13TH ANNUAL
MIT SLOAN
SPORTS ANALYTICS CONFERENCE
MARCH 1 - 2, 2019 | BOSTON, MA
presented by ESPN

DeepQB: Deep Learning with Player Tracking to Quantify Quarterback Decision-Making & Performance

Brian Burke, ESPN Analytics, brian.j.burke@espn.com

DeepQB is a proposed application of deep neural networks to player tracking data from over two full seasons of American professional football. This novel approach demonstrates the ability to successfully understand complex aspects of the passing game, most notably quarterback decision-making. It can assess and compare individual quarterback pass target selection based on a snapshot presented to the passer by the receivers and defenders. Assessments of quarterback decision-making are made by comparing actual target selection to that predicted by our model. The model performs well, correctly identifying the targeted receiver in 60% of cross-validated cases. When passers target the predicted receiver, passes are completed 74% of the time, compared to 55% when the QB targets any other receiver. This performance is surprisingly strong, given that the offense often conceals its intent by design, while defenses try not to allow any single receiver to be open. Further, quarterback passing skills separate and apart from his receivers and defense are isolated and assessed by comparing metrics of actual play success to the metrics of success predicted by the situation presented to the passer. This approach represents a new way for teams, media, and fans to understand and quantitatively assess quarterback decision-making, an aspect of the sport which has previously been opaque and inaccessible.

1. Introduction

Perhaps the most enigmatic and yet most important player attribute in all of American football is the decision-making abilities of quarterbacks. Although mental abilities are important for every position, quarterback is unique in that psycho-cognitive abilities rival physical abilities in regards to successful performance. Measurable and physical attributes are easily observed through the scouting process, but a professional quarterback's ability to process and exploit highly dynamic information during the course of a play is not well understood. Previously only relatively crude aggregate statistical methods - that cannot fully separate the quarterback's individual impact from those of his teammates, play design, and opponents - have existed to quantify this skill.

Early efforts to exploit football tracking data only scratched the surface of what is possible with such rich information. Typical applications of tracking data merely involved measuring the maximum speeds of the fastest players, or totaling the distance traveled by a player on a play.

DeepQB is a deep learning approach to evaluate quarterback decision-making and performance. This approach provides a suite of models that predict and evaluate pass decisions and outcomes at the play level. These models are relatively straightforward neural networks that are capable of producing useful analysis and insights in near real-time during live games.

This paper demonstrates that such an approach is a viable and promising way to analyze the passing game, the most critical aspect of professional football. Although there remain limitations with this approach, DeepQB proves that some of the most complex aspects of football can be understood though a deep learning approach to multi-agent features. It offers a fully quantitative

ability to measure individual performance for the most important position, and offers fans, media, and teams an innovative new way to evaluate the sport.

2. Related Research

Due to the comparatively late advent of tracking data in football, this the first major project to develop methods of understanding football player tracking using advanced machine learning techniques. However, several publicized research efforts that exploit player tracking have applied neural networks and other advanced techniques to data from basketball, soccer, and hockey. Cervone et al [1] built a framework based on a Markov model using player-tracking data to assign an expected point value to each time segment of a basketball possession. They compared a player's expected contribution to actual outcomes to assess player performance. Wang and Zemel [2] used both a feed-forward (FFN) and a recurrent neural network (RNN) based on optical tracking data for play type recognition and classification in basketball. Le et al [3] used "deep imitation learning" based on Long Short Term Memory (LSTM) networks to mimic soccer defenders' movements based on the dynamics of the other players. The authors stressed the importance of meaningful role representations for each player in the network. The model allowed individual team-specific playing styles to be represented as well. Harmon et al [4] transformed player tracking data from basketball into multi-channel pictorial representations that a convolutional neural network (CNN) could classify. The authors combined the CNN with a feed-forward network to predict shot making. Lucey et al [5] used a conditional random field model with soccer player tracking data to estimate the probability of a shot's success. This was used to assess team performance at both the season- and game-level. Mehrasa et al [6] applied one-dimensional convolutional networks to model player trajectories on both basketball and hockey, for play recognition and team classification.

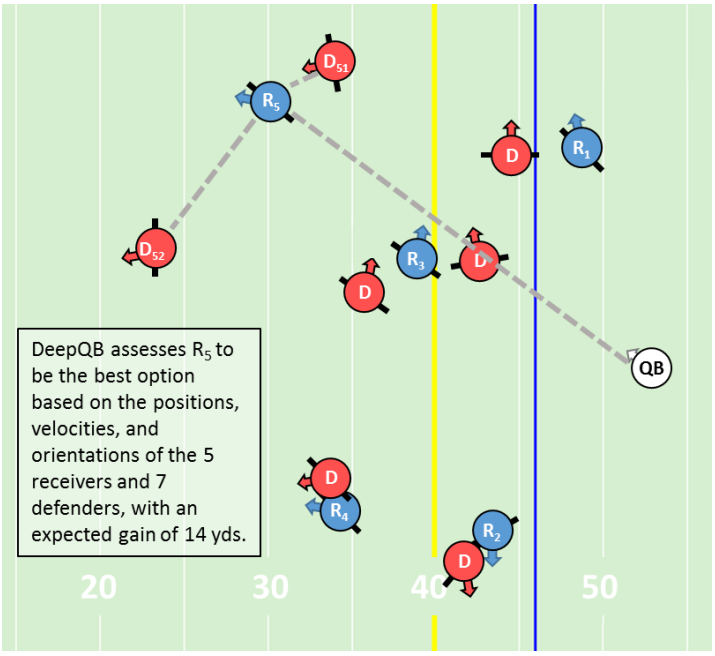
The single relevant paper on football attempted to assess quarterback decision-making using Voronoi tessellation. Hochsteler's [7] technique partitioned the area of the playing field according to which player is closest to each point of the field and assign exclusive ownership of each area. The size of each receiver's owned area and proximity to the nearest defender was used to assess the openness of a receiver. Although the research had only 224 plays to analyze and used relatively rudimentary techniques, it hinted at the potential of more powerful methods to understand and assess quarterback play.

3. Approach

The DeepQB suite of models takes advantage of a comprehensive representation of tracking data as presented to a quarterback to predict receiver target selection, completion-incompletion-interception outcome probabilities, and yardage expectation. These predictions are then compared to actual outcomes to assess quarterback target selection and overall performance.

Model input is a snapshot of player configuration at the time of pass release, including receiver position, velocity, acceleration, and orientation, as well as those of the secondary. Hand-built features, specifically relative positions and angles of receivers with respect to other players, were engineered and used as inputs. Additionally, play metadata, (down, distance, and yards to the end zone) were fed to the network. Lastly, data from ESPN's Video Analysis Tracking (VAT) project was used as inputs, specifically whether the quarterback was under duress from the pass rush and whether there was a play-action fake following the snap.

The heart of DeepQB is a modular feed-forward artificial neural network (FFN) architecture. It is modular in that the inputs and the architecture of the hidden layers can remain identical regardless of what type of target variable the model is being asked to predict. The final output layer is modified to accommodate whether we are asking the model to estimate probabilities for discrete outcomes (such as which receiver should be targeted or what the pass outcome would be), or we are asking the model to estimate a continuous value (such as the expected yardage gained by a pass play).



There are four variants of the general model discussed in this paper:

- Variant 1 estimates the probability the quarterback will target each of the five eligible receivers given each play's configuration of the receivers and defense.
- Variant 2 estimates the expected yards gained for each of the five eligible receivers on each play.
- Variant 3 estimates the probabilities of the three types of outcomes on the play: complete, incomplete, or interception.
- Variant 4 is an experimental model that produces target predictions like Variant 1, but uses transfer learning to capture the decision-making of individual quarterbacks.

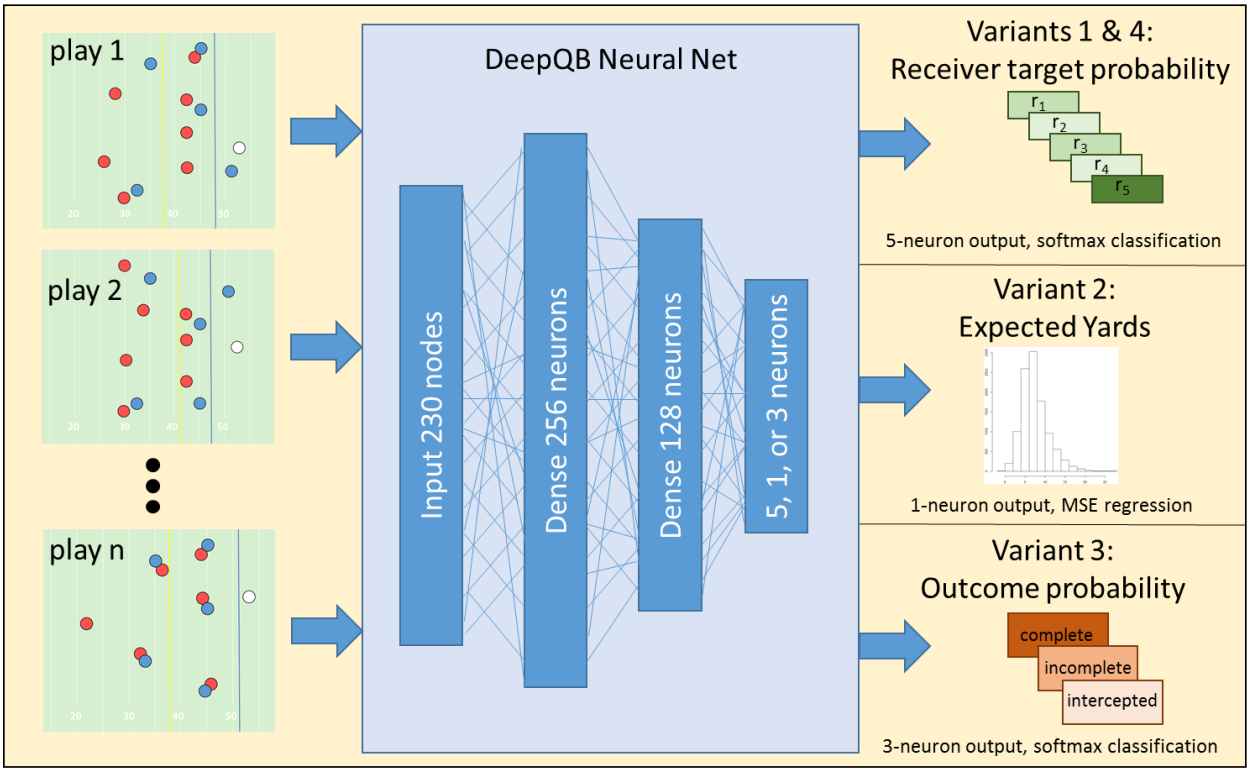
3.1. Data

Training and validation data consists of every targeted pass attempt with from the 2016 and 2017 NFL regular and post-seasons. Player tracking information comes from the *NFL's Next Gen Stats* system, which uses two electronic RFID chips in each of the players' shoulder pads. Compared to the optical method of tracking currently used for basketball, hockey, and soccer, the RFID-based method has the advantage of recording player shoulder orientation, which is helpful for understanding the dynamics of a pass play. Player position, velocity, acceleration, and orientation is provided at 10Hz, and is available in near-real time via an API.

Overall, 45,501 pass attempts were used to train, and validate the models. The data was split between training, validation, and test sets to accurately measure model performance and minimize overfitting. Training (n=24,414) and validation (n=10,464) data were built as a 70%/30% split of passes from the 2016 and 2017 NFL regular and post-seasons. The test data was built from the 2018 regular season through week 12 (n=10,623). All results stated here are from the 2018 test set unless otherwise specified.

3.2. Architecture

The general model consists of a four-layer network as depicted in figure 2:
-An input layer of 230 nodes
-A dense hidden layer of 256 neurons.
-A second dense hidden layer of 128 neurons.
-A dense output layer, with activation dependent on the type of target variable the model is being asked to estimate.



Each dense layer, aside from the output layer, uses rectilinear units (relu) for nonlinear activation. Batch normalization and dropout are employed between each dense layer to prevent overfitting and improve performance when generalizing with out-of-sample data.

Based on the insights of [5], the order of inputs to the model were given meaning and kept consistent. For each play, the the set of features associated with each of the five potential target receivers (position, velocity, acceleration), were ordered from shallowest to deepest downfield. Within each set of features for each receiver, the features associated with the two nearest defenders were included.[1] Within the set of features for each receiver, the features of the nearest defender to

[1] Note that DeepQB is intended to analyze plays primarily from the perspective of the quarterback. Passes classified as "drops" by ESPN's VAT analysis are counted as successful completions from the passer's point of view. Additionally, all inputs are as of the time of pass release, so factors such as

that receiver appears first in the input vector, and the features of the next-nearest defender appears second. An additional set of features specific to the quarterback was concatenated to the inputs. These features include position and velocity information. The orientation of the quarterback was deliberately excluded from the inputs. The orientation of the passer's shoulders at the time of pass release would be "cheating" for this model's purposes, as it would very often correlate strongly with the direction of the intended target.

Lastly, play information was concatenated to the input vector. This information included down, distance to gain, and yard line of the play.

The training set was augmented with its mirror image plays, rotated about the longitudinal axis of the field of play, effectively increasing sample size. The assumption is that target selection and other outcomes of a pass play are invariant with respect to lateral symmetry. A receiver open on the left side of the field would be just as open if he were on the right side. The drawback is quarterback handedness, especially if he throws while on the run. Despite this limitation, augmentation improved out-of-sample performance.

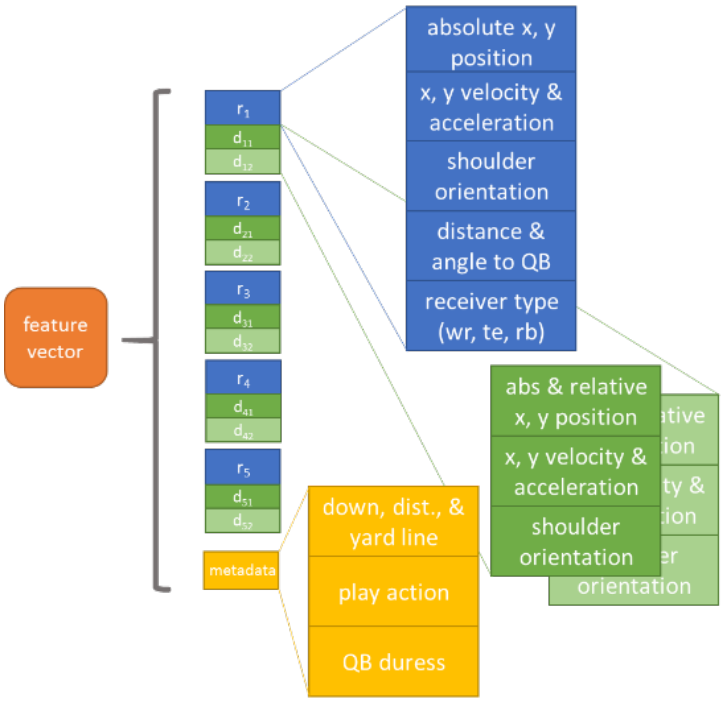


Figure 3 The feature vector used as input for the general model.

3.3. Learning

Each variant of the model was trained using Keras [8] with a TensorFlow backend. Training was stopped at the optimum performance on the validation set. Model hyper-parameters were selected using a grid search procedure.

4. Results

Results from each variant of the model are presented here, along with general insights.

4.1. Variant 1: Target Probability

The primary purpose of this variant is to verify that this type of model can truly make sense of the tracking data of a pass play. In essence, the model answers the question, "Given the picture presented to a typical quarterback, who would he choose to target?" The model predicts the actual

---

the proximity of defenders *at the time of pass arrival* are intentionally excluded, even if they would improve prediction accuracy.

target with an accuracy of 59.8% – three times the naïve estimate for a one-in-five proposition. This performance should be considered in light of actual football tactics. Offenses go to great lengths to conceal the intent of their play designs, while defenses endeavor to not allow any single receiver from being a disproportionately rewarding pass target.

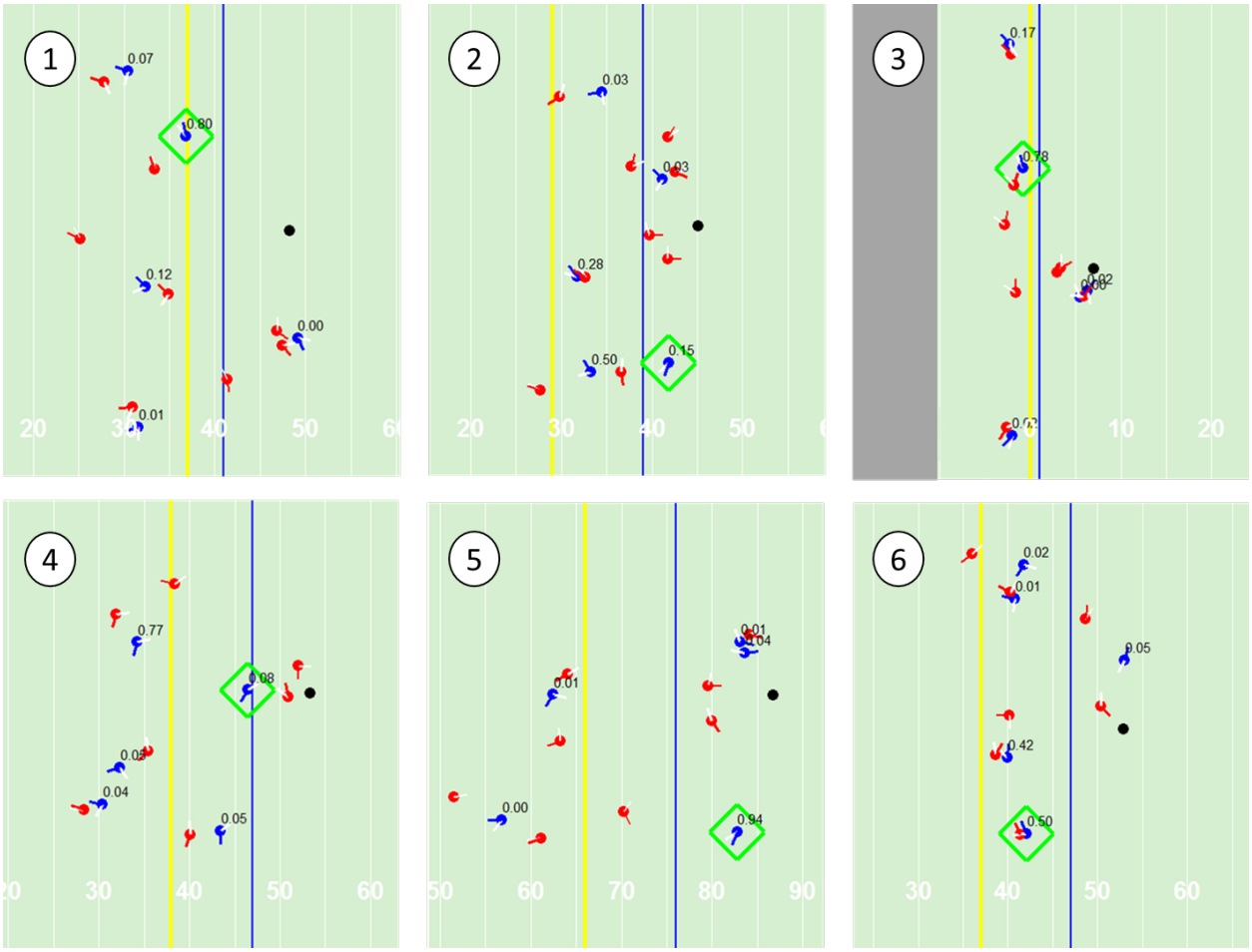


Figure 4 Actual plays from 2018 with associated receiver target probability. The white vectors show orientation. Actual targets are indicated by green diamonds. Play (1) illustrates an obvious target with a high probability. Play (2) shows a check-down, although DeepQB recommends the deeper emerging target with p=.50. Play (3) demonstrates the model's ability to see the proper target on goal-line plays. Play (4) suggests another targeting error by the QB. Play (5) shows a logical check-down, as all other receivers are smothered. Play (6) illustrates that defender proximity alone does not determine if a receiver is "open." Relative position and velocity are key factors.

Further, when quarterbacks target the receiver as predicted by DeepQB, the completion rate is 74%. This compares to a completion rate of 55% when quarterbacks target any other receiver. These results indicate that the model can successfully make inferences using the tracking data, and is capturing some degree of understanding of each pass play.

However, the Yards Per Attempt (YPA) when quarterbacks target the predicted receiver is significantly lower than when he targets another receiver (7.8 vs 8.0 YPA). This is a surprising and counterintuitive result. If completion percentages are higher, how could YPA be lower? This leads

to the first major insight: the typical (league average) quarterback is most likely too cautious in his reads. The theory that explains both results is that the typical quarterback favors a safer choice at the expense of optimum efficiency.

4.2. Variant 2: Expected Yards

The second variant of the DeepQB model directly estimates the expected yards given the overall picture that is presented to the quarterback. In addition it estimates the expected yards given that the quarterback targets each of the five eligible receivers. This allows us to evaluate individual quarterbacks, as well as their team's receiving corps and passing schemes.

Overall quarterback performance can be assessed by measuring the difference between their actual YPA and the expected YPA produced by the overall picture of receivers and defenders presented to the quarterback on each pass play. This reveals an estimate of the value added by the quarterback, over and above the capabilities that his receiving corps and scheme provide. For example, a quarterback may have a low YPA, but his binding constraint may be scheme or receivers.

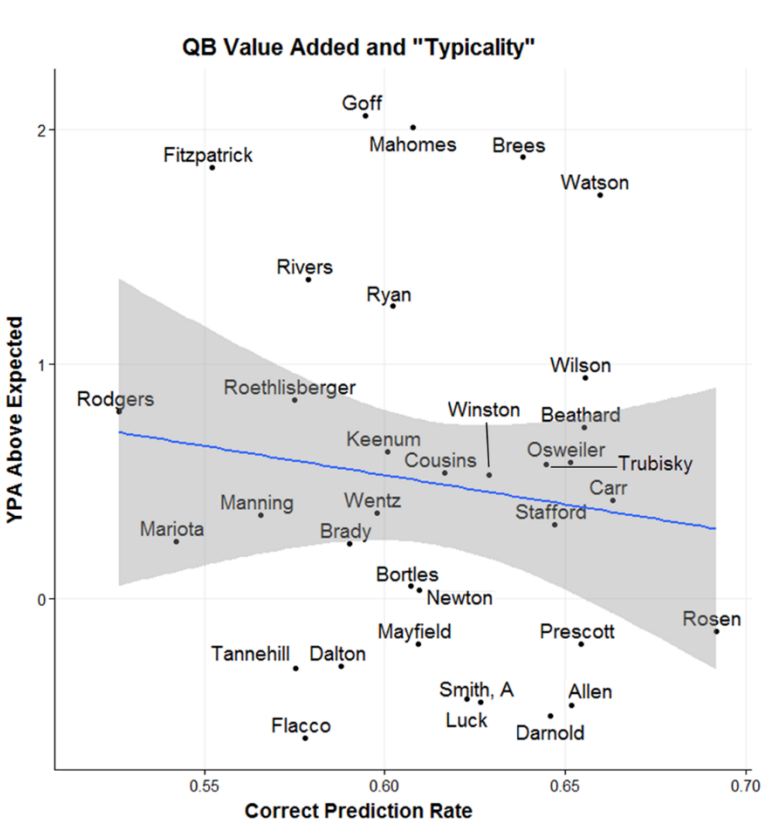
For each play, quarterback decision-making can be assessed by determining whether the quarterback targeted the receiver with the maximum expected yardage. Perhaps a better way to measure the same idea in aggregate is to calculate the proportion of the maximum available expected yardage represented by the expected yardage of the receiver the quarterback actually targeted. (For example, if the maximum available expected yardage on a play was 9.0 yards, and the quarterback targeted a receiver with 7.5 expected yards, the proportion would be 7.5/9.0 = 83%.)

Table 1 lists the results for all quarterbacks in 2018 (through week 12) with at least 80 pass attempts, along with the expected YPA predicted by the model. The *YPA above/below expected* column is the difference between actual and expected YPA. The proportion of maximum expected YPA targeted for each quarterback is also listed.

Table 1 Comparison of Expected YPA to actual YPA for 2018 quarterbacks with at least 80 targeted pass attempts through week 12. The top and bottom 5 passers are listed.

Rank	QB	Expected YPA	YPA above/ below expected	Proportion of max YPA available		Rank	QB	Expected YPA	YPA above/ below expected	Proportion of max YPA available
1	Goff	7.8	+2.1	70.3%		33	Smith, A	7.6	-0.4	66.7
2	Mahomes	7.9	+2.0	65.6		34	Allen	7.0	-0.5	65.5
3	Brees	7.1	+1.9	66.7		35	Darnold	7.1	-0.5	63.0
4	Fitzpatrick	7.6	+1.8	70.1		36	Flacco	7.7	-0.6	69.5
5	Watson	7.2	+1.7	69.8		37	Taylor	7.6	-1.5	70.4

Figure 5 plots the YPA Above Expected metric against the predictability of each quarterback, which is measured by how often variant 1 (target prediction) is correct. This is essentially a measure of how "typical" a quarterback is in his targeting decisions. Notice that many of the rookies (Allen, Darnold, Rosen) are toward the right of the chart (highly typical), with Mayfield as the exception. Some of the more aggressive passers, such as Aaron Rodgers, are toward the left. As discussed with Variant 1, there is a weak negative correlation between the "typicality" of a quarterback's targeting decisions and the value he adds above the potential provided by his receivers(r=-0.30).



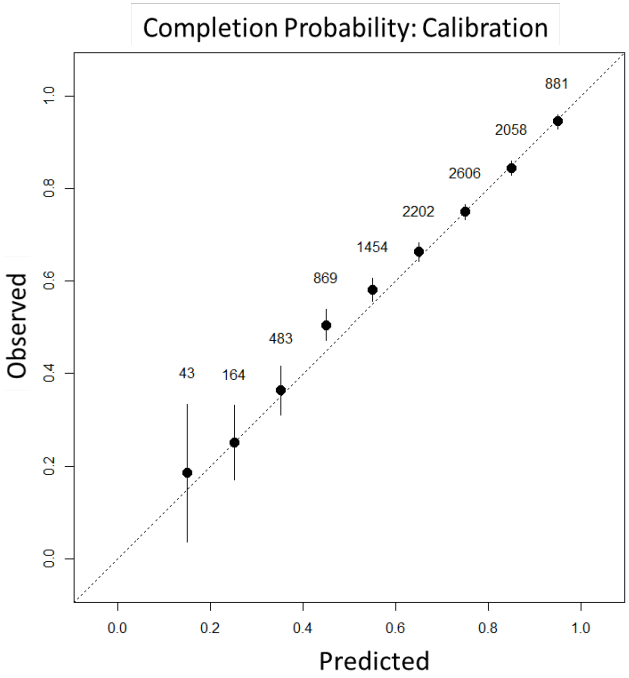
Unfortunately, we cannot escape the possibility of selection bias. For example, Aaron Rodgers may attempt more unorthodox, aggressive targeting *because he can*, while other quarterbacks may not have the skills to do so reliably.

4.3. Variant 3: Pass Outcomes

The third variant of our model estimates the probabilities of completion, incompletion, and interception for each play.

Considering just completions and incompletions, the model appears well-calibrated (figure 6), although it rarely considers pass completions as highly improbable. This is substantially due to the exclusion of deliberate "throw-away" passes from the data, which are technically attempts but did not have a targeted receiver.

Table 2 Modeled outcome probabilities by actual outcome types.



Passing accuracy is a widely misunderstood attribute, primarily because completion percentage is used ubiquitously as a proxy for accuracy. Completion percentage does not account for the difficulty of the throw, nor does it exclude "throw-away" pass attempts and receiver drops. This variant of DeepQB can better isolate and assess the accuracy component of passing by calculating a quarterback's expected completion rate based on each play's dynamics and which receiver was targeted, and then comparing that to his actual completion rate. Note that these numbers are on targeted passes only, so they will be higher than common completion percentage statistics.

Actual Outcome	Predicted Completion Probability	Predicted Interception Probability
Complete	79%	1.70%
Incomplete	56%	3.40%
Intercepted	57%	4.40%

Of particular interest are interceptions, which have large impacts on the game, and doubtlessly convey a good deal of information about quarterback decision-making. Interceptions are events with a very low base rate and often involve tipped balls or other random factors. The absolute number of interceptions, or even interceptions per pass, will therefore be very noisy measures of a quarterback's true propensity to throw them. An accurate estimate of interception probability would be a truer measure.

The play snapshots in figure 7 portray examples of how the model assesses interception probability. Notice that the first two examples show a risky configuration of defenders, and the interception probabilities are accordingly very high. The third example shows a safe and open pass, with a correspondingly lower probability of interception.

Table 3 Passing accuracy as assessed by each quarterback's actual completion rate above the expected rate estimated. The top 5 and bottom 5 qualified passers of 2018 through week 12 are listed here.

Rank	QB	Expected Compl Rate	Compl Rate Abv Exp		Rank	QB	Expected Compl Rate	Compl Rate Abv Exp
1	Brees	67.0%	15.0%		33	Bortles	67.6%	-0.2%
2	Watson	63.0	8.9		34	Rosen	68.1	-0.8
3	Cousins	67.8	8.6		35	Beathard	71.2	-2.1
4	Winston	66.8	8.2		36	Darnold	68.1	-5.7
5	Mahomes	68.4	7.6		37	Taylor	69.3	-10.5

Mean interception probability for the top and bottom five individual quarterbacks are listed in table 4. Keep in mind that game situation, primarily time and score, has a large influence on the optimum

risk level, and that is not accounted for in the aggregate rates. However, the model's estimates can provide even more valuable insight on a pass-by-pass basis.

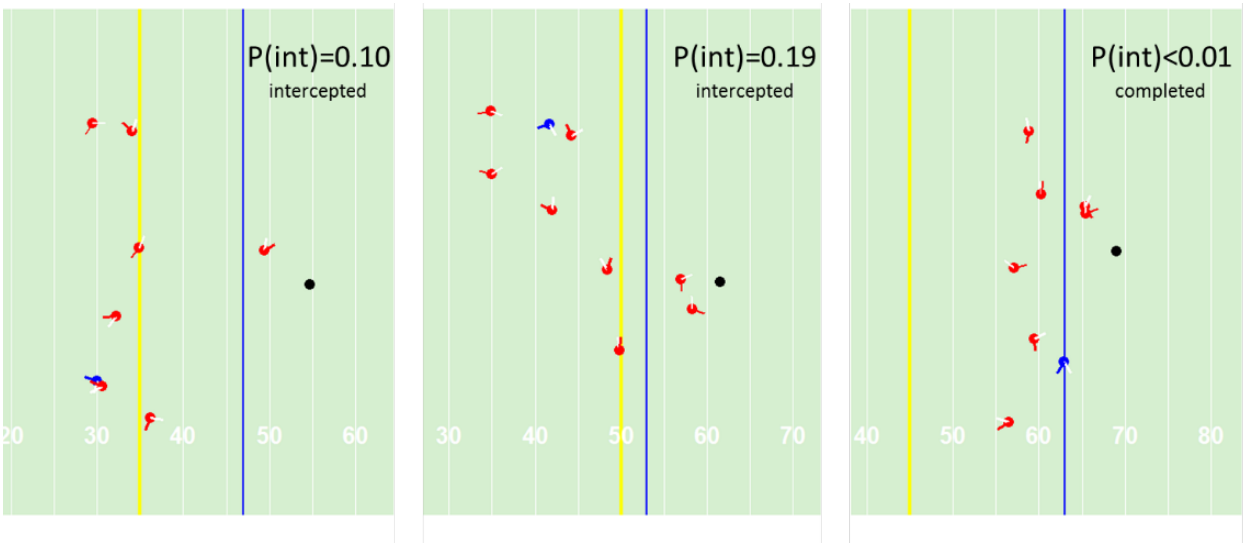


Table 4 Mean interception probabilities, along with actual rates, for the top 5 and bottom 5 qualified passers of 2018 through week 12.

Rank	QB	Mean Int Prob	Actual Int rate		Rank	QB	Mean Int Prob	Actual Int Rate
1	Foles	1.23%	1.2%		33	Mayfield	1.97%	2.3%
2	Trubisky	1.31	2.8		34	Stafford	1.99	2.5
3	Beathard	1.34	4.1		35	Fitzpatrick	2.00	4.9
4	Keenum	1.37	2.6		36	Watson	2.03	2.7
5	Wentz	1.38	1.8		37	Winston	2.28	5.4

Assuming the pass outcome estimates from the model are reasonable, an important insight is revealed. Because expected interception rates, based on the geometries of the receivers and defense, are lower than the actual interception rates of high-interception quarterbacks, it follows that interceptions in the NFL are primarily a result of inaccurate throws and random sample error rather than bad targeting decisions.

4.4. Variant 4: Individual Quarterback Models

The final variant of the DeepQB suite is a highly experimental version intended to capture the decision-making tendencies of individual quarterbacks. The sample size of pass attempts for individual quarterbacks is not sufficient to adequately train a model to accurately generalize on out-of-sample cases. However, it is possible to capture individual decision-making using transfer learning.

Transfer learning is a technique that trains lower layers of a neural network on a broad training set, but trains the higher levels of the network on a more specific set of cases. In this context, the

weights of the lower levels of the neural network are trained on plays from the full set of quarterbacks. The base level of the network represent the fundamental relationships among positions and velocities—understanding and projecting the geometries of the players on the field, while the higher levels of the network represent the executive functions—choosing the passing target.

This is analogous to a biological brain in that most athletes share similar understandings of instinctive, intuitive physics—basics like gravity, distance, and motion. But individuals may differ in their dispositions toward various decisions despite having common intuitive physics. This variant of the model exploits such instinctive commonalities by training its lowest hidden layer using a very large sample of passes by all quarterbacks, and then freezing the resulting weights of that layer. Then, the network's higher decision-level layers were trained on the smaller sample of pass attempts of a single quarterback. The resulting network is a model that combines common intuitive physics shared by all quarterbacks with the executive-level decision-making of an individual passer.

Just one example of such a model is presented here, that of Saints' quarterback Drew Brees. The two higher layers of Variant 1 of the DeepQB model (target selection & prediction) were retrained using just Brees' passes. Overall accuracy specific to Brees' 2018 pass attempts rose to 64%.

It is interesting to see how other quarterbacks might compare to a cyber-Brees. Table 5 lists the passers of 2018 whose target selections align the best and worst with the Brees model.

Table 5 How passers' targeting tendencies align with "cyber-Brees" based on model variant 4.

Rank	Quarterback	Target Prediction Accuracy		Rank	Quarterback	Target Prediction Accuracy
1	Brees	64.3%		33	Goff	49.7%
2	Smith, A	57.5		34	Mahomes	49.2
3	Allen	57.3		35	Wilson	48.7
4	Carr	57.1		36	Ryan	47.7
5	Darnold	56.3		37	Brady	46.1

Some of the worst performing quarterbacks of the season are most aligned with the Brees model, while most of the best quarterbacks of the season are the least aligned. This should make one skeptical of these results. This is an admittedly experimental application, but one plausible interpretation of the results is that Brees has frequently targeted his running backs, similar to how poor offenses rely often on shallow check-down routes to complete passes. In Brees' case, this is likely by design due to the exceptional receiving skills of his backs, but for many others this tactic tends to be a last resort.

5. Limitations and Future Development

All variants of the model currently rely solely on a binary "QB duress" variable to account for the constraints on the QB due to the pass rush. This modeling decision was made at the outset of the project to ensure tractability of the model, but since then it became clear the model can accept a larger set of inputs without causing a prohibitively long training time. Future iterations of the

model will use a more complete picture of pass protection based on advances in modeling pass protection [9]. This should enhance accuracy, as quarterbacks are sometimes not able to target open receivers due to being obscured by pass rushers.

Limiting the picture to the two nearest defenders to each receiver was another modeling decision made for the same reason. Review of hundreds of output plays indicated the model could still piece together a relatively complete picture from this structure—understanding the presence of all defenders to all receivers in the feature vector. However, future versions of the model can include up to all eleven defenders and their relationships to each receiver.

Receiver talents and skills are not fully accounted for. Although aspects of receiving such as route running and speed can be captured by the model, other aspects such as the size and strength of receivers is a factor in the probability of reception and the yardage estimate of each play.

Interceptions are a critical aspect to decision-making. Although Variant 3 addresses interception probability directly, Variant 2 addresses each pass play's expected yards separately from interception chances. This may make some passing targets appear more lucrative on net than is the case due to a possibly high chance of interception. Future versions of this approach will incorporate concepts such as Expected Points Added (EPA) [10], which accurately measures the complex interactions of down, distance, field position, and possession, and is the consensus utility function of football events.

The general model is based on a snapshot of the play at the time of pass release. Open receivers earlier in the development of a play may later become covered by the time of release if the quarterback fails to recognize them. Indeed, this is a natural consequence of how the position is taught. Quarterbacks are typically instructed to make their reads in a progression, from one primary receiver to a secondary, and then to an alternate. This model may be penalizing a quarterback who makes the correct read in the planned progression, but because a read becomes open after his step in the progression, it appears that the quarterback has missed an open receiver. Future variants of the DeepQB model may use techniques such as an LSTM network to assess passing outcomes on a continuous basis throughout the play.

6. Summary

This paper proposes a new approach to examining passing in professional football. Despite the chaos and deception intrinsic to the sport, a neural network approach to player tracking data performs surprisingly well, and has been demonstrated as an effective tool for analysis. The model demonstrates how quarterback effectiveness and decision-making can be assessed both for individual plays and in aggregate. Variants of the DeepQB suite of models focus on target selection, expected yardage, and pass outcomes. A fourth experimental variant explores the possibility of tailoring the model to individual quarterbacks using transfer learning. This overall approach offers the promise of making a previously opaque aspect of the sport accessible to teams, media, and fans.

References

[1] 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.

[2] Kuan-Chieh Wang, Richard Zemel. Classifying NBA Offensive Plays Using Neural Networks. *MIT Sloan Sports Analytics Conference 2016*. 2016.

[3] Le, H., Carr, P., Yue, Y., Lucey, P. Data-Driven Ghosting using Deep Imitation Learning. *MIT Sloan Sports Analytics Conference 2017*. 2017.

[4] Harmon, M., Lucey, P., Klabjan, D. Predicting Shot Making in Basketball using Convolutional Neural Networks Learnt from Adversarial Multiagent Trajectories. *Association for the Advancement of Artificial Intelligence,* 2016.

[5] Lucey, P., Bialkowski, A., Monfort, M., Carr, P., Matthews, I. "Quality vs Quantity": Improved Shot Prediction in Soccer using Strategic Features from Spatiotemporal Data. *MIT Sloan Sports Analytics Conference 2015.* 2015.

[6] Mehrasa, M., Zhong, Y., Tung, F., Bornn, L., Mori, G. "Deep Learning of Player Trajectory Representations for Team Activity Analysis. *MIT Sloan Sports Analytics Conference 2017.* 2017.

[7] Hochstedler, Jeremy. Finding the Open Receiver: Quantitative Geospatial Analysis of Quarterback Decision-Making. *MIT Sloan Sports Analytics Conference 2016.* 2016.

[8] Chollet, F. (2015) Keras, GitHub. https://github.com/fchollet/keras

[9] Burke, Brian. "We created better pass-rusher and pass-blocker stats: How they work." espn.com/nfl/story/_/id/24892208/creating-better-nfl-pass-blocking-pass-rushing-stats-analytics-explainer-faq-how-work. 2018.

[10] Burke, Brian. "Expected Points and Expected Points Added." Advanced Football Analytics. *archive.advancedfootballanalytics.com/2010/01/expected-points-ep-and-expected-points.html*. 2010.