

Big Data in Sports: The Downside of Surveillance and a Player's Right to Privacy

Michael Kim

Cypress High School
Cypress, California

Abstract: Big data in the sports industry which has led to many concerns over a player's right to privacy. The use of wearable technology has made it possible to collect video surveillance and biometric data in vast amounts. The excessive video and biometric monitoring of athletes has led players to voice their concerns while revealing the legal limitations in protecting their right to privacy. In an age of social media and technological advancements, the law cannot keep up with big data. To gain more insight into this topic, this paper aims to explain how big data is used in sports while highlighting the dangers of big data usage. HIPAA laws are not enough to protect players due to gray areas in definition and bureaucratic processes. In reviewing the current literature, it has been revealed that there is a grave need for new laws to address privacy concerns in the world of big data and sports.

Keywords: big data, sports, wearable technology, privacy laws, biometrics.

I. INTRODUCTION

In recent years, many of the biggest sports leagues have taken on big data to improve their teams' performance. Through the use of biometrics via wearable technology, analysts are able to predetermine the ability of athletes based on statistics from current and previous seasons. The data for biometrics is found using Zephyr Bio-Harness, a device used as a marker of training intensity, detecting a player's heart rate, heart rate variability, movement, breathing rate, core temperature, and acceleration.¹ The trauma-monitoring stickers help determine if a player is at risk of having a concussion. Motus, MotusBASEBALL, and MotusQB, are also some of the many devices used in training which provide information on injury prevention, effective training methods, and injury rehabilitation¹. The goal is to find the best possible way of training athletes so that they can improve themselves in areas of weakness allowing for teams to be better prepared for games including in-game situations that they are training for.

Big data has also allowed for technologies such as Video-Assistant Referee in soccer which is used to help referees make the right calls for plays that may be difficult to judge, including overcalls on the field.² A closer look will determine the final call through video replays. This however, has left a negative impression on the audience who believe that the technology has ruined the flow of the game due to an increased number of penalty calls. In soccer, the technology made its first World Cup debut this past summer as the most penalties were ever called in the history of the World Cup.² In baseball, Pitchf/x which has been installed in every MLB stadium to help determine if a pitch was a strike or a ball. In basketball, there are six cameras all around the court in order to capture each and every move of a player. This allows the team to analyze each player's moves and determine the weakness and strengths of all players. For example, the Rockets from the NBA have used the new technology to their advantage to shoot many three-pointers as well as run many fast breaks.

II. WHAT IS BIG DATA?

It is important to note that big data consists of four main characteristics known as the four V's. The first V is volume, which is the amount of storage that the data takes up. The second V is velocity which is the speed at which data is being created. The third V is variety or the amount of sources used to create that data. And last but not least, veracity, is the fourth V, which refers to the accuracy of the data found, thus, how valuable the data is. Big Data in sports is collected not

only through wearable technologies but also through cameras. The velocity of big data is the mass of the amount of data being created as well as how fast the data is being processed, and in soccer clubs such as Arsenal, there cameras that records data at speeds of ten data plots per player per second which is 1.4 million data points per game.³ This means that coaches no longer have to rely solely on game replays when analyzing each play; instead, wearable technology with a GPS tracking system tracks every movement the players make.

III. THE CHALLENGE OF BIG DATA MANAGEMENT

Big data is not merely used for sports but is currently being used in a variety of fields from medicine to education.³ The overarching issues related to big data are as follows: data degree, data pace, data range, data charge, data volume, and data veracity. The degree of the data refers to the amount of data processed which is higher in the case of unstructured data. Data pace refers to the amount of data being produced at an alarming rate which traditional power structures cannot handle. With information constantly in motion due to social media and e-commerce, new power structures must be able to handle at the rate in which data is produced. Data range refers to the type of data from text to video to audio to images with 80 percent of data being unstructured.³ Data charge refers to the high power data organizations such as Google and Apple, organized that tend to have a large gap in understanding between the management and tech specialists. The volume of the data has moved from Exabytes to Zettabytes to Yottabytes.³ Lastly, the validity of the data, also known as veracity, is a huge area of concern since a user who is male can easily identify as a female online. Users can easily manipulate the data behind a computer screen.

To address such challenges regarding big data Hadoop was developed by Doug Cutting who worked for Yahoo at the time. Hadoop was spun off of Nutch, an open source project, to provide data that provides meaningful data for analytical purposes.³ Hadoop stems from a Java framework capable of managing and sorting through immense amounts of data. It runs on thousands of nodes that process Terabytes of data in an efficient manner even when a node fails to operate.³ Map-Reduce allows Hadoop to breakdown the data into smaller components and only major companies such as Google, IBM, and Yahoo can handle this software whereas smaller companies cannot. There exists several machines that help run Hadoop which do not share memory with each other but is able to organize all the data presented and distribute them onto different servers.³ Map-Reduce, which can be written in any language, is used to integrate results from various servers and sent back collectively. However, challenges regarding low efficiency and energy usage is a common problem due to frequent checkpoints and energy costs totalling 23 percent of total operating costs.

IV. TACTICAL ANALYSIS IN SOCCER

In soccer, a tactic is used to outplay an opposing team and ultimately, win the game. This means teams have to change the way their players play each game against the numerous opposing teams. A shift in the game such as the substitution of a player, no matter the team, will also impact how each strategy is made. Prior to big data, to study game tactics, coaches or scouts must manually watch all games from the beginning to end. The data is often found to be very unreliable since much of the data stored is based on personal experiences and must be frequently changed out since opposing teams are always adjusting their tactics.⁴ Manually analyzing the data is very time consuming and the needed data is often mixed in with significant amounts of irrelevant data.

In soccer, monitoring technologies were not allowed but FIFA has recently changed this rule to not only allow big data during training but also during games.³ Germany's success at the 2014 World Cup can be attributed to this. The team was able to cut down their possession time average by 2.3 seconds which allowed them to run faster and figure out the defense more efficiently.³ The German Football Association work with SAP to create Match Insights, a technology powered by Intel processors to sort through immense amounts of data to understand key moments in their gameplay. To note how much data is produced, the Germany coach cited that in just 10 minutes, playing with three balls, 10 players can create 7 million data points for analysis.³ The technology allowed the team to adjust their training understanding each players' strengths and weakness to prepare for each match.

Some points of analysis in soccer includes the control of space or the amount of surface area that a team uses to attack the opponent. Analysts help determine the spacing between each player on the field to figure out which passes would be the most effective. Space control can also be determined with a voronoi diagram in which the distance between each individual player is measured to calculate the controlled amount of space. In addition, team centroid measures how the geometric center of the positions of a player can be analyzed to determine the behavior of a team. Some possible factors

that can affect the team’s behavior are changes of inter-centroid distance due to pitch size variations and goal shots.⁴ Teams use Approximate Entropy or ApEn, a non-linear time-series measurement technique to quantify the regularity in time series data to understand the centroid of their group when compared to other groups.⁴ ApEn, has also shown that a team’s centroid increases during a situation of inferiority in which one team has an advantage over the other in terms of the number of players.

Currently, there exists an incredible amount of datasets encoded using Extensible Markup Language (XML) ranging between 86 and 300 megabytes.⁴ Video data stored amount to 400 gigabytes of tracking data in one complete Bundesliga soccer season. This does not include physiological data nor event data. The issue is how to organize such large amounts of data which is later categorized as structured, semi-structured, or unstructured data. Video data and text messages are placed in the unstructured data category. In soccer data variety includes position, video, skill performance, fitness, health records, and even crowd data. Big data makes it possible to combine all the data present into points of analysis. The velocity or the speed at which new data is being generated can vary from real-time streams, physiological and positional data, and data which requires more time such as notational analysis during the training including games.⁴ Figure 2 depicts how the soccer data should be organized which begins with data collection then storing the data and finally, processing the data. Extracting relevant information and merging the provided datasets into a predictive model is the goal in soccer analysis.

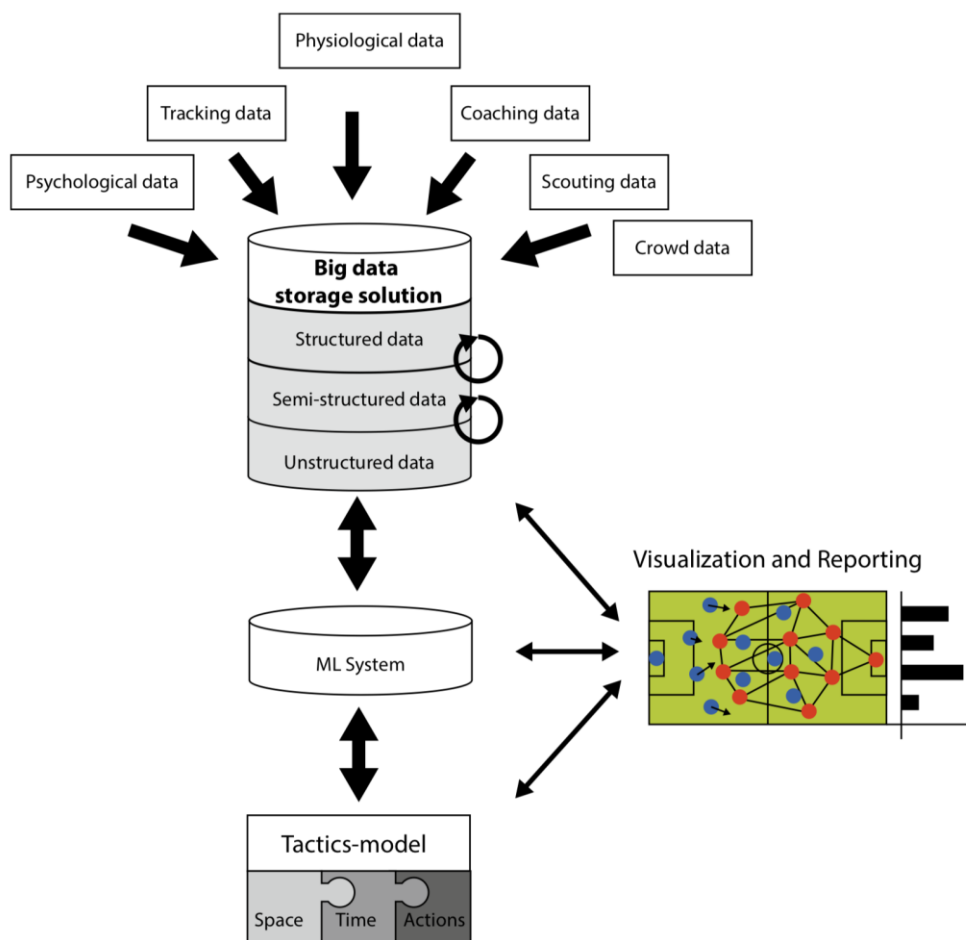


Figure 2.

Source: Technological diagram of big data in elite soccer adapted from “Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science”

V. TRACKING PHYSICAL CAPABILITIES THROUGH GPA IN THE NBA

In the NBA, players are monitored at all times during training sessions as well as during each game. Each step made by the players are tracked which adds up to large amounts of data that are collected for each player which allows for trainers

to see how fast the athletes can accelerate or decelerate as well as how forcefully they can jump and then land, afterwards. This practice helps trainers to find a way to prevent injuries among athletes and determine whether or not they are giving their full effort when practicing.⁵ Players are tracked closely using GPS as every movement the player makes is recorded 25 times in one second and 72,000 times during each game.⁵ For instance, with such close monitoring, coaches are able to tell when a player is fatigued prompting them to train the athlete to work on increasing their stamina. Players' movements are also recorded off the court, tracking their sleeping schedule and eating habits, to find correlations between a player's habits and their positive or negative performance on the court.⁵

With the constant tracking of players off of the court, some players see this as a violation of their privacy because they have no freedom outside of the playing field. They feel immense pressure as a result of such close scrutiny. Another reason for concern is that their personal data could be leaked not only to the public but other teams, hurting their reputation and strategy for success in the league. More specifically, health issues of a player could prevent players from receiving contract negotiations and scouting opportunities which can ultimately hurt their status in the league. Due to these concerns, many NBA players have started to disregard big data analytics offered by the coach and have instead hired their own analysts to get the information they want and need. As a result, an increasing number of professional athletes have joined the movement called "The Quantified Self Movement" wherein players hire their own personal analysts to control for privacy.⁵ For instance, Kevin Durant hired a personal data analyst and statistician named Justin Zomelo in 2009 to improve on areas of his techniques which he personally wanted to work on. This, ended up in Kevin Durant being able to reach phenomenal numbers as he was able to shoot 50% from the field, 40% from the three point line, and 90% from the free throw line, showing how Durant had optimized his best shooting spots to become a lethal offensive player by focusing on what he wanted to improve on instead of what his coaches recommended.⁵ And although these data analysts have helped many players improve, team executives are calling their data "a total waste."⁵

VI. THE IMPLICATION OF BIOMETRICS

Biometric data is at the core of big data in the world of sports. However, like all other technology, extensive data tracking can be compromised when in the wrong hands. To accommodate these problems, teams have sought to receive the consent of players in order to extract appropriate data from athletes.

Currently, there are no federal regulations to justify the use of big data of athletes although HIPAA regulates some of the biometric data of players. Due to waivers offered to sports teams, players are often exempt from abiding by government regulations as biometric and biomechanical data are not considered personal health information (PHI).¹ However, the unprecedented amounts of data available today raises concerns over the privacy of players. Routine health check-ups such as heart rate, blood pressure, and ECG readings are considered PHI, but metrics such as speed and distance are not considered "personal" data by law. In this regard, HIPAA, also known as the Health Insurance and Portability Accountability Act of 1996 (HIPAA) exists to protect patients from misuse of PHI while addressing limitations on healthcare insurance coverage.

The greatest challenge for athletes is that not all athletes' biometric data (ABD) are considered PHI. An ABD is only protected depending on whether or not it is for "medical" use. Some ABD are considered to be "intellectual property" and not medical data because it is used by coaches and trainers to prepare for a game. However, in another light, much of the ABD can be considered personal medical information which can be later used in case an athlete is injured. Also, injury prevention can be considered information used for medical purposes but it is still vague in definition to qualify as "medical" information. When passing HIPAA, Congress appointed the Department of Health and Human Services (DHHS) to implement regulations that providers must abide by when submitting healthcare information electronically.¹ Currently, the waivers placed on teams by DHHS means that sports teams are not considered healthcare providers. This means that teams can only be partially accountable to HIPAA depending on how the medical information is handled by the medical staff and whether or not doctors are hired by teams for medical purposes.

In this regard, teams may change how they employ medical staff and transmit and store the data in order to avoid HIPAA laws and have full access to player information. This is why many players' are taking matters into their own hands to protect their privacy. As for DHHS, they fear that HIPAA will be treated as a disclosure law and not as a protection law, which is what it was originally intended for - to protect patient rights. As big data and ABD continues to grow, new laws must be created to so that HIPAA is no longer misunderstood nor misused and thus become ineffective in the future.

VII. HOW PLAYERS AND TEAMS PROTECT THEIR PRIVACY

The irony of data protection in sports is that it benefits all parties involved. Teams and coaches want to protect ABD from getting out to ensure that other teams don't ever find out about their athletes which would minimize the team's competitive edge.¹ Teams have also started to encrypt data and de-identify themselves but this does not mean that teams cannot be hacked.¹ For example, in baseball, one of the employees of the Cardinals hacked the Astros database which allowed them to gain information on every one of their players which impacted both of the team's gameplay. Teams now change the passwords to the files storing the ABD, having other third party experts check the security and encrypting the data so that if it is hacked, the hacker cannot identify one player from another unless it is a very small team.

Aside from game strategy, players must protect their information to protect their data from release to the media but also because it is used against players during contract negotiations. In effect, players hire account administrators to handle all their personal data to gain more control over internal permissions and for the players themselves to decide which members of the team can use the data. This includes the extent to which the data can be used to control whether or not managers re-sign a player depending on the data they have stored about them. In regards to the wearable tech industries itself, they need to protect ABD to prevent lawsuits and sustain their business models where teams and players rely on privacy to gain a competitive edge.

VIII. CONCLUSION

With the rise of big data, the world of sports will never be the same. It has completely changed the way the game is played and the business is run. Unfortunately, with such potent technologies, the law has not caught up with the potential risks in big data use, most notably the player's right to privacy. The present literature has revealed that HIPAA alone cannot properly handle the issue anymore, a law intended for protecting patient's rights not player's rights. In addition, the issue of privacy is not merely a business one that affects transfers or game strategies, but how far is too far when surveilling players? ABD collection means constant surveillance in daily activities, even off the field, in many instances, young players are not fully aware what they are agreeing to when they join a team. With regards to big data and sports, it is critical for federal agencies to determine what is legal and what is not before the issue becomes a bigger problem than it is today. In our technologically-driven society, privacy issues will constantly draw concern, no matter the industry, therefore this is a matter that cannot be overlooked but addressed with greater efforts made by all parties involved.

REFERENCES

- [1] Osborne, B and Cunningham JL. "Legal and Ethical Implications of Athletes' Biometric Data Collection in Professional Sport." *Marquette Sports Law*, vol. 28, no. 1, (2017).
- [2] Lalin A. "How soccer's video assistant referee (VAR) influences belief in human referees' competence." *Behavioral Scientist*. July 31, 2018. Accessed November 21, 2018. <http://behavioralscientist.org/how-soccers-video-assistant-referee-var-influences-belief-in-human-referees-competence/>
- [3] Gagandeep J and Gurpreet S, "Big data diagnosis enhances innovative winning formula in the world of sports," *Indian Journal of Science and Technology*, vol. 10, (September 2017), 1-10, DOI: 10.17485/ijst/2017/v10i35/118953
- [4] Robert R and Daniel M, "Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science," *SpringerPlus*, vol. 5, (2016) DOI: 10.1186/s40064-016-3108-2
- [5] Andrew B. "Big data, sport, and the digital divide: theorizing how athletes might respond to big data monitoring," *Journal of Sport and Social Issues*, vol. 41, (2017): 3-20, DOI: 10.1177/0193723516673409