Automated Identification of a Defensive Basketball Strategy Using Whole Team Player Tracking Data

TILEMACHOS MARKOS BAZAKAS

DIPLOMA THESIS

Department of Computer Science and Engineering

School of Engineering

University of Ioannina

October 2023

TABLE OF CONTENTS

ABSTRACT

Tilemachos Markos Bazakas, Department of Computer Science and Engineering, School of Engineering, University of Ioannina, Greece, October 2023.
Automated Identification of Defensive Basketball Strategies Using Whole Team Player Tracking Data.
Advisor: Konstantinos Blekas, Professor.

In the realm of sports analytics, the utilization of spatio-temporal data has revolutionized the understanding of team dynamics and strategic decision-making in games like basketball. This thesis delves into the intricacies of defensive strategies in basketball, focusing on the recognition and analysis of these strategies through advanced spatio-temporal modeling techniques. Most research to date on basketball strategy learning has focused on offensive effectiveness and is based on the interaction between the on-ball player and principle on-ball defender, thereby ignoring the contribution of the remaining players. Furthermore, most sports analytical systems that provide play-by-play data is heavily biased towards offensive metrics such as passes, dribbles, and shots. The aim of the current study was to create a defensive strategy recognition model to classify the different defensive strategies basketball players adopt when deviating from their initial defensive action. An analytical model was developed to recognise the one-on-one relationships of the players, which is utilised to automatically identify any change of defensive strategy. The model is developed based on a player and ball tracking dataset from National Basketball Association (NBA) game play to classify the adopted defensive strategy against pick-and-roll play. The models results indicate that the proposed technique for automatic defensive strategy identification can achieve up to 76.3% accuracy of labeling. In conclusion, this thesis advances our understanding of basketball defensive strategies by harnessing the power of spatio-temporal data and sophisticated modeling techniques. The proposed model not only accurately recognizes diverse defensive plays but also provides valuable insights for optimizing team performance. This research marks a significant step towards a data-driven approach to sports strategy analysis, with implications extending to various other team sports beyond basketball.

1.Introduction

1.1 Background and Context

In the ever-evolving landscape of sports analytics, the integration of advanced technologies has propelled our understanding of team dynamics, player interactions, and strategic decision-making processes in athletic contests. Particularly, the advent of player tracking data has ushered in a new era of granular analysis, enabling researchers and coaches alike to delve deep into the intricate patterns of player movements on the field. In the realm of team sports, such as basketball, this data provides a wealth of information that can be harnessed to unlock valuable insights into the strategies employed by both offensive and defensive units.

1.2 Importance of Player Tracking Data in Sports Analytics

The National Basketball Association (NBA) now captures optical data from cameras positioned directly over the playing floor in every arena. The NBA is the most competitive basketball league in the world, comprising 82 games for each team and in total 1230 games spanning across approximately 24 weeks [1]. A basketball game is modelled as a process by which the players that form the dyads(defender-attacker) attract to and repel from each other to produce the unique patterns that characterise player behaviours [2]. The team behaviour in basketball can be characterised by the space creation dynamics that relate to offensive behaviour [3] and the space protection dynamics that work to counteract space creation dynamics with defensive play [4]. Pick-and-Roll (PNR) has the highest frequency of occurrence of all space creation dynamics in basketball [3]. Understanding game strategies from past events (e.g., historical match analysis) may enable teams to gain a competitive advantage by knowing their opponents and by coming up with novel strategies to mitigate the perceived strengths of the opposing team. In the past, researchers have predominantly used two types of data to analyse basketball game strategies, i.e., play-by-play data that describe different events that happen on court such as shots, passes, dribbles, and fouls and player and ball tracking data. For example, play-by-play data can be used to learn the effectiveness of different types of PNR plays [5], or the factors that influence the effectiveness of inside-passes [6].

1.3 Focus of the Study: Exploring Team Sports Dynamics

This thesis embarks on a pioneering exploration into the realm of sports analytics, aiming to create a comprehensive model leveraging player tracking data to unravel the complexities of team sports dynamics. The focus of this study is manifold, encompassing player localization, team segmentation, possession determination, defensive assignment probabilities, and temporal pattern recognition. By meticulously analyzing player positions and movements, this research endeavors to develop a sophisticated algorithm that accurately calculates distances between players, identifies team affiliations, recognizes attacking teams based on ball possession, computes defender probabilities concerning every attacker, and classifies distinct timesteps and temporal windows based on dynamic changes in defender-to-attacker assignments.

1.4 Objectives and Significance of the Research

The significance of this research lies not only in its technical intricacy but also in its potential to revolutionize how we perceive and strategize team sports. The proposed model is designed to be versatile, with applications spanning across various sports where player tracking data is available. Through this innovative approach, the thesis aims to enhance our comprehension of the tactical nuances inherent in team sports, offering actionable insights for coaches, analysts, and players.

Furthermore, this study undertakes a rigorous evaluation process, measuring the accuracy of the model's classifications against a set of ground truth labels. This meticulous validation procedure ensures the reliability and robustness of the proposed algorithm, providing a foundation for its practical application in real-world scenarios. By systematically comparing the model's labeling with the established ground truth, this research establishes a framework for assessing the efficacy of similar models in the future.

1.5 Contribution of the Thesis

In summary, this thesis not only endeavors to develop an advanced analytical tool but also aims to contribute significantly to the evolving landscape of sports analytics. By harnessing the power of player tracking data and sophisticated algorithms, this study strives to offer a paradigm shift in how we analyze and understand team sports dynamics, paving the way for more informed decision-making processes and strategic advancements in the world of athletics.

1.6 Structure of the Thesis

The contributions of this thesis are as follows:

• Development of an analytical model to match the one-on-one relationship between players in both teams based on the players' location data. The process of matching is based on the underlying factors behind the trajectory data of players, such as distance between players, distance of players from the ball and whole team distance and probability thresholds.

•Manual classification of events as ground truth, to compare the results of the models classification in order to get the success rate. The identification specifies on a change of defensive tactic mechanism against PNR: 'switch'. The current study focuses on all players in the team (on-ball players and off-ball players).

The rest of this thesis is organized as follows:

- Section 2.
  describes the most used offensive strategy in basketball and it's corresponding defensive strategies. It also presents the work associated with data-driven algorithms for learning strategies in basketball.
- Section 3.
  introduces us to the most technical part of this thesis where we dive into the proposed methodology for classifying defensive strategies in basketball utilising player tracking data.
- Section 4.
  discusses the experimental setup and results

- Section 5.

  resents a summary and conclusion of the work done in the thesis and comprises indications of future work.

2.Background on Basketball Strategies and Related Research

2.1 Historical Evolution of Basketball Strategies

Basketball, originating in 1891, has transformed significantly since its inception. The evolution of basketball strategies, particularly in defense, reflects a dynamic interplay between rule changes, coaching philosophies, and the evolving skill sets of players.

2.1.1 Origins and Early Strategies (Late 19th Century - Early 20th Century)

In the nascent stages, basketball was characterized by basic strategies. Defenders employed man-to-man marking, attempting to nullify opponents individually. The absence of a shot clock led to deliberate gameplay, emphasizing ball control and selective shooting.

2.1.2 Rule Changes and Defensive Innovations (1920s - 1960s)

The introduction of the shot clock in 1954 revolutionized defensive tactics. Defenders began utilizing zone defenses, marking areas rather than specific players, pressurizing opponents to make hurried shots. Coaches started emphasizing teamwork, introducing complex defensive rotations and trapping techniques to disrupt opponents' offensive flow.

2.1.3 Era of Modern Basketball (1970s - 1990s)

The 1970s witnessed a shift towards more physical defense. With the advent of iconic players like Bill Russell and Wilt Chamberlain, shot-blocking and rebounding became pivotal defensive skills. Coaches like Pat Riley popularized the concept of 'Showtime' defense, focusing on fast breaks and transitional play, while others like Phil Jackson emphasized team-oriented, strategic zone defenses.

2.1.4 Analytical Revolution and Defensive Adaptations (2000s - 2010s)

The 2000s marked the analytical revolution in basketball. Coaches, armed with advanced analytics, devised intricate defensive schemes. Zone defenses evolved into dynamic systems, adjusting based on opponents' shooting tendencies. Help defense became prevalent, emphasizing players' abilities to cover multiple positions, while individual skills such as steals and deflections gained prominence.

2.1.5 Present Day and Future Trends (2020s - Present)

In the contemporary era, basketball strategies have become more fluid and adaptable. The emphasis on three-point shooting has led to defensive strategies that focus on closing out shooters swiftly. Agile big men are now vital for defending both the paint and the perimeter. Moreover, the rise of positionless basketball demands defenders to be versatile, capable of guarding multiple positions effectively.

Conclusion

The evolution of basketball strategies showcases the sport's adaptability. Rule changes, coaching innovations, and player skill sets have collectively shaped defensive approaches, moving from rudimentary man-to-man marking to sophisticated, data-driven defensive systems. As basketball continues to evolve, these historical developments serve as the foundation for future innovations, promising a sport that remains compelling and dynamic.

2.2 Impact of Technology in Basketball Strategies

Advancements in technology have significantly reshaped the landscape of basketball, revolutionizing how strategies are developed, analyzed, and executed. From player tracking systems and video analysis software to sophisticated simulation tools, these technological innovations have profoundly influenced coaching methods and player performance analysis.

2.2.1. Player Tracking Systems

Player tracking systems, equipped with sensors and cameras, provide real-time data on players' movements, speed, shot accuracy, and positioning. This wealth of information enables coaches to assess player performance objectively. By analyzing this data, coaches can identify players' strengths and weaknesses, optimize player positioning on the court, and design strategies that capitalize on their unique skills. Defensive strategies, such as identifying player tendencies and improving defensive rotations, benefit immensely from the insights garnered through tracking systems.

2.2.2 Video Analysis Software

Video analysis software allows coaches and players to review game footage, offering a detailed perspective on player actions, team dynamics, and opponent strategies. Coaches can dissect offensive and defensive plays frame by frame, enabling in-depth analysis of decision-making, shot selection, and defensive positioning. This granular understanding facilitates targeted coaching, helping players refine their techniques and make better strategic decisions during games. Video analysis also aids in scouting opponents, allowing teams to anticipate their adversaries' moves and adapt their strategies accordingly.

2.2.3 Simulation Tools

Simulation tools, often powered by artificial intelligence, enable coaches to simulate game scenarios, test different strategies, and assess their potential outcomes. Coaches can create virtual environments where they experiment with various player rotations, offensive plays, and defensive formations. By simulating different scenarios, teams can anticipate opponents' responses, strategize for crucial moments in the game, and make data-driven decisions. This enhances strategic planning, allowing teams to devise and refine their game plans with a high degree of precision.

2.2.4 Enhancing Coaching Methods

Technology has democratized access to coaching resources. Coaches can now compile vast databases of plays, strategies, and training techniques. They can share instructional videos and personalized feedback with players, fostering continuous improvement. Virtual reality training programs further immerse players in realistic game situations, allowing them to practice decision-making and hone their skills in a controlled, immersive environment.

2.2.5 Improving Player Performance Analysis

Beyond games, technology enables continuous player performance analysis during practices and training sessions. Wearable devices, such as smart vests and wristbands, collect real-time data on players' physical exertion, heart rate, and fatigue levels. This information helps coaches optimize training regimens, prevent injuries, and ensure players are in peak condition during games.

Conclusion

The integration of technology into basketball has ushered in an era of unprecedented strategic sophistication. Player tracking systems, video analysis software, and simulation tools empower coaches and players with actionable insights, enhancing both individual skills and team dynamics. As technology continues to advance, basketball strategies will evolve further, fostering a new era of strategic brilliance and athletic achievement on the court.

2.3 Introduction to Basketball Strategies

In this research, the ‘pick-and-roll’ is focused upon and more specifically the defensive strategy “switch” that is used against the “pick-and-roll”. To differentiate between the strategies, it is necessary to describe related basketball terms, such as one-on-one defense, zone defense, triangle-and-two, trap and switch, and help defense, as follows:

2.3.1 One-on-One Defence and Zone Defence

One-on-one defense: Each defensive player defends one offensive player. As offensive players move, the defensive players who are assigned to specific offensive players move accordingly. Zone defense: In contrast to one-on-one defensive strategies where defenders are assigned to specific opposing players, defenders in a zone defense focus first on defending specific areas (zones) of the floor [7]. When an offensive player moves into a defender’s assigned area, the defensive player traditionally defends that player with one-on-one principles (until the offensive player vacates). While 1-on-1 defense is popular in senior teams, zone defense is popular among youth teams such as Under-16 or Under-18 [8]. Combination or hybrid defense where a mix of man-on-man and zone defense is also utilised in some basketball settings. There are different types of combination defense systems such as Box and one, Diamond and one, and triangle and one. For example, in triangle and one method, two players are left to match up man-to-man, while the remaining three defenders protect against penetration by forming a triangle. On the other hand, full court press is where a defensive team applies pressure on the offensive team the entire length of the court, through man-on-man or zone defense. Half court press allows the offense to arrive halfway down the court before applying defensive pressure.

2.3.2 Pick-and-Roll

The general pick-and-roll, also known as on-ball screen [7], involves one offensive player setting a screen (pick) for a teammate in possession of the ball (on-ball player). As shown in Figures 2 and 3, it contains three steps: Firstly, the off-ball player sets a screen for a ball holder player; Secondly, the ball holder player reads defensive strategies and uses the screen to create an open-shot opportunity for their teammates or themselves; Thirdly, after screening, the off-ball player reads the defense and creates an open-shot opportunity for their teammates or themselves.

The pick-and-pop is a variation of the pick-and-roll strategy, where the player does not roll to the basket, instead popping out to the perimeter. On perimeter the player gets an open look at the basket to shoot, when the pass is received from the guard.

2.3.3 Switch and Trap

For defensive players, two options named ‘switch’ and ‘trap’ are often deployed to defend pick-and-roll. According to McIntyre [9], ‘switch’ denotes an on-ball defender and off-ball defender(s) switching their original matchups as shown in Figure 2 [10], while ‘trap’ means both on-ball defender and off-ball defender(s) double the on-ball player (i.e., two or, very occasionally, more, players directly defending the player with the ball) as shown in Figure 3.

Frame 100 Frame 103 Frame 107 Frame 110

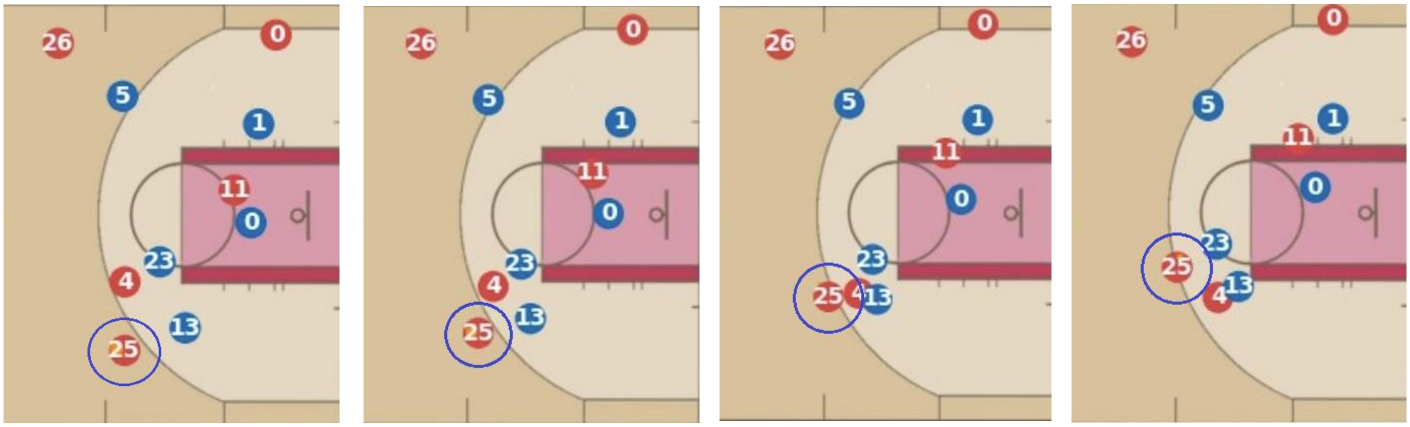

Figure 2. In these four timesteps, 25 and 4 play pick-and-roll. After that, 23, which is the original defender of 4, becomes the defender of 25, and 13 becomes the defender of 4. This process is called ‘switch’.

Frame 168 Frame 171 Frame 174 Frame 177

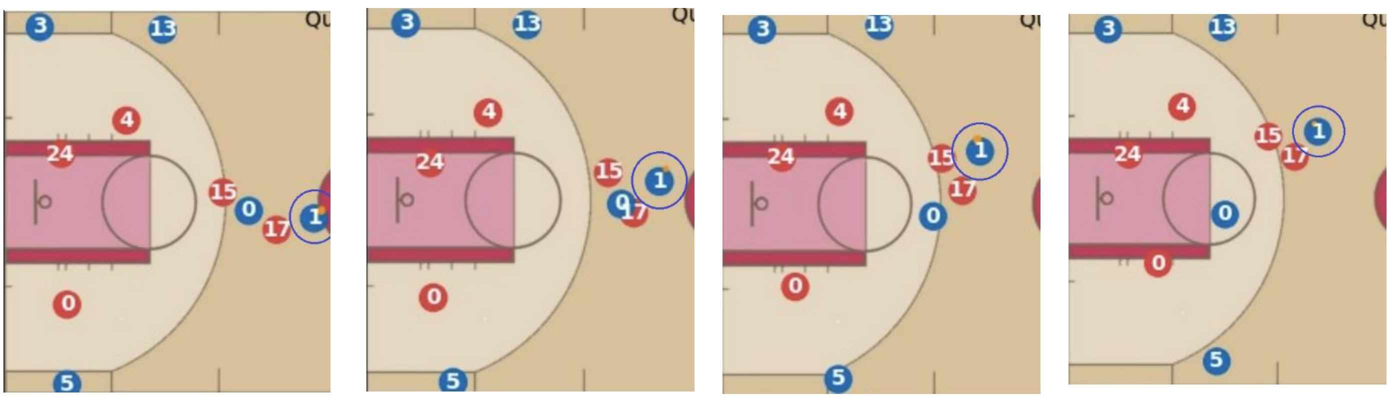

Figure 3. In this case, 1 and 0 play pick-and-roll. After that, both 15 and 17 try to defend 1 (ball holder), and at the same time, 0 (off-ball player) make space for themselves. This process is called ‘trap’.

2.4 Data Driven Performance Analysis in Basketball

Evaluating basketball performance is intricate, involving dynamic interactions among players' technical, tactical, fitness, and anthropometric traits. Player tracking data unveils intricate connections within player pairs, both longitudinally and laterally, creating unique behavioral patterns. Teams attract and repel, forming distinctive spatial arrangements.

Researchers explore spatial dynamics, revealing in-phase relations between teams' spatial centers longitudinally and laterally, while the stretch index, indicating players' deviation from the center, demonstrates in-phase relationships longitudinally. Space creation dynamics, offensive tactics generating scoring opportunities, intertwine with space protection dynamics, defensive strategies countering offensive moves.

Crucial offensive maneuvers like the Pick-and-Roll involve a screener creating opportunities for a teammate, analyzed based on factors like time, space, and player actions. Effective screens, combined with strategic passes, enhance offensive effectiveness. Contextual factors like game duration affect offensive strategies. Early games exhibit faster pace and shorter possessions, while later stages involve more players and longer possessions, increasing scoring chances. Ball screens are an important offensive tool used by teams in basketball, where a screener sets the screen generating an advantageous situation to the dribbler who will pass it on to a teammate who is in an open field-goal situation or shoot without the defensive pressure. In [11], authors investigated the predictors of success of ball screens such as time, space, players, and tasks performed, and illustrated the relationships of ball screens and offensive success with the orientation of the ball screen and actions of the dribbler after the screen. By considering a sample of 668 ball screens from 17 games of the Spanish basketball league, the contextual factors that affect the ball screens such as, score-line, offensive system, duration and game quarter, were analysed in [12]. Through statistical analysis of play-by-play data related to the ball screens, it was concluded that the effectiveness of the ball screens was greater when there was high time pressure, especially towards the final few seconds of a ball possession. A total of 12,376 pick-and-rolls in European basketball league were analysed in [13], by considering the play-by-play data from Synergy sports systems. Different types of PNR plays are analysed for their effectiveness and the authors conclude that possessions that end with the screener's rolling in the shot and those that end with 2 passes following the PNR are the most effective uses of the PNR and the least successful type is when the ball handler shoots [13].

Scientific studies often focus on match events, overlooking behavioral patterns and team strategies. Predicting play outcomes involves generative probability models capturing insights from performance analyses. Possession-based Markov models and graphical representations incorporating player actions, events, and moves offer avenues for predicting match progress. Developing accurate team dynamic models proves challenging due to basketball's complexity. Integrating diverse contextual factors enhances these models' precision

2.5 Use of Machine Learning in Basketball Analytics

Wang and Zemel [14] developed a machine learning model to process player tracking data to identify offensive plays in basketball. The trajectory dataset utilised in [14] was provided by SportVU [15]. Each data point within this dataset is stored as a sequence which has the coordinates of the ball and all players on the basketball court. A sequence is recorded every 0.16 s. The authors devised variants of neural network algorithms to model a classifier to recognise basketball tactics based on unlabelled historical data. This study implemented the recurrent neural network (RNN) which can deal with sequential data of variable length. The inputs of the RNN comprise the player coordinates and the outputs are the labels for the different offensive tactics. The authors also describe the development of an autoencoder neural network based on an extra dataset provided by NBA team, the Toronto Raptors, to enhance the model to become a player-specific model. The researchers demonstrated that the model could be used to predict the movement of players, i.e., to forecast future player position [14]. This recurrent neural network achieved a classification performance of 80.6% accuracy. Nevertheless, due to the limited number of training data, the predictive capability of the recurrent network was relatively weak (less than 60%).

Mclntyre et al. [15] also developed a supervised machine learning algorithm utilising NBA player tracking data from SportVU. The authors focussed on recognising and analysing ball screen defence (pick-and-roll). Their system takes the unlabelled trajectory data as input, and then derives the time of all ball screens that occur during each game. Finally, classifying how the ball screen was guarded by defence players. In [9], the researchers first defined four players who formed the ball screen tactics: a ball handler, a ball defender, a screener and a screener defender. Secondly, they defined four different categories for the defensive tactics based on the trajectories of these four players. In [9] authors trained a logistic regression machine learning model that uses the distance between four defined characters as features. Finally, through 5-fold cross-validation they achieved an overall average classification accuracy of 69%. In addition, this thesis analysed the trends of different players and different teams in the face of the ball screen.

Learning long-term behaviour for multi-agent spatiotemporal trajectories is a key challenge in many learning problems [16]. Zhan et al., proposed Multi-Agent Generative Behavioural Cloning, referred to as MAGnet, based on a deep learning algorithm. The authors describe MAGnet as a flexible class of generative models that can generate rich multi-agent spatiotemporal trajectories over a long-time horizon [16]. Moreover, the main advantage of MAGnet is that it has a shared structure between agents, and it has a hierarchical latent structure to jointly represent long-term (macro) and short-term (micro) temporal dependencies. Furthermore, the researchers compared MAGnet with a Variational Recurrent Neural Network algorithm (VRNN) [17]. The VRNN algorithm proposed by Chung et al. [17] is built by the combination of a recurrent neural network with variational autoencoder. According to [16], the MAGnet algorithm has significantly higher performance than the VRNN model [17]. The concept of macro-goals allows the observer to analyse the long-term goals of a player and how they change their objectives during game play. Zhan et al. suggested that exploring a more powerful probabilistic structure to handle more agents under a complex condition is an avenue worth pursuing.

Nearly all basketball tactical analysis studies only analyse the status of a section of players and not the whole team. For example, the pick and roll strategy typically comprises four players—two players on each team—the remaining six players are not analysed. Another example is where the study of shooting outcome predictions only takes into account the tracking data of one player. However, basketball is a team game consisting of ten players, therefore, if only the status of some players is analysed, the results of the analysis will be compromised by the inevitable interference from undefined factors.

2.6 Challenges and Future Direction in Basketball Strategy

2.6.1 Challenges in Implementing Advanced Strategies

a. Player Adaptation: One of the primary challenges faced by coaches is ensuring players adapt to advanced strategies seamlessly. Implementing complex defensive rotations or intricate offensive plays requires players to understand and execute these strategies effectively, which demands extensive practice and player coordination.
b. Injury Management: The high-paced nature of modern basketball, with an emphasis on speed and agility, poses a challenge in managing player injuries. Coaches must balance rigorous training regimes with player rest and recovery to prevent injuries that could hinder the implementation of advanced strategies.

c. Opposition Analysis: Teams must continually adapt their strategies based on the opposition's strengths and weaknesses. Analyzing and countering the strategies of diverse opponents demand extensive scouting and analysis, requiring significant time and resources.

2.6.2 Future Direction in Basketball Strategy Research

a. Integration of AI and Machine Learning: The use of artificial intelligence and machine learning algorithms is on the rise. Predictive analytics can help anticipate opponents' moves, optimize player rotations, and simulate various game scenarios. Further research in AI-driven decision support systems can revolutionize strategic planning in basketball.
b. Biomechanical Analysis: Advancements in biomechanics can provide in-depth insights into players' movements, helping coaches optimize techniques and prevent injuries. Understanding the biomechanics of shooting, passing, and defensive maneuvers can enhance skill development programs, leading to more effective strategies on the court.
c. Virtual Reality and Augmented Reality Training: Virtual reality (VR) and augmented reality (AR) technologies offer immersive training experiences. Players can practice scenarios, analyze defenses, and refine strategies in virtual environments. Future research might focus on enhancing the realism and effectiveness of these technologies, making them integral components of team training sessions.
d. Interdisciplinary Approaches: Collaborations between basketball experts, data scientists, psychologists, and physiologists can provide a holistic understanding of player performance. Integrating psychological aspects, such as decision-making under pressure, with physiological data and game statistics can offer comprehensive insights into player behavior, leading to more nuanced and effective strategies.
e. Ethical and Social Implications: As technology advances, ethical considerations surrounding player privacy, data security, and fair play become crucial. Future research should address these ethical concerns and develop guidelines to ensure the responsible use of advanced technologies in basketball strategy development.
f. Environmental and Sustainability Considerations: Research in sustainable sports technologies can explore how energy-efficient arena designs, eco-friendly equipment, and reduced carbon footprints can impact player performance and influence strategic decisions,

contributing to a greener sports industry.

Conclusion

Addressing the challenges in implementing advanced strategies while exploring innovative research avenues is essential for the future of basketball. By embracing emerging technologies, interdisciplinary collaborations, and ethical considerations, the basketball community can navigate the evolving landscape of the sport, ensuring that strategies continue to evolve while upholding the integrity and fairness of the game.

3.Methodology

This section describes the proposed methodology for dynamic (automated) recognition of defensive strategies in basketball. This section is organized in two subsections: the data extraction, the analytical model for defensive relationships between players ball possession, distance probabilities, defensive assignments, distance probability thresholds, whole team distance thresholds.

3.1 Dataset

SportVU tracking data, provided by STATS LLC [18], was utilised. The dataset contains basketball player trajectory data obtained from the NBA. Each data file represents a whole basketball game. Each game is a sequence of events where every event represents a possession during which one team retains the ball and ends when the ball is captured by the opposing team, or if a shot at the basket is made [16]. Possession, is an essential efficiency statistic because it allows statistical analysis based on a per-possession basis. The dataset obtained from SportVU consists of possessions of variable length: each possession is a sequence of tracking coordinates (si t = x i t , yi t) for each player i, recorded at 25 Hz, where one team has continuous possession of the ball. The possessions last between 50 and 1000 frames. The dataset obtained contains 32,377 possessions from the 2012–2013 NBA season. These data have come from approximately 630 games. It should be noted that in the original study that used this data [19], 80,000 possessions were analysed along with the event data, of which only the 36,330 possessions are made publicly available, and excludes the associated event data. Of the 32,377 possessions used in our study, 469 of them were manually labeled, of the defensive strategy used against pick-and-roll, by our own analysis. Our analyst has over 20 years of experience in basketball and over 10 years in professional basketball and is reliable to recognize the defensive strategies used in every possession available.

The objective of the thesis was to use an automatic defensive strategy labeling model to classify “switch” by analysing the player and ball tracking data, and for this the labeled data need to be provided. An analyst should ideally go through every individual frames of a possession in sequence to recognise if there was a switch in the possession. However, this was a very time-consuming task to label all the possessions. Therefore, we labeled 469 possesions from which 120 contain a “switch”. These manually labeled possessions are used for identifying the reliability of the analytical model. The analytical model selects 10 frames at a time from a possession and analyse if there is a switch in the possession. Not all the possessions in the dataset will contain a switch, and notice that some possessions will contain more than one occurrence of switch. Finally, a total of 42,865 plays of switches and traps were identified from 32,377 possessions from which only 469 could eventually be confirmed.

3.2 Analytical Model for Labeling the Dataset

Objective of this model is to develop a labeled set of switch plays identified from the tracking data to be compared with the ground truth labels given manually to the possessions. For this purpose, the analytical model encodes contextual knowledge about basketball defense as described in Section 2.

A basketball game consists of two teams each with five players. The team that possesses the ball at any given time is considered the offensive team, and the team without the ball is considered the defensive team. In our model, to determine which team is the offensive team we check the first 30 timesteps of the possession. In each timestep we have in our possession the position of every player on the field and the position of the ball. From here we simply check which player has the ball in his possession by calculating the euclidean distance between each player and the ball:

$$\text{player\_ball\_distance} = \sqrt{[(x_{ball} - x_{player})^2 + (y_{ball} - y_{player})^2]}$$

The player that has the smallest distance from the ball is the ball holder at that timestep. By the end of the first 30 timesteps, the ball may have changed many “hands” but we keep track which team’s players had the ball in their possession. After this process we determine the defending team’s players and the offensive team’s players. Typically, each defensive player selects one matched offensive player to defend. An experienced player, coach or spectator can identify the correspondence between players (who is defending whom) based on the position of players. The objective of this algorithm is to understand the defensive relationship between the players of the two teams based on the player tracking data. This model underpins the overarching aim of this study by enabling the classification of defensive strategies in basketball.

The analysis of the underlying relationship between in-game players needs to be a team-level analysis rather than a specific player-level analysis. Determining the one-on-one defensive relationship between players will act as the foundation of the proceeding automated identification model building. The analytical model is based on the following attributes:

- the data extraction and transformation from SportVU tracking data, provided by STATS LLC
- the creation of windows of 10 timesteps
- the recognition of the ball possessing team
- the calculation of distances between players
- the conversion of player distances into probabilities
- the assignment of defenders to attackers (coupling) based on player distances
- the recognition of change in defender-attacker assignments
  - the threshold creation for average player distances
  - the recognition of change in defender-attacker assignment when the change occurs on a player who holds the ball, which means a "switch" on a pick-and-roll
  - the confirmation of a "switch" if a change remained throughout the timesteps of a window
  - the threshold creation for probability changes between timesteps of defender assignments.
- the labeling of windows and events as 0 or 1 based on whether the "switch" action in defender-attacker assignment was observed
- the comparison of the ground truth labeled possessions with the models labeled possessions

The structure of the proposed analytical model is illustrated in Figure 5.

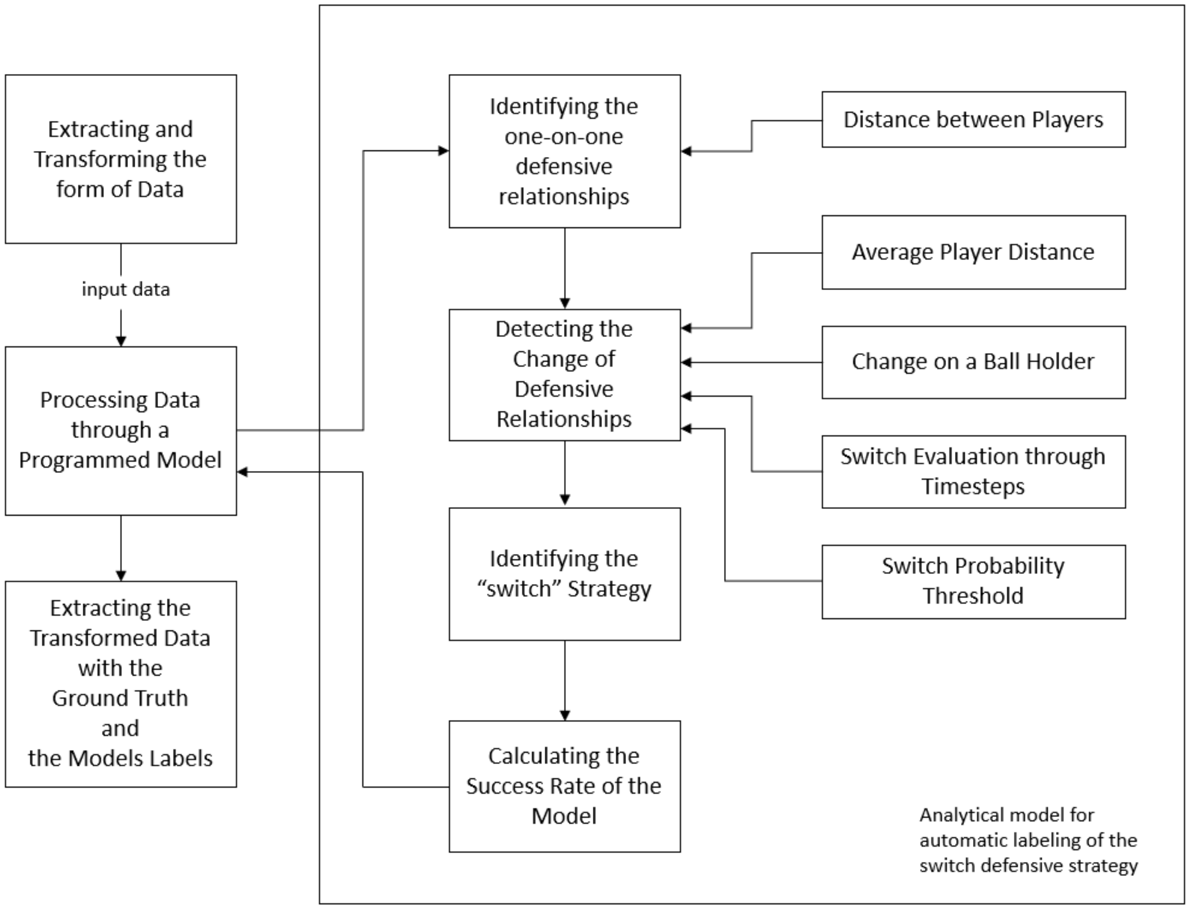

Figure 5. Analytical model for automatic labeling of the switch defensive strategy

3.2.1 Data Extraction and Transformation

Regarding the extraction of the data we need from the tracking data, we have in our possession roughly 630 NBA games from raw SportVU data logs. Each game has the structure of a dictionary. A dictionary is a collection of key-value pairs. The keys in a dictionary must be unique, and they are used to retrieve their corresponding values. The structure is:

- A dictionary that holds the keys: { gameid: , gamedate: , events: } and their values
- The "events" value is a list of dictionaries where some of the keys have dictionaries and lists of dictionaries as values too. So, the value of the "events" key is:
  - [{eventid: , visitor:{ name: , teamid: , abbreviation: , players: [{}]} , moments: [[, [[ ]]], … [, [[ ]]] … [, [[ ]]]]
  - The "players" value is also a list of dictionaries where each dictionary in the list contains the following:

        - [{lastname: , firstname: , playerid: , jersey: , position: },
    - After the “players” list of dictionaries ends, the “visitor” dictionary ends too and “moments” dictionary continues as follows:
        - [List that contains the following lists:
            - [quarter , game_clock , shot_clock ,
            - [List that contains the following lists:
            - [ball_position,],[player_1 position],…,[player_10 position]]]
- After the “moments” dictionary ends, the “events” list dictionaries continues to the next “eventid” and after all the “eventid”s end, finally the whole dictionary of keys and values ends.

The above structure shows one game’s dictionaries of data from where we extract the position of the ball (x,y) at each timestep, the position of every player (x.y) at each timestep and for each timestep that we extract positions, we keep the “eventid” number in order for us to recognize which event’s timesteps we are analyzing later on. In Figure 4 we visualize the extracted and transformed data. Players 1 to 5 are the on the same team and players 6 to 10 on the same opposing team.

Ball Position	Player 1 Position	Player 2 Position	Player 3 Position	Player 4 Position	Player 5 Position	Player 6 Position	Player 7 Position	Player 8 Position	Player 9 Position	Player 10 Position	EventID
(x,y)	(x,y)	(x,y)	(x,y)	(x,y)	(x,y)	(x,y)	(x,y)	(x,y)	(x,y)	(x,y)	i

Figure 4. The data extracted and transformed from raw SportVU data logs.

3.2.2 Separation of Timesteps into Windows

In terms of the location at each timestep, a data example means a possession and a possession consists of many consecutive time-steps. In addition, the interval between each time step is 0.16 s (6.25 Hz). Through processing the raw data, we obtain the location of all the ten players at a given timestep. The separation of timesteps into windows of 10 timesteps happens at the beginning but, it’s utilized later on as the last step of the model which decides whether a “switch” has actually occurred. It allows us to keep track of any assignment change and confirm if it remained. In the case where a change in a defender-attacker assignment was observed, we keep the couple of the players that switched assignments and if the coupling remains throughout the window’s timesteps, then we decide that a “switch” actually happened.

3.2.3 Recognition of the Ball Possessing Team

Finding which team has the possession of the ball is very simple. As we mentioned above, in one possession the ball changes many "hands" so we can't just keep track of one player for the possession of the ball. For 30 timesteps we keep track which of the players 1 to 5 or 6 to 10 keep the ball on them and the team with the most timestep counts is the ball possessing team. It is very important to know which team is the defending and which is the attacking. By dividing the players into attackers and defenders allows us to have an epicenter. The epicenter gives us footing on which players we want to start calculating distances between. In our model we decided to have the defenders as the epicenter and find all the distances of each defender with every attacker.

3.2.4 Calculation of Player Distances

By dividing the players into attackers and defenders we use the defenders as the epicenter. From the extracted and transformed data, we possess the location of all the players at each timestep and the eventid that we are focusing on. We divide the data into groups of events(possessions) in order to analyze each possession at a time. For each eventid, for each timestep of the event and for each defender we calculate the distances of that defender with every attacker by using the Euclidean equation.

$$\text{defender\_attacker\_distance} = \sqrt{[(x_{defender} - x_{attacker})^2 + (y_{defender} - y_{attacker})^2]}$$

From here on, we now have every distance for each defender with every attacker stored. That means we have 5 defender where every defender has 1 distance for every attacker. In total we have 5*5= 25 distances stored for every timestep.

3.2.5 Probability Conversion

Converting player distances into probabilities is an extra step which gives as a better view on the defender-attacker assignment (coupling) that takes place later on. The procedure conversion is as follows:

probability = 100 * (1 - (distance_to_attacker / total_distances_to_attackers))

This simple calculation leaves us with 25 probabilities stored for each timestep. The 25 probabilities are in the form of a list of dictionaries. The dictionaries have the following structure:

defender_probabilities = [{attacker_1: probability , attacker_2: probability , attacker_3: probability , attacker_4: probability , attacker_5: probability} , {…} , {…} , {…} , {…}]

Every timestep has the above list of dictionaries where each dictionary in the list contains the probabilities to every attacker of the defender of the position of the dictionary in the list.

3.2.6 Assignment of Defenders to Attackers

In this subsection, the process of assigning defenders to attackers is detailed through a systematic algorithmic approach. The primary objective of this assignment is to optimize defensive strategies based on the calculated probabilities derived from the distances between attackers and defenders. The code snippet provided illustrates the iterative procedure employed for this assignment process.

Algorithm Overview:
The algorithm iteratively evaluates potential defender assignments for each attacker, considering their respective probabilities of successful defense based on proximity. The process unfolds as follows:

- For each iteration, the algorithm initializes variables to keep track of the maximum probability (max_probability), the attacker with the highest probability (max_attacker), and the corresponding defender (max_defender).

- The algorithm loops through all attacking players (attacking_players), attempting to find suitable defenders for each attacker.

- For each attacker, the algorithm checks if the attacker has not been assigned a defender (attacker not in assigned_defenders). If unassigned, the algorithm evaluates the available defenders based on the probabilities provided in defender_probabilities.

- For each potential defender, the algorithm compares their probability of successful defense (probability) against the current maximum probability (max_probability). If the new probability is higher, the algorithm updates max_probability, max_attacker, and max_defender with the values of the current attacker and defender under consideration.

- Once the best defender is identified for the given attacker, the algorithm assigns the defender to the attacker in the assigned_defenders dictionary.

- The process continues until all attackers are either assigned a defender or evaluated for potential assignments. This iterative optimization ensures that each attacker is paired with the defender that offers the highest probability of successful defense based on their respective distances.

Outcome: Through this systematic approach, the algorithm optimizes the assignment of defenders to attackers, maximizing the overall defensive capabilities of the team. By leveraging the calculated probabilities derived from distances, the algorithm enhances the strategic positioning of defenders, contributing significantly to the team's defensive prowess during gameplay.

3.2.7 Recognition of Coupling Change

In a game of basketball, there are many occurrences where defenders switch the attackers that they are guarding in the action of a possession. Especially, when we are analyzing timesteps and the condition of a defender to be assigned on an attacker is the distance between them, there are cases where players pass between a coupled assignment, cases where players cluster or are dispersed across the court, cases where assignment changes occur on players away from the ball. Excluding these type of situations in a possession, allows us to filter out the unimportant coupling changes and lets us focus on the pick-and-roll action. To recognize a coupling (defender-attacker assignment) change, we simply compare the assigned defenders to attackers of the current timestep, with the previous timestep’s assigned defenders. After the recognition of a change in a defender-attacker assignment, which means two defenders changed the attacker they are guarding with each other, we start to exclude situations where a "switch" happens inside the game of basketball that are not in the act of a pick-and-roll. So, when a change in a coupling is observed, we pass the timesteps, that the change occurred, through a filter of rules and thresholds to confirm the "switch" action.

3.2.7.1 Average Player Distance Threshold

In this subsection, an essential refinement to the defender-attacker assignment strategy is introduced through the implementation of average player distance thresholds. Recognizing the dynamic nature of basketball gameplay, where players frequently cluster or disperse across the court, these thresholds serve as critical discriminators, ensuring the accuracy and legitimacy of defender assignments across multiple timesteps.

Algorithm Integration:

- Two thresholds are defined: min_distance_threshold and max_distance_threshold. These values represent the minimum and maximum allowable average distances between players for a legitimate defender-attacker assignment to be considered valid.

- The algorithm, which monitors defender assignments across timesteps, verifies if there have been any changes in assignments (assignments_changed function). This function is essential to discern between real strategic changes and momentary fluctuations due to player proximity.

- For each timestep, the algorithm calculates the average distance between all defending players and attacking players on the court. This calculation involves summing up the distances between all possible defender-attacker pairs and dividing by the total number of such pairs.

```
average_player_distance = sum(distances_by_timestep[defender_id][attacker_id] for
defender_id in range(len(defending_players)) for attacker_id in
range(len(attacking_players))) / (len(defending_players) * len(attacking_players))
```

- Once the average player distance is computed, the algorithm evaluates whether it falls within the predefined threshold range (min_distance_threshold to max_distance_threshold). This evaluation serves as a pivotal decision point.

- If the average player distance surpasses the min_distance_threshold and is below the max_distance_threshold, the algorithm proceeds with the subsequent logic, considering the defender assignments as valid and representative of the actual gameplay dynamics. This validation step ensures that only assignments made during periods of reasonable player spacing are considered, enhancing the reliability of the algorithm's decisions.

The integration of average player distance thresholds critically enhances the robustness and accuracy of the defender-attacker assignment algorithm. By discerning between genuine strategic shifts and temporary distortions caused by player proximity, the algorithm becomes more adept at capturing the nuanced gameplay scenarios of basketball. This ensures that the defender assignments derived through the algorithm align with the spatial dynamics of the game, contributing significantly to the precision and authenticity of the defensive strategies devised by the system.

3.2.7.2 Recognition of Coupling Change on a Ball Holder

In this subsection, an advanced mechanism for recognizing crucial changes in the defender-attacker assignments is presented. The algorithm, designed to enhance the precision of assignment detection, focuses specifically on situations where the player possessing the ball (the "ball holder") undergoes a defensive coupling change. This optimization ensures the system's acute responsiveness to strategic shifts centered around the player in possession, a pivotal aspect of basketball gameplay.

Algorithmic Enhancement:

- The algorithm meticulously tracks changes in defender-attacker assignments across timesteps. This monitoring is vital for discerning alterations in defensive strategies during gameplay.

- A distinctive feature of this mechanism is its exclusive focus on situations where the attacker with the ball experiences a change in their defensive counterpart. This specialized evaluation is triggered when any defender-attacker pairing is altered.

- The algorithm identifies the attacker involved in the assignment change and determines their unique identifier (attacker_id). This identifier is crucial for subsequent computations.

- To validate the change's significance, the algorithm calculates the distance between the ball's position and the identified attacker's location on the court. This computation employs precise geometric calculations, ensuring accuracy in assessing the proximity of the ball to the relevant attacker.

- A critical threshold is employed to gauge the closeness of the ball to the identified attacker. If the distance between the ball and the attacker falls below this threshold, it signifies the attacker's active possession of the ball during the assignment change.

- If the calculated distance between the ball and the identified attacker is less than the threshold employed, the algorithm recognizes a coupling change specifically related to the ball holder. This recognition triggers the boolean variable ball_defender_changed and captures the identity of the ball holder (ball_holder_attacker).

By focusing on assignment changes involving the ball holder, this enhanced recognition mechanism adds a layer of strategic depth to the algorithm. It ensures that alterations in defensive couplings associated with the player in possession are accurately identified and distinguished from other assignment modifications. This heightened sensitivity to the ball holder's defensive dynamics provides the system with a nuanced understanding of pivotal gameplay moments, enabling more precise defensive strategies tailored to the evolving possession scenarios in basketball.

3.2.7.3 Switch Evaluation through Window Timesteps

In this subsection, a meticulous evaluation process is introduced to validate the legitimacy and consistency of defender-attacker assignment changes, specifically focusing on scenarios involving the ball holder. Recognizing the temporal dynamics of basketball gameplay, the algorithm employs a window-based approach, scrutinizing sequential timesteps in groups (here, window size is set to 10 timesteps). This methodology ensures that detected assignment changes remain persistent and coherent over a defined duration, enhancing the algorithm’s ability to differentiate strategic shifts from momentary fluctuations.

Algorithmic Verification Process:

- The algorithm processes the recorded gameplay data in distinct windows, each comprising 10 sequential timesteps. This segmentation enables a granular analysis of defender-attacker assignments, ensuring that changes observed within each window are assessed for their continuity.

- Within each window, the algorithm reconstructs defender-attacker assignments based on calculated probabilities and maximization logic. For every attacker, the defender with the highest probability of successful defense is selected, considering constraints and previous assignments. This step provides a coherent snapshot of assignments within the window.

- The algorithm specifically focuses on the ball holder's assignments, identified during the previous step. It compares the defender assigned to the ball holder in the current window (win_assigned_defenders.get(ball_holder_attacker)) with the defender assigned to the same attacker in the previous window (assigned_defenders.get(ball_holder_attacker)).

- If the defender assigned to the ball holder remains consistent throughout the entire window, the boolean variable change_remained is set to True, signifying a sustained assignment change for the ball holder across the window's timesteps.

- This verification process is essential for temporal validation, ensuring that the detected changes are not momentary anomalies but represent genuine strategic shifts sustained over a meaningful duration of gameplay. It safeguards against misinterpretation of transient events, thereby enhancing the reliability of the algorithm's decision-making process regarding defender-attacker assignments.

By implementing this window-based evaluation, the algorithm gains a comprehensive understanding of the stability and consistency of detected assignment changes. This meticulous temporal validation method is vital for distinguishing genuine strategic adaptations from sporadic, short-lived alterations, reinforcing the algorithm's ability to capture the nuanced and evolving dynamics of basketball gameplay. This approach significantly refines the system's decision-making capabilities, ensuring that the identified changes in defender-attacker assignments are both accurate and contextually relevant within the broader gameplay context.

3.2.7.4 Probability Threshold Between Window Timesteps

In this subsection, we introduce a critical enhancement to our defender-attacker assignment algorithm the Probability Threshold Between Windows' Timesteps method. The method is an extension of the previous algorithm we analyzed above which checks if a coupling change remains throughout all the windows' timesteps. This approach is pivotal in ensuring the stability and reliability of defender-attacker assignments over consecutive timesteps of a window. As outlined in our earlier methodology, for every defender, we calculate the probabilities of being assigned to different attackers based on their respective distances.

Algorithm Overview:

- The algorithm further incorporates a crucial check based on probability thresholds. After ensuring the initial assignment change consistency, the code examines the probabilities of defender-attacker pairings. If, at any point within the window, the probability of a defender being assigned to a new attacker falls below a predefined threshold, the change_remained variable is set to False.

- At the end of the window analysis, the change_remained variable indicates whether the assignment changes, including the probability threshold check, remained consistent throughout the entire window. If change_remained is True, it signifies that the assignment changes, including the probability threshold condition, were stable and reliable for the given window.

This methodology ensures that not only are the defender-attacker assignments consistent within a window of timesteps but also that these assignments are supported by sufficiently high probabilities, thereby adding an additional layer of reliability to our algorithm, The Probability Threshold Between Windows' Timesteps method significantly enhances the robustness of our defender-attacker assignment system.

3.2.8 Window and Event Labeling

At the end of the recognition of coupling change algorithm, we delve into the crucial aspect of labeling windows within the events based on defender-attacker assignment changes and the ball-handler's involvement. Below follows a complete overview of the algorithm that leads to the window and event labeling.

Algorithm Overview:

- The algorithm begins by checking if there has been a change in defender-attacker assignments from the previous timestep. This is accomplished by comparing the current assignments (assigned_defenders) with the previous assignments (prev_assignments). If a change is detected, the algorithm proceeds to evaluate the nature and consistency of this change.

- Before further analysis, the code calculates the average distance between all players on the court for the current timestep. This average distance is then compared against predefined thresholds (min_distance_threshold and max_distance_threshold). If the average distance falls within the specified range, the algorithm continues to evaluate the change. This step ensures that the players are neither too clumped together nor too dispersed.

- The algorithm then checks if the change in assignments involves the player currently holding the ball. It iterates through the assigned defenders and identifies if the ball-holder's defender has changed. If so, the ball_holder_attacker variable is set, signifying the attacker holding the ball in the current timestep.

- If the ball_holder_attacker is confirmed to be included in the assignment change, the algorithm proceeds to check the consistency of the assignment change throughout a window of 10 timesteps. It iterates through the timesteps in the window, ensuring that the assignment changes involving the ball-holder attacker remain stable across all these timesteps. If any inconsistency is found, the change_remained variable is set to False.

- Additionally, the algorithm includes a check for probability changes within the window. It compares the current probabilities of defender-attacker pairings with the probabilities from the previous timestep. If the difference in probabilities exceeds a certain threshold (probability_threshold), the change_remained variable is set to False.

- Based on the analysis results, the window_label variable is assigned a value: If a consistent and stable change in defender-attacker assignments involving the ball-holder attacker is detected (change_remained is True), window_label is set to 1, indicating a significant change in defensive strategy, “switch”. If there is no consistent change in defender-attacker assignments (change_remained is False), or if the players are too clumped together (average_player_distance falls outside the specified range), window_label is set to 0, indicating no notable change in defensive assignments. Finally, the whole possession is labeled as 0 or 1 if a change was involved in any of the windows of timesteps of the possession.

This meticulous process of window and event labeling ensures that only substantial and consistent changes in defender-attacker assignments, particularly involving the ball-holder attacker, are identified and labeled.

3.2.9 Extraction of the Models Results

In this subsection, we outline the critical process of evaluating the model's performance by comparing its generated labels with the ground truth labels provided by an analyst. This evaluation is essential to validate the accuracy and reliability of the model's predictions in identifying significant defensive events during basketball games.

- The algorithm iterates through the corresponding elements of both lists (ground_truth_labels and model_labels). For each pair of labels, it checks if the model's prediction matches the ground truth label.

- After comparing all labels, the algorithm calculates the percentage of matching labels by dividing the matching labels by the total number of labels.

- The algorithm showcases the percentage of matching labels, providing a clear and quantifiable measure of the model's accuracy in capturing the “switch” action. This value serves as a crucial metric for evaluating the model's performance against the analyst's annotations.

- Finally, the algorithm extracts all the available data and the generated data into a file in order for it to be used for future research. The data file contains information about the position of the ball and the 10 active players on the court at each timestep, the possession’s number at each timestep, the label of each window of 10 timesteps, the model’s label of each possession and the ground truth label given by our analysis.

By executing this process, the model's ability to accurately identify and classify the “switch” in a pick-and-roll is quantified, offering valuable insights into its effectiveness in understanding complex basketball gameplay dynamics. The calculated percentage of matching labels serves as a key result, demonstrating the model's capability to align with human expert judgments in identifying strategic defensive maneuvers during basketball games.

4.Experimental Results

The proposed model for the automated identification of defensive strategies in basketball has been evaluated for every rule that is used to filter the recognition of a change in a coupling of a defender-attacker assignment. Each rule has been evaluated on it's own, then a combination of each rule and finally all together to forge the final result.

4.1 Application of Rules and Performance

The Model's goal is to isolate the situations in a possession where a pick-and-roll takes place and recognize if the "switch" defensive strategy occurs. The elimination algorithm that excludes situations where a pick-and-roll is not observed follows the steps that we analyzed in the subsection (Recognition of Coupling Change) of the previous section (3). The rules applied are:

- Compare the previous Timestep's Defensive Assignments with the Current Timestep's Defensive Assignments (assignments_changed)
- Average Player Distance Threshold (average_player_distance)
- Recognition of a Coupling Change on a Ball Holder (ball_holder_attacker)
- Switch Evaluation through Window Timesteps (change_remained)
- Probability Threshold Between Window Timesteps (probability_threshold)

Before we analyze any results we need to mention that the number of "switches" in pick_and_roll plays counted by our analysis is 120 in 469 possessions of the basketball game. which means (120/469) *100 = 25.586% of the possessions are labeled accordingly that they contain a "switch" defensive strategy.

The core where every rule is applied on is the assignments_changed rule and from that point, we test each rule application on the assignments_changed rule, all the combinations of the rules and finally all of them together. The assignments_changed rule on it's own, without any other rules applied has a percentage match of 25.586%, exactly as the percentage of switches found by our analysis. This means it labels every possession of the game as a positive of "switch" occurrence. From that percentage we conclude that in every possession in the very complex game of basketball, players change distances from their corresponding assigned player. So, just by observing changing of assignments can lead us to nowhere but, as we highlighted, it is the core of our elimination procedure.

Below, in figure 6, we present the percentage of matching labels of each rule (alone) of the model's results with the ground truth labels of a basketball game that consists of 469 possessions (events).

Rule	Percentage Match
average_player_distance	27.931%
ball_holder_attacker	52.238%
change_remained	28.144%
probability_threshold	28.571%

Figure 6. Percentage Match of Distinct "switch" Recognition Rules.

Before we go into further investigation to combine rules together, we can already distinguish that the ball_holder_attacker's percentage match protrudes. With further analysis, the pick-and-roll action always involves the attacking player that holds the ball. So, our next step is to distinguish the case where in a pick-and-roll situation the defensive strategy applied is "switch" and not other strategies, e.g trap, help, cover, heads-out, push, over, under, check. The most important feature of the "switch" defensive strategy is that the defenders switch attackers that they guard. In contrast, all the other strategies involve temporary change in coupling assignments but the final coupling ends up the same as before the pick-and-roll action. That said, we are ready to proceed to the application of combination of rules and mostly focus on the combinations with the ball_holder_attacker rule.

In the following figures (Figure 7, Figure 8, Figure 9) we present the percentage of matching labels of the combinations of the 4 rules. The number of combinations is 6

Rule 1	Rule 2	Percentage Match
average_player_distance	ball_holder_attacker	64.392%
average_player_distance	change_remained	30.916%
average_player_distance	probability_threshold	33.688%

Figure 7. Percentage Match of the Combinations of the average_player_distance Rule with all the Rules.

Rule 1	Rule 2	Percentage Match
ball_holder_attacker	change_remained	59.914%
ball_holder_attacker	probability_threshold	65.884%

Figure 8. Percentage Match of the Combinations of the ball_holder_attacker Rule with the rest of the Rules.

Rule 1	Rule 2	Percentage Match
change_remained	probability_threshold	32.835%

Figure 9. Percentage Match of the Combinations of the change_remained Rule with the Last Combination Possible.

By combining in dyads the situation isolating rules, we already observed an increase in percentage match in every possible combination. As expected, the most notable combinations include the ball_holder_attacker rule.

Continuing, in Figure 10 we present the percentage match of the combinations of 3 rules.

Rule 1	Rule 2	Rule 3	Percentage Match
ball_holder_attacker	change_remained	probability_threshold	71.428%
average_player_distance	ball_holder_attacker	probability_threshold	74.626%
average_player_distance	ball_holder_attacker	change_remained	72.281%
average_player_distance	change_remained	probability_threshold	44.136%

Figure 10. Percentage Match of the Combinations of 3 Rules.

Concerning the combination of the rules that do not contain the ball_holder_attacker, we can notice slight increase in percentage match just by combining them. About the important combinations that contain the ball_holder_attacker we observe a smaller percentage match increase than the other combination and this is due to the fact that the number of changes of defender-attacker assignments on the ball holder is greatly smaller than the rest of the changes that don't involve the ball holder. So, as the number of cases we focus on has shrunk, the percentage match increases by less and less as we add more rules that eliminate the occasions we are not interested in.

The final result of all the rules combined is presented below in Figure 11.

Rule 1	Rule 2	Rule 3	Rule 4	Percentage Match
average_player_distance	ball_holder_attacker	change_remained	probability_threshold	76.332%

Figure 11. Percentage Match of All the Rules of the Automated Identification of the Defensive Strategy Model.

Combining all the rules together confirms the previous statement that the more rules added the smaller the increase of the percentage match we observe due to the fact that the number of situations we try to discern the "switch" action, keeps narrowing down for each rule we add up. The final percentage match of our model for an NBA game of basketball of 469 possessions and a total count of possessions with "switch" occurrences of 120 is 76.332%.

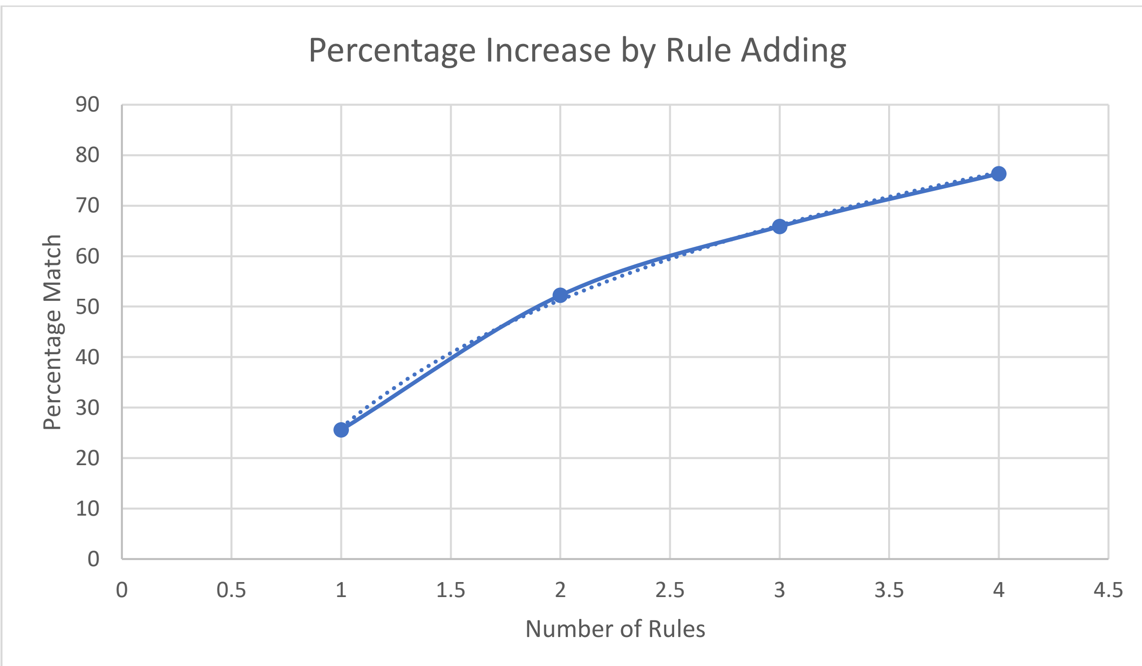

Figure 12. Line Chart of (Number of Rules/Percentage Match).

An extra detail, worth of mention, that was tested on, was the size of windows of timesteps used to quantize the timesteps of each possession. We used windows of 10 timesteps for out model and that number was decided by testing. We observed that using too small or too large sized timestep windows, was decreasing the percentage match of the model's "switch" identification due to the fact that we used a rule (change_remained) that if a change is observed it is checking the following timesteps of the window's size if that change remained. We also used a rule (probability_threshold) which checks if a change in assignment was legitimate enough if it passed a certain threshold of probability assignment by also checking the following timesteps of the window's size. These two rules work against each other for too small or too large sized windows of timesteps so, the size of the windows of timesteps narrowed down to 10.

4.2 Application of the Proposed Method for Basketball Analytics

Team sport strategy enables individual and group actions to be organized in order to produce collective, creative and unpredictable execution of the team's actions [20]. Development of analytical tools that can be used to predict the outcomes of different game strategies is a trending field of study in sports analysis. In the past, extracting the player and ball location was a cumbersome process that involved extensive video analysis. Availability of player and ball tracking data is enabling the researchers to develop such analytical tools.

Due to the ease of recording points, assists, and related goal-scoring statistics, most play-by-play datasets often focus on offensive analytics [21]. Therefore, the presented model in this thesis focused on the defensive tactics. Specifically, we focus on the defensive action against pick-and-roll (PNR), which is the most commonly used offensive strategy in basketball [22]. The proposed automatic identification model can identify a defensive tactic against PNR, known as switch, by considering features derived from player and ball tracking data. The model achieved an overall accuracy of 76.3%. Although, the current model was adjusted to identify only the switch defensive strategy, the methodology can be replicated to learn other tactics, provided labeled data.

The proposed model can be used for different applications of basketball analytics. For example, it can be used for match analysis to automatically retrieve a given defensive strategy from player and ball tracking data. After retrieval, the analysts can see how different opponents defend PNR, and analyse the relative effectiveness of different defensive strategies. Such information can be very useful to design strategies before games to thus providing competitive advantage through analytics. Data driven ghosting schemes are quite a useful and emerging tool for analysing the defensive behaviour of a team [23]. However, as of present such schemes are strictly based on tracking information and some basic contexts such as team roles, but lack crucial important contextual information such as defensive types and game period. However, contextual information regarding different defensive moves cannot directly be accessed from play-by-play data available. Hence, our proposal can be used as a method to embed additional contextual knowledge for player tracking data.

There has been a significant interest in the sports analytics research community generative probability models for predicting the outcomes of games. These models are predominantly based on the available play-by-play data [24] and encode the knowledge of analytical studies on basketball performance [26]. The play-by-play data that are collected for basketball matches do not encode many offensive or defensive strategies, but only the high-level events such as shots, passes, dribbles, chances and fouls. The presented tool to recognise defensive moves against pick and roll from player-tracking data is an example of a model that represents dynamics of a subset of players, can enrich game outcome prediction models that represent the match situations as a continuous dynamical system [12] or a graphical model [26].

In summary, the presented automatic identification methodology is an important step toward the provision of additional contextual information that cannot easily be accessed through play-by-play data. The presented model to identify defensive strategies can augment such play-by-play data, especially if coupled with further analysis on the outcomes of the plays to provide additional metrics about defensive effectiveness.

4.3 Limitations of the Current Methodology

There are many offensive plays in basketball other than Pick-and-roll, and many defensive strategies against PNR other than switch. The objective of this study is to illustrate the possibility to identify a specific defensive strategy in a highly dynamical team setting, where the team roles may change over the course of a given possession, by utilising the player and ball tracking data. This model is not an exhaustive representation of all possible offensive or defensive strategies, but an important contribution especially towards the community of basketball analytics researchers who have taken an active role to develop models of game progression dynamics.

Basketball player and ball tracking data are not available for free, but play by play data can be obtained from NBA.com. Our study was based on a limited player-ball tracking dataset (36,330 basketball possessions), which is freely available from STATS LLC, without any corresponding contextual information such as play-by-play data associated with the tracking data. We illustrate the possibilities that are present if data is made available freely. Availability of more data, along with the contextual information will expedite the progress of this field and widen the interest of academic researchers in the field.

Furthermore, the number of manually annotated data points (1000 out of 32,377 total data points) considered within this study are also limited. It should be noted that the data annotation is an extremely time-consuming task, which requires the analyst to go through the tracking data multiple times. This is worsened when corresponding video footage is not available nor when the contextual information is missing. Ideally, the dataset has to be labelled by an analyst, and availability of such labeled datasets is crucial for development of such automatic labeling algorithms.

5.Conclusions And Future Research

The objective of this thesis was to identify a defensive strategy in the game of basketball by utilising spatio-temporal pattern recognition using player and ball tracking data. A tracking dataset of player and ball trajectories in 32,377 possessions from nearly 630 basketball games in 2012/13 NBA season from STATS LLC was used in the study. The objective of this study was to identify one common defensive strategy (known as switch) used against a popular offensive strategy known as pick-and-roll, by considering different features from the play-and-ball tracking data. To develop a substantial annotated dataset, an analytical model was developed with a capability to automatically identify and label windows of timesteps that contain switch defensive plays. This analytical model that recognised defensive plays was based on five features: the distance between players, the average player distance, the coupling change on the ball holder, the "switch" evaluation through timesteps and a switch probability threshold. The model extracted raw location data and features derived from the location data such as (defensive and offensive players, ball possessing teams, player distances, player and ball distances, player defending probabilites) to construct the automatic identification model, which was able to identify "switch" strategies in basketball. The model produced a identification accuracy of 76.3%. While the current work considered only the "switch" strategy, there are many more defensive strategies involved with Basketball. Future research may also wish to consider alternative strategies for defending strategies such as pick-and-roll, and perhaps consider defensive moves against other offensive strategies such as pick-and-pop. Furthermore, alternative methods to label large spatio-temporal datasets would also lead to better outcomes, as compared to analytical method proposed in this thesis.

This study was inspired by the following paper [27] where Machine Learning to Automate the Identification of Basketball Strategies Using Whole Team Player Tracking Data was used. This paper used the same tracking data as our study for their research and created an automated identification model to label the data in order for it to be used for machine learning algorithms for classification. Their machine learning algorithm classification reached a maximum of 68,9% of accuracy. In contrast, our study reached an accuracy of 76,3% without using machine learning algorithms but only using various advanced, identification of basketball defensive strategies, rules to label the dataset. This outcome illustrates the possibilities that could arise by combining deeper knowledge into basketball strategies and the ability to interpret them into more complex identification models with machine learning algorithms to reach levels of accuracy that could be dependable to label datasets without the need of manual confirmation.

The outputs of this study illustrate the suitability of player tracking data to learn competitive game-related strategies employed in team sports such as basketball. Data-driven methods such as the automatic identification methodology presented in this thesis, can provide useful insights into game play. Furthermore, such models may enable the improvement of analytical software by providing additional contextual information related to defensive plays, in addition to the play-by-play data that are commonly available. Thus, facilitating advanced analytical solutions that has the potential to inform player development, coaching strategies and game specific tactics.

Future work may consist of further investigation of more dynamic and robust rules of isolation of the defensive strategies as well as more precise and informative data, also testing the method's accuracy on multiple different analyst labeled spatio-temporal datasets.

Our approach is a scalable method, it uses a relatively low amount of rules according to the complexity of the game of basketball. Adding advanced rules, such as player field position, player trajectory, shot clock time and game clock time good skyrocket the model's percentage matchings. After, reaching high amounts of accuracy, machine learning models could be applied to further analyze the capabilities spatio-temporal data analysis.

BIBLIOGRAPHY

[1] Sampaio, J.; McGarry, T.; Calleja-González, J.; Sáiz, S.J.; i del Alcázar, X.S.; Balciunas, M. Exploring Game Performance in the National Basketball Association Using Player Tracking Data. PLoS ONE 2015, 10, e0132894. [CrossRef] [PubMed]

[2] Bourbousson, J.; Sève, C.; McGarry, T. Space–time coordination dynamics in basketball: Part 1. Intra- and inter-couplings among player dyads. J. Sports Sci. 2010, 28, 339–347. [CrossRef] [PubMed]

[3] Lamas, L.; Junior, D.D.R.; Santana, F.; Rostaiser, E.; Negretti, L.; Ugrinowitsch, C. Space creation dynamics in basketball offence: Validation and evaluation of elite teams. Int. J. Perform. Anal. Sport 2011, 11, 71–84. [CrossRef]

[4] Santana, F.L.; Rostaiser, E.; Sherzer, E.; Ugrinowitsch, C.; Barrera, J.; Lamas, L.; Santana, F.L.; Rostaiser, E.; Sherzer, E.; Ugrinowitsch, C.; et al. Space protection dynamics in basketball: Validation and application to the evaluation of offense-defense patterns. Mot. Rev. Educ. Física 2015, 21, 34–44. [CrossRef]

[5] Marmarinos, C.; Apostolidis, N.; Kostopoulos, N.; Apostolidis, A. Efficacy of the “pick and roll” offense in top level European basketball teams. J. Hum. Kinet. 2016, 51, 121–129. [CrossRef] [PubMed]

[6] Courel-Ibáñez, J.; McRobert, A.P.; Toro, E.O.; Vélez, D.C. Inside game effectiveness in NBA basketball: Analysis of collective interactions. Kinesiology 2018, 50, 218–227. [CrossRef]

[7] Basketball for Coaches | Basketball Drills, Plays, and Coaching Tips. Available online: https://www. basketballforcoaches.com/ (accessed on 5 October 2019).

[8] Ángel, G.M.; Evangelos, T.; Alberto, L. Defensive systems in basketball ball possessions. Int. J. Perform. Anal. Sport 2006, 6, 98–107. [CrossRef]

[9] McIntyre, A.; Brooks, J.; Guttag, J.; Wiens, J. Recognizing and Analyzing Ball Screen Defense in the NBA. In Proceedings of the MIT Sloan Sports Analytics Conference, Boston, MA, USA, 11–12 March 2016; Volume 1001, p. 48104

[10] The 9 Rules of Switching Defense. Available online: http://team.fastmodelsports.com/2019/08/14/georgiosbartzokas-switching-defense/ (accessed on 18 December 2019)

[11] Gómez, M.-Á.; Battaglia, O.; Lorenzo, A.; Lorenzo, J.; Jiménez, S.; Sampaio, J. Effectiveness during ball screens in elite basketball games. J. Sports Sci. 2015, 33, 1844–1852. [CrossRef] [PubMed]

[12] Vaquera, A.; García-Tormo, J.V.; Ruano, M.A.G.; Morante, J.C. An exploration of ball screen effectiveness on elite basketball teams. Int. J. Perform. Anal. Sport 2016, 16, 475–485. [CrossRef]

[13] Marmarinos, C.; Apostolidis, N.; Kostopoulos, N.; Apostolidis, A. Efficacy of the "pick and roll" offense in top level European basketball teams. J. Hum. Kinet. 2016, 51, 121–129. [CrossRef] [PubMed]

[14] Wang, K.; Zemel, R.S. Classifying NBA Offensive Plays Using Neural Networks-MIT Sloan Analytics Conference. In Proceedings of the MIT Sloan Sports Analytics Conference, Boston, MA, USA, 11–12 March 2016.

[15] Football Player Tracking Football Tracking System. Available online: https://www.stats.com/sportvu-football/ (accessed on 28 November 2018).

[16] Zhan, E.; Zheng, S.; Yue, Y.; Sha, L.; Lucey, P. Generative Multi-Agent Behavioral Cloning. arXiv 2018, arXiv:abs/1803.07612.

[17] Chung, J.; Kastner, K.; Dinh, L.; Goel, K.; Courville, A.C.; Bengio, Y. A Recurrent Latent Variable Model for Sequential Data. In Advances in Neural Information Processing Systems 28; Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2015; pp. 2980–2988.

[18] Yue, Y.; Lucey, P.; Carr, P.; Bialkowski, A.; Matthews, I. Learning Fine-Grained Spatial Models for Dynamic Sports Play Prediction. In Proceedings of the 2014 IEEE International Conference on Data Mining, Shenzhen, China, 14–17 December 2014; pp. 670–679.

[19] Zheng, S.; Yue, Y.; Hobbs, J. Generating Long-term Trajectories Using Deep Hierarchical Networks. In Advances in Neural Information Processing Systems 29; Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2016; pp. 1543–1551.

[20] Lamas, L.; Barrera, J.; Otranto, G.; Ugrinowitsch, C. Invasion team sports: Strategy and match modeling. Int. J. Perform. Anal. Sport 2014, 14, 307–329. [CrossRef]

[21] Franks, A.; Miller, A.; Bronn, L.; Goldsberry, K. Counterpoints: Advanced Defensive Metrics for NBA Basketball. In Proceedings of the MIT Sloan Sports Analytics Conference, Boston, MA, USA, 27–28 February 2015.

[22] Lamas, L.; Junior, D.D.R.; Santana, F.; Rostaiser, E.; Negretti, L.; Ugrinowitsch, C. Space creation dynamics in basketball offence: Validation and evaluation of elite teams. Int. J. Perform. Anal. Sport 2011, 11, 71–84. [CrossRef]

[23] Le, H.M.; Carr, P.; Yue, Y.; Lucey, P. Data-Driven Ghosting using Deep Imitation

Learning. In Proceedings of the MIT Sloan Sports Analytics Conference, Boston, MA, USA, 3–4 March 2017.

[24] Štrumbelj, E.; Vraˇcar, P. Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. Int. J. Forecast. 2012, 28, 532–542. [CrossRef]

[25] Basketball for Coaches | Basketball Drills, Plays, and Coaching Tips. Available online: https://www. basketballforcoaches.com/ (accessed on 5 October 2019).

[26] Oh, M.; Keshri, S.; Iyengar, G. Graphical Model for Basketball Match Simulation. In Proceedings of the MIT Sloan Sports Analytics Conference, Boston, MA, USA, 27–28 February 2015.

[27] Changjia Tian.; Varuna De Silva.; Michael Caine.; and Steve Swanson. Use of Machine Learning to Automate the Identification of Basketball Strategies Using Whole Team Player Tracking Data. Available online: https://www.mdpi.com/2076-3417/10/1/24 (accessed on 18 December 2019).