

## Big data analytics for smart sports using apache spark

Jacob Reece, *Carson-Newman University, jwreece@cn.edu*

SeongYong Hong, *Carson-Newman University, shong@cn.edu*

### Abstract

On most collegiate sports teams' injuries are plaguing rosters. The issue with the injuries on a football team is the amount of them that come from training overloads and lack of recovery from their training sessions. Coaches are not being able to take their team to their full potential and athletes are not being able to perform at their full potential as well. We are getting direct real-time data from wearable technology called Whoop. It collects the strain that the athlete goes through, their Heart Rate Variability (HRV), Respiration efficiency, and so many more things that could help improve the health of the athletes. Therefore, this paper will present the research proposal that the use of Big Data Analysis will be able to incorporate predictive analysis methods. With this proposal system in place coaches, athletes, and athletic trainers will be able to help prevent injuries and improve the overall performance of the team. By retrieving the real-time data from the Whoop band, we will be able to make real-time decisions and create correlations between the collected data and the performance of the athletes. The purpose of this proposed research is to create a system that will be beneficial to all types of sports teams, athletes, and university athletics budgets with big data analysis.

**Keywords:** Big Data Analytics, Smart Sports, Apache Spark, Metadata Analysis

### Introduction

Big Data can be defined as a field to be used to analyze, and extract information from complex data sets. The way that these data fields are interpreted, or dealt with, is through data-processing applications like SQL or in our case, Spark. On bigger schools' campuses, and most sports teams across the country, are already implementing the use of Smart Sports to improve the competitiveness of their sports teams. It also is being used to help improve the health and performance of their athletes. There are many different branches that pertain to Smart Sports. For example, there is the use of GPS systems that display just the distance traveled by whoever is wearing the device. There are applications that can be downloaded that track the movements of the athletes and the user must input their own daily routines, rather than just letting the application track it yourself. Collegiate sports, in my case the Carson-Newman University football team, is a perfect example of how Smart Sports could be implemented. Injuries from training overloads and lack of recovery between training sessions can be detrimental to the athlete's body without proper supervision or in our case, Smart Sports. This can become a hassle on the training staff and the strength and conditioning staff when it comes to preparing the athletes for competition/practice. In a specific case, one of the athletes on the football team pulled his hamstring during a week of practice right before a big game. This injury kept him from being a deciding factor in the game that weekend and you could possibly make the argument that if he would have been able to play then the outcome of the game would have been in favor of his team. What if I were to tell you that his injury could have been prevented? The purpose of this research project is to describe the implementation of predictive big data analysis methods to a collegiate sports team, making it "Smart" Sports. Athletes will be wearing an IoT device that tracks information that is important to the

day-to-day life of a collegiate athlete. This IoT device, Whoop band, can track your recovery based on the amount of sleep you get on a nightly basis, and your daily strain which is based on your daily activity/workout. The Whoop band tracks all of these by the electrocardiogram (EKG) that is implemented into the band itself. The most essential thing we can take away with the data being collected is not creating something out of the data that is collected but more so as gaining a deeper understanding of the why behind these types of injuries are plaguing the Carson Newman football roster. The major components of the design for the implementation of Smart Sports include the software used to make correlations between the collected data via the Whoop band. Using Spark software to turn the collected data into virtualizations so that the correlations can be seen on graphs, and on a database to store the information collected from the athletes. The transition of the data from the Whoop band will come from the application that the Whoop band constantly uploads to the Cloud server. From there we will be able to take the collected data in a PDF document format and enter it into an Excel sheet so that the transition from the Whoop application is easy into the database. Then we will be able to use predictive methods and analysis methods of the big data to make predictions when an athlete is near injury or if there is any correlation with the type of injury that the athlete has sustained from training. Spark open-source software being used will be available to the training staff and strength and conditioning staff, allowing them to be able to access and see which athletes are near the injury thresholds. The software and development of our project have real-time been done before but never to the scale of which I am conducting currently. I am connecting different software tools to illustrate to the sports team staff that it could be possible to help improve collegiate sports right here on campus and to improve the health of the athletes.

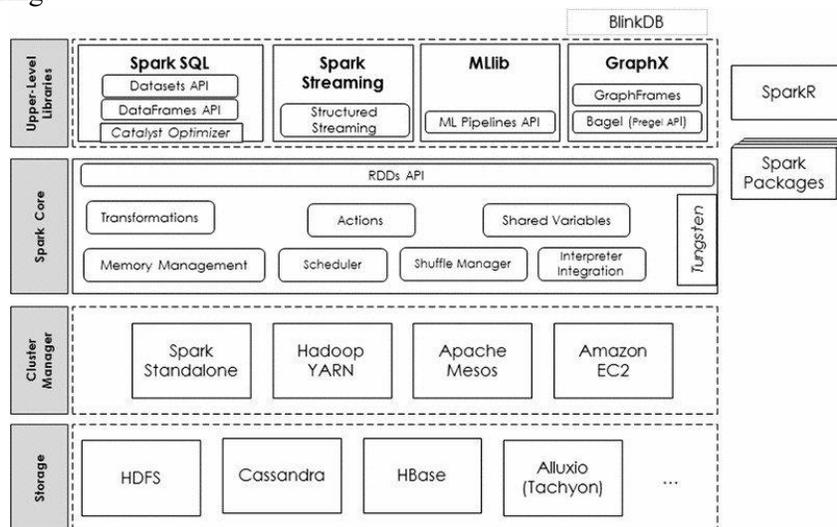
## Problem Motivation

I am currently a senior here at Carson Newman University and a member of the football team. In my four years as a member of the football team, I have experienced my own injuries as well as seen a large number of my teammates experience injuries as well. These injuries have sabotaged our season leaving us with a sense of disappointment that we may leave the field with regrets and unfulfilled goals. Now, I do know that football is a contact sport and that you cannot really prevent injuries when each practice is filled with intense physical contact. However, you can prevent non-contact injuries from a predictive analysis method fueled by Whoop real-time data. This past football season in the spring of 2021, we only played one game out of a five-game schedule, due to injury, and we also had to opt out of the rest of the season. Our head coach made a decision that would hurt some people's feelings but, in the end, he saved our team from an imminent demise. The current roster of the football team consists of up to 127 guys, and during this past season, we had fifty-one guys who were hurt and could not play. Roughly forty percent of the roster was injured and to play collegiate football at a high level you cannot have roughly forty percent of your roster inactive. After witnessing this event occur, it registered to me there must be a way to protect us. Recently, wearable technology is something we have incorporated into our day-to-day operations as a football program. The coaching staff has chosen to go with another wearable technology, GPS, to try and gather more data on what a Carson-Newman's football player's body goes through during typical practices and games. GPS is only worn by the athlete during practice and games. The GPS technology we currently use does not even provide real-time data. The current technology we use is charging the university \$100,000 for each team we want to analyze the data. To access the data, we need to upload it to the GPS Catapult server and that is after the fact. So, relatively speaking the data being gathered is not entirely useful when trying to make predictions, correlations and to show trends amongst athletes. For example, this past spring if we were incorporating the predictive analysis during our football game, we would not have been able to prevent any injury with the data we were collecting. This is where the wearable IoT device, Whoop band, comes into play. As far as gathering data goes the Whoop band never leaves the athlete, to where its counterpart GPS is only worn during games and practice. Whoop provides more useful real-time data and is the perfect candidate to apply this real-time data into a predictive analysis method to get results that can prevent us

from being in the same situation as we were in the Spring of 2021. The motivation behind this research is fueled by personal experience on both sides as an athlete and a data analyst. With the implementation of Whoop and predictive analysis, it will not only lower the number of non-contact injuries but also improve the overall performance of the athlete. In our opinion, to solve this problem of noncontact-injuries we need to incorporate a predictive analysis method using the K-Means algorithm. By using this algorithm, we will be able to successfully group together and show a direct visualization of the correlation between the over-training and the injuries occurring on the football team. With our experiences as current collegiate athletes, I think that this problem is something I have experienced on a firsthand basis and that with our knowledge of data analysis as well as the capabilities of Whoop technology that we could indeed implement Smart Sports on the Carson-Newman campus. If a system were like this in place before this past spring football season, then maybe I could have been able to play my senior year.

## Literature Reviews

We introduce an overview of Apache Spark project and Spark’s main components. Apache Spark is a data processing framework that can quickly perform processing tasks on very large data sets and can also distribute data processing tasks across multiple computers, either on its own or in tandem with other distributed computing tools. Apache Spark consists of many high-level libraries that can be compiled on many cluster managers. In figure 1, it shows in detail just how these high-level libraries are broken up into cluster managers, upper-level libraries, storage, and even the Spark core. For Spark, the Core category lays the foundation of what makes up Apache Spark (Salloum, et al, 2016). It provides the basics like large datasets and even includes a simple programming interface for the purpose of compiling massive datasets. This has been purposely built on top of Spark’s core for us to be able to implement libraries that are capable of machine learning.



**Figure 1: Architecture of Apache Spark Stack**

Through the research, we have conducted revolving around Smart Sports was quite surprising in how little there was to find about the topic or its implementation. While extensively researching smart applications, we discovered a substantial number pertaining to the different categories of research on Smart applications and how they implemented predictive analysis methods within those applications. Smart Sports applications will allow you to do the following: predict injuries, prevent training overloads, and improve the overall performance of the athlete on the field. Carey, et al. (2017) looked at the predictive modeling of the training

loads and their correlations to the type of injury. They found two correlations within their research; one being that the performance of their athletes tended to improve as more data was being processed and that the performance of their players did seem to improve as they used the predictive modeling to monitor the health of the athletes closer. One of the challenges that they faced was that there seemed to be limitations on the implementation of predictive modeling on a single sports team. They also were limited by the wearable technology that they were using and were unable to collect real-time data. In their case, they brought to my attention that we could see a correlation between injury and if the athletes were being overtrained. Inside their research, they also made it obvious that with the type of data we are collecting that we could not predict injuries that occurred from direct contact. We will only be able to prevent injuries from non-contact injuries, like pulled hamstring and pulled quad injuries. This is something I will explore later in my own research but a very interesting point of their research. Harms (2018) looked at this from the perspective of measuring the performance based on the data being collected by the Whoop band. Harms looks at it from the perspective of a Coach or Strength coach to show how important wearable IoT devices specifically Whoop are when trying to improve the performance of the athlete/team. He also discusses the Impact of Whoop technology on the collegiate sports world from a Coach or competitive standpoint. This project is direct visualization of how we can show this predictive modeling technique can be used in a real-world setting. He used the data to show correlations between the performance of the player on the playing field and the data collected by the Whoop band. With my research, I will take this a step further and contribute to their study by showing not only does the data being collected correlate to the performance of the player but also the health of the player. Most of these injuries that are plaguing Carson-Newman's football team can be prevented if the real-time data is used to its full potential. In, Thornton, et al. (2019) describe how important it is we as data scientists and people apart of the computer science industry must communicate the necessary conclusions we find to, in my research case, the coaching staff and medical staff of the Carson-Newman football team. In this case, it could be that we are monitoring the data being collected during a game and that we are confident in the data being collected we can give an appropriate suggestion to the football team staff to decide. In a separate implementation with Asri, et al. (2019) they lay the groundworks for the prediction system being used by a collection of real-time data. They also discuss the importance of how that while using predictive modeling we can prevent people from injuries and perhaps even sickness which leads to the overall success of the team. Another point that their research makes is where they also implement the K-means algorithm that can help you cluster the results of the real-time data being collected. They also describe how to have conceptual data modeling and implementation by using big data management tools like Spark. I would like to expand the models they used to so that we could help predict specific injuries based on the data collected. I would also use the way they applied the data that they gathered to their predictive algorithms. Rajšp, et al. (2020) discuss the importance of Gamification to Smart Sports. In their case, the version of gamification they use is creating a mobile application that is a virtual sports trainer to encourage people who are not collegiate athletes to stay fit and motivated. I soon learned within our research that to replace the old wearable technology, that it was going to take the use of Gamification. Gamification is a marketing technique to encourage a style of engagement with this new product, Whoop. Sikka, et al. (2019) introduces how that Big Data Analytics is taking over the Sports Medicine world and that they can use the real-time data being collected to make game-time decisions or cost-efficient decisions when pertaining to the athlete or program. The analysis of this real-time data is critical when pertaining to the asses if the person is injury prone and can even measure the performance of the athlete. Dhar (2014) presents the argument for a need the need to incorporate this type of predictive style system. In his research, it shows the capabilities of having a "predictive analytics" system from the health care point of view. We live in such a data-intensive environment that enables the health care and sports teams the advantages to help prevent and protect their patients/athletes. His research is leaning more towards health care only situations, but we can take the same type of system to incorporate it into sports. In most cases, sports teams are working closely with health care teams, so it would be beneficial to both (Armbrust, et al. 2015). We learned from their research paper how to look at relational data in

Apache Spark. They describe the architecture of Spark's API and how we can use it to implement machine learning algorithms. For my research, the research that they conducted is incredibly important because my research needs to be continued. The machine learning capabilities that can be implemented through Apache Spark is going to help benefit my research in finding a clear solution to my problem. The analysis that you can conduct inside spark is also discussed within their research and helps my research along the lines when trying to see correlations amongst metadata. Salloum, et al. (2016) discuss how we can use data management tools such as Apache Spark to conduct Big Data analytics with our own data. Within their research, we can see how capable Spark's engine is when we are preprocessing our information amongst the clusters. In addition, Shoro and Soomro (2015) also discuss the perspective of we can use Apache Spark to incorporate data analysis into our own research. Both previous literature reviews show that Apache Spark is the future of Big Data analysis methods. Ray (2015) is an incredibly interesting viewpoint of the architecture of IoT (Internet of Things) for smart sports. It is here that we can see the architecture in which IoT devices fall when collecting their information. For my research, the Whoop band falls into the category of the "Sports Application Layer" and with that being established hopefully we are able to link the application layer to the "Sports Management Layer". In the management layer, it is where the analysis is conducted and where all the virtualizations are located as well. Chen, et al. (2017) and Ismail, et al. (2020) both discuss topics that are different than mine but what we can take away from their research is their prediction analysis and algorithms used when making decisions. They might have different uses for their analysis methods or algorithms, but we can still implement their algorithms and analysis methods into my research. Unfortunately, the data we are trying to process is limited, however, this paper will supply information and evidence that this system is made for the future. Liu, et al. (2017) discuss innovative applications working toward a Smart Campus, and from this is where I came up with the idea to incorporate my version of Smart Sports application to be implemented on collegiate campuses. Smart Sports would just be another branch to a Smart Campus which would help give the ability for colleges to mine their athletes' data inevitably improving their performance. In Gaber, et al. (2019) have found certain data mining techniques that we are going to be able to implement to our research soon. IoT devices are providing data that can be used for the better good of the Athlete's health, and with the techniques used along with in their research, we are able to implement them into our own collected data. Hong (2009) has tried to support various access paths on web images with customized feedback according to Internet user's preferences. They have applied web mining along with data mining techniques. Also, Hong, et al. (2016) have tried to analyze data collected from current IT learners who have already studied various IT subjects to find the IT learners' learning patterns. Kalantari (2017) have found that with the future trends that the implementation of some form of wearable devices, such as Whoop, will be implemented eventually. So why not get ahead of the competition and implement a system as such so that we can improve our athletes on all fronts. Figure 2 is a representation of how the model will look after it has been fully implemented into a campus environment. It shows you how the real time data has been collected by the Whoop technology and is transferring the data to the memory of the mobile device to all the way to Spark's management tools.

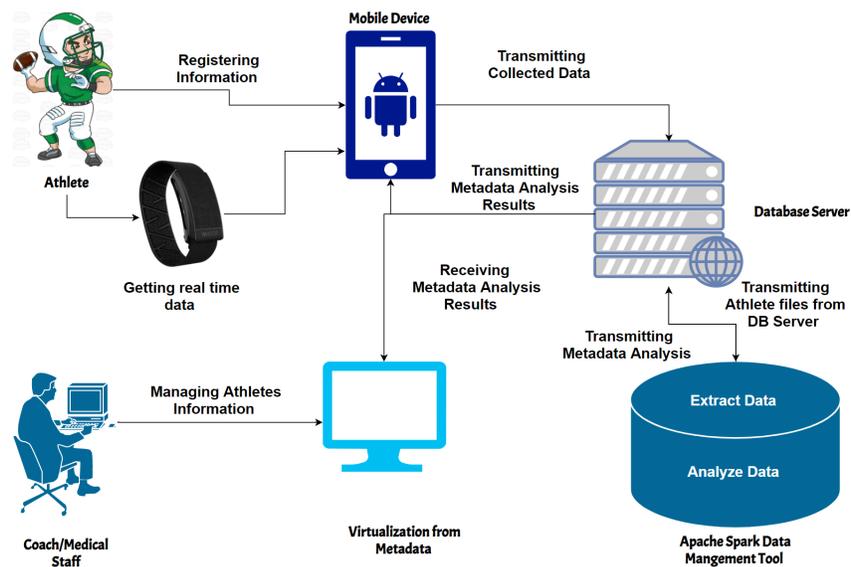


Figure 2: A Design of Technology for Smart Sports

### Research Methodology

In 2010, the University of Alabama spent \$1.96 million, and the University of Auburn spent \$852,477 on medical expenses for their athletes. Most colleges across the world cannot even afford a fraction of what these two schools paid for their athletes’ surgeries. The solution is not trying to eradicate injuries from sports completely but more so trying to use data analysis to help prevent some of these injuries from happening. Now, I know that not all injuries are preventable but for the ones that are, preventing some injuries that could protect the athlete and save the school money is worth pursuing. Implementing Smart Sports on college campuses would help improve the overall performance of the team and eliminate injuries that could have been prevented. Most college athletic departments must set a budget for each sports team, which could be used to purchase the necessary equipment for the steps of implementing wearable technology like the Whoop band. Carson-Newman University and the athletic department are currently already incorporating wearable technology into the sports team’s budget which means there would not be any exclusion or requirement for the athlete to make the purchase on his or her own accord. According to sports team information, Carson-Newman University is incorporating \$100,000 already into sports teams’ budgets for wearable technology. Student-athletes are our target group here, so they would be required to always wear the Whoop technology. That includes that the athlete is to wear it during class, to practice, to sleep, or to weights. In figure 3, describes a Big Data analysis for Smart Sports on collegiate campuses that allow us to help improve the athlete’s day-to-day life and help prevent injuries from occurring based on the real-time data being gathered. The athlete must always wear the Whoop band and send their monthly PDF document from their Whoop application (iOS or Android) to the sports medical team email so that the data can be implemented into the database that will eventually be extracted and analyzed using predictive modeling. The proposed prediction system is using wearable technology and your mobile device to upload and gather real-time data from the athletes’ activities throughout the day/night. The application will gather the main contribution from the athlete such as their height, weight, and position to use later in the advanced database server. While the Whoop band itself will record the real-time data such as Strain, Recovery, and Strain levels. It will then supply the information to the database server. After all the necessary information

has been recorded, we will then incorporate the database server into the Spark management tool to start the predictive modeling outcomes. In this research paper, we will also be looking at this through the lens of the predictive analysis method to help improve the overall performance of a sports team to reduce the overall risk of getting injured and check to see if there are any correlations between the collected data to the injuries.

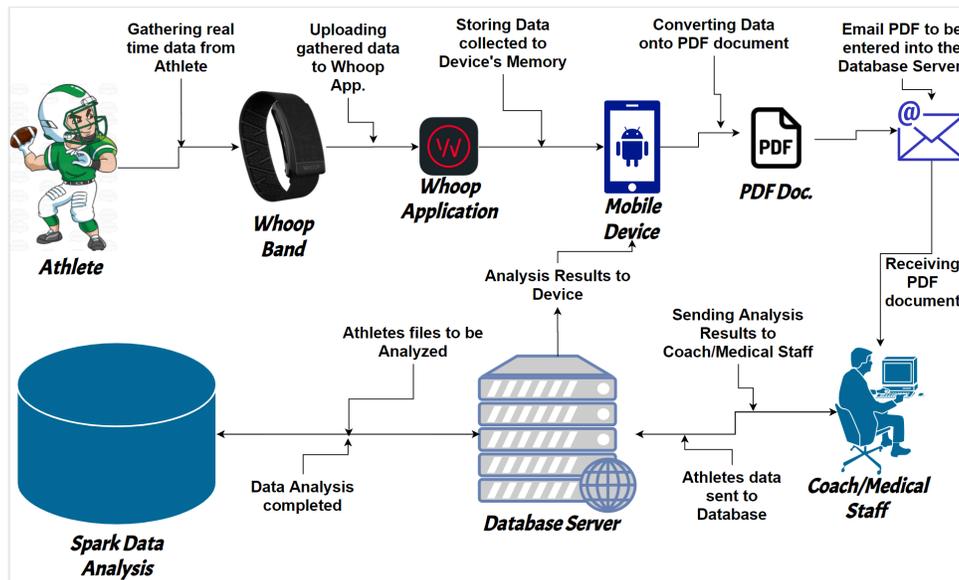


Figure 3: The Architecture of Big Data Analytics for Smart Sports Environment

## Gathering Data Process

After the data has been successfully gathered, it will be simple in the fact that the Whoop band is doing all the gathering of real-time data for us. The data that the Whoop band and the application are gathering is of great importance because this will determine the results after all the data has been compiled through the prediction model. Users will be required to have the application downloaded on their mobile device and of course, participate in wearing the Whoop band itself. The application will be collecting all the data to where it will store on the local mobile device's memory. Then the application will be able to transfer this information onto a PDF document to where the user will then need to supply a version of that to the coaching/medical staff. From the document gathered it will then be transferred to the database and then later to the data analysis. After the completion of the analysis, it will then forward the results back to the user using the same route it took to get there.

## Prediction Process

In our prediction process we are going to obtain the predicted result from Spark's management tools and then forward the predicted results back to the user. The first step is gathering the data from the Whoop band and its application. After we have successfully gathered all the data required, we will then store this data into a database for storage. There we will obtain the information needed to make a successful prediction and then extract this data via Spark's management tools so that it will be categorized/treated. Next, we will have to make some of the data into specific data types because the algorithm we chose only will accept certain data types. This is called transforming the collected real-time data. Once the data has been transformed, we will then apply the K-means algorithm to the data so that it can properly be analyzed to

the needs of this research project. After the analysis is completed, we then will be able to view the results of the prediction and in so many ways. I want to be able to show the results through virtualization so that we can prove our research hypothesis

## Contributions

This research is presenting that when Big Data Analysis is implemented on Carson-Newman's campus (or other universities' athletic programs), it will indeed create Smart Sports. Not only will Smart Sports improve the health of the athletes at Carson-Newman University but help improve the overall level of competition amongst the sports teams. We also must look at how this could benefit the campus in bringing forth a ton of new athletes wanting to attend this University because of the competition level being increased from Smart Sports. It is a remarkably simple concept of when more people attend this University it will generate the school profit and could lead to expansions that before were not possible. This implementation of smart Sports would help the Strength Coaches, Athletic Trainers, and coaches access player's information to closely monitor when a player might be nearing injury or training overload. Monthly or whatever time you set up for your predictive analysis scheduling it will send the athlete and whoever oversees that player a report on the prediction results. Once they both received the report, they will then both meet and discuss the best course of action to either change their current course of action or to continue what they are doing. The capabilities of the Whoop being able to supply real-time data and supplying that to a database to which is fed into software like Apache Spark's framework that will then be able to predict or prevent negative outcomes. The main contributions of using a predictive system like this are that the medical staff overseeing the athlete will then be able to forecast what the athlete is currently going through. It will also give the athletes context from a data analysis standpoint to why their body is feeling the way it is.

## Materials

For our project to be implemented successfully, the only hardware we will need will be the Whoop band itself. The reason we need the Whoop band is that we need to acquire the real-time data in a timely manner and this IoT device is the best you can get your hands on today. For the software for this project to be successfully compiled, we will need to use Apache Spark Open-Source. Apache Spark is an Open-Source software, so it is indeed free. Another software feature needed for this project to run successfully is that you need to download the Whoop application off your preferred operating system's (iOS or Android) app store. The Whoop application is also free and connects to your Whoop via Bluetooth sync, after downloading all you need to do is log in or create a new account if you are a first-time user. Once you have completed the previous steps, everything inside the Whoop application should become accessible. The software is free, however, the Whoop band itself is not free. The set price for the Whoop band itself is \$25 for the generic black band and \$30 a month for the membership. It gives you three payment options, 30\$ monthly, \$24 yearly, and \$18 for 18-month membership. This may seem like an expensive hardware piece but compare to its counterparts and competitors it is the best out of all of them.

## Survey Results

During our research, I wanted the collective census from the football team on their thoughts about their body's health during their time as collegiate athletes. I conducted this survey by creating a link and QR code to where then I posted it all throughout the football fieldhouse for the players to complete during the allotted time for completion. I created the survey through Google Forms, and it consisted of multiple-choice questions that regarded their health and thoughts of implementing Smart Sports here on their campus. From March 12th to March 28th, the survey was open and completed by 120 members of the Carson-Newman football team. The results of the survey will either enable or deny the implementation of

Smart Sports on a collegiate campus. These results will be relaying the attitudes of the athletes on the current wearable technologies and the desire of implementing a better technology that could help improve and prevent them from injury. The following figures are the survey results that were conducted with the Carson-Newman University football team. Figure 4 (left) illustrates that eighty-seven percent of the football team either strongly disagrees or disagrees that the current implementation does not improve their performance as an athlete, while the nine percent remain neutral, and six percent agree. These results indicate that the current technology we are using is ineffective and that it needs a dire improvement. Figure 4 (right) displays the evidence that ninety-five percent of the athletes on the Carson-Newman football team agree or strongly agree that they would prefer to wear and use Whoop technology. This justifies and implies that the athletes would prefer the new implementation of Smart Sports technology.

## Whoop Band Implementation to the team

The Whoop band is a wearable piece of technology that is superior to all wearable technologies. This band sits on your wrist, where it analyzes the following: strain, recovery, and sleep. All of these are tracked through the EKG monitor that is placed within the strap. Each Whoop band is completely unique to every individual user, meaning that people cannot just switch out their Whoop's with someone else in a trade deal and continue to get accurate results. To implement to the Carson-Newman football team I was able to acquire data by allowing all the players currently on the football team to wear the whoop band for at least 1 full week. In my specific case, there were a lot of players who currently already own their own Whoop band and were kind enough to send me their information when wearing the Whoop band. Overall. We were able to gather data from every player on the current roster. We asked the players who participated in my study had nothing but positive thoughts when describing their time with the device. Compared to its counterpart, the current wearable technology we use is not very comfortable for athletes to put on and wear.

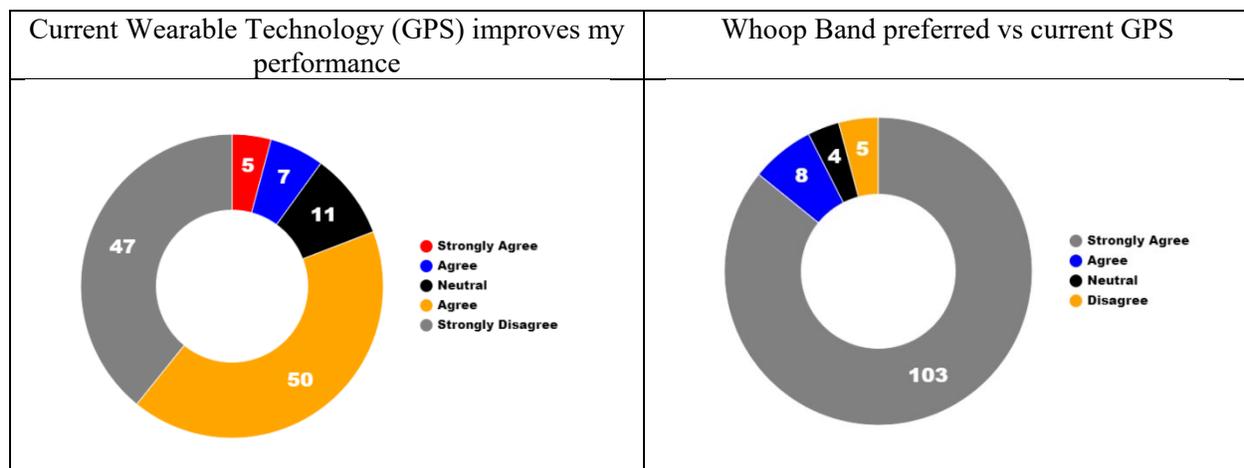


Figure 4: Survey of Results

## Whoop Application for iOS and Android

The application for this project is already available for all users to download from their preferred operating system's application store. Your Whoop band will connect via Bluetooth sync to this mobile application, this connection is how the application will gather your information to display inside the application's interface. The application itself has its own way of interpreting data and displaying it in its own form of

virtualization, however; they have not yet implemented a feature to where you could have predictive modeling. The application uses the operating system’s memory to store all the real-time data is collected. When the user downloads the application, it will prompt him or her to make an account and to connect to their Whoop via Bluetooth. Constantly throughout the day, the Whoop will be in sync with your mobile device uploading your real-time data. On the application, you can download a pdf document that has your monthly report of data. In our specific case, the user will then send that PDF document to an email address to where the football team’s medical staff will be able to upload it to our own database to store important information collected from the Whoop band.

## Results

When the system implementation has been successfully implemented into a collegiate setting, the expected results would be a positive impact amongst all the university’s athletes and sports teams. When implemented, this will improve the performance of the athletes and the university would not have to spend as much money on surgeries. Another benefit to the implementation of Smart Sports would it also save the university money. By getting rid of old and expensive wearable IoT devices that provided no data that the coaching staff or medical staff could use in a real-time situation. This system would increase the success of the sports teams on campus because we will have a greater understanding of the athlete’s body performance when they are playing at their best. By collecting athlete’s data via their Whoop bands and using Big Data analysis techniques we will be able to show an overall increase in the performance of the individual. Figure 5 explains an example of big data analysis based on the Apache Spark platform. Figure 5 is a virtualization of the attributes selected from the database we have created for our proposed system. With this research using the powerful Spark data management tool, we can plot the points onto a graph for furthermore investigation. From the graph above, we can dictate and start to categorize the position groups into groups based on their recovery percentages as low and high-risk groups. This is an important starting point when analyzing the collected data so that we can start breaking down the needs from position groups, all the way down to the individual athlete. Recovery is an important measuring factor when trying to predict and improve the performance of the athlete. For example, when athlete X’s recovery percentages are less than or equal to forty-five percent, X’s likely not going to be able to compete at his/her best. With their recovery number being low as well, they could become more susceptible to injuring themselves when in competitive play. Below, there is a pseudocode example of this being implemented into a real-life situation within the sport of football.

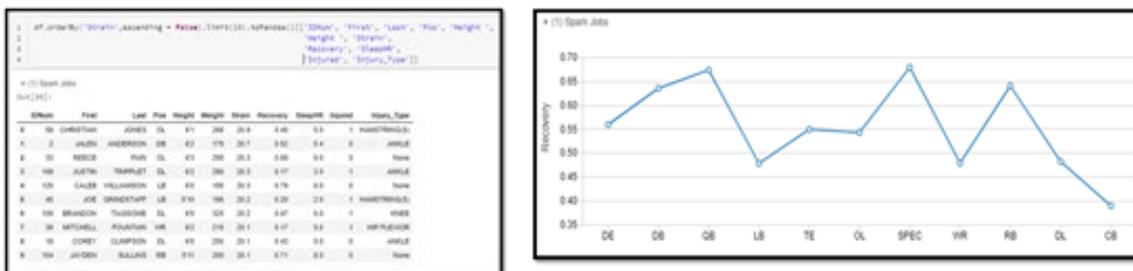


Figure 5: An example of big data analysis based on the Apache Spark platform

Pseudocode for Improving Athlete’s Performance and preventing the Athlete being in situations where they are more likely to get injured. In a sport as physically demanding as football, one misplaced step could determine you injuring yourself or not. This also goes for pulling hamstrings, and other muscle groups. The athlete’s recovery percentage will be a direct indication of the current state of the athlete’s body is in.

```
X represents a specific, SINGLE player inside the dataset
X = Athlete's Recovery percentage.
DO
    X == null
    if ( X <= .49)
        then
            print ("Athlete is not fully recovered.")
            print ("Athlete is susceptible to injury, or performing below avg.")
    else if ( X = .50)
        then
            print ("Athlete can perform at half of their capacity.")
            Print ("Injury is not likely.")
    else
        Print ("Athlete is fully recovered and can compete at their best.")
end
```

**Figure 6: Pseudocode Example of Analyzing Athlete's Recovery %**

Figure 6, the pseudocode is an example of what analyzing the athlete's recovery percentage would look like. In this specific example, the medical staff will be able to check this daily throughout the week as practice and training occur. With this type of simulation being compiled daily, they will be able to keep a closer eye on a specific player before an important competition. It will also encourage the players to take the proper steps to improve their recovery so that they are not held out of competition, training, and practice in precaution to prevent injuries. Another real-life example could be that the athlete's recovery percentage is at half of what it usually is, and from there we will be able to change the athlete's workout level, and practice involvement to make sure they are at their best when competing on the field. This would give an advantage to the team implementing this system because they would be able to make sure their athletes are able to perform at their best ability. In doing so, it will lead them to a higher level of success.

## Conclusion

After further research has been conducted on using Big Data Analysis for Smart Sports, we can see that Smart Sports present a numerous number of benefits to collegiate campuses of any size. Although universities will have different issues with their athletic teams, injuries are a common theme in collegiate sports no matter the campus. University coaches and medical staff spend each year trying to improve the overall performance of their athletes. Teams and athletic departments currently spend money on IoT devices to uphold their commitment to improving their athletes. Then there is the issue of how the types of data being collected by these IoT devices are not entirely useful nor beneficial to the athletes in any way. The device and analysis system that I have presented to replace this device and the old system provides the coaching/medical staff with the proper information to uphold their initial commitment to their athletes. Not only will the athletes, coaches, and medical staff benefit from this, but the university will benefit from it as well. The list of benefits for the university is endless but just to name a few would be the following: increase the overall notoriety of their athletic programs, increase the number of prospective athletes who want to attend their university, and they will also benefit from it financially. This study was limited since this study was only conducted on one sports team on a collegiate campus. This was a preliminary study and will allow future research to implement this Smart Sports system. This study had some shortcomings. There was a limited number of sports teams used in this research and it was also conducted on a single collegiate campus. Soon, it could be used to look at all sports teams on a collegiate campus and even look at colleges across the nation. It is assumed that all collegiate sports teams are dealing with the same issues that pertain to their

athletes being injured, however, there needs to be a deeper dive into the specifics of how each sport's injuries and training differ from the other, and we need to acknowledge that each sport is faced with issues that might not be faced in another. The Smart Sports system has the potential to become a system that can be incorporated globally to all collegiate campuses. To continue or expand, the implementation of using it campus-wide would eliminate the idea of this implementation only being applicable to a specific sport. Another limitation I faced during our research was that the size of the data being used was small. With smaller datasets do come less accurate results, however, we are confident in our system will uphold with bigger datasets. The true purpose was for it to be a system that would stand the test of time and be more efficient when larger datasets come into play. In this research, we are using Big Data analysis techniques and applying them to a 126-man roster. This could impose some setbacks when trying to analyze the data, however, we do think this is a great place for the project to be expanded on in the future. This project was the stepping stone to get the conversation started and how we could use Big Data analysis to implement Smart Sports on collegiate campuses. With bigger sample data we will be able to make more accurate decisions and even apply more data analysis techniques to make more accurate predictions.

## Acknowledgements

We gratefully thank the Carson-Newman football coaches, medical and strength staff for allowing us to conduct my senior capstone research project on the players of the current football team.

## References

- Armbrust, M., Xin, R. S., Lian, C., Huai, Y., Liu, D., Bradley, J. K., & Zaharia, M. (2015). Spark sql: Relational data processing in spark. In Proceedings of the 2015 ACM SIGMOD international conference on management of data, 1383-1394.
- Asri, H., Mousannif, H., & Al Moatassime, H. (2019). Reality mining and predictive analytics for building smart applications. *Journal of Big Data*, 6(1), 1-25.
- Carey, D. L., Ong, K. L., Whiteley, R., Crossley, K. M., Crow, J., & Morris, M. E. (2017). Predictive modelling of training loads and injury in Australian football. <https://arxiv.org/abs/1706.04336>
- Chen, Y., Crespi, N., Ortiz, A. M., & Shu, L. (2017). Reality mining: A prediction algorithm for disease dynamics based on mobile big data. *Information Sciences*, 379, 82-93.
- Dhar, V. (2014). Big data and predictive analytics in health care. 113-116
- Gaber, M. M., Aneiba, A., Basurra, S., Batty, O., Elmisery, A. M., Kovalchuk, Y., & Rehman, M. H. U. (2019). Internet of Things and data mining: From applications to techniques and systems. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3), e1292.
- Harms, N. (2018). The Impact of WHOOP Technology on Sleep, Recovery, and Performance in NAIA Baseball Players. <https://digitalcommons.unl.edu/teachlearnstudent/97/>
- Hong, S. Y. (2009). Web Image Information Mining System Using UUP for E-Business Intelligence. *Issues in Information Systems*, 10(2), 444-450.

- Hong, S. Y., Cho, J., & Hwang, Y. (2016). Prediction And Recommendations on The It Learners' Learning Path as A Collective Intelligence Using a Data Mining Technique. *Journal of International Technology and Information Management*, 25(3), 99-136.
- Ismail, A., Abdlerazek, S., & El-Henawy, I. M. (2020). Big data analytics in heart diseases prediction. *Journal of Theoretical and Applied Information Technology*, 98(11).
- Kalantari, M. (2017). Consumers' adoption of wearable technologies: Literature review, synthesis, and future research agenda. *International Journal of Technology Marketing*, 12(3), 274-307.
- Liu, Y., Shou, G., Hu, Y., Guo, Z., Li, H., Peng, F., & Seah, H. S. (2017). Towards a smart campus: Innovative applications with WiCloud platform based on mobile edge computing. In 2017 12th International Conference on Computer Science and Education (ICCSE), 133-138.
- Ray, P. P. (2015). Generic Internet of Things architecture for smart sports. In 2015 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 405-410.
- Rajšp, A., Heričko, M., & Fister Jr, I. (2020). The use of Gamification in Smart Sport Training. 405-410.
- Sikka, R. S., Baer, M., Raja, A., Stuart, M., & Tompkins, M. (2019). Analytics in sports medicine: implications and responsibilities that accompany the era of big data. *JBJS*, 101(3), 276-283.
- Salloum, S., Dautov, R., Chen, X., Peng, P. X., & Huang, J. Z. (2016). Big data analytics on Apache Spark. *International Journal of Data Science and Analytics*, 1(3), 145-164.  
<https://doi.org/10.1007/s41060-016-0027-9>
- Shoro, A. G., & Soomro, T. R. (2015). Big data analysis: Apache spark perspective. *Global Journal of Computer Science and Technology*. 7-14.
- Thornton, H. R., Delaney, J. A., Duthie, G. M., & Dascombe, B. J. (2019). Developing athlete monitoring systems in team sports: data analysis and visualization. *International journal of sports physiology and performance*, 14(6), 698-705.