SMU Data Science Review

Volume 7 | Number 2 Article 3

2023

Using Geographic Information to Explore Player-Specific Movement and its Effects on Play Success in the NFL

Hayley Horn
Southern Methodist University, hayleyhorn@gmail.com

Eric Laigaie
Southern Methodist University, ericlaigaie@gmail.com

Alexander Lopez
Southern Methodist University, alexanderlopez@mail.smu.edu

Shravan Reddy
Southern Methodist University, skreddy@mail.smu.edu

Using Geographic Information to Explore Player-Specific Movement and its Effects on Play Success in the NFL

Eric Laigaie[1], Alexander Lopez[1], Shravan Reddy[1], Hayley Horn[2]
[1]Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
{elaigaie, alopez, skreddy}@smu.edu
[2]Sr. Solutions Architect, Databricks,
160 Spear St 15th floor, San Francisco, CA 94105 USA
hayley.horn@smu.edu

Abstract. American Football is a billion-dollar industry in the United States. The analytical aspect of the sport is an ever-growing domain, with open-source competitions like the NFL Big Data Bowl accelerating this growth. With the amount of player movement during each play, tracking data can prove valuable in many areas of football analytics. While concussion detection, catch recognition, and completion percentage prediction are all existing use cases for this data, player-specific movement attributes, such as speed and agility, may be helpful in predicting play success. This research calculates player-specific speed and agility attributes from tracking data and supplements them with descriptive factors to produce a quality data set that, with machine learning models, can lead to accurate predictions of success on a play-by-play basis. A neural network was trained to predict play success with an F1 score of 40%. Therefore, the true effect of the inclusion of player movement attributes in predicting play success appears to have a minimal effect, but additional data and future research may be needed to confirm that.

1 Introduction

As far as American sports go, the National Football League (NFL) is king. To give an idea of the league's popularity, 18.5 million Americans tuned in to Sunday Night Football weekly in 2021 and 32 million played fantasy football in 2023 (Statista & Fox Sports, 2023). According to Forbes, the NFL accounted for an impressive 84 of the 100 most watched telecasts in the US during 2022 (Forbes, 2023). Finally, the league signed an 11-year contract worth $110 billion in 2021, with channels including FOX, NBC, and the streaming service Amazon Prime (New York Times, 2021). The league and its teams have no shortage of capital or exposure, even outside of the United States. In 2022, a matchup of the Buccaneers and Seahawks in Munich, Germany averaged 5.8 million viewers, breaking the NFL's record for most watched international game (Front Office Sports, 2022). This record was a 63% increase from the viewership of the Falcons versus Jets game in London during the 2021 season. While some of this increase may be attributed to the presence of a now-retired Tom Brady, the growth of worldwide viewership is significant.

The NFL is utilizing data to drive improvements to player safety, advance player progression, and help create a more immersive experience for their millions of fans worldwide. In 2022, the league announced a partnership with eight Division 1 universities to equip athletes with mouthguards that collect data on head impacts (NFL 2022). Additionally, the league used tracking data to observe kickoff return speeds and the resulting concussions. The pattern that was found contributed to the NFL's decision to extend the fair catch boundary to the 25-yard line in 2018 (CBS, 2021). On another front, the lead analyst of the Super Bowl 52 champion Philadelphia Eagles, Ryan Paganetti, used statistics such as expected points and expected points added to judge not only a player's impact on the game, but the risk of attempting a fourth down or two-point conversion. In fact, Doug Pederson, the coach of the Eagles, was graded as the most aggressive play-caller in the league when it came to these risky plays during Paganetti's tenure (Forbes, 2021). According to CBS, the Tennessee Titans were the only franchise without a staffer with the word "analyst" in their title in 2021 (CBS, 2021).

Since 2014, nickel-sized RFID (Radio Frequency Identification) tags have been embedded into not only the football itself, but player's shoes and pads, pylons, and first down markers (CBS, 2021). These sensors then record data and send them to receivers within each stadium. The data collected contains many factors that are used to track each player's movement across the field (Kaggle, 2023). On a basic level, a single frame of tracking data, which is collected about ten times a second, contains the x and y coordinates of all players and the football. Player tracking data has now evolved to capture more complicated movement data. From speed and acceleration to total distance traveled and player orientation, this tracking data provides a comprehensive picture of the play. In fact, even the RPM of the football is recorded for each throw during a game. Interestingly enough, more statistics are on the way. The NFL's VP of emerging products and technology, Matt Swensson, mentioned that new football-focused statistics were in the works (CBS, 2021).

In addition to these internal uses of NFL data, the fans are beginning to be included in this adoption of analytics. The NFL Next Gen Stats platform, powered by Amazon Web Services, provides the average fan with data-backed insights that blend seamlessly into the live broadcast of any game (Amazon Web Services, 2023). Additionally, the NFL Big Data Bowl is a popular data science competition held annually on Kaggle, drawing hundreds of submissions and rewarding over $100,000 in prize money (Kaggle, 2023). Fortunately, this competition provides large swaths of NFL tracking data, organized in neat and processed datasets. From this data, those that are interested can work to answer a specific question posed by the league, such as evaluating linemen on passing plays, or use it to research another question of interest. According to Matt Swensson, the uses of this data are hard to predict. The VP claims that the use cases for this data can get "pretty broad pretty quickly" (CBS, 2021).

With the 2018 Supreme Court decision to strike down the national ban on sports betting, the industry looks to be on the rise. As of 2023, almost half of the US states allow mobile sports betting (Legal Sports Report, 2023). In a recent report from Data Bridge Market Research, the growth of the sports betting market is shown to have a 9.2% CAGR (Compound Annual Growth Rate) through 2030, resulting in a market size of nearly $300 million USD (Yahoo, 2023). This growth increases the motivation for researchers to introduce reliable predictive models to help stack the odds in their favor. One study focused on predicting NFL game results using a hybrid of a linear regression model and an artificial neural network. The researchers reached a prediction accuracy of just over 90% (Anyama & Igiri, 2015). This result is impressive, but later works seem inconsistent. A 2020 study used similar data and nine classifiers to classify match outcomes but reached a maximum of approximately 67% accuracy and F1 score with a Naïve Bayes model (Beal et al., 2020).

Offensive play calling has been viewed as a key to efficiently utilizing the skills of their team's roster. All the while, teams are trying to predict the opposing offense's plays. Even before calling for a specific play, coaches must consider personnel combinations, formations, the number of downs, distance, ball location, and game time. All of which needs to be communicated to the players on the field and begin the play within 40 seconds. Most research has focused on simple methods to predict run versus pass plays or predicting the result of a specific play. Lee, Chen, and Lakshman (2012) developed an approach to classify offensive plays, but do not utilize any personnel information in their models (Lee, et al., 2012). Teich, Lutz, and Kassarning (2016) use machine learning to, among many applications, predict the best play given certain characteristics (Teich, et al., 2016).

After observing the existing contributions in the field of predicting NFL success, there is a noticeable gap: player-by-player impact on a single play. Vince Lombardi, considered one of the greatest football coaches of all time, is credited with saying "Football is a game of inches and inches make the champion." During any one football play, individual matchups have the potential to disrupt the game, whether it is through overwhelming speed or efficient route running. These effects can be hard to quantify, but player-specific speed and agility metrics might be helpful in predicting outcomes during games, essentially predicting which team has players that will take those extra couple inches. This research aims to use NFL tracking data and machine learning algorithms to solve the question of "Can player-specific movement capabilities be useful in predicting offensive success on a play-by-play basis?".

2 Literature Review

The literature review focuses on three main areas: Geographic data in non-NFL domains, machine learning applications in the NFL, and NFL-based tracking data applications.

2.1 Tracking Data in Non-NFL Domains

While this study focuses on spatial data uses in American Football, investigating the use of tracking data in other sports can establish the domain of this data and how it has been used before. According to a 2017 study, the interest in leveraging spatial data recently increased significantly, from six papers between 2000 and 2005 to 58 from 2010 to 2015 (Gudmundsson & Horton, 2017). These works, amongst others, provide research into player and team activity patterns, activity recognition, and performance evaluation metrics.

Soccer is what is considered an "invasion sport", where one team with possession of the ball must invade the other team's side of the playing field to score (Gudmundsson, & Horton, 2017). The nature of this sport allows researchers to construct mathematically backed structures to investigate the flow of the sport. For example, researchers can use player positions and trajectories to segment the pitch into "dominant regions" (Gudmundsson, & Horton, 2017). This study references previous work that created bisectors between all pairs of players. A player's area inside these bisectors represents the area they can reach before their closest competitors (Gudmundsson, & Wolle, 2014). Expanding on this idea of dominant regions, Gudmundsson and Wolle (2014) were able to define successful passing lanes given a fixed pass velocity. This measure was evolved in a later study that sampled 54,000 passes of varying directions and velocities to determine the "receivable pass variation", which can be used to determine the safest passing options (Taki & Hasegawa, 2000).

Comprehensive team networks are another useful tool in observing the play style of a team. Passos et al. (2010) defined these networks as having vertices representing players, and edges representing passes with weights being calculated as the number of successful passes on that edge. From these networks, PageRank can be used to find the probability of a player having possession of the ball after a number of passes (Gudmundsson, & Horton, 2017). From these ranks, researchers observed the top four teams from the 2010 FIFA World Cup and found that the Dutch and Uruguay teams had a much more uniform distribution of ranks than Spain and Germany (Peña & Touchette, 2012). Therefore, the former two teams do not rely on a sole player to carry the offense, but rather distribute the ball in a much more equitable fashion than the latter pair of teams (Peña & Touchette, 2012).

In addition to the observation of team activity patterns, activity recognition is the focus of other studies. First, the various passes performed in ultimate frisbee were observed and classified using a neural network. Link et al. (2022) used wrist sensors to gather movement data, which was then processed to weed out routine actions, such as running and clapping. Once a neural network was trained on the nine types of frisbee throws, an F1 score of 52.3% was accomplished (Link et al., 2022). However, since many of the misclassifications were between similar pass types, the researchers recalculated performance metrics on a consolidated set of the five major pass groups, yielding a new F1 score of 88.4% (Link et al., 2022). Another study, focused on the Uruguayan sport of futsal, used tracking devices to monitor player movement and EMG outputs (Rodrigues et al., 2020). Various models, ranging from artificial neural networks and a long short-term memory model, were then used to classify activities such as running, shooting, and jumping (Rodrigues et al., 2020) After comparing the metrics of the different models, it was concluded that the dynamic Bayesian mixture model was the optimal model with an accuracy of 96% and F1 score of 80% (Rodrigues et al., 2020).

An analysis of the use cases for tracking data in the National Basketball Association (NBA) explored the methodology for player-value metrics and how they can be applied. Cervone et al. (2014) calculated the expected possession value (EPV) – the expected points that will be scored at the end of the play – by observing the positions and movements of all players. From this metric, players can be graded by how much their actions affect the EPV (Cervone et al., 2014). These actions, deemed "micro-actions" by Sicilia et al. (2019), were further investigated by calculating the probability of four terminal events, a shooting foul, non-shooting foul, field goal, and turnover. Each of these events have an associated probability, and these probabilities are limited to the next five seconds of the play (Sicilia et al., 2019). The limiting of time to five seconds helps remove any actions that likely do not influence the outcome of the play, such as dribbling the ball past midcourt.

It is clear that spatio-temporal data has many uses, from team networks and page ranks to classifying frisbee throws. While many of these studies focus on popular sports, such as basketball, the marginal sports of ultimate frisbee and futsal still have avenues to take advantage of this data.

2.2 Machine Learning Applications in NFL

Current research in predictive modeling using machine learning algorithms is abundant in all sporting domains. However, there is an apparent push for researchers, as well as professional organizations, to implement more machine learning techniques into the NFL, likely because of the betting implications and viewership involved with the league. With the NFL making large portions of its data available to all through its yearly Kaggle Big Data Bowl Competition, the applications of machine learning in the NFL are beginning to yield promising results.

One common application of machine learning in sports, especially in the NFL, is predicting the outcome of sporting events. While this goal has been pursued since long before any machine learning model, or computer for that matter, no clear consensus has emerged, even with the newer, more advanced models. For example, a recent study (Beal, T.J.N. et. al, 2020) compared the performance of nine different machine learning techniques in predicting NFL game outcomes based on various features, such as total passing yards and rushing touchdowns. Surprisingly, the simpler Naïve Bayes model outperformed more sophisticated techniques in terms of overall and year-over-year accuracy. Although the study (Beal et. al, 2020) focused on a much more general research question, it demonstrated how machine learning can be used to analyze game features and metrics to make measurable predictions in the NFL.

Although machine learning methods have gained attention for predicting the outcome of NFL games, it seems that NFL teams may be more interested in predicting the opposing team's individual game plays. One recent study by (Ötting, 2021) utilized play-by-play data from the NFL's Kaggle Big Data Bowl to predict a team's next offensive play, specifically whether it would be a pass or run play. This prediction would only benefit a defensive game plan, but it could also be used by an offensive coordinator to review the predictability of their own offensive scheme. Ötting (2021) used time series data and derived features to model a team's propensity for pass versus run plays, which was found to have a strong correlation with the types of plays called based on the team's state during a game. For example, losing teams with many yards to travel in late-game situations typically rely on passing plays to maximize yardage gained per play. By modeling the state sequence using a Markov chain, Ötting (2021) accounted for correlation in observations and time-series data. The study found that incorporating the team's state improved game play prediction accuracy compared to previous studies using basic modeling techniques. While this research does not account for player position or ability metrics, it highlights the potential impact of using machine learning to analyze play-by-play data and suggests that accounting for the state of the team may also affect individual player effort.

2.3 NFL Based Tracking Data with Machine Learning Applications

The use of tracking data and machine learning algorithms for NFL-Based questions is not a new phenomenon. While player-specific movement attributes are not the focus on many studies, they still have provided an existing background on how to use NFL tracking data to engineer features, train models, and evaluate player metrics.

There is little doubt that the quarterback is the most important position on a football team, as evident by their average earnings relative to other positions, making their evaluation critical to a team's decision to sign a particular player. Reyers and Swartz (2021) used location data to evaluate quarterback play by considering wide receiver separation and observing how often the quarterback threw to the optimal option (Reyers, M. et al., 2021). After calculating covariates such as receiver separation and rush separation, the researchers Reyers and Swartz used a stacking algorithm constructed of base models ranging from random forests to neural networks and general linear models to predict completion probability (Reyers, M. et al., 2021). Reyers and Swartz provided a wealth of knowledge not only on feature engineering from tracking data, but also how the unscripted facets of NFL plays affect the modeling process (I.e., unscripted runs).

No matter the quarterback, a football team will need wide receivers to be able to become open for potential passes. Football wide receivers study route trees which are efficient ways of classifying receiver options during a play. Chu, Reyers, Thomson, and Wu (2020) utilized tracking data in hopes of clustering receiver movements into specific routes. Although model performance metrics were not able to be calculated, the value provided by this research is held in the data processing. The researchers recognized that the coordinate system used for tracking players can be simplified by normalizing all values to a common line of scrimmage (Chu, D. et al., 2020). These methods can provide a valuable resource to the current study as data processing and engineering is integral to reducing the full dataset down to its most significant attributes.

Models using long short-term memory (LSTM) have been developed using tracking data to study football topics such as defensive pass interference (DPI), expected points, and probability of winning. DPI penalties occur between a single offensive and defensive player, so Skoki, Lerga, and Stajduhar (2022) took that into account and only used tracking data for those pairs of players. Their study opted for distance measures between the receiver and defensive back over player movement, as DPI requires contact between the players that might be harder to capture in coordinate values. Since DPI requires a certain timeline of events, the researchers opted for time-series models with binary classification capabilities such as LSTM and gated recurrent units (GRU) models, amongst others (Skoki, A. et al., 2022). Exploring the distance between receivers and their matching defensive back pre-play may be valuable in predicting successful plays.

LSTM was also used to illustrate the continuous changes in the probability of winning and expected points during an NFL game. Researchers (Yurko, R. et al., 2020) utilized player tracking data to study continuous-time valuation of game outcomes. Their work using tracking data provided by the NFL produced a general structure for continuous-time within-play valuation of game outcomes. In addition, they introduced a ball-carrier model that estimated expected yards gained from a ball-carrier's current position considering locations and trajectories of the ball-carrier, their teammates, and their opponents. Their approach combined sub models for quarterback decisions, ball carrier, target probability, and catch probabilities. Their models were able to illustrate continuous-time predictions and player evaluation metrics using the NFL-player tracking data. Using a LSTM recurrent neural network, they provided an expected yard line metric and its change during significant moments of a running play such as a hand of or first contact of the ball carrier. Their LSTM model also provided insight into individual athlete contributions over the course of a play.

With all the focus on quarterbacks and wide receivers, one area lacking in research involved defensive players. Introducing an unsupervised approach to categorizing the coverage type of defensive backs using player-tracking data, Dutta, et al. (2020) used a Gaussian mixture modeling (GMM) clustering to provide a model that could differentiate zone coverage from man coverage during any given play. They noted the advantage of using a mixture model 1) providing probabilistic labels for each cluster and 2) being a density-based, statistical model the GMM allowed the researchers to provide probability data for each player during a play (Dutta, R. et al., 2020). The research also demonstrated how the mixture models allowed for even deeper insight such as how pass coverage schemes change and become clearer throughout a play (Dutta, R. et al., 2020). The researchers also found GMM's flexibility, interpretability, and soft cluster assignments a good approach to using player tracking data to label specific play schemes (Dutta, R. et al., 2020).

Overall, previous studies into NFL tracking data have helped establish a pipeline for processing the data, engineering new features, and training models. Attributes such as pass rush separation and receiver separation may prove significant in the current study, and normalizing the coordinate grid for all plays can help decrease the time required to process the data.

The notion that the success of an individual play in football is often determined by one specific read or matchup, such as a wide receiver and cornerback, is consistent with previous research on the importance of player performance in determining the outcome of football games. Therefore, incorporating player speed and agility into quantifying player matchups has the potential to improve the accuracy of success predictions and play-calling. This approach is consistent with the current trend in sports analytics, which emphasizes the use of data and advanced statistical techniques to gain a competitive advantage. By leveraging player performance data to make more informed decisions about player matchups and play-calling, teams can increase their chances of success on the field.

3 Methods

The data for this analysis was sourced directly from the NFL via its 2023 Big Data Bowl Kaggle Competition. Eight weeks of play information and player locations were provided to the Kaggle competitors. The data was collected using a tracking system called Next Gen Stats (NFL Football Operations), that includes RFID chips in player equipment, as well as in the football itself, and a complex camera system that captures multiple angles of the field. Using the Next Gen Stats system, the NFL can provide extensive data, from basic player and game information to player speed and acceleration at every tenth of a second. In Table 1 below, the datasets included in this analysis are detailed.

Table 1. Description of the datasets provided in the NFL Big Data Bowl.

Dataset	Description
Games	Basic information about each game, such as date, time, and home and away team.
PFF Scouting Data	Play-by-play information on each player's role and statistics recorded.
Plays	Information on each specific play. Personnel options, penalties, and play outcomes are included.
Players	Height, weight, college, and position of each player in the NFL.
Weeks 1 through 8	Frame-by-frame information on each player's position, movement, and activity (pass thrown, caught, etc.).

The modeling set will be structured so that each row represents one play with columns containing basic descriptions (i.e., down and yards-to-go), each position's speed and acceleration, and basic route classification. This structure is far from the original format of the data. The restructuring is detailed below in the Data Processing section.

3.1 Data Processing

The data set was designed for a data competition, so the required cleaning and preprocessing was minimal. Upon an initial pass over the data, missing values were present but sparse. After review, it is apparent that many of the missing values were intentional. For example, quarterbacks have missing values in the column denoting sacks, as quarterbacks cannot record a sack. The rest of the missing values were irrelevant to the study, such as missing values in the "college / university attended" column in the player information data frame. This feature will not be used in the model, so the missing values can be ignored. The next step in data processing was restructuring the data into a modeling set. This involved two major parts: determining a single speed and acceleration value for each player and classifying all players into roles or position groups.

The tracking data was originally organized into a tall and narrow dataset with rows detailing a single frame's data for one player. Therefore, each player had many speed and acceleration values associated with them that needed to be summarized into one number. The initial step for this process was observing the count of records for each frame. In football, plays last for varying lengths of time and longer plays are typically "broken down", where players tend to improvise more, leading to unpredictable play. In Figure 1 below, the proportion of plays that contain a specific frame number is shown. This can be interpreted as "100% of the plays last for at least one frame, but only about 20% of the plays in the dataset last for at least 50 frames".

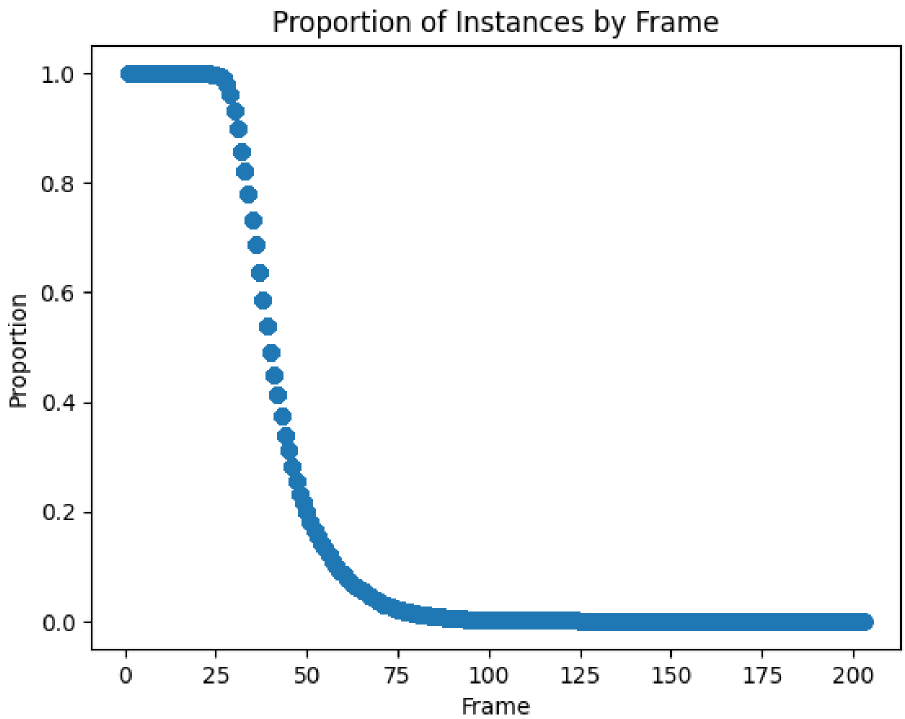

Fig 1: Proportion of plays containing each frame.

To continue rolling up the speed and acceleration values, any frames larger than 35 were removed as they were not common in the data set. From here, positions were classified into larger groups, such as "Offensive Skill" and "Defensive Backfield". In Figures 2 and 3 below, the average speed and acceleration for each position is plotted over each frame.

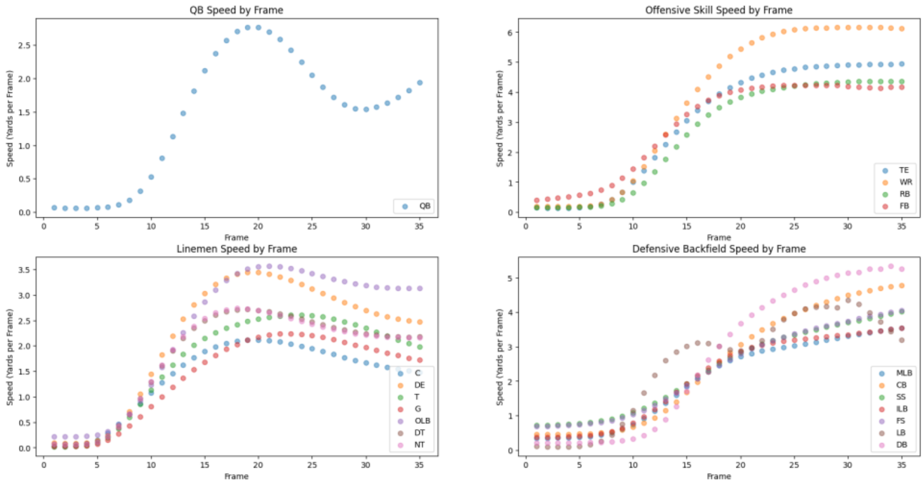

Fig 2: Average position speed by frame.

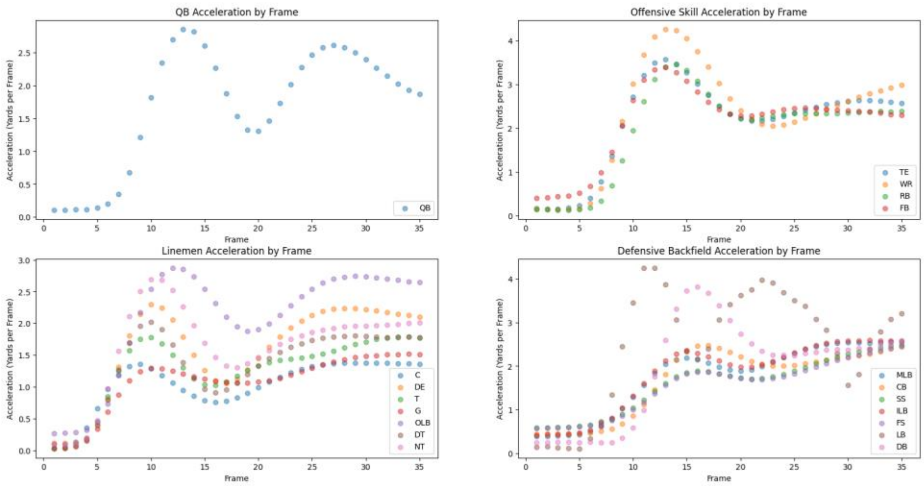

Fig 3: Average position acceleration by frame.

In cases where an entire position group follows a similar speed or acceleration trend (i.e., Linemen Speed), a single value will be used as the optimal point to calculate playing speed or acceleration from. In cases where specific positions act differently than others in their group (i.e., LB Acceleration), a more specific frame value is used for only that position. In Table 2, the specific frame for each position is denoted. Specific positions are only broken out when necessary.

Table 2. The optimal speed and acceleration frames for each position group.

Position	Optimal Speed Frame	Optimal Acceleration Frame
QB	20	13

Offensive Skill	25	12
Linemen	20	10
Defensive Backfield	35	30
LB	30	12
OLB	20	12
DB	35	15

From the optimal frames listed above, average speed and acceleration values were found for each player. The next step in data processing is to classify each player into roles that can be generalized for each play. Across the dataset, different personnel options are used on both offense and defense. Therefore, each player needs to be assigned a specific, but generalizable role.

Initially, offensive personnel options were considered. Of the various personnel options, a base structure of one quarterback, five offensive linemen, two wide receivers, one running back, one tight end, and one flex position (RB/ WR/ TE) was the most generalizable. Only 10% of the plays fall outside these positions counts, so they were removed. Defensive personnel options are harder to generalize, so players will be sorted into five buckets based on position and the role assigned to them by Pro Football Focus (PFF). The PFF role data is available from the NFL's Kaggle data and contains the defensive groups "Pass Rush" and "Coverage". From these two groups, players will be sorted into "Left Pass Rush", "Right Pass Rush", "Left Outside Pass Coverage", "Inside Pass Coverage", and "Right Outside Pass Coverage". In Figure 4 below, an example play is shown with players shaded by their position group.

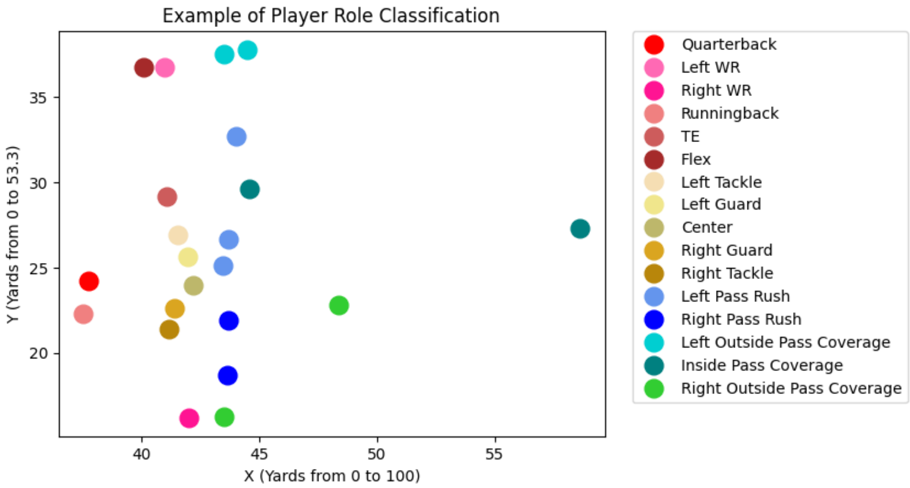

Fig 4: An example play with each position group labeled.

Once all plays were fully classified, the speed and acceleration columns were created by averaging the speed and acceleration values of each role group for each play. The next step was building a basic route classification algorithm. Receivers are tasked with running a full "route-tree" consisting of nine or more routes. Using tracking data to classify these routes has been the focus of previous studies, as mentioned in the literature review above, but the classification process is typically unsupervised, as comprehensive data on route assignment on a play-by-play basis is hard to find. In the case of this study, the NFL-provided data does not contain information on routes, but the player coordinates can be used to classify them. A simpler classification structure was decided upon, which would split routes into six different outcomes described below in Figure 5.

Left Deep 10+ Yards Vertically, 5+ Yards Left	***Middle Deep*** 10+ Yards Vertically, <5 Yard Horizontally	***Left Deep*** 10+ Yards Vertically, 5+ Yards Right
Left Shallow <10 Yards Vertically, 5+ Yards Left	***Left Deep*** <10 Yards Vertically, <5 Yard Horizontally	***Left Deep*** <10 Yards Vertically, 5+ Yards Right
	(PLAYER)	

Fig 5: Structure of the route classification algorithm.

To compare the first and last locations of each receiver, a common "end" frame needed to be decided upon. As mentioned above, plays that run exceedingly long typically include more improvisation which might result in obscured and inaccurate route assignments. Therefore, the maximum "end" frame will be set to frame 35, like the speed and assignment calculations. From there, the position of each receiver at the first and last frame of each play will be compared and assigned accordingly from the figure above. Finally, descriptive information about plays was joined into the dataset, such as the number of defenders in the box and the down in which the play occurred. These columns are detailed in Table 3 below.

Table 3. Description of variables included in the modeling set.

Variable	Description
"Role"_S	A specific role's average speed.
"Role"_A	A specific role's average acceleration.
"Role"_route	A receiver's route assignment (applicable to pass-catching roles)
offenseDeficit	Score deficit that the offense has at the beginning of the play.
quarter	Game quarter (1 through 4).
down	Down (1 through 4).
gameClock	Describes the amount of time left in the play clock.
yardsToEndzone	Number of yards needed to score a touchdown.
defendersInBox	Number of defenders in close proximity to the line-of-scrimmage.

playDirection	Direction the play is heading (left or right)
field	Which side of the field is the "field" versus the "boundary" (top or bottom).
possessionTeam / defensiveTeam	Denotes which teams are on offense and defense.
yardlineSide / yardlineNumber	Denotes which team's side of the field the play is on and the line of scrimmage.
offenseFormation / personnelO	Describes the offensive personnel numbers and formation.

Before modeling began, the categorical columns were one hot encoded and the numerical columns were scaled using the standard scaler from Scikit-learn. Figure 6 below displays the data frame sizes throughout the processing steps.

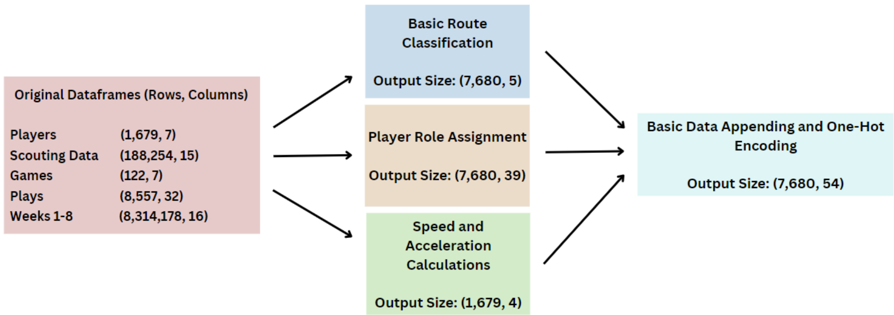

Fig 6: Diagram of data preprocessing.

3.2 Modeling and Evaluation Methods

The response variable in question has three classes and is derived from the variables "prePenaltyPlayResult" and "yardsToGo". The classes contain "Zero or Negative Yards", "Positive Yards: No First Down", and "First Down". To judge the performance of the model, a weighted F1 score will be used. The dataset is balanced with class shares of 42%, 33%, and 25%, but the F1 score combines both precision and recall, which is important to consider in a multiclass problem. After reviewing potential model candidates, it was determined that a deep neural network would be most suitable for the problem at hand.

The neural network used for this research is a deep, dense, feed-forward network with five hidden layers. The number of nodes in each layer decreases from 512 to 64 as we go deeper into the network, in attempts to help the model gradually abstract and generalize the input data across layers. This structure is common in feed-forward networks as it can effectively learn complex patterns in the data. The model makes use of the rectified linear unit (ReLU) activation function for the hidden layers. ReLU is a popular choice due to its simplicity and efficiency. It helps a network learn and introduce non-linearity into the model, which means it can learn complex patterns from the input data. Batch normalization is applied after each layer. This is a technique to improve the speed, performance, and stability of the neural network. It normalizes the inputs of each layer, thereby improving the efficiency of backpropagation and generally leading to faster learning rates. To mitigate the risk of overfitting, dropout is applied after each layer of the network. Dropout randomly ignores a fraction of nodes, in this case 50%, during training, which helps prevent the model from relying too heavily on any individual nodes and therefore improves generalization to unseen data. The final layer of the network uses a softmax activation function, which is commonly used for multi-class categorical output variables. It gives the output as a probability distribution across the three classes, which in this case represents the different outcomes of a play in the game. The Adam optimizer is used with a learning rate that decays exponentially (ExponentialDecay). This learning rate technique is beneficial because it allows the model to make large adjustments to the weights early in the training process and smaller adjustments later, allowing the model to fine-tune the weights and converge to the optimal solution more effectively. The model is trained for 100 epochs with a batch size of 32, and early stopping is used to prevent overfitting by stopping the training process if the model's performance on the validation data doesn't improve for a set number of epochs, in this case, 10.

To evaluate the model's performance, the weighted F1-score is used. This metric considers both precision, which is the proportion of true positive predictions out of all positive predictions, and recall, which is the proportion of true positive predictions out of all actual positive instances. giving a balanced measure of the model's performance. The F1-score is particularly useful in multi-class problems with imbalanced datasets, as it helps to ensure that the model performs well across all classes, not just the majority class. In this problem, it's weighted to account for the different number of instances in each play result class.

4 Results

The overall performance of the final neural network model was evaluated using the F1-score, but to have a complete understanding of our results, precision, recall, and overall accuracy were also examined. In Tables 4 and 5 below, the performance statistics and confusion matrix are displayed.

Table 4. Classification report for neural network model.

Outcome Category	Precision	Recall	F1-Score
First Down	0.39	0.28	0.33
Positive Yards, No First Down	0.32	0.37	0.35
Zero or Negative Yards	0.44	0.5	0.47
Macro Average	0.39	0.39	0.38
Weighted Average	0.4	0.4	0.39
Overall Accuracy	-	-	0.4

Table 5. Confusion matrix for test set predictions.

	Predicted Label

	Outcome Category	First Down	Positive Yards, No First Down	Zero or Negative Yards
True Label	First Down	126	189	454
	Positive Yards, No First Down	74	224	265
	Zero or Negative Yards	116	237	619

Overall, the neural network model does not appear to be overly successful. With a weighted F1 score of 40%, the model likely does not have sufficient performance to justify adoption into NFL play-calling practices. However, additional data and future research may help improve the performance of tracking data in a similar model. Those thoughts are detailed below.

5 Discussion

5.1 Limitations

This study contains limitations that may affect the final performance of the model. The data used only included passing plays from the first eight weeks of the 2022 NFL season, and some plays were cut short in the frame-by-frame data. For example, a play that resulted in a long after-catch run by a receiver does not include all of the frames between the catch and the tackle, but only the frames up until right after the pass is thrown. This is because the original Kaggle competition is only interested in offensive line pass blocking. However, additional data might have led to better predictive power.

One potential data source is the NFL Combine. Hundreds of draft-eligible players showcase their athleticism through multiple drills that focus on speed, agility, and strength. This study's performance might improve if that data is introduced to the model in a complementary matter. Additionally, this data provides an interesting idea for future research; how do Combine results compare to the actual movement of NFL players during a real game? If there is a noticeable difference in the two data sets, a couple of questions can be raised. First, which set is better for predicting play-by-by success, and second, what is changing? Are NFL players training exclusively for the Combine in hopes of improving their stock and providing an inaccurate depiction of their athletic abilities, or are those drills holding players' true abilities back? It is evident that draftees spend significant time perfecting the specific drills featured in the Combine. Exos, a "comprehensive sports science program", is an institute that trains athletes specifically for the NFL Combine drills and boasts an impressive eight No. 1 picks since 1999 (Sports Business Journal, 2023). Depending on the answer to these questions, the NFL Combine may be viewed in a significantly different light, and scouting college prospects may change drastically.

Route patterns and concepts are another interesting source of possible data. While fast enough receivers can just run past the pass coverage, receivers are typically grouped into route concepts that aim to confuse and pull apart the pass coverage. Categorical data on the route assignment of each receiver is used in this study, but an expansion of this data could be beneficial. Routes are typically difficult to classify based only on tracking data due to the unsupervised nature of the approach. This study used a rather simple categorization method to avoid misclassifications, but a more complex assignment algorithm or NFL-licensed data could help improve results.

5.2 Implications

While the model built in this study has suboptimal performance, researchers can still look to this work as a stepping point towards more significant findings. Tracking data provides many useful features, but many of these features are buried within the dataset. Without comprehensive data processing steps, these features can be difficult to uncover, and a modeling set can be difficult to form. Therefore, future researchers can seek out the techniques used in this study to aid in feature creation and modeling set formation.

The model built in this study was primarily used to predict the outcome of a single play, but the use-cases of it expand much farther, and future research can focus on these additional possibilities that may yield higher utility. By creating a modeling set that allows the user to swap players in and out of the lineup, a researcher can observe the differences in projection when swapping “Player A” for a “median” replacement at their position. The difference in predictions based on this swap could indicate whether “Player A” has a generally positive or negative effect on the play. Additionally, the model could yield valuable implications in the areas of player training and development. By consistently updating the data behind the model, teams can quantify the impact of a change in player movement. For instance, if the tracking data reveals a player's speed or agility decreases significantly in the later quarters of a game, they could use the model to quantify how much this could limit their on-field success. If a team finds that model projections are consistently decreasing towards the end of the game, it may be indicative that there is a need for improved stamina and conditioning training. The team can also use this model to learn more about player features that are not included in the data. If a team finds that the model is constantly over-projecting on plays where the quarterback faces any kind of pass rushing pressure, the coaching staff could then develop a specific training module that is designed to enhance performance under defensive pressure. Therefore, the integration of player tracking data into a predictive model could shape training regimes and allow for a more personalized, data-driven approach to player development.

Although sports can be difficult to quantify, similar work from this study exists in other sports already, and the specific use-cases detailed in this study can also apply to these new fields. The most famous example of a similar methodology to this study is Bill James’ “Moneyball”, also known as Sabermetrics. James’ Moneyball theory suggests that putting players on base at a higher rate can lead to more runs, which in turn, leads to more wins. Similarly, this study’s approach to analyzing tracking data could benefit NFL teams in increasing the success of their plays by isolating players at positions that need improvement to increase the likelihood of a successful play. Based on that insight, teams could train players to improve their performance, or even add a player to their roster that would provide that specific metric improvement through trade, draft, or free agency signing. One advantage that baseball has over football is their sample size. However, NBA basketball might be able to replicate that sample size advantage by observing field goal attempts. By utilizing tracking data and the large samples of associated field goal attempts, teams could build a model to quantify how well a player shoots in a specific area of the court and under a specific amount of defensive pressure. The team's management can then build a team that provides the precise shooting capabilities desired to increase performance.

5.2 Ethics

The ethical concerns of data collection are nonexistent in this study because it is sourced straight from the NFL. However, there are still ethical implications of work in this field. First, player safety must be considered and prioritized above all else. While playing football, players constantly put their body and mind on the line for their team. This can lead to repeated head impacts at high speeds, resulting in players developing Chronic Traumatic Encephalopathy (CTE). According to US News, an autopsy of almost 400 former NFL players diagnosed 91.7% of them with CTE. Although player tracking data provides exciting opportunities to quantify the sport of football, all implementations of new models or algorithms should undergo a thorough assessment of risk to ensure that it is not putting players into any additional harm's way. For example, an algorithm that consistently recommends aggressive kickoff returns or repeated forced passes through the middle of the defense where receivers are constantly under threat of a blindside hit must not be allowed full access to decision making.

Another point of ethical concern is sports betting. While an advanced machine learning model like the one produced in this research can predict the outcomes of individual plays with relative accuracy, it's important to remember that such models are based on historical data and statistical patterns, and they can never guarantee future outcomes. Hence, advertising these models as a surefire way to win bets would be misleading and unethical. Moreover, it is crucial to consider the ethics of using models like this to influence betting behaviors. Gambling can lead to significant financial harm and addiction, and using predictive models to encourage people to gamble more could exacerbate these issues.

Like personal data, the use of player tracking data can begin with good intentions, but those intentions have the potential to shift to questionable uses that can negatively impact players. Tracking data can be the cornerstone for a team's performance management approach that can optimize physical performance while minimizing risk of injury. As beneficial as tracking data has become with improving player performance and even better, more effective business decisions, it also introduces gray areas around the data's ultimate ownership and its uses. In an effort called Project Red Card, players across the English and Scottish soccer leagues argued that player performance data is personal and is being exploited without their consent (Holden, J., 2022). The future of player-team relationships may be altered with the advancement of tracking data as teams are exploring data science applications and their surveillance of players' athletic well-being and performance is growing. Some leagues have already addressed possible ethical boundaries within their respective collective bargaining agreements (CBA). In their 2017 CBA, NBA teams are prohibited from utilizing data collected via wearable technology against player contract negotiations. Similarly, Major League Baseball prohibits the data from use in salary arbitration discussions. The latest NFL CBA established a "Joint Sensors Committee to review and approve Sensors for NFL and Club use." As players in these performance-based leagues become more apprehensive about the idea that sensor data might be used against them, leagues need to confront risks and develop policies that address concerns regarding ownership, privacy, confidentially, and security.

6 Conclusion

The NFL, and sports as a whole, present an opportunity for researchers to quantify and predict the outcomes of passes, plays, games, and more. The creation and expansion of tracking data can be paired with traditional data science methods to approximate player movement abilities and use them to analyze, model, and predict. This study highlights the potential of using complex modeling in sports analytics to provide a deeper understanding of in-game dynamics. Yet, despite the efforts of this research, limitations were encountered that may have provided difficulties for a model to predict play success at a high level. Continued research, perhaps incorporating data from the NFL combine or exploring complex route assignment algorithms, could enhance the model developed in this study. Despite its limitations, this study contributes to the broader conversation about the use of predictive models in sports and their potential impacts, both positive and potentially negative.

Acknowledgments. Jacquelyn Cheun, PhD. – Capstone Professor, Hayley Horn – Capstone Advisor

References

1. Hollaus, B., Reiter, B., & Volmer, J. C. (2023). Catch Recognition in Automated American Football Training Using Machine Learning. *Sensors (Basel, Switzerland); Sensors (Basel), 23*(2), 840. 10.3390/s23020840
2. Gudmundsson, J., & Horton, M. (2017). Spatio-Temporal Analysis of Team Sports. *ACM Computing Surveys, 50*(2), 1-34. 10.1145/3054132
3. Beal, R., Norman, T. J., & Ramchurn, S. D. (2020). A Critical Comparison of Machine Learning Classifiers to Predict Match Outcomes in the NFL. *International Journal of Computer Science in Sport, 19*(2), 36-50. 10.2478/ijcss-2020-0009

4. Gustavo Pompeu, d. S., & de Andrade Moral, R. (2021). *Frame by frame completion probability of an NFL pass*. Cornell University Library, arXiv.org.
5. Reyers, M., & Swartz, T. B. (2021). Quarterback evaluation in the national football league using tracking data. *Advances in Statistical Analysis: AStA : A Journal of the German Statistical Society*, 10.1007/s10182-021-00406-8
6. Chu, D., Reyers, M., Thomson, J., & Wu, L. Y. (2020). Route identification in the National Football League: An application of model-based curve clustering using the EM algorithm. *Journal of Quantitative Analysis in Sports, 16*(2), 121-132. 10.1515/jqas-2019-0047
7. Dutta, R., Yurko, R., & Ventura, S. L. (2020). Unsupervised methods for identifying pass coverage among defensive backs with NFL player tracking data. *Journal of Quantitative Analysis in Sports, 16*(2), 143-161. 10.1515/jqas-2020-0017
8. Ötting, M. (2021). Predicting play calls in the National Football League using hidden Markov models. IMA Journal of Management Mathematics. https://doi.org/10.1093/imaman/dpab005
9. Liu, H., Aderon, C., Wagon, N., Liu, H., MacCall, S., & Gan, Y. (2022). Deep Learning-based Automatic Player Identification and Logging in American Football Videos. ArXiv:2204.13809 [Cs, Eess]. https://arxiv.org/abs/2204.13809v1
10. Skoki, A., Lerga, J., & Štajduhar, I. (2022). ML-Based Approach for NFL Defensive Pass Interference Prediction Using GPS Tracking Data. Paper presented at the 10.23919/MIPRO52101.2021.9596877
11. Mallepalle, S., Yurko, R., Pelechrinis, K., & Ventura, S. L. (2020). Extracting NFL tracking data from images to evaluate quarterbacks and pass defenses. Journal of Quantitative Analysis in Sports, 16(2), 95-120. 10.1515/jqas-2019-0052
12. Ronald Yurko, Francesca Matano, Lee F. Richardson, Nicholas Granered, Taylor Pospisil, Konstantinos Pelechrinis, and Samuel L. Ventura. Going deep: models for continuous-time within-play valuation of game outcomes in American football with tracking data. *Journal of Quantitative Analysis in Sports*, 0(0), January 2020.
13. Deshpande, S. K., & Evans, K. (2020). Expected hypothetical completion probability. Journal of Quantitative Analysis in Sports, 16(2), 85-94. 10.1515/jqas-2019-0050
14. Schmid M, Blauberger P, Lames M. Simulating Defensive Trajectories in American Football for Predicting League Average Defensive Movements. Front Sports Act Living. 2021 Jul 22;3:669845.
15. Anyama, Oscar Uzoma, Igiri, Chinwe Peace. (2015). An Application of Linear Regression & Artificial Neural Network Model in the NFL Result Prediction. INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT), Volume 04, Issue 01 (January 2015)

16. Joash Fernandes, C., Yakubov, R., Li, Y., Prasad, A. K., & Chan, T. C. Y. (2020). Predicting plays in the National Football League. Journal of Sports Analytics, 6(1), 35–43. https://doi.org/10.3233/JSA-190348
17. Tian, C., De Silva, V., Caine, M., & Swanson, S. (2019). Use of Machine Learning to Automate the Identification of Basketball Strategies Using Whole Team Player Tracking Data. Applied Sciences, 10(1), 24. 10.3390/app10010024
18. Link, J., Perst, T., Stoeve, M., & Eskofier, B. M. (2022). Wearable Sensors for Activity Recognition in Ultimate Frisbee Using Convolutional Neural Networks and Transfer Learning. Sensors (Basel, Switzerland); Sensors (Basel), 22(7), 2560. 10.3390/s22072560
19. Macdonald, B. (2020). Recreating the Game: Using Player Tracking Data to Analyze Dynamics in Basketball and Football. *Issue 2.4, Fall 2020*, *2*(4). https://doi.org/10.1162/99608f92.6e25c7ee
20. Rodrigues, A. C. N., Pereira, A. S., Mendes, R. M. S., Araújo, A. G., Couceiro, M. S., & Figueiredo, A. J. (2020). Using Artificial Intelligence for Pattern Recognition in a Sports Context. Sensors, 20(11), 3040. MDPI AG. Retrieved from http://dx.doi.org/10.3390/s20113040
21. Cervone, D., D'Amour, A., Bornn, L., Goldsberry, K. POINTWISE: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data. *MIT Sloan Sports Analytics Conference 2014*. 2014.
22. Gudmundsson, J., & Wolle, T. (2014). Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, 47, 16-27. 10.1016/j.compenvurbsys.2013.09.004
23. Taki, T., & Hasegawa, J. (2000). Visualization of dominant region in team games and its application to teamwork analysis. Paper presented at 227-235. 10.1109/CGI.2000.852338
24. Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., & Mendes, J. (2010). Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport*; J Sci Med Sport, 14(2), 170-176. 10.1016/j.jsams.2010.10.459
25. Peña, J. L., & Touchette, H. (2012). A network theory analysis of football strategies.10.48550/arxiv.1206.6904
26. Sicilia, A., Pelechrinis, K., & Goldsberry, K. (2019). DeepHoops: Evaluating Micro-Actions in Basketball Using Deep Feature Representations of Spatio-Temporal Data.10.48550/arxiv.1902.08081
27. Adgate, B. (2023). The Audience Disparity Of NFL Games And Everything Else Widens. Forbes. https://www.forbes.com/sites/bradadgate/2023/01/12/the-audience-disparity-of-nfl-games-and-everything-else-widens/?sh=6177bbb9113a
28. Amazon Web Services: NFL Advanced Stats. Amazon. https://aws.amazon.com/sports/nfl/
29. Fox, L. (2021). How the NFL Uses Analytics. Forbes. https://www.forbes.com/sites/liamfox/2021/08/12/how-the-nfl-uses-

analytics-according-to-the-lead-analyst-of-a-super-bowl-champion/?sh=ac82914424e6
30. Jones, J. (2021). The Future of NFL Computerized Player Tracking and Data Collection. CBS Sports. https://www.cbssports.com/nfl/news/the-future-of-nfl-computerized-player-tracking-and-data-collection-its-rapid-evolution-and-what-comes-next/
31. McCarthy, M. (2022). NFL Achieves Pair Of TV Ratings Milestones. Front Office Sports. https://frontofficesports.com/nfl-tv-ratings-milestones/
32. NFL Big Data Bowl 2023. Kaggle. https://www.kaggle.com/competitions/nfl-big-data-bowl-2023/data
33. NFL Data Collaboration: Building a Safer Game. (2022). NFL. https://www.nfl.com/playerhealthandsafety/equipment-and-innovation/equipment-testing/more-data-collaboration-building-a-safer-game-in-the-nfl-and-beyond
34. Sports Betting Market. (2023). Yahoo. https://www.yahoo.com/now/sports-betting-market-prospering-usd-084500892.html
35. Waters, M. (2023). Legislative Tracker: Sports Betting. Legal Sports Report. https://www.legalsportsreport.com/sportsbetting-bill-tracker/#:~:text=In%202018%2C%20the%20Supreme%20Court,at%20some%20point%20since%202018
36. Gough, C. (2022, September 23). NFL Sunday night football viewership 2021. Statista. Retrieved February 16, 2023, from https://www.statista.com/statistics/819106/nfl-sunday-night-football-viewers/
37. Fantasy NFL is a $70 billion industry - yes that's a 7 followed by ten zeros (2014, September 1). Fox Sports. Retrieved February 16, 2023, from https://www.foxsports.com.au/what-the-fox/fantasy-nfl-is-a-70-billion-industry-yes-thats-a-7-followed-by-ten-zeros/news-story/8521b28c0929d36d17b0203dd5d2786b
38. NFL Next Gen Stats | NFL Football Operations. (n.d.). Operations.nfl.com. Retrieved March 4, 2023, from https://operations.nfl.com/gameday/technology/nfl-next-gen-stats/#:~:text=A%20tracking%20system%20is%20installed
39. Lee, P., Chen, R., & Lakshman, V. (2016). Predicting Offensive Play Types in the National Football League.
40. Brendan Teich, Roman Lutz, and Valentin Kassarnig. NFL Play Prediction, 2016.
41. HealthDay. (2023). In Autopsy Study, Over 90% of Former NFL Players Showed Signs of Brain Disease CTE. U.S. News. https://www.usnews.com/news/health-news/articles/2023-02-09/in-autopsy-study-over-90-of-former-nfl-players-showed-signs-of-brain-disease~cte#:~:text=Researchers%20at%20the%20Boston%20University,%25)%20of%20NFL%20players%20studied.
42. Joe Lemire. (2023). NFL Draft prospects are finding success with Exos' comprehensive Combine training programs.

https://www.sportsbusinessjournal.com/Journal/Issues/2023/02/27/Technology/nfl-combine-exos.aspx

43. Holden, J. (2022, May 13). *How project red card data fight could affect US sports betting*. Legal Sports Report. https://www.legalsportsreport.com/67676/project-red-card-data-battle-us-implications/