PassIQ: A Metric for Grading Quarterback Decision-Making

Dan Schrage (dschrag2@nd.edu), Sean Dougherty (sdoughe5@nd.edu)

1 Introduction

In 2024, the National Football League (NFL) dominated American television, accounting for 70 of the top 100 broadcasts and an average of 17.5 million viewers per broadcast [7]. The 2025 Super Bowl drew an all-time record of 127.7 million viewers, reaffirming the NFL's dominant media presence. The NFL has traditionally been a huge draw for American television viewers, and this is unlikely to change anytime soon. In recent years, we have seen the rise of data-based sports analytics being applied everywhere, including during games and practices. The NFL is no exception, visible in the NFL Big Data Bowl. Hosted on Kaggle, this competition offers $100,000 for statistical models that offer new approaches that coaches and staff can use to improve everything from play calling to player performance, and much more. We have used the provided datasets to analyze the performance of quarterbacks and receivers. The quarterback is the most important man on the field, as he directly controls the rest of the offense. The quarterback's most potent offensive weapon is the ability to pass, which requires both physical and mental skills to accurately target and pass to an open receiver. Receivers must also get "open" so that passes can be made quickly and securely to avoid interceptions and capitalize on the high risk/reward that pass plays offer. We decided to look at quarterback decision-making when passing, as failed passes can be devastating to an offensive drive.

Evaluating quarterback performance is the key to developing and maintaining a strong offensive presence within the NFL. Quarterbacks are the highest paid position in the sport, with top salaries reaching over $50 million per year [3]. Teams are often built around quarterbacks, and drafting or trading for a good quarterback can change a franchise's playoff hopes. As a result, significant time and money is spent on finding and developing quarterbacks. Teams have the option of either drafting a young, untested quarterback from college or trading millions of dollars or valuable current players for a more experienced quarterback already in the league. Regardless of the approach taken, it is crucial for team owners to have the maximum amount of knowledge of the new quarterback's abilities and shortcomings. Mistakes in the evaluation process can cripple a team's performance for years, given that quarterbacks receive a large portion of a team's salary cap, plus significant training is required for new quarterbacks to adapt to a team's offense.

Beyond quarterback acquisition, evaluation is also key to training and preparing a quarterback for games on a week-to-week basis. Coaches rely on accurate performance data to tailor their plays for each week's matchup. With more advanced evaluation methods, coaches can focus their quarterback on certain strategies to counter the opponent's defensive strengths. Due to the high level of play, these slight adjustments mean the difference between a catch and an interception. At the same time, quarterback evaluation is crucial to ensure performance is up to par. Older quarterbacks are often respected for their experience and knowledge, but begin to falter in real-time game scenarios. Coaches need to know how their quarterback performs through events like injuries and personal changes in the off-season.

In this paper, we discuss how to evaluate the decision-making ability of a quarterback. We designed a method to quantify this ability by looking at how "open" the target receiver is for each play. In the following sections, we describe our methodology, results, and conclusions.

2 Related Work

In the following section, we will discuss some literature related to player analysis in the NFL. With the expansion of NFL's Next Gen Stats, highly detailed statistics have become available to the public for analysis, greatly improving the quality and quantity of statistical analysis in the NFL. Previous methods were rudimentary and lacked the depth and breadth of knowledge available today. One attempt to analyze the impact of the NFL Combine revealed that very few aspects of the combine actually helped to predict the success of future NFL quarterbacks, and these aspects could be simply a result of random chance [4]. With the explosion of data given by Next Gen Stats, statistical analysis methods began to improve significantly. Beal, Norman, and Ramchurn [1] gives an overview of global sports analytics, including the NFL. This is one of the first papers to directly discuss the use of machine learning in the sports analytics domain. He touches upon DeepQB, which applies deep learning to quarterback decision-making [2]. This approach utilizes the new data available from NFL's Next Gen Stats and provides a range of outputs, including predicted receiver, expected play outcomes, and expected yardage gain. This approach is similar to our method, and it notably discusses the difficulty of prediction because each team attempts to appear as unpredictable as possible to the other team. Burke [2] uses the model's results to perform preliminary quarterback analysis, showing the potential of this type of analysis. Reyers and Swartz [6] takes a similar approach, but refines some of the metrics used and expands the variety of assessable plays, in addition to separating individual receiver and defender characteristics from quarterback rating. Notably, they use expected points instead of expected yards in an attempt to integrate real game characteristics.

3 Methodology

This methodology is organized into data collection, preprocessing, and the construction, training, and testing of our model. The full code is available at `https://github.com/Dschrag2/NFLSimulation`.

3.1 Data

We used data from the National Football League Big Data Bowl 2025 on Kaggle [5], which includes detailed information from the 2022 NFL season, covering weeks 1 through 9. The dataset provides game, player, play, and frame-level granularity. While all available data was valuable, our focus was primarily on play-level and frame-level data.

Play-level data includes detailed information about each play, such as outcomes, formations, and contextual factors within the game. Additionally, binary indicators are available for each player's actions, such as whether a player was running a route on a given play.

The most important data for our study, however, came from the frame-level dataset. This provided detailed, frame-by-frame (0.1-second interval) information for each player on the field, including position, speed, acceleration, and direction. Key play events, such as when the ball was snapped, thrown, or caught, are also marked within this data. This level of detail allowed for a better understanding of the dynamics of a play.

3.2 Data Preprocessing

Despite the wealth of information in the dataset, significant preprocessing was required. First, we filtered the plays to include only those with frames containing the "pass_forward" or "pass_shovel" features and a "pass_arrived" feature. This ensured we only analyzed passing plays where the ball was thrown and arrived at a target (completion, incompletion, or interception). We also excluded plays where the quarterback scrambled, as we chose to focus solely on passing decisions, not play decisions. The goal was to answer the question: When quarterbacks throw the ball, are they making the correct decision?

After filtering the plays, we structured the data into inputs for our graph model. Each graph node represented an eligible receiver or defender in pass coverage. On average, each graph included around four receivers and seven defenders. Each node contained the following attributes: x and y position, speed, acceleration, orientation, direction, boolean indicators for offense/defense, and whether the receiver was targeted. We constructed edges between all offensive and defensive players, with no edges between players on the same team. Edge weights were determined by the Euclidean distance between the connected nodes. Each graph represents the state of the play at a single time step.

To train the Graph Attention Network (GAT), we provided a sequence of graphs over time. For each play, we took frames from the moment of the throw release and backtracked 3 seconds in 0.1-second intervals, resulting in 30 frames of data. If the time from the snap to the throw was less than 3 seconds, we used only the frames from the snap to the throw. This yielded sequences of 30 or fewer graphs. The target variable for the model was the openness score of the targeted receiver, calculated as the Euclidean distance to the nearest defender.

3.3 Graph Attention Network Model

We chose a Graph Attention Network (GAT) for our model. It consists of two graph attention layers followed by a fully connected layer for the final output prediction. The model has 8 input channels, 64 hidden channels, and 1 output channel. The 8 input channels correspond to the 8 features of each node, and the 64 hidden channels were found to perform well. The output layer predicts the openness score for the targeted receiver. Although the model calculates an openness score for every node, we mask out non-target nodes, focusing the prediction only on the targeted receiver. This is because the openness score is known for the targeted receiver as the play develops, while other receivers may stop running if they realize the ball is not coming their way, and defenders may alter their coverage in response.

Given the large dataset, we used CUDA with PyTorch for model training. We employed the Adam optimizer with a learning rate of 0.001 and a batch size of 32. The model was trained for 100 epochs, with mean-squared error as the loss function.

For quarterback evaluation, we froze the trained model and ran it on all plays from weeks 1 through 9 of the 2022 NFL season. The targeted receiver mask was removed, and the model predicted the openness score for all receivers on the field for each play. With these openness scores, we could assess a quarterback's decision-making by determining whether the targeted receiver was the most open option according to the model.

4 Results

Over the course of 100 training epochs, we reduced the loss from 8.766 to 6.371. It's important to note that the error was significantly higher when training on the entire 9-week dataset compared to a single week, due to the larger size of the dataset.

Quarterbacks were evaluated using two metrics: average error and accuracy. Average error measures the average distance in openness between the targeted receiver and the most open receiver on the play. If the most open receiver is targeted, the error is zero. Accuracy reflects the percentage of times a quarterback selects the most open receiver. As shown in Figure 1, there is a relationship between average error and accuracy, though they do not perfectly correlate. The table also includes attempts to provide context on the sample size; players with fewer than 100 pass attempts were excluded.

As seen in Figure 1, Joe Flacco exhibited the lowest average error, while Baker Mayfield had the highest accuracy. However, these results do not necessarily indicate the top quarterbacks in the league, as the analysis is based on a small 9-week sample. For instance, Flacco only played in three games during this period. While the rankings in the table are not fully representative, they do provide a snapshot of quarterback performance over this stretch.

Rank	QB_Name	Average_Error	Accuracy	Attempts
1	Joe Flacco	1.0629	0.4128	109
2	Baker Mayfield	1.1172	0.4336	113
3	Aaron Rodgers	1.1392	0.3682	220
4	Matthew Stafford	1.1574	0.4055	217
5	Justin Herbert	1.1675	0.3992	248
6	Kirk Cousins	1.1783	0.3623	207
7	Davis Mills	1.1885	0.3846	182
8	Daniel Jones	1.1965	0.3725	153
9	Lamar Jackson	1.2194	0.3883	188
10	Ryan Tannehill	1.2366	0.3679	106
11	Trevor Lawrence	1.2548	0.3822	225
12	Geno Smith	1.2681	0.3413	208
13	Jared Goff	1.2734	0.3232	198
14	Cooper Rush	1.2763	0.3772	114
15	Mac Jones	1.2822	0.3902	123
16	Zach Wilson	1.2898	0.3613	119
17	Jimmy Garoppolo	1.2975	0.3873	173
18	Joe Burrow	1.2997	0.3209	268
19	Justin Fields	1.3012	0.3556	135
20	Tom Brady	1.3026	0.339	295
21	Russell Wilson	1.3035	0.4142	169
22	Carson Wentz	1.3183	0.4011	177
23	Tua Tagovailoa	1.3287	0.3352	182
24	Patrick Mahomes	1.3424	0.3173	249
25	Kenny Pickett	1.3438	0.304	125
26	Josh Allen	1.3487	0.3565	216
27	Matt Ryan	1.3593	0.3506	231
28	Kyler Murray	1.3725	0.3295	264
29	Marcus Mariota	1.4422	0.2517	151
30	Derek Carr	1.4682	0.3738	206
31	Andy Dalton	1.5251	0.3061	147
32	Jalen Hurts	1.5617	0.3135	185
33	Jacoby Brissett	1.5855	0.2697	178

Figure 1: Quarterback Accuracy and Error Metrics

5 Conclusion

The Graph Attention Network (GAT) model demonstrated strong performance in identifying receiver openness, achieving satisfactory error rates and overall effectiveness. The model's rankings closely aligned with established quarterback assessments, with only a few notable outliers. However, it is important to recognize that the analysis was conducted over a limited nine-week period. Extending this analysis across multiple seasons would likely yield a more comprehensive understanding of quarterback performance.

While the model shows promise as a scouting tool for NFL teams evaluating talent in the draft or throughout the league, it also holds potential as a training aid for quarterbacks, helping them improve their decision-making. However, the model does have limitations. Notably, it does not account for pass-rushing pressure, a critical factor that can significantly influence a quarterback's decisions during a play. This limitation will be addressed in future iterations, where a more nuanced model could better capture the dynamics of the game.

6 Future Direction

Football is an inherently complex and dynamic sport, making it impossible to capture every nuance through data alone. While this work simplifies certain aspects—such as the relationship between quarterbacks and receivers—our aim was to create an interpretable and foundational model. That said, there remains significant room for expanding the model's scope to better reflect real-world decision-making.

One promising direction for future development is the inclusion of all defensive players on the field. Incorporating the actions and positions of defensive linemen, for instance, would allow the model to account for pressure on the quarterback, forced early throws, or disrupted passing lanes due to batted balls. Additionally, the quarterback's spatial context can influence which throws are feasible or advisable, and in some situations, scrambling may represent the optimal decision over any passing option.

Another promising development is the addition of new metrics accounting for specific game situations, such as overtime or end of game scenarios. These situations necessitate different play calling and styles, affecting the quarterback's decision-making at key points in the game. Standard receiver openness may be less of a factor when the game is on the line with little time left, and the quarterback's risk tolerance changes greatly. Additionally, trick plays and specific quarterback strengths may affect standard metrics, and these could be explored to see what specific plays fit a quarterback's play style.

These examples illustrate just a few of the many interactions unfolding simultaneously during a play. Modeling them all presents a considerable challenge, but doing so could lead to a more comprehensive understanding of quarterback behavior. Enhancing the model in this way may offer deeper insights into the situational choices quarterbacks face and the factors that influence their decision-making under pressure.

References

[1] Ryan Beal, Timothy J Norman, and Sarvapali D Ramchurn. "Artificial intelligence for team sports: a survey". In: *The Knowledge Engineering Review* 34 (2019), e28.

[2] Brian Burke. "DeepQB: deep learning with player tracking to quantify quarterback decision-making & performance". In: *Proceedings of the 2019 MIT Sloan Sports Analytics Conference.* 2019.

[3] ESPN Staff. *Highest-paid NFL players: Most guaranteed money at every position.* Accessed: 2025-04-14. Apr. 2025. URL: `https://www.espn.com/nfl/story/_/id/34096853/highest-paid-nfl-players-tracking-most-money-guaranteed-per-year-every-position`.

[4] Frank E Kuzmits and Arthur J Adams. "The NFL combine: does it predict performance in the National Football League?" In: *The Journal of strength & conditioning research* 22.6 (2008), pp. 1721–1727.

[5] National Football League. *NFL Big Data Bowl 2025.* `https://www.kaggle.com/competitions/nfl-big-data-bowl-2025`. Accessed: 2025-04-13. 2025.

[6] Matthew Reyers and Tim B Swartz. "Quarterback evaluation in the national football league using tracking data". In: *AStA Advances in Statistical Analysis* 107.1 (2023), pp. 327–342.

[7] Amy Tennery. *NFL dominates 2024 ratings despite 2.2% dip*. Accessed: 2025-04-14. 2025. URL: https://www.reuters.com/sports/nfl-dominates-2024-ratings-despite-22-dip-2025-01-09/.