monitored with wearables and compared amongst players to get insights of players strength levels. Muscular endurance could also be monitored through the same technology or through other physical attributes such as endurance as they are strongly linked together. Furthermore, game events and statistics could be used to monitor physical metrics. As discussed in the previous research, video technology could be used to automatically track game events and performance such as sprints made, fast changes in direction, tackles and so on (Shah et al. 2015; Xing et al. 2010).

According to the interviewees, players technical attributes was considered to be one of the most influential categories of the overall player performance. The most important technical primary metrics highlighted were passing, ball control, shooting, finishing, and dribbling. However, monitoring of the technical metrics was still mainly carried out manually watching video material or live performance.

Reflecting to previous research, data could give objective insights into evaluating players technical attributes. Various data sources can be utilized including location data, video data, and statistics. Depending on team's objective, most relevant metrics could be turned into quantifiable values which could then be assessed and compared more easily together. (Carling 2019; Sacha et al. 2014; Davenport 2014b.)

There are multiple ways of using data to evaluate players technical metrics such as passing skill, ball control, finishing, and playing under pressure. For example, with descriptive information and statistics one could assess for example the number of successful passes (pass completion percentage), number of key passes for example passes which lead to a shot on goal or other criteria and passing accuracy for example percentage of passes which accurately reaches their intended destination. On the other hand, ball controlling could be assessed through number of times a player loses or obtains a possession of the ball, the percentage of successful dribbles, and average touches taken in a game. Finishing and shooting could be evaluated through shooting accuracy and technique - the percentage of shots on the goal, average goals in a game, and percentage of shots taken per goal, shooting power, accuracy, and placement. Shooting power could be optionally measured with location data, calculating the speed of the ball. Moreover, ability to play under pressure could be identified with secondary metrics such as successful passes under pressure and ability to keep and obtain the ball under pressure. In overall, there are many ways to use descriptive data to evaluate players technical attributes. The criterion used for certain metric depends on the objectives and needs of a team. (Davenport 2014a.)

According to the interviewees the most relevant tactical primary metrics included game reading skills, decision-making skills, positioning, and adaptability. However, quantifying these skills for measurement was perceived challenging. Currently, common theme in the interviews was that the monitoring and evaluating players tactical attributes mainly leaned on experience, game, and training performance, and watching video material. Additionally, players could be compared by using statistics, which however often lacked. Interviewees highlighted that it would be good to be able to get objective insights of players tactical skills with the help of data, however it was currently seen as difficult to implement.

In line with the previous research, evaluating players tactical attributes with data could be done taking advantage of descriptive data, statistics (Davenport 2014a). Chosen statistics would depend on the objective of the team. Game reading skills could be evaluated with data by monitoring players secondary metrics such as interceptions, clearances, and opportunities created for example. Metrics could include for example number of times a player intercepts opponents passes or attacks, number of times a player saves the team by clearing the ball, avoiding a danger, and number of key passes or created scoring opportunities. On the other hand, statistics could be used to evaluate players decision-making skills. Depending on the playing position, relevant metrics could be monitored such as passes, key passes, and pass accuracy for midfielders. Also, number of shots taken in opponents goal keepers zone compared to passes made in the same zone could indicate strikers decision-making skills on whether to shoot or pass.

Furthermore, interviewees felt that data was already somehow used to bring objectivity to the tactical side and formulation of different strategies. However, it was common that the interviewees could not explain more deeply how the data was used. Only one interviewee with more technical background highlighted that positional data gathered through GPS technology was used to evaluate players positions. In addition, video material was used to evaluate players positioning and teams' tactical performance, thus manually by watching previous games.

This is in line with the previous research, as Meng et al. (2020) discussed that positioning could be monitored through location data. Additionally, video technology could be used to provide tactical information (Wang & Parameswaran 2004). With video and location data, players positioning in different game situations and times could be compared to optimal positions. In addition, players positioning related to teammates, opponents and the ball would give insights of players positioning skills. However, this would require experience as common theme in the interviews was that a threshold for good positioning is a subjective and dependent on team's strategy and objective.

Finally, according to the interviewees the most relevant mental primary metrics were motivation, self-confidence, concentration, mental toughness, resilience, and commitment. However, measuring and evaluating the mental attributes were found difficult.

"There are some areas that are more challenging to evaluate with data, such as a player's mental state or their ability to work well within a team"

Reflecting to the previous research, one way of measuring these metrics could be found from the statistics, biometric and video data. Data could be utilized to recognize trends and patterns indicating mental attributes in player's game. For example, related to self-confidence through decision-making: if a winger always returned the ball back with first touch without once trying to challenge opposing defenders, this could be a sign of a lack of self-confidence. On the other hand, other quantifiable metrics could be tried to utilize in identifying mental qualities. For example, physical and technical data could point out if a player would always lower the intensity in a losing situation or for example start making mistakes in a tied situation. This would indicate on the levels of motivation and self-confidence.

In overall, combining the interview answers with the previous research strongly indicates that physical and technical metrics are currently more easy to evaluate with data. Tactical and mental metrics are seen as relevant as physical and technical metrics, thus more difficult to quantify. However, reflecting to the answers and previous research same data used for physical and technical metrics, such as statistics and video data could be suitable for tactical and mental attributes if the desired outcomes and metrics are clearly defined.

One insight from the interviews was that video technology is still mainly used manually. The introduction of video technology and its applications was seen as attractive and likely option in the future. Other insight was that automatic monitoring of the ball contributed to the many beneficial metrics to monitor for evaluating player performance. This affects to the choice of technology, as currently most of the tracking technologies are not able to track the ball in Finland. Finally, most of the presented analytical tools and means in the literature were descriptive, and related to the production and utilization of descriptive information, i.e. statistics. For the future, most common view was that it is expected that descriptive analytics will most likely have a greater impact in Finland. Furthermore, different predictive models were seen to become more common in the longer-term in the future.

Finally, most of the answers in the interview began with a sentence “it depends on…”. This highlights the subjectivity of football. Various factors such as coaches’ opinions, team’s needs and objectives, strategies, playing positions, league rules and so on have to be considered and thus makes it challenging to create an accurate framework suitable for every team. Moreover, subjectivity means that data will not replace traditional methods, but at best will help and bring an objective view.

“…it is important to note that data will most likely support in player performance evaluation and decision-making more widely in the near future, however subjective evaluations of players' performance by coaches and scouts will remain an essential part of the decision-making process…”

8 Conclusion

The purpose of this thesis was to create understanding and practices for taking advance of data for evaluation of player performance in football. The main research question was:

How can data be used in evaluating the performance of football players?

The main research question was approached by identifying and combining technological possibilities with sports knowledge. This was addressed through various support research questions which were examined by reviewing previous research. The support research questions were the following:

- Technology: How can data be collected in football?
- Data: What kind of data can be collected in football?
- Analytics: How can data be used in football?
- Decision-making: What factors affect the players' performance?

The first support research question was approached by reviewing previous research on most common technologies in football. Most used technologies in sports were GPS technology, video technology, radar and microwave based technologies such as UWB and Bluetooth, and wearable technologies. All these technologies were identified to have their own strengths and weaknesses in collecting data in football. When choosing a suitable technology, team should consider its objectives and needs and reflect these to choose the most suitable technology or a combination of them. Answering to the first support research question, there are many technologies already used in football to collect data.

The second support research question was answered by reviewing previous research on data types used in football. Previous research had identified that common data sources in football included football team's internal and external sources. Internal source consisted of data types such as location data gathered with tracking technologies, biometric data collected with wearables, and video data collected with video technology. These data types provide different kind of use cases and can be used to create useful cross-disciplinary knowledge. The choice of data types depends on the needs and objectives of a team.

The third support research question was approached by reviewing previous research on how data can be used in football. Previous findings suggested that data can be used through different analytical methods generally described as sports analytics. Common framework of sports analytics included three basic components which were information gathering (technology), data management (data), and data analysis (analytics). These components enable the decision-makers to do better decisions. To answer the third support research question, data can be generally used in football through different analytical areas. These areas are descriptive, predictive, and prescriptive analytics. Applications from all these areas were used in the sub-areas of sports analytics, which were, based on the previous research, performance analytics and health analytics. Performance analytics uses data to evaluate players physical, technical, and tactical attributes. Currently, this is mainly done with descriptive analytics using statistics on performance. Health analytics used various types of data, emphasizing the biometric data, to promote players health and to predict and prevent injuries. Descriptive analytics was used the most also in this area, however, in both areas predictive analytics was seen to be a trend of the future, and already being used by the pioneers.

Final support research question was answered by reviewing previous research on factors contributing to players performance. Identified external factors affecting the players performance in football were related to team performance. These external factors included teams' strategy such as formation, team's structure, and dynamics. Main focus was on the players internal factors contributing to player performance. Previous research had identified that player attributes consisted of physical, technical, tactical, and mental attributes. These together formed player's overall skills which were further studied in the empirical section of the thesis to answer the main research question of how data can be used in evaluating the performance of football players.

Finally, the main research question was addressed based on the answers of the support research questions and through the interviews. Interviews provided crucial sports knowledge to be able to answer the main research question. Additionally, at the latest the interviews proved how subjective sport football is. Interviewees provided important information by jointly recognizing the same most relevant player attributes on to which base the evaluation of player performance. For the most relevant primary metrics such as speed, passing, positioning, or motivation there was already scatter in the answers. Finally, the interviewees proved that discussing secondary metrics such as whether to use speed in meters per second or top-speed or average speed, or which is the best way to monitor ball controlling, is not relevant for the objective of this study, as every coach has their own opinions. Due to these reasons, answering the main research question has been scoped to the players attributes and primary metrics level.

Data can be used in evaluating the performance of football players by using data to evaluate players physical, technical, tactical, and mental attributes. These attributes have primary metrics which vary depending on the needs and objectives of a team. Following attributes and their primary metrics were identified based on the previous research and interviews (Table 13).

Table 13 Player attributes and primary metrics

	Physical attributes	**Technical attributes**	**Tactical attributes**	**Mental attributes**
Primary metrics	Agility	Passing	Game reading	Motivation
	Strength	Ball control	Decision-making	Confidence
	Speed	Shooting	Positioning	Concentration
	Endurance	Dribbling	Adaptability	Mental toughness

Furthermore, the selection of desired attributes and primary metrics guides the user to consider which data is needed to evaluate them, which then then can be linked to certain technology based on the findings from the theoretical section of the thesis. These technologies ultimately provide the required data which can be processed with analytical tools. This process is presented on a high-level on the figure below (Figure 5). In the figure, player attributes could be further divided into primary metrics and secondary metrics.

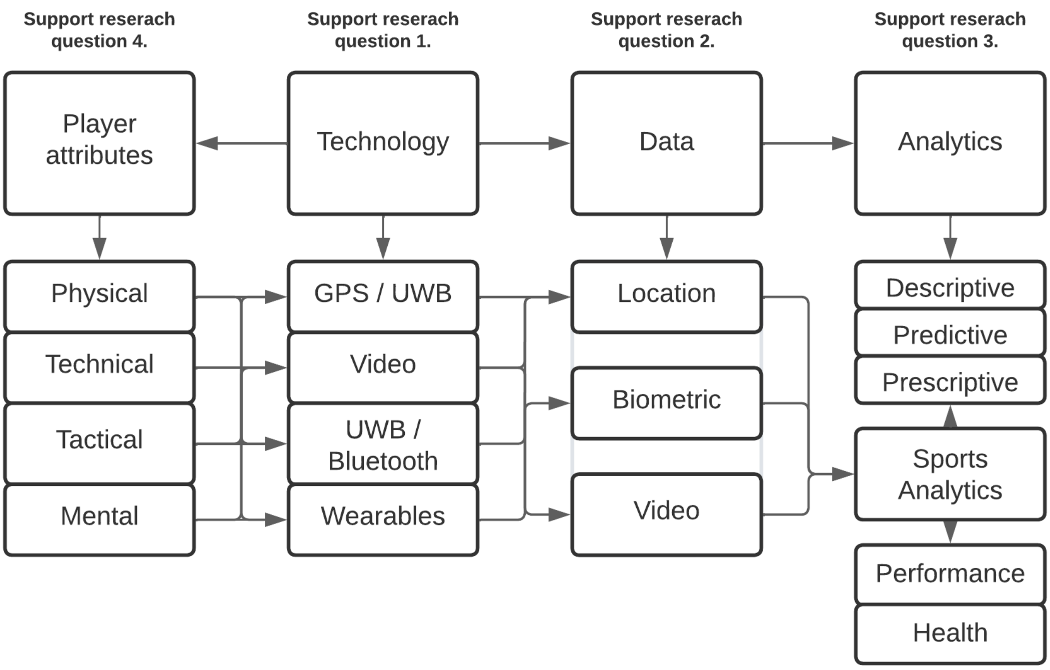

Figure 5 Data-driven evaluation of player performance

Theoretically, the level of quality of each attribute's primary metrics, all together would form a metric for the overall quality of a given attribute (Figure 6). Metrics could be given certain weighting factors according to their relative importance. However, this is difficult to generalize due to subjectivity.

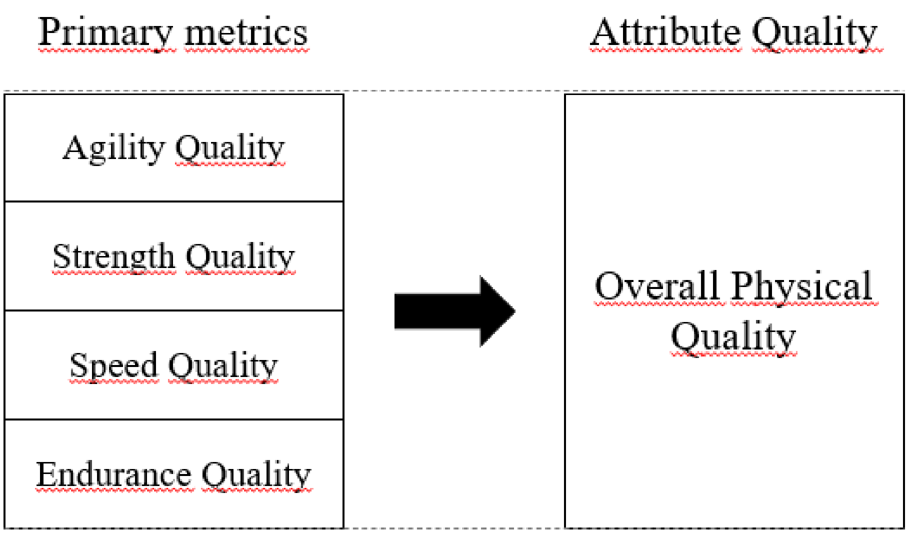

Figure 6 Illustration of overall attribute quality determination

Moreover, combining the overall qualities of all the attributes together would form a metric for the overall performance of a player (Figure 6).

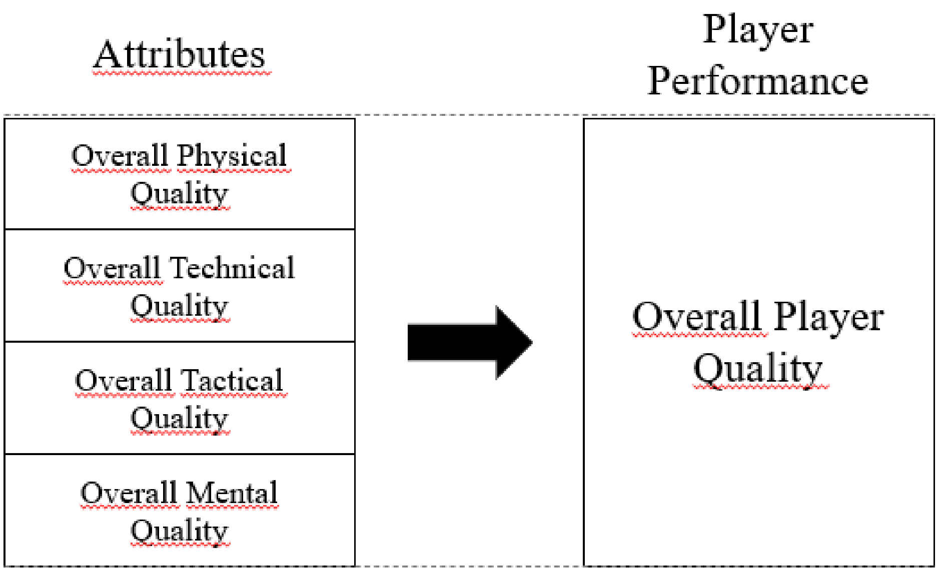

Figure 7 Illustration of overall player quality

However, to be able to reliably assess the primary metrics, teams should select their secondary metrics based on their needs and objectives. This would also guide in more detail the choosing of the most optimal data types and analytical tools. Finally, to reasonably compare a player's overall performance metrics between different players and teams, the selected attributes, primary metrics, and secondary metrics should all be the same. As mentioned, level of secondary metrics is beyond the scope of this research, but there would be a great opportunity for future research.

The theoretical contribution of this thesis is that it combines scattered previous research together and fills a research gap by suggesting a qualitative approach on evaluating players performance. By doing so, the thesis serves as a basis for future research. Future research should continue to study the sports analytics applications applicable for the performance evaluation. Thereafter, suggested primary and secondary metrics provide an opportunity to be used for further research.

The practical implications of this thesis are that it provides sports teams, coaches, and players a manual to sports analytics and data-driven performance evaluation. It can support in evaluating different technologies, data types, and analytical solutions.

Moreover, it suggests a high-level approach for data-driven player performance evaluation which can be implied into different environments. This approach can be complemented with team's own prioritized attributes and metrics.

The limitations of the research are that it examines a very subjective topic for which there is no right or wrong answer. This means that empiric has been implemented by combining the most prominent issues with previous theory. In addition, the research is limited so that it does not take a position on data quality, data ownership or information security. However, these are particularly important to take into account, especially when working with personal data. Moreover, Finally, the study ignores the actual finite resources of the real world. However, the approaches presented in the study are generalizations that can be applied according to the team's resources.

References

Abdullah, M. R., Musa, R. M., Maliki, A. B. H. M. B., Kosni, N. A., & Suppiah, P. K. (2016). Role of psychological factors on the performance of elite soccer players. Journal of Physical Education and Sport, 16(1), 170.

AC Milan. (2018) FIFA.COM INTERVIEWS OUR MILAN LAB DIRECTOR. <https://www.acmilan.com/en/news/interview/2018-02-03/fifacom-interviews-our-milan-lab-director>, retrieved 5.11.2022.

Adesida, Y., Papi, E., & McGregor, A. H. (2019). Exploring the role of wearable technology in sport kinematics and kinetics: A systematic review. *Sensors*, *19*(7), 1597.

Arnold, J. F., & Sade, R. M. (2017). Wearable technologies in collegiate sports: the ethics of collecting biometric data from student-athletes. *The American Journal of Bioethics*, *17*(1), 67-70.

Aroganam, G., Manivannan, N., & Harrison, D. (2019). Review on wearable technology sensors used in consumer sport applications. Sensors, 19(9), 1983.

Aughey, R. J. – Falloon, C. (2010) Real-time versus post-game GPS data in team sports. Journal of Science and Medicine in Sport, 13(3), 348-349.

Aughey, R. J. (2011) Applications of GPS technologies to field sports. International journal of sports physiology and performance, 6(3), 295-310.

Awan, U., Bhatti, S. H., Shamim, S., Khan, Z., Akhtar, P., & Balta, M. E. (2022). The role of big data analytics in manufacturing agility and performance: moderation–mediation analysis of organizational creativity and of the involvement of customers as data analysts. British Journal of Management, 33(3), 1200-1220.

Bajaj, R., Ranaweera, S. L., & Agrawal, D. P. (2002). GPS: location-tracking technology. Computer, 35(4), 92-94.

Balafoutas, L., Chowdhury, S. M., & Plessner, H. (2019). Applications of sports data to study decision making. Journal of Economic Psychology, 75, 102153.

Bates, M. (2020). The rise of biometrics in sports. IEEE pulse, 11(3), 25-28.

Bojanova, I. (2014). It enhances football at world cup 2014. IT professional, 16(4), 12-17.

Botwicz, M., Klembowski, W., Kulpa, K., Samczyński, P., Misiurewicz, J., & Gromek, D. (2017, September). The concept of an RF system for the detection and tracking of the ball and players in ball sports: An extended abstract for poster presentation. In 2017 Signal Processing Symposium (SPSympo) (pp. 1-3). IEEE.

Bush, M. – Barnes, C. – Archer, D. T. – Hogg, B. – Bradley, P. S. (2015) Evolution of match performance parameters for various playing positions in the English Premier League. Human movement science, 39, 1-11.

Carling, C. – Bloomfield, J. – Nelsen, L. – Reilly, T. (2008) The role of motion analysis in elite soccer. Sports medicine, 38(10), 839-862.

Carling, C. (2019). Performance analysis: Working in football. In Routledge Handbook of Elite Sport Performance (pp. 99-113). Routledge.

Castellano, J. – Alvarez-Pastor, D. – Bradley, P. S. (2014) Evaluation of research using computerised tracking systems (Amisco® and Prozone®) to analyse physical performance in elite soccer: A systematic review. Sports medicine, 44(5), 701-712.

Caya, O. – Bourdon, A. (2016) A framework of value creation from business intelligence and analytics in competitive sports. In 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 1061-1071). IEEE.

Chambers, R., Gabbett, T. J., Cole, M. H., & Beard, A. (2015). The use of wearable microsensors to quantify sport-specific movements. Sports medicine, 45, 1065-1081.

Chen, C. P. – Zhang, C. Y. (2014) Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information sciences, 275, 314-347.

Chen, M. – Mao, S. – Liu, Y. (2014) Big data: A survey. Mobile networks and applications, 19(2), 171-209.

Cintia, P., Rinzivillo, S., & Pappalardo, L. (2015, September). A network-based approach to evaluate the performance of football teams. In Machine learning and data mining for sports analytics workshop, Porto, Portugal.

Cuevas, C., Quilón, D., & García, N. (2020). Techniques and applications for soccer video analysis: A survey. Multimedia Tools and Applications, 79(39-40), 29685-29721.

D’Orazio, T. – Leo, M. (2010) A review of vision-based systems for soccer video analysis. Pattern recognition, 43(8), 2911-2926.

Davenport, T. H. – Harris, J. G. (2007) Competing on analytics: the new science of winning. Harvard Business School Press, Boston.

Davenport, T. H. (2014a) Analytics in sports: The new science of winning. International Institute for Analytics, 2, 1-28.

Davenport, T. H. (2014b) What businesses can learn from sports analytics. MIT Sloan Management Review, 55(4), 10.

Davis, N. (2013) Scoring With Soccer: Analytics Come to MLS. Grantland. <https://grantland.com/the-triangle/scoring-with-soccer-analytics-come-to-mls/>, retrieved 6.11.2022.

de la Torre, R., Calvet, L. O., Lopez-Lopez, D., Juan, A. A., & Hatami, S. (2022). Business Analytics in Sport Talent Acquisition: Methods, Experiences, and Open Research Opportunities. International Journal of Business Analytics (IJBAN), 9(1), 1-20.

Di Salvo, V. – Modonutti, M. (2009) Integration of different technology systems for the development of football training. J Sports Sci Med, 11, 3.

Dixon, M. J. – Coles, S. G. (1997) Modelling association football scores and inefficiencies in the football betting market. Journal of the Royal Statistical Society: Series C (Applied Statistics), 46(2), 265-280.

Dobre, C. – Xhafa, F. (2014) Intelligent services for big data science. Future Generation Computer Systems, 37, 267-281.

Duarte, R., Araújo, D., Correia, V., Davids, K., Marques, P., & Richardson, M. J. (2013). Competing together: Assessing the dynamics of team–team and player–team synchrony in professional association football. Human movement science, 32(4), 555-566.

Ehrmann, F. E. – Duncan, C. S. – Sindhusake, D. – Franzsen, W. N. – Greene, D. A. (2016) GPS and injury prevention in professional soccer. The Journal of Strength & Conditioning Research, 30(2), 360-367.

Elliot, C. A., Hamlin, M. J., & Lizamore, C. A. (2019). Validity and reliability of the hexoskin wearable biometric vest during maximal aerobic power testing in elite cyclists. The Journal of Strength & Conditioning Research, 33(5), 1437-1444.

Eriksson, P. – Kovalainen, A. (2008) Qualitative Methods in Business Research. SAGE

Evans, J. R. – Lindner, C. H. (2012) Business analytics: the next frontier for decision sciences. Decision Line, 43(2), 4-6.

FIFA VAR 2018. Fédération Internationale de Football Association. (2018). Laws of the Game.

https://digitalhub.fifa.com/m/50518593a0941079/original/khhloe2xoigyna8juxw3-pdf.pdf, retrieved 14.11.2022.

Figueira, B., Gonçalves, B., Folgado, H., Masiulis, N., Calleja-González, J., & Sampaio, J. (2018). Accuracy of a basketball indoor tracking system based on standard bluetooth low energy channels (NBN23®). Sensors, 18(6), 1940.

Flick, U. (2009). Qualitative Methoden in der Evaluationsforschung. Zeitschrift für qualitative Forschung, 10(1), 9-18.

Frencken, W. – Lemmink, K. – Delleman, N. – Visscher, C. (2011) Oscillations of centroid position and surface area of soccer teams in small-sided games. European Journal of Sport Science, 11(4), 215-223.

Frizzo-Barker, J. – Chow-White, P. A. – Mozafari, M. – Ha, D. (2016) An empirical study of the rise of big data in business scholarship. International Journal of Information Management, 36(3), 403-413.

Gandomi, A. – Haider, M. (2015) Beyond the hype: Big data concepts, methods, and analytics. International journal of information management, 35(2), 137-144.

Ghasemaghaei, M. – Calic, G. (2019) Does big data enhance firm innovation competency? The mediating role of data-driven insights. Journal of Business Research, 104, 69-84.

Ghasemaghaei, M. – Hassanein, K. (2016) A macro model of online information quality perceptions: A review and synthesis of the literature. Computers in Human Behavior, 55, 972-991.

Ghoniem, A. – Ali, A. I. – Al-Salem, M. – Khallouli, W. (2017) Prescriptive analytics for FIFA World Cup lodging capacity planning. Journal of the Operational Research Society, 68(10), 1183-1194.

Godfrey, A., Hetherington, V., Shum, H., Bonato, P., Lovell, N. H., & Stuart, S. (2018). From A to Z: Wearable technology explained. *Maturitas*, *113*, 40-47.

Goes, F. R., Meerhoff, L. A., Bueno, M. J. O., Rodrigues, D. M., Moura, F. A., Brink, M. S., ... & Lemmink, K. A. P. M. (2021). Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. European Journal of Sport Science, 21(4), 481-496.

Grand View Research. (2021) Sports Analytics Market Size Worth $6.34 Billion By 2030: Sports Analytics Market Growth & Trends. < https://www.grandviewresearch.com/press-release/global-sports-analytics-market>, retrieved 14.6.2022.

Haghighat, M. – Rastegari, H. – Nourafza, N. (2013) A review of data mining techniques for result prediction in sports. Advances in Computer Science: an International Journal, 2(5), 7-12.

Hagström, M. I. K. A. E. L., & GILL, N. (2012). The wisdom of the cloud: hyperconnectivity, big data, and real-time analytics. The global information technology report, 97-103.

Hanisch, B. – Hanisch, J. (2007) Injury management: The development and implementation of innovative software in an elite sporting club. PACIS 2007 Proceedings, 8.

Hatze, H. (1983) Computerized optimization of sports motions: An overview of possibilities, methods and recent developments. Journal of Sports Sciences, 1(1), 3-12.

Helsinki.

Hui, Q. (2019). Motion video tracking technology in sports training based on Mean-Shift algorithm. The Journal of Supercomputing, 75, 6021-6037.

Johnson, J. E. (2012) Big data+ big analytics= big opportunity: big data is dominating the strategy discussion for many financial executives. As these market dynamics continue to evolve, expectations will continue to shift about what should be disclosed, when and to whom. Financial Executive, 28(6), 50-54.

Jossey–Bass, Hoboken. Park, CA.

Kelly, A. L., & Williams, C. A. (2020). Physical characteristics and the talent identification and development processes in male youth soccer: A narrative review. Strength & Conditioning Journal, 42(6), 15-34.

KMD. (2018) Vinder på viden: Sådan brugte FC Midtjylland data på vej mod DM-guldet.
< https://www.kmd.dk/indsigter/vinder-paa-viden-fc-midtjylland-data-paa-vej-mod-dm-guldet>, retrieved 14.10.2022.

Kos, M. – Kramberger, I. (2017) A wearable device and system for movement and biometric data acquisition for sports applications. IEEE Access, 5, 6411-6420.

Kubayi, A., & Larkin, P. (2020). Technical performance of soccer teams according to match outcome at the 2019 FIFA Women's World Cup. International Journal of Performance Analysis in Sport, 20(5), 908-916.

Kwon, O. – Lee, N. – Shin, B. (2014) Data quality management, data usage experience and acquisition intention of big data analytics. International journal of information management, 34(3), 387-394.

Lam, S. K. – Sleep, S. – Hennig-Thurau, T. – Sridhar, S. – Saboo, A. R. (2017) Leveraging frontline employees' small data and firm-level big data in frontline management: An absorptive capacity perspective. Journal of Service Research, 20(1), 12-28.

Li, H., Manickam, A., & Samuel, R. D. J. (2022). Automatic detection technology for sports players based on image recognition technology: the significance of big data technology in China's sports field. Annals of Operations Research, 1-18.

Lichtenthaler, U. (2022). Mixing data analytics with intuition: Liverpool Football Club scores with integrated intelligence. Journal of Business Strategy, 43(1), 10-16.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. sage.

Linke, D., Link, D., & Lames, M. (2020). Football-specific validity of TRACAB's optical video tracking systems. PloS one, 15(3), e0230179.

Liti, C. – Piccialli, V. – Sciandrone, M. (2017) Predicting soccer match outcome using machine learning algorithms. In Proceedings: MathSport International 2017 Conference (p. 229).

Malone, J. J., Barrett, S., Barnes, C., Twist, C., & Drust, B. (2020). To infinity and beyond: the use of GPS devices within the football codes. Science and medicine in football, 4(1), 82-84.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Martins, F., Przednowek, K., França, C., Lopes, H., de Maio Nascimento, M., Sarmento, H., ... & 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.

McLean, S., Salmon, P. M., Gorman, A. D., Wickham, J., Berber, E., & Solomon, C. (2018). The effect of playing formation on the passing network characteristics of a professional football team. Human Movement Special Issues, 2018(5), 14-22.

Memmert, D., & Raabe, D. (2018). Data analytics in football: Positional data collection, modelling and analysis. Routledge.

Meng, X., Deng, X., Zhu, S., Zhang, X., & Zeng, B. (2020). A robust quality enhancement method based on joint spatial-temporal priors for video coding. IEEE Transactions on Circuits and Systems for Video Technology, 31(6), 2401-2414.

Merriam, S. B. (2014) Qualitative Research: A Guide to Design and Implementation.

Millington, B., & Millington, R. (2015). 'The datafication of everything': Toward a sociology of sport and big data. Sociology of Sport Journal, 32(2), 140-160.

MIT Sloan Sports Analytics Conference. (2019) Past conferences. <http://www.sloansportsconference.com/archive/>, retrieved 22.9.2022.

Mondello, M. – Kamke, C. (2014) The introduction and application of sports analytics in professional sport organizations. Journal of Applied Sport Management, 6(2).

Morgulev, E. – Azar, O. H. – Lidor, R. (2018) Sports analytics and the big-data era. International Journal of Data Science and Analytics, 5(4), 213-222.

Mortenson, M. J. – Doherty, N. F. – Robinson, S. (2015) Operational research from Taylorism to Terabytes: A research agenda for the analytics age. European Journal of Operational Research, 241(3), 583-595.

Müller, O. – Junglas, I. – Brocke, J. V. – Debortoli, S. (2016) Utilizing big data analytics for information systems research: challenges, promises and guidelines. European Journal of Information Systems, 25(4), 289-302.

Muniz, M., & Flamand, T. (2022). Sports analytics for balanced team-building decisions. Journal of the Operational Research Society, 1-18.

Najafabadi, M. M. – Villanustre, F. – Khoshgoftaar, T. M. – Seliya, N. – Wald, R. – Muharemagic, E. (2015) Deep learning applications and challenges in big data analytics. Journal of Big Data, 2(1), 1.

Optasports. (2023) The Opta difference. <https://www.optasports.com/about/the-opta-difference/>, retrieved 5.4.2022.

Paavilainen-Mäntymäki, E. (2017) Tutkimusprosessit ja Kvalitatiiviset Tutkimusmenetelmät, harjoitukset. YSM lectures, spring 2021. Turku School of Economics.

Patel, D., Shah, D., & Shah, M. (2020). The intertwine of brain and body: a quantitative analysis on how big data influences the system of sports. Annals of Data Science, 7, 1-16.

Pino-Ortega, J., Gantois, P., Méndez, A., & Rico-González, M. (2021). The influence of the setup shape of a portable UWB system's antennas in sport. Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology, 17543371211041885.

Pons, E., García-Calvo, T., Resta, R., Blanco, H., López del Campo, R., Díaz García, J., & Pulido, J. J. (2019). A comparison of a GPS device and a multi-camera video technology during official soccer matches: Agreement between systems. PloS one, 14(8), e0220729.

PR Newswire. (2015) Global Sports Analytics Market 2015: Market Shares, Strategies, and Forecasts to 2021 for the $4.7 Billion Market. <https://www.prnewswire.com/news-releases/global-sports-analytics-market-2015-market-shares-strategies-and-forecasts-to-2021-for-the-47-billion-market-300090911.html>, retrieved 2.6.2022.

Rago, V., Brito, J., Figueiredo, P., Costa, J., Barreira, D., Krustrup, P., & Rebelo, A. (2020). Methods to collect and interpret external training load using microtechnology incorporating GPS in professional football: a systematic review. Research in Sports Medicine, 28(3), 437-458.

Ratten, V. (2019). Sports technology and innovation. Cham: Springer Books.

Rein, R. – Memmert, D. (2016) Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. SpringerPlus, 5(1), 1-13.

Ruan, L., Ge, H., Gómez, M. Á., Shen, Y., Gong, B., & Cui, Y. (2022). Analysis of defensive playing styles in the professional Chinese Football Super League. Science and Medicine in Football, 1-9.

Ryoo, M., Kim, N., & Park, K. (2018). Visual analysis of soccer players and a team. Multimedia Tools and Applications, 77, 15603-15623.

Saaranen-Kauppinen, A. & Puusniekka, A. (2006) KvaliMOTV – Menetelmäopetuksen Tietovaranto.

Sacha, D. – Stein, M. – Schreck, T. – Keim, D. A. – Deussen, O. (2014) Feature-driven visual analytics of soccer data. In 2014 IEEE conference on visual analytics science and technology (VAST) (pp. 13-22). IEEE.

Sarajärvi, A. & Tuomi, J. (2009) Laadullinen Tutkimus ja Sisällönanalyysi. Tammi, Publications, Lontoo.

Schulze, E., Julian, R., & Skorski, S. (2021). The accuracy of a low-cost GPS system during football-specific movements. Journal of sports science & medicine, 20(1), 126.

Schumaker, R. P. – Solieman, O. K. – Chen, H. (2010) Sports knowledge management and data mining. ARIST, 44(1), 115-157.

Shah, F. A. – Kretzer, M. – Mädche, A. (2015) Designing an analytics platform for professional sports teams.

Sharma, R. – Mithas, S. – Kankanhalli, A. (2014) Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations.

Sikka, R. S., Baer, M., Raja, A., Stuart, M., & Tompkins, M. (2019). Analytics in sports medicine: implications and responsibilities that accompany the era of big data. jbjs, 101(3), 276-283.

Sjöberg, L. (2003). Intuitive vs. analytical decision making: which is preferred?. Scandinavian Journal of Management, 19(1), 17-29.

Spitz, J., Wagemans, J., Memmert, D., Williams, A. M., & Helsen, W. F. (2021). Video assistant referees (VAR): The impact of technology on decision making in association football referees. Journal of Sports Sciences, 39(2), 147-153.

Statista 2022. (2022) Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. < https://www.statista.com/statistics/871513/worldwide-data-created/>, retrieved 17.9.2022.

Steijlen, A., Burgers, B., Wilmes, E., Bastemeijer, J., Bastiaansen, B., French, P., ... & Jansen, K. (2021). Smart sensor tights: Movement tracking of the lower limbs in football. Wearable Technologies, 2, e17.

Steinberg, L. (2015) CHANGING THE GAME: The Rise of Sports Analytics. Forbes. <https://www.forbes.com/sites/leighsteinberg/2015/08/18/changing-the-game-the-rise-of-sports-analytics/#1f89b9144c1f>, retrieved 18.9.2022.

Tan, F. – Hedman, J. – Xiao, X. (2017) Beyond 'Moneyball'to Analytics Leadership in Sports: An Ecological Analysis of Fc Bayern Munich's Digital Transformation.

Tankard, C. (2012) Big data security. Network security, 2012(7), 5-8.

Thudumu, S. (2021). Anomaly Detection in High-dimensional Big Data (Doctoral dissertation, Swinburne University of Technology Australia).

Torres-Ronda, N., Bartlett, J. D., O'Connor, F., Pitchford, L., & Robertson, S. J. (2017). Relationships between internal and external training load in team-sport athletes: evidence for an individualized approach. International journal of sports physiology and performance, 12(2), 230-234.

Travassos, B. – Davids, K. – Araújo, D. – Esteves, T. P. (2013) Performance analysis in team sports: Advances from an Ecological Dynamics approach. International Journal of Performance Analysis in Sport, 13(1), 83-95.

Trequattrini, R., Lombardi, R., & Battista, M. (2015). Network analysis and football team performance: a first application. Team Performance Management.

Twellaar, M. – Verstappen, F. T. – Huson, A. (1996) Is prevention of sports injuries a realistic goal? A four-year prospective investigation of sports injuries among physical education students. The American Journal of Sports Medicine, 24(4), 528-534.

Wang, J. R. – Parameswaran, N. (2004) Survey of sports video analysis: research issues and applications. In Proceedings of the Pan-Sydney area workshop on Visual information processing (pp. 87-90). Australian Computer Society, Inc.

Waqar, A., Ahmad, I., Habibi, D., & Phung, Q. V. (2021). Analysis of GPS and UWB positioning system for athlete tracking. Measurement: Sensors, 14, 100036.

Ward, J. S. – Barker, A. (2013) Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821.

Wilder, C. R. – Ozgur, C. O. (2015) Business analytics curriculum for undergraduate majors. INFORMS Transactions on Education, 15(2), 180-187.

Wilkerson, G. B. – Gupta, A. – Colston, M. A. (2018) Mitigating sports injury risks using internet of things and analytics approaches. Risk analysis, 38(7), 1348-1360.

Williams, A. M., & Ford, P. R. (2013). ‘Game intelligence’: anticipation and decision making. In Science and soccer (pp. 117-133). Routledge.

Williams, A. M., & Reilly, T. (2000). Talent identification and development in soccer. Journal of sports sciences, 18(9), 657-667.

Windt, J., MacDonald, K., Taylor, D., Zumbo, B. D., Sporer, B. C., & Martin, D. T. (2020). “To tech or not to tech?” A critical decision-making framework for implementing technology in sport. Journal of Athletic Training, 55(9), 902-910.

Wright, M. B. (2009) 50 years of OR in sport. Journal of the Operational Research Society, 60(sup1), S161-S168.

Wu, X. – Zhu, X. – Wu, G. Q. – Ding, W. (2013) Data mining with big data. IEEE transactions on knowledge and data engineering, 26(1), 97-107.

Wu, Y., Wang, W., Xu, H., & Kang, H. (2021, October). Research on Indoor Sports Positioning Algorithm Based on UWB. In 2021 International Conference on Control, Automation and Information Sciences (ICCAIS) (pp. 638-643). IEEE.

Xing, J., Ai, H., Liu, L., & Lao, S. (2010). Multiple player tracking in sports video: A dual-mode two-way bayesian inference approach with progressive observation modeling. IEEE Transactions on Image Processing, 20(6), 1652-1667.

Yang, G., Leicht, A. S., Lago, C., & Gómez, M. Á. (2018). Key team physical and technical performance indicators indicative of team quality in the soccer Chinese super league. Research in Sports Medicine, 26(2), 158-167.

Yaqoob, I. – Hashem, I. A. T. – Gani, A. – Mokhtar, S. – Ahmed, E. – Anuar, N. B. – Vasilakos, A. V. (2016) Big data: From beginning to future. International Journal of Information Management, 36(6), 1231-1247.

Yi, T. H., Li, H. N., & Gu, M. (2013). Recent research and applications of GPS-based monitoring technology for high-rise structures. Structural control and health monitoring, 20(5), 649-670.

Ying, S., Gang, W., & Yaojun, W. (2011, August). The Application of Information Technologyin Sports Training. In 2011 International Conference on Future Computer Science and Education (pp. 210-212). IEEE.

Young, C. M., Luo, W., Gastin, P. B., & Dwyer, D. B. (2020). Understanding the relative contribution of technical and tactical performance to match outcome in Australian Football. Journal of Sports Sciences, 38(6), 676-681.

Yu, X., & Farin, D. (2005, July). Current and emerging topics in sports video processing. In 2005 IEEE International Conference on Multimedia and Expo (pp. 526-529). IEEE.

Zafari, F., Gkelias, A., & Leung, K. K. (2019). A survey of indoor localization systems and technologies. IEEE Communications Surveys & Tutorials, 21(3), 2568-2599.

Appendices

Appendix 1 – Interview questions

Theme 1: Team performance

1. What do you think are the most important elements and qualities of a good team and why?
2. How can you evaluate these elements of a good team?
3. How can you evaluate team performance?
4. How can you evaluate if certain player fits to the team?
5. How do you try to evaluate or recognize the adaptation of different players when building a team?

Theme 2: Player performance

6. What do you consider to be the most important elements and qualities of a good player and why?
7. In your opinion, what is the order of importance of these qualities?
8. How can these qualities be identified / measured / compared among different players?
9. What qualities have the greatest influence on the acquisition of new players or, for example, the selection of players for the team or the starting line-up or for the overall performance evaluation of a player?
10. How to identify a potential player?

For each identified quality:

11. What criteria do you consider the most important when evaluating a player's X qualities and Y attributes?
12. How can data help determine X qualities of a player required for a certain position?
13. Can you describe a situation where data played a key role in developing, identifying, and comparing players' X qualities and Y attributes?

Theme 3: Data

14. How do you think data generally affects and helps football at the moment?
15. Can you describe your experiences, how you have utilized data in football?
16. In your experience, what kind of data is most useful for a football team?
17. How do you see the use of data affecting football in general in the future?
18. How do you see the use of data affecting players performance evaluation in the future?

Appendix 2 – Data management plan

1. Research data

Research data refers to all the material with which the analysis and results of the research can be verified and reproduced. It may be, for example, various measurement results, data from surveys or interviews, recordings or videos, notes, software, source codes, biological samples, text samples, or collection data.

In the table below, list all the research data you use in your research. Note that the data may consist of several different types of data, so please remember to list all the different data types. List both digital and physical research data.

Research data type	Contains personal details/information*	I will gather/produce the data myself	Someone else has gathered/produced the data	Other notes
Example, Data type 1: *Interviews*	x	x		
Example, Data type 2: *Recordings*	x	x		

* Personal details/information are all information based on which a person can be identified directly or indirectly, for example by connecting a specific piece of data to another, which makes identification possible. For more information about what data is considered personal go to the Office of the Finnish Data Protection Ombudsman's website

2. Processing personal data in research

If your data contains personal details/information, you are obliged to comply with the EU's General Data Protection Regulation (GDPR) and the Finnish Data Protection Act. For data that contains personal details, you must prepare a Data Protection Notice for your research participants and determine who is the controller for the research data.

I will prepare a Data Protection Notice** and give it to the research participants before collecting data ☒

The controller** for the personal details is the student themself ☒ the university ☐

My data does not contain any personal data ☐

** More information at the university's intranet page, [Data Protection Guideline for Thesis Research]

3. Permissions and rights related to the use of data

Find out what permissions and rights are involved in the use of the data. Consult your thesis supervisor, if necessary. Describe the use permissions and rights for each data type. You can add more data types to the list, if necessary.

3.1. Self-collected data

You may need separate permissions to use the data you collect or produce, both in research and in publishing the results. If you are archiving your data, remember to ask the research participants for the necessary permissions for archiving and further use of the data. Also, find out if the repository/archive you have selected requires written permissions from the participants.

Necessary permissions and how they are acquired

Data type 1: Interviews and recordings

In the interviews, I have clearly explained and received approval for that:

- Participation in the interview is completely voluntary.
- The interview answers are used as part of the thesis research material; however, all the answers and personal opinions are completely anonymized.
- Interview documents and transcripts do not contain the interviewee's personal information, such as his/her name or other identifiable information.
- The interviews are recorded.
- The interview answers are used only for this thesis and for research purposes.

3.2 Data collected by someone else

This thesis does not use any data collected by someone else.

4. Storing the data during the research process

Where will you store your data during the research process?

In the university's network drive ☐
In the university-provided Seafile Cloud Service ☐
Other location, please specify: ☒

The data collected during the research process is stored as recordings in the personal memory of my phone. Unlocking my phone requires a password and recognizing my face. The transcriptions written from the recordings are stored in the memory of my personal computer. My computer is protected with a password, anti-virus software and a firewall. All the data will be stored, in accordance with the instructions of the University, for five years.

5. Documenting the data and metadata

How would you describe your research data so that even an outsider or a person unfamiliar with it will understand what the data is? How would you help yourself recall years later what your data consists of?

5.1 Data documentation

Can you describe what has happened to your research data during the research process? Data documentation is essential when you try to track any changes made to the data.

To document the data, I will use:

A field/research journal ☐

A separate document where I will record the main points of the data, such as changes made, phases of analysis, and significance of variables ☒

A readme file linked to the data that describes the main points of the data ☐

Other, please specify: ☐

5.2 Data arrangement and integrity

How will you keep your data in order and intact, as well as prevent any accidental changes to it?

I will keep the original data files separate from the data I am using in the research process, so that I can always revert back to the original, if need be. ☒

Version control: I will plan before starting the research how I will name the different data versions and I will adhere to the plan consistently. ☒

I recognise the life span of the data from the beginning of the research and am already prepared for situations, where the data can alter unnoticed, for example while recording, transcribing, downloading, or in data conversions from one file format to another, etc. ☒

5.3 Metadata

Metadata is a description of you research data. Based on metadata someone unfamiliar with your data will understand what it consists of. Metadata should include, among others, the file name, location, file size, and information about the producer of the data. Will you require metadata?

I will save my data into an archive or a repository that will take care of the metadata for me. ☐

I will have to create the metadata myself, because the archive/repository where I am uploading the data requires it. ☐

I will not store my data into a public archive/repository, and therefore I will not need to create any metadata. ☒

6. Data after completing the research

You are responsible for the data even after the research process has ended. Make sure you will handle the data according to the agreements you have made. The university recommends a general retention period of five (5) years, with an exception for medical research data, where the retention period is 15 years. Personal data can only be stored as long as it is necessary. If you have agreed to destroy the data after a set time period, you are responsible for destroying the data, even if you no longer are a student at the university. Likewise, when using the university's online storage services, destroying the data is your responsibility.

What happens to your research data, when the research is completed?

I will store all data for 5 years on my personal computer in accordance with the security measures mentioned above. After this, I delete the data permanently.