

# Analysis of NFL Combine Metrics ('04-'23)

Ian Hill

December 16, 2024

## 1 Introduction

The NFL Scouting Combine is a pivotal event in the evaluation of prospective players, allowing teams to assess physical and athletic capabilities before the draft. This project examines the relationship between combine performance metrics and draft outcomes, seeking to identify patterns that could guide future scouting and drafting strategies and predict player performance in the NFL Kuzmits and Adams [4]. By analyzing historical data, this study aims to answer three key research questions:

1. How do combine event performances correlate with draft round?
2. What is the relationship between players who were drafted and their combine performance?
3. Can positions be predicted based on combine performance, and, which positions excel in specific events?

The dataset used in this study contains NFL Scouting Combine results from 2004-2023. It was collected publicly from Sports Reference StatHead Football Hawk, Shuckers, and Lock [1]. It's records represent individual players and Relevant variables include round drafted ('Round') with rounds 1-7 indicated accordingly and undrafted players being represented by (8), position ('position'): QB, and performance metrics from the combine such as 40-yard dash ('forty'): 4.79, vertical jump ('vertical'): 30.5, bench press repetitions ('bench\_reps'): 21.0, broad jump ('broad\_jump'): 110.0, three-cone drill ('three\_cone'): 7.66 and shuttle drill ('shuttle'): 4.41. NaN values in combine events were removed according to model needs.

The analysis focuses on exploring the predictive values of these variables. Metrics like the 40-yard dash and vertical jump are hypothesized to be strong indicators for speed and explosiveness, particularly for skill positions such as wide receivers and runningbacks Hedlund [2]. Conversely, the bench press and shuttle drill are expected to be more relevant for strength and agility in positions like offensive lineman Hedlund [2].

Through data visualization, statistical analysis, and machine learning techniques, this project aims to provide actionable insights into the scouting process. By understanding how combine metrics relate to draft outcomes and player positions, teams can optimize their draft strategies to build stronger, more competitive rosters.

## 2 Methods

This study utilized NFL Scouting Combine data from 2004 to 2023 to examine the relationship between player performance metrics and draft outcomes. The dataset was cleaned to address missing values: undrafted players were assigned a draft round value of 8, while missing combine performance values were retained until necessary for specific analyses.

To explore relationships between variables, visualizations such as scatterplots, correlation matrices, histograms, and violin plots were created using 'seaborn' and 'matplotlib'. These visualizations provided insights into trends like differences in combine performance by position and draft outcomes. Machine learning models from 'scikit-learn' were applied to predict outcomes, including: Random Forest Classifier and Decision Tree Classifier for predictive modeling and feature importance analysis. Randomized Search CV for hyperparameter optimization. Hist Gradient Boosting Classifier for efficient modeling with large datasets.

Model evaluation used metrics such as accuracy, precision, recall, and ROC curves. Prediction tests validated observed differences between drafted and undrafted players across combine metrics. These analyses provided actionable insights into the predictive power of combine metrics, supporting NFL teams in refining their scouting and draft strategies.

### 3 Results

#### 3.1 Combine Performance Related to Round Drafted

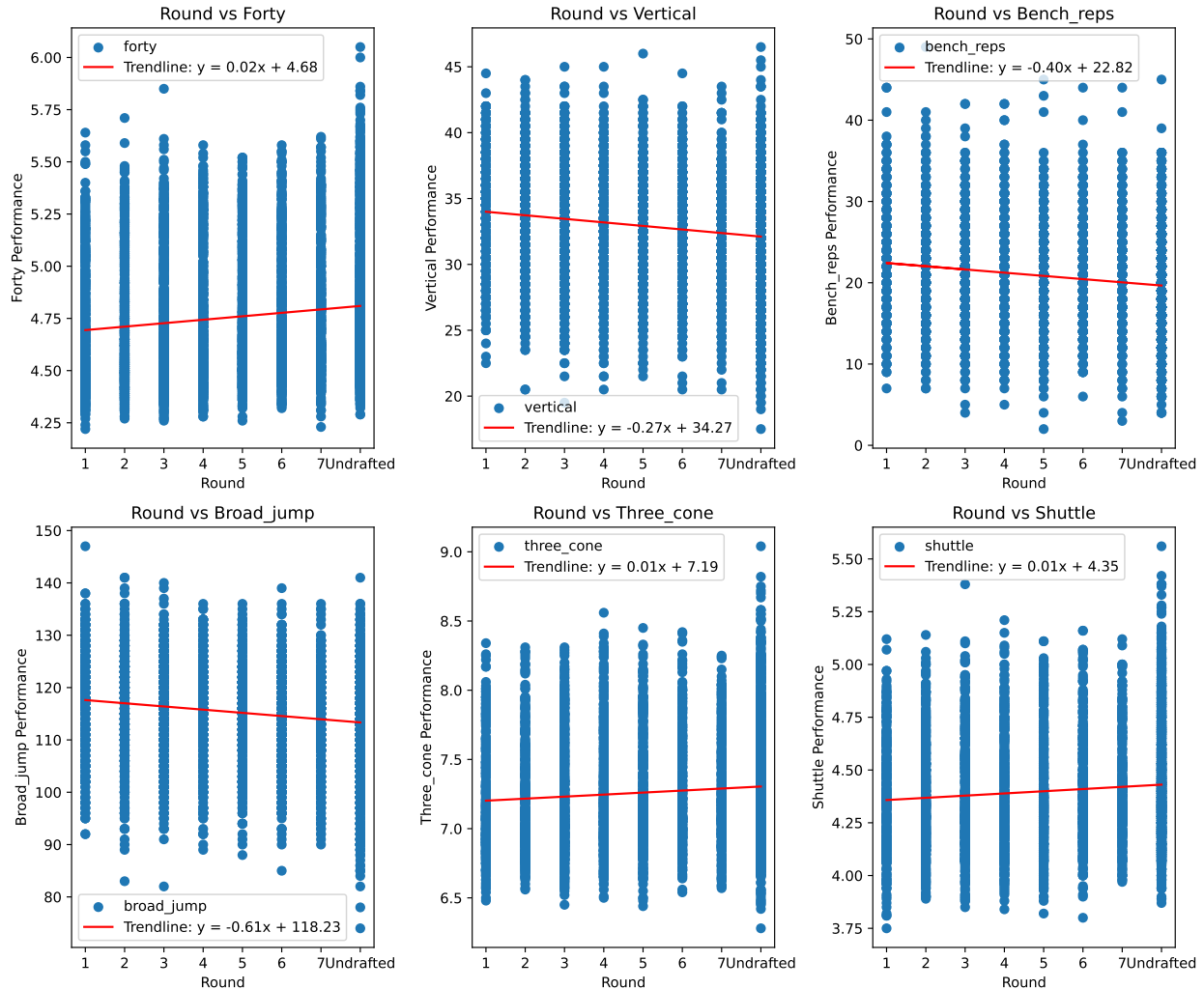


Figure 1: Relationship of Round Drafted and Resulting Combine Performance

When it comes to the combine, the results of a player's performance can ultimately cause a change in the round they are drafted in. Because of this, it was important to explore the change of combine performance over draft rounds.

In Figure 1, six scatterplots were created to examine the relationship between each combine event and the performance by round drafted. As is expected faster and stronger players tend to get drafted in earlier rounds while weaker and slower players tend to fall and ultimately can become undrafted.

At first glance, the trends between each graph can be confusing if not familiar with each combine events scoring design. The 'forty', 'three\_cone' and 'shuttle' drills all measure on a timed assessment. Therefore, a "higher" time results in a slower player. As we can see in graphs 1, 5 and 6,

the correlation between round drafted and time in these events is positively correlated. As time increases, so does a player's potential draft round. The trendline formulas show this relationship with positive slopes:

- 'forty' |  $y = 0.02x + 4.68$
- 'three\_cone' |  $y = 0.01x + 7.19$
- 'shuttle' |  $y = 0.01x + 4.35$

The opposite occurs for 'vertical', 'bench\_reps' and 'broad\_jump'. These events are scored off of purely physical measurements. For example, the height of a player's vertical jump or the amount of reps a player can perform on the bench press. In this case, a larger score results in a more athletic or stronger individual. It can be seen in plots, 2, 3 and 4 that the correlation between round drafted and physical event score is negatively correlated. As score decreases, round drafted increases. The trendline formulas show this relationship with negative slopes:

- 'vertical' |  $y = -0.27x + 34.27$
- 'bench\_reps' |  $y = -0.40x + 22.82$
- 'broad\_jump' |  $y = -0.61x + 118.23$

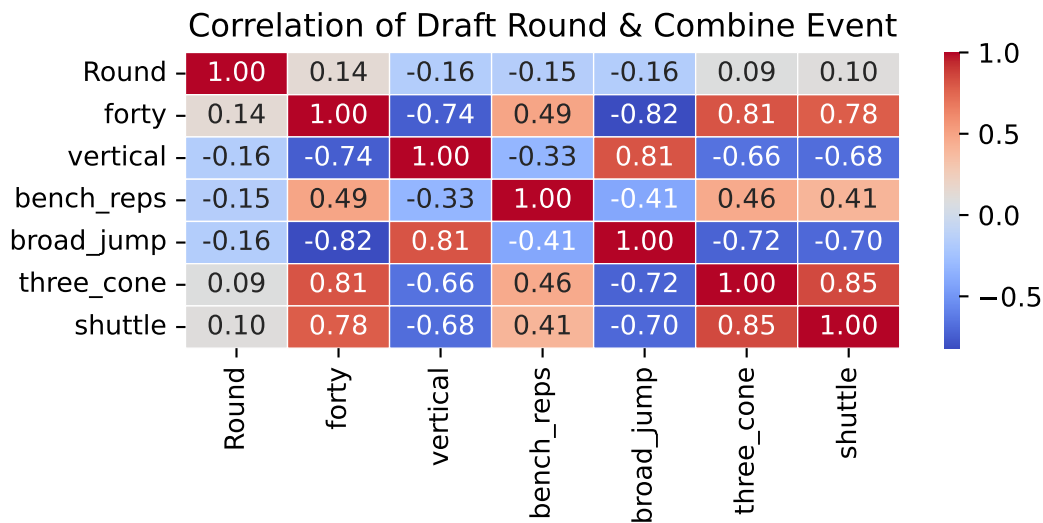


Figure 2: Correlation Matrix Observing Round Drafted and Combine Performance

In Figure 2, the top row is to be observed to see correlation between Round and combine events. Red is seen as positively correlated, while blue is seen as negatively correlated indicated by the legend to the right on a 0-1 scale. As can be seen Round vs ('forty', 'three\_cone' 'shuttle') are all positive and shaded in a faint red shade to indicate positive correlation. This is just as was seen in the scatterplots seen in Figure 1. Alternatively, ('vertical', 'bench\_reps', 'braod\_jump') are shaded in a faint blue and include negative values -0.16, 0.15 and -0.16 respectively, indicating a negative correlation between Round and these events.

To reiterate, it was seen as a player's timed event score increases so does their potential draft round indicated a positive relationship. In the player's case, this is not to be sought after. Teams value quick and agile players to perform at the highest level against the fastest competition. Meanwhile,

Table 1: Performance of GridSearchCV Tuned Decision Tree Predicting Draft Status Based on Combine Performance

Train Accuracy with Tuning: 0.6998706338939198

Test Accuracy with Tuning: 0.6758620689655173

it can also be seen that as a player's physical event score increases, there potential draft round decreases indicated a negative relationship. In this case a player should focus on achieving a higher score as it indicates a stronger and more athletic player which teams look for in the NFL draft. This relationship is utilized by draft scouts to determine where a player may get drafted according to there performance and can be scouted accordingly based on draft capital held by the team.

## 3.2 Predicting Draft Status Based on Combine Performance

There are 7 available rounds for a player to be drafted in the NFL draft. The chances of being selected are narrow to begin. However, several players go undrafted every year as teams continue evaluating a grabbing potential practice squad players to create a complete organization capable employing squad depth throughout the season

The distribution of drafted and undrafted players from 2004-2023 was 4023 drafted players and 2105 undrafted players in this dataset. Around one-third of players between 2004-2023 went undrafted. Many of those players will get a chance to perform individually for a team's roster spot, and many will have finally completed their football careers. Teams will often wait for players to fall out of the draft for them to be signed to a contract later on as to not waste a draft pick. Therefore, many teams employ classification techniques as to determine which players are likely to be drafted and thus should be taken early and not be given the chance to fall out of the draft as the chances of them being taken by another organization can increase based on their performance. Murran [6]

### 3.2.1 Tuned Decision Tree

First, a simple decision tree model was employed to predict undrafted(0) and drafted(1) based on combine performances from player records in the data. This model provided sub par results due to overfitting on the training data and was not included in this report so a hyperparametized tuned Random Forest Model was utilized to increase performance.

After hyper tuning the simple decision tree model the performance did increase. Seen in Table 1, training accuracy fell to ~0.67, while testing accuracy increased to ~0.66. This is much more consistant of a model than before as testing appears to perform as well as predictions on the training set indicating the model will perform consistently on all data in the dataset and new data.

However, this performance is still poor in terms of target accuracy. Good to great models fall within an accuracy of ~0.7-0.8 consistently in the testing set and training set. This model may be suffering from lack of data as the dataset is relatively small.

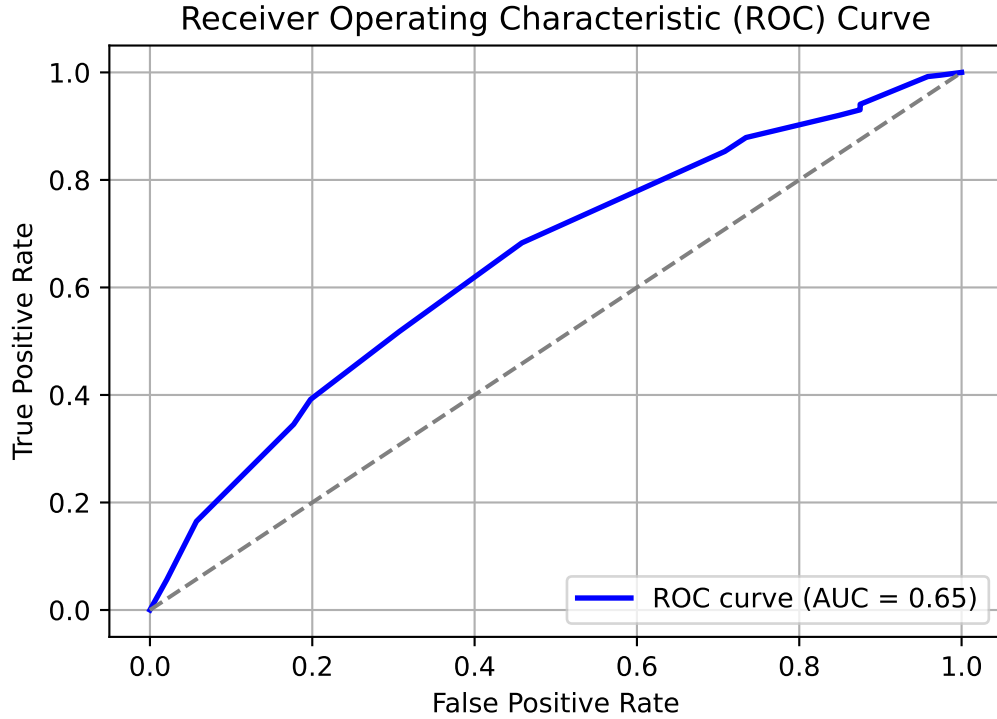


Figure 3: ROC curve for tuned decision tree predictions of draft status

The ROC curve produced in Figure 3 by this model did increase from the first generation decision tree model performing at ~0.6-0.7 showing the increase in performance using hyperparameters.

### 3.2.2 ROC Curve Calculations

The ROC curve is a measure of the correct positively predicted cases when testing. These values are calculated at several thresholds of the model and are then plotted to create the ROC curve seen. A curve with points more NW in the graph indicates a stronger performing model while a curve close to  $x=y$  or the linear diagonal from (0,0) to (1,1) indicates a model similar to random guessing or a poor performing model. The Area Under the Curve is calculated as the integral of the ROC curve. This value falls on a range of 0-1. With closer to 0 indicated worse than random guessing, close to 0.5 as similar to random guessing and close to 1 as perfect prediction performance on the testing data Hoo, Candlish, and Teare [3].

### 3.2.3 RANDOM FORESTS

A random forest classifier was deployed to see if the model performance could be increased. Based on the results, in Table 2, the precision of this model predicting drafted players was robust and efficacious. With a recall of 0.94, this model correctly predicted 94% of players to be drafted based on their combine results. However, this model does not perform well on undrafted players. With a recall of 0.19, it only correctly predicted 19% of players to be undrafted. The f1 score for both, which is a harmonic mean between precision and recall also supports this. Predicting

Table 2: Randomized Search Random Forest Classification Report

Randomized Search Random Forest Classification Report:				
	precision	recall	f1-score	support
0	0.58	0.20	0.30	427
1	0.68	0.92	0.79	799
accuracy			0.67	1226
macro avg	0.63	0.56	0.54	1226
weighted avg	0.65	0.67	0.62	1226

drafted players = 0.77, predicting undrafted = 0.29. This could possibly be due to lack of data or potentially lack of combine results for undrafted players making it difficult for the model to base it's undrafted predictions.

It was seen that the 40-yard dash was the most relevant feature for predicting draft status with a relevance score of 0.2875. This makes sense as the most infamous and televised combine event is the 40 yard dash and speed is one of the most important features an NFL player can have. Kuzmits and Adams [4] The bench press came in second with 0.2100. For larger players this is reasonable as they will excel in the strength portions of the combine. The draft status for offensive and defensive lineman would most likely be influenced by this metric.

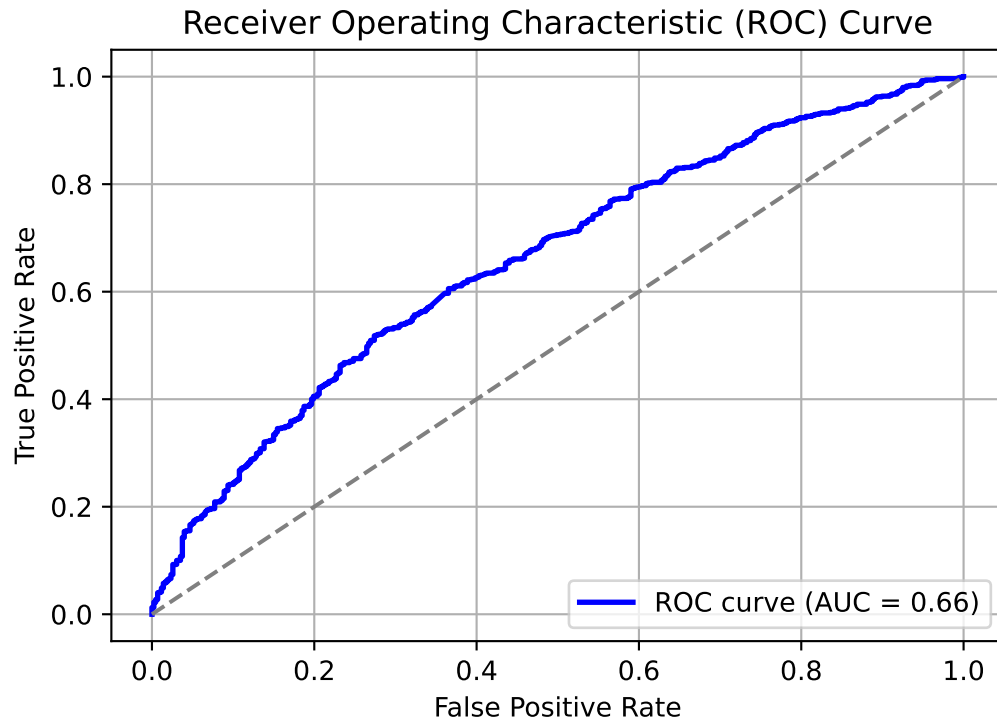


Figure 4: ROC curve for Tuned Random Forest model predicting player draft status on combine results

This model's performance was much better at predicting draft status than the decision tree of Figure 3. In Figure 4, the AUC comes out to  $\sim 0.7$ . This average performance shows increase in Random Forest classifying over decision tree classifying. The model could be improved in the future with more robust and complete data.

## 4 Predicting Player Position Based on Combine Performance

The question arises now of how well does each position group do at the combine. It would be expected that smaller players like wide receivers and defensive backs will excel in timed events such as the 40-yard dash and three-cone drills. While larger most physically imposing players like offensive and defensive lineman would predominate in strength events such as the bench press. First, the positional groups were created to simplify model generation. Offensive tackle, guard and center were grouped into offensive lineman. Defensive tackle, end, and edge were grouped into defensive lineman. Cornerbacks and safeties were grouped into defensive back. Outside and inside linebackers were grouped into linebackers.

### 4.1 Plotting Player Performance Based on Playing Position

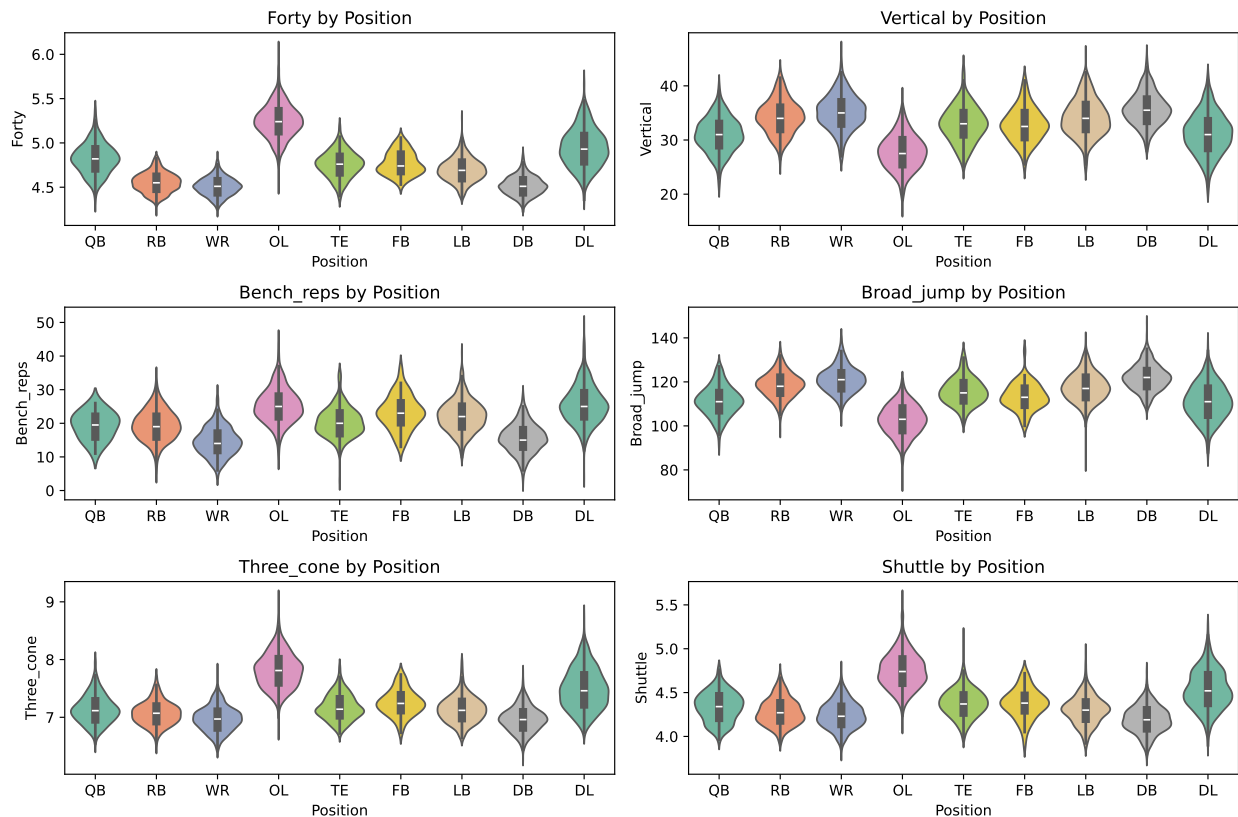


Figure 5: Distribution of players combine performance at each position



Table 3: Hist Gradient Boost Classification Report for predicting playing position

Classification Report:				
	precision	recall	f1-score	support
DB	0.42	0.51	0.46	211
DL	0.45	0.40	0.42	204
LB	0.36	0.39	0.37	155
OL	0.66	0.81	0.73	197
QB	0.60	0.47	0.53	72
RB	0.41	0.23	0.30	129
TE	0.37	0.29	0.33	69
WR	0.45	0.46	0.45	172
accuracy			0.47	1209
macro avg	0.46	0.45	0.45	1209
weighted avg	0.47	0.47	0.46	1209

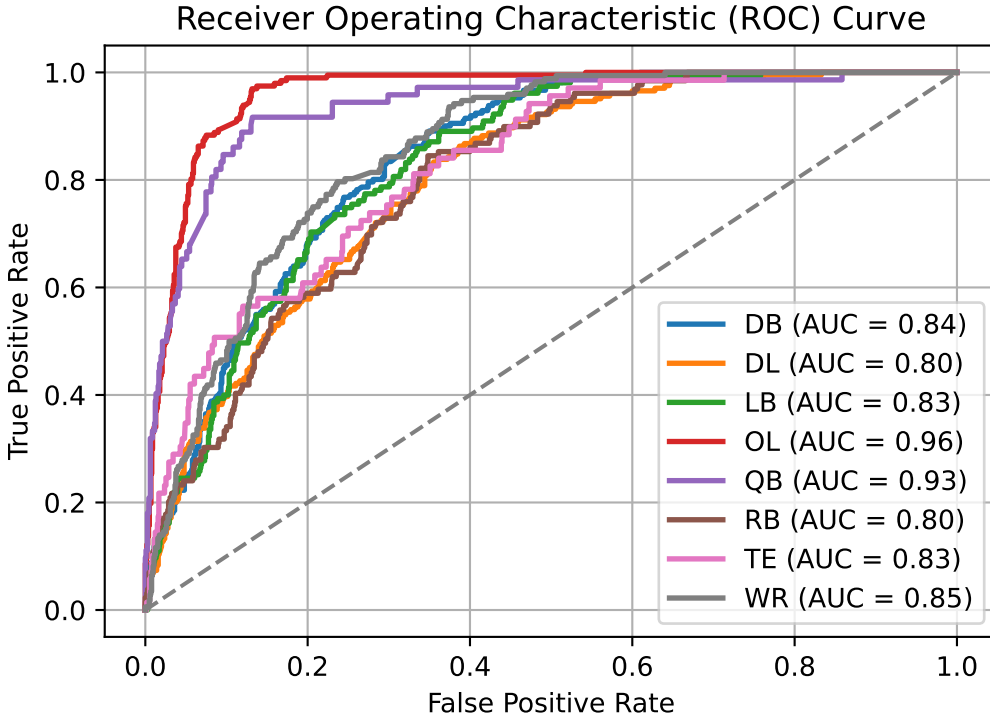
In Figure 5, the distribution of players combine performance based on playing position can be seen. The results align with the prediction that wide receivers and defensive backs excelled in the forty yard dash, three cone and shuttle drill. Again, lower scores indicate faster players. In the vertical and broad jump plot, it can also be seen that all players besides offensive and defensive lineman do well. It appears the extra weight carried by lineman prevent better scores in these categories. In the bench press however, lineman excel in this category. The strength needed to play lineman is one of the most important features of these players draft potential.

## 4.2 Model, Predicting Positions

For greater model generation and for its ability to work based incomplete values in data. The Hist-GradientBoost Classifier was utilized to predict playing position based on combine performance.

In Table 3, the precision and recall metrics for each position was poor however, when predicting offensive lineman, this model performed very well, correctly predicting the position of offensive lineman 81% of the time. This was most likely due to the drastic difference in bench press scores than other positions seen in Figure 5. Every other combine event had much overlap for each position most likely preventing the model from distinguishing the player's position. More variables would need to be introduced to increase this performance to distinguish other positions from each other.

Below is the ROC curve for this model:



The ROC curve shows the exceptional performance for predicting player position. The Hist Gradient Boost Classifier performed well when positively predicting a player's position. Notably it predicted quarterbacks (0.93) almost as well as it did with offensive lineman (0.96). This will be something to keep in mind for future research and modeling because of the quarterback's premier position. Teams could use this model to evaluate QB metrics and determine whether or not they performed well at the combine.

## 5 Discussion

The analysis of NFL Scouting Combine data revealed distinct relationships between combine events and draft outcomes. Timed drills such as the 40-yard dash, shuttle drill, and three-cone drill were positively correlated with an increased draft round, indicating that slower times in these events were associated with later draft selections or going undrafted. Conversely, power and explosiveness metrics such as bench press repetitions, vertical jump, and broad jump showed negative correlations with increased draft round, suggesting that stronger performances in these events were linked to earlier draft rounds.

In modeling draft status, the Random Forest Classifier with optimized hyperparameters outperformed simpler models like Decision Trees, showcasing its ability to handle complex interactions between variables and improve predictive accuracy. However, the overlap of player positions in combine metrics posed challenges in distinguishing between groups. Despite this, specific patterns emerged: wide receivers and defensive backs excelled in timed drills as well as leaping drills, while linemen and larger players dominated strength-based events such as the bench press.

When predicting player position based on combine performance, the HistGradientBoost Classifier performed exceptionally well. It was able to effectively predict offensive lineman and quarterbacks

specifically well. This classifier excelled even when data was missing by identifying splits during the training process which in turn modify and create the best performing model Mayr et al. [5].

The NFL draft occurs in April. It may be interesting to deploy these models onto each player's combine performance to perhaps understand their draft potential as well as what positions they may be most fit to play. Data from the 2024 draft as well as the 2025 can be obtained and used to refine the models used in this study to continue to improve this report's validity.

In conclusion, this study highlights the importance of combine performance metrics in evaluating draft prospects and player positions. By leveraging machine learning techniques, NFL teams can gain deeper insights into player potential and upgrade their scouting strategies, enhancing draft decision-making and roster building.

## References

- [1] Shane Hauk, Michael Shuckers, and Robin Lock. "NFL Draft Combine". In: (Aug. 2023).
- [2] David P Hedlund. "Performance of future elite players at the National Football League scouting combine". In: *The Journal of Strength & Conditioning Research* 32.11 (2018), pp. 3112–3118.
- [3] Zhe Hui Hoo, Jane Candlish, and Dawn Teare. *What is an ROC curve?* 2017.
- [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] Andreas Mayr et al. "The evolution of boosting algorithms". In: *Methods of information in medicine* 53.06 (2014), pp. 419–427.
- [6] Daniel Murran. "Ahead of the Curve, Analysis of the NFL Draft: Technical Report". PhD thesis. Dublin, National College of Ireland, 2018.