A Disturbance in the Force? Modeling QB Pressure with Force-based Metrics

Annie Liang, Harrison Lian, Jingchun Quan, Joshua Choi
Project Advisor: Ron Yurko

Background & Introduction

Quarterbacks often get all of the attention, but a key to their success often lie in the linemen in front of them. At the beginning of a play, the offensive and defensive line up against each other at rest. As soon as the play is in motion, the defensive line attempts to break through to pressure the quarterback while the offensive line works to stave them off. Over the course of the play, both teams consistently face off to push for their desired outcome. We are interested in seeing how the position and acceleration of offensive and defensive linemen impact the outcome of the quarterback getting hurried, hit, or sacked. To do so, we analyzed player, play, game, scouting, and tracking data from NFL and Pro Football Focus in order to create features that would be predictive of a negative outcome and reveal new insights into how football coaches can integrate this information into plays.

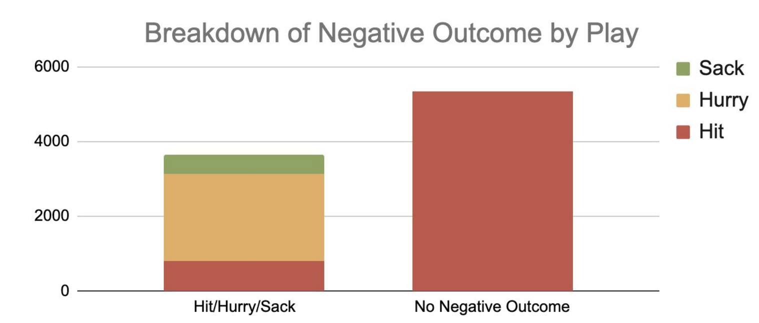


Figure 1: 38% of plays have a negative outcome, divided among hit, hurry, and sack

Figure 2: Negative correlation between weight and acceleration of pass blockers and rushers

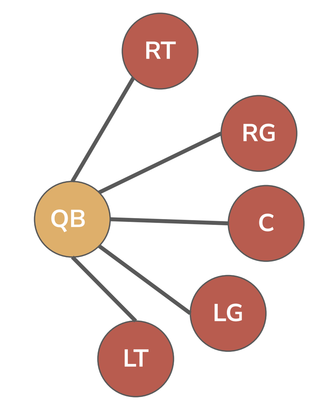
Overall Goal: How does the position and acceleration of offensive and defensive linemen in a play impact the outcome of the quarterback getting hit, hurried, or sacked?

Motivation behind using force features: as the weight of a player increases, there is an inverse relationship with the maximum acceleration that the player can exert. Furthermore, we also see that pass blockers or the offensive linemen are generally heavier and have lower acceleration than pass rushers or the defensive linemen. This relationship motivated us to look at force as a predictive feature since it takes both acceleration and weight into account.

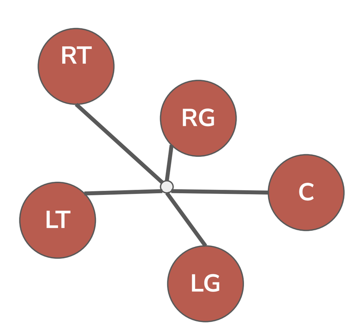
Feature Engineering

1. Distance/Area	2. Forces exerted by pass blockers and rushers		
	Net X & Y Force	Distance Weighted	Partitioned
• Linemen distance to QB • Distance between linemen • Area formed by linemen	• Forces exerted by offense and defense linemen	• Net forces weighted by inverse distance to QB	• Partitioning the field into three areas based on position and the forces exerted by linemen in each partition

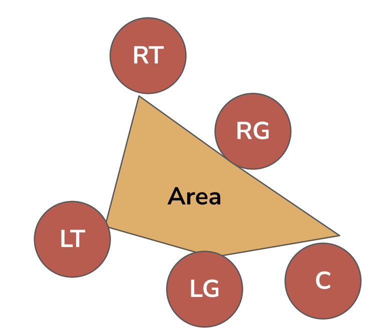
Average distance to the QB:



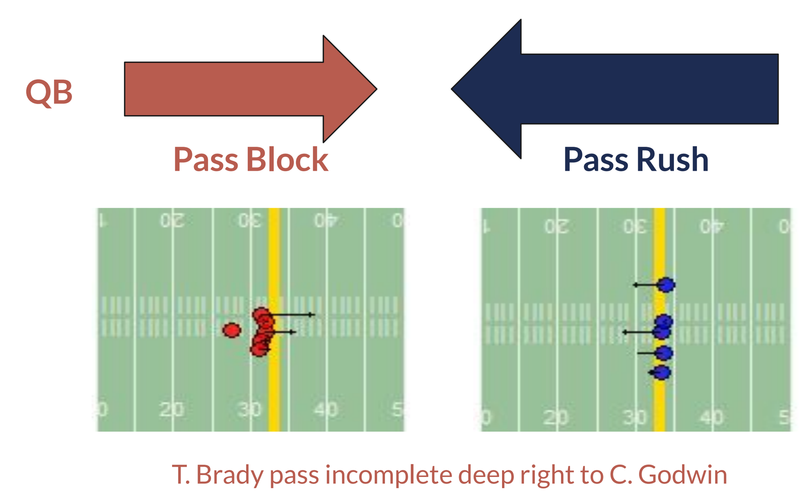
Average distance to the center point of the linemen:



Area enclosed by linemen through shoelace method:



Hypothesis: If defense exerts greater force, higher chance of negative outcome for QB

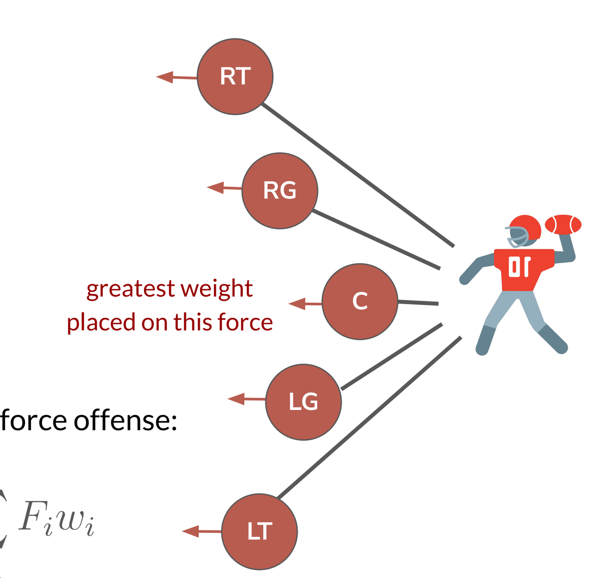


1. Calculated force exerted by player
2. Determined x and y forces exerted by direction for pass rushers and pass blockers
3. Force exerted was summed together to get net force
   a. Net force > 0: offense exerted more force
   b. Net force < 0: defense exerted more force

Distance weighted net forces

$$w_i = \frac{1/d_i}{\sum_{j=1}^{n_o} 1/d_i}$$

- n_o = number of pass blockers
- n_d = number of pass rushers
- d_i = distance of player i from QB (yards)
- w_i = weight of player i (0-1)
- F_i = force of player i



- Weighted force defense: $n_d \sum_{i=1}^{n_d} F_i w_i$
- Weighted force offense: $n_o \sum_{i=1}^{n_o} F_i w_i$

Partitioned net forces

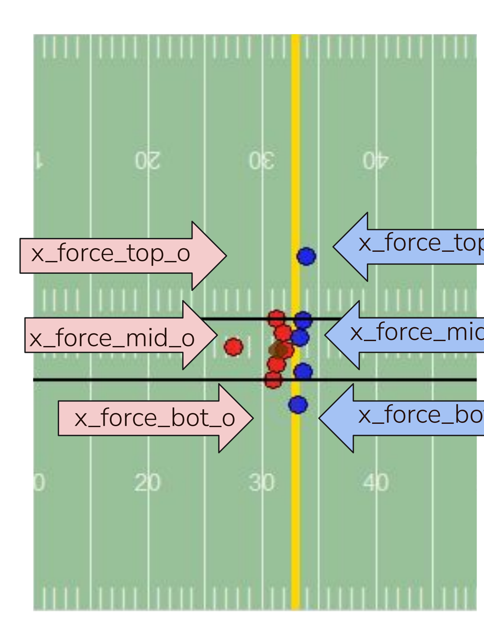


Figure 3: Net force features partitioned by guard locations at start of play

Analysis & Modeling

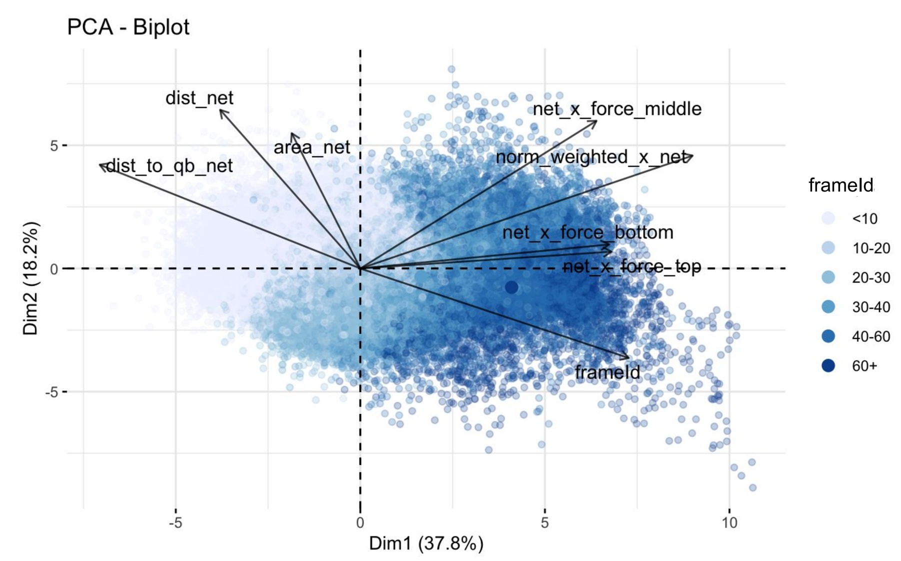


Figure 4: PCA analysis of engineered features

- Force features and distance features are orthogonal, revealing different sources of variability in the data
- Dist_to_qb and frame_Id have strong negative correlation, since as plays go on, players tend to move closer to the QB, giving more opportunities to hit/hurry/sack

1. Fit XGBoost and GLM as baseline models on the entire dataset using frameId + all engineered features as covariates

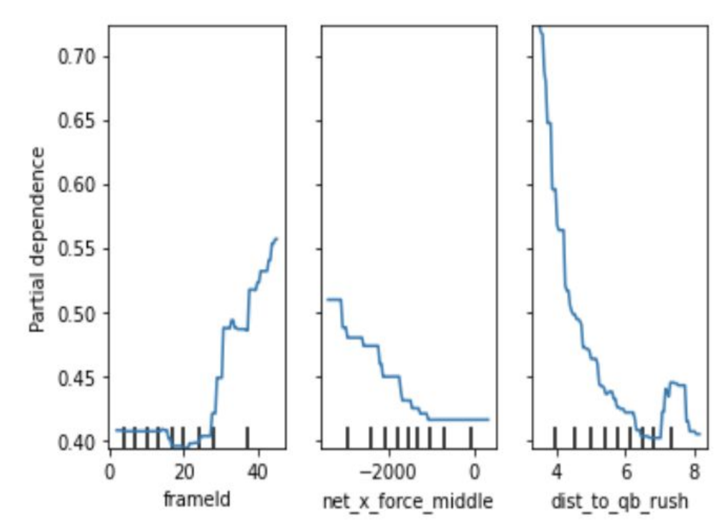


Figure 5: Partial dependence plots of top 3 variables

- In a Oct 7th, 2021 Rams vs Seahawks sample play, as the defensive line starts to exert greater force and they move closer to Wilson, the probability of a negative outcome increases
- When Lewis ultimately sacks Wilson, our features indicate the chances of the negative outcome/sack to be at 95%.

- Top 3 features in terms of information gain: frameId, net_x_force_middle and dist_to_qb_rush
- The likelihood of a negative outcome occurring rises as frame ID increases and defensive distance to quarterback decreases, confirming our hypothesis from PCA analysis
- As the net force in the x direction exerted in the middle of the field becomes more negative, the defensive linemen exerts more net force than the offensive, increasing the likelihood for a negative QB outcome

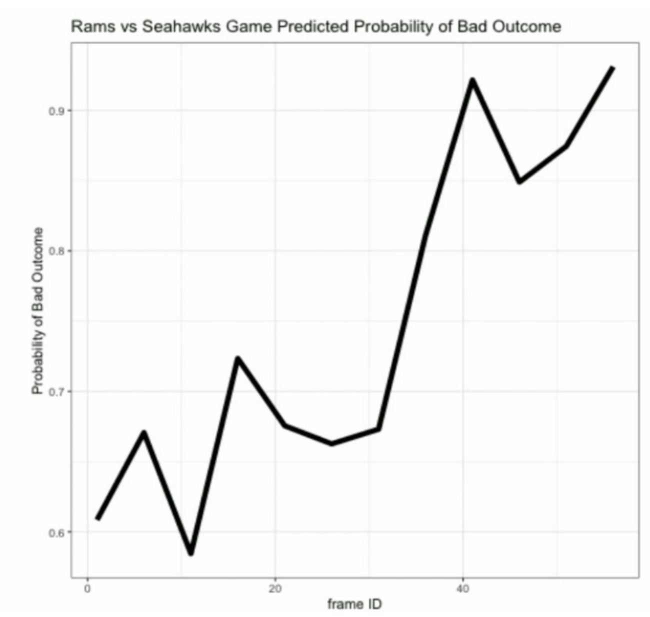


Figure 6: Changes in probabilities of bad outcomes over time in the Rams vs Seahawks sample play

2. Fit separate models (logistic regression and random forest) for each frameId
   a. **Football intuition:** FrameId is most predictive variable but not something that a coach/player can control
      i. We can take away the influence of frame ID by having separate models trained for each frame to better observe other features
   b. **Statistical intuition:** Response is at the play level but our observations are at the frameId level
      i. If we fit model on all frame IDs, we will be adding unnecessary error terms

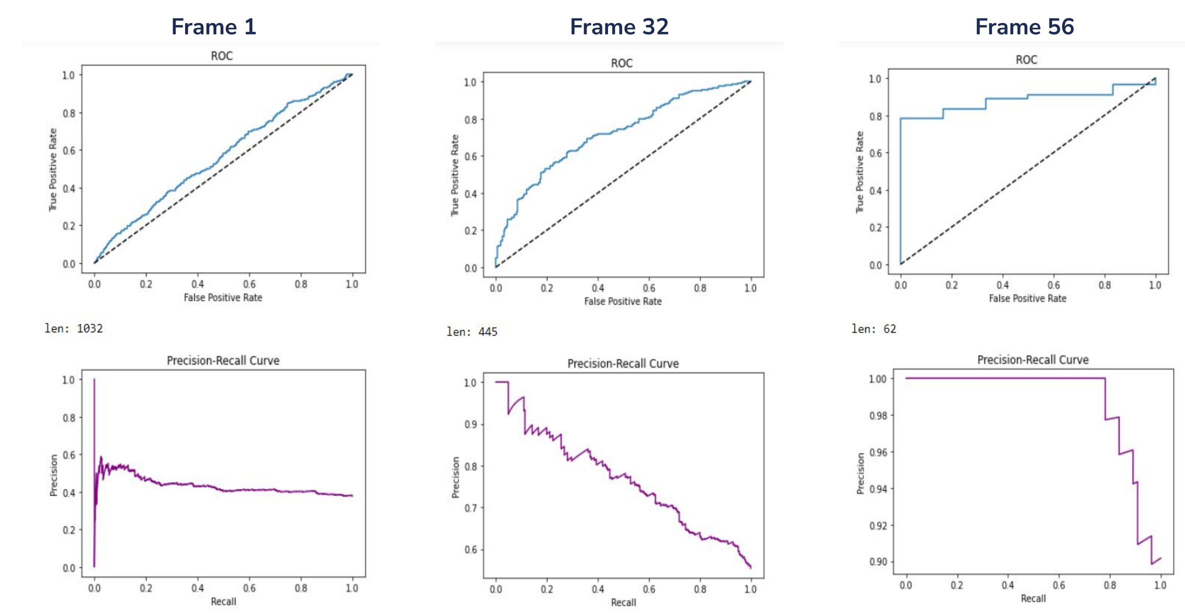


Figure 7: ROC and Precision-Recall Curves for selected frames from the random forest model

- Predictive accuracy ranges between 57 - 83% and increases with time elapsed in a play, along with the variance in predictions
- Models with all features had similar recall/accuracy to those with only distance or force features, pointing to the predictive ability of our force features
- Force features are measurable aspects of individual players (their mass + ability to accelerate), while distance to the QB within the play is a result of such traits

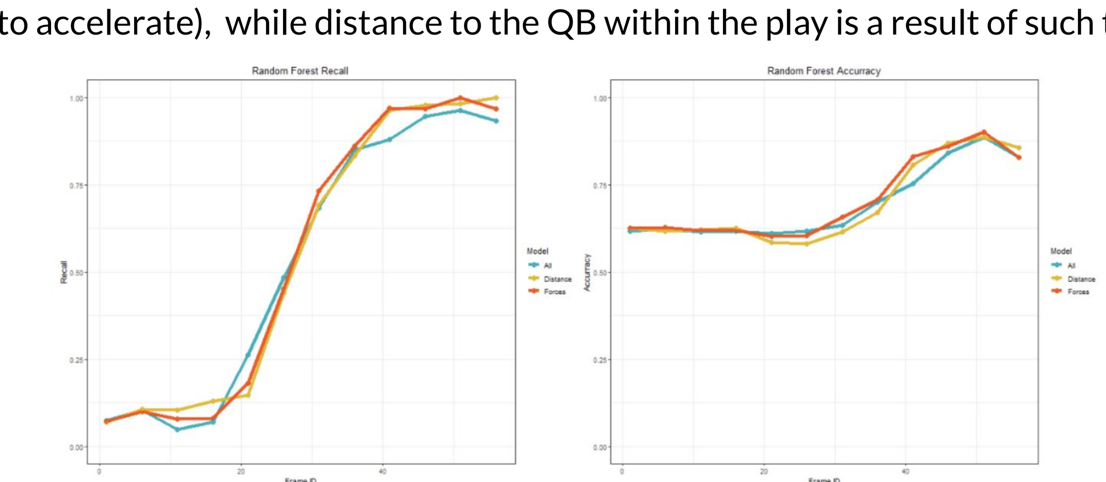


Figure 8: Tradeoff of accuracy and recall between models with all, distance, and just force features

Conclusion & Takeaways

To evaluate how well a team's defensive linemen worked together to inflict a bad outcome on the QB, we looked at the average maximum net force that the defense exerted along with the total number of hits, hurries, and sacks that were inflicted in the season. We found that defensive lines with a more negative max net force correlate to inflicting a greater number of bad outcomes, giving credit to the force feature approach. **The dataset we used included plays from the 2021 season, with the Rams winning. The Rams can be seen at the far left, exerting the highest average max defensive force out of any team and inflicting >120 bad outcomes.**

To conclude, our distance and force features provide an actionable technique for players & coaches to control occurrences of hits, hurries, and sacks. We made connections between measurable player-level attributes such as weight and force, which teams can assess, to distance and time attributes, which are result-based measurements that tend to evolve over time. Player-level insights can be helpful in evaluating where top force exerting defensive linemen should be located on the field to break through and provide a warning to the offensive linemen about which defensive player could hit, hurry, or sack the QB.

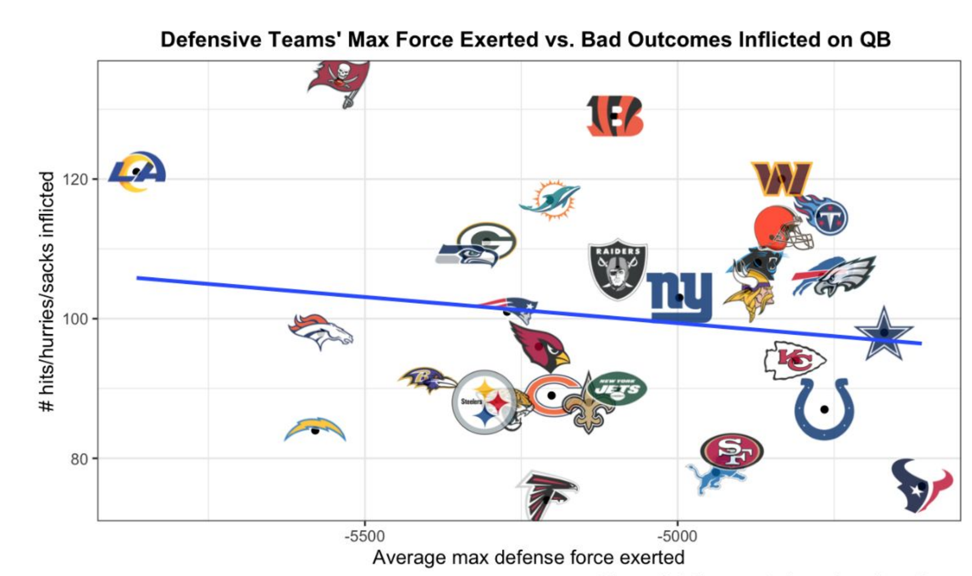


Figure 9: Relationship between forces and negative outcomes at the team level

Next Steps

There are several future improvements that could be made to our work:

1. We grouped hits, hurries, and sacks together due to the class imbalance of each negative outcome. Future work would focus on the impact of our force features on each specific play outcome. For instance, coaches and players are often more interested in the occurrence of sacks and how they can prevent or push for that outcome
2. Our work focuses more holistically on the force displayed by a team. Future research can focus on the occurrence of a bad outcome given when certain players are matched together or face off
3. We can explore models that better capture the autocorrelation between frames such as modeling the occurrence of a bad outcome in next 10 frames instead of at the end of the entire play