NFL Tackle Probability Model

By Ethan Kawahara & Antonio Olivas

Vince Lombardi

2× Super Bowl champion (I, II)
5× NFL champion (1961, 1962, 1965, 1966, 1967)
2× NFL Coach of the Year (1959, 1961)

“Football is two things. It’s blocking and **tackling**.”

What Is The Problem?

- **Current Tackling Evaluations Are Subjective**
  - Limited quantitative insights into **how and why** tackles happen or fail.
- **Need for a Data-Driven Approach**
  - How can we **quantify** the likelihood of a tackle happening?
  - Can we **predict** tackling outcomes based on player position, movement, and surrounding defenders?
  - How can teams **use this information** to improve defensive schemes and player positioning?
- **Gaps in Existing Models**

Problem: SEVERE Class Imbalance

- **Filter Out Distance Defenders**
  - Only want defensive players within 1 Yard of the ball carrier. (Median distance all tackles are made in)
- **Balance the dataset**
  - Take a smaller count of tackles and non tackle frames
- **Stratify the samping of tackles and non tackles across plays**
  - Ensure each play is accurately represented in the data

Before:

Tackle Frames: 686,578

Non-Tackle Frames: **5,142,091**

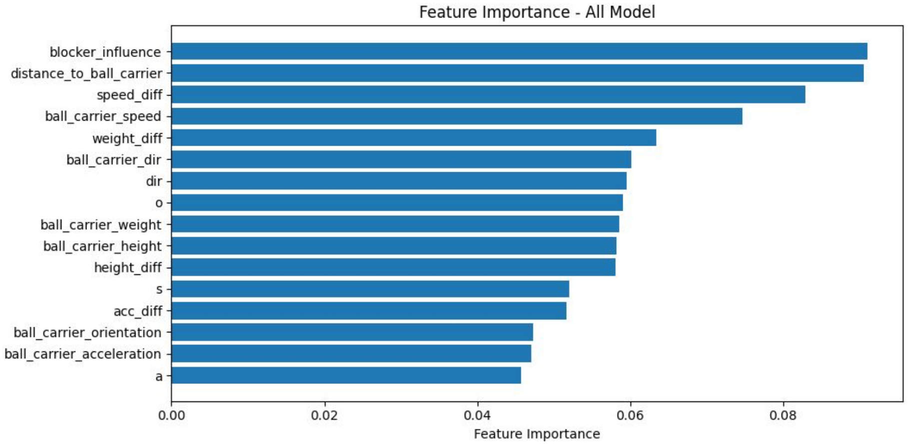
After:

Tackle Frames: **176,578**

Non-Tackle Frames: **106,788**

Feature Engineering

1. **Distance To Ball Carrier**
    a. Find the **Euclidean Distance** between the defender and the ball carrier.
2. **Blocker Influence (Unique)**
    a. Uses **exponential distance decay**, where closer blockers have a stronger effect on limiting a defender’s movement.
3. **Height & Weight Difference (Unique)**
    a. Modeled as a **scaling factor**, where larger size mismatches impact tackle success probability.

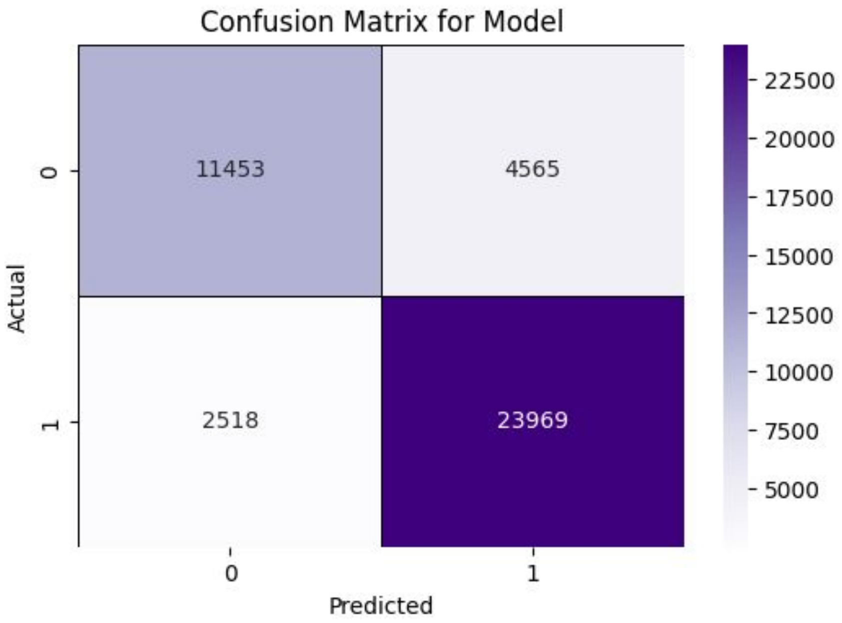


Josh Norman: 6’0, 194 lbs
Derrick Henry: 6’3, 247 lbs

Model Architecture

- **XGBoost Model** – Gradient boosting with decision trees, optimizing errors iteratively.
- **Bayesian Optimization** – Hyperparameter tuning via **BayesSearchCV** for efficient search.
- **Regularization & Sampling** – Lasso and Ridge regression and feature subsampling to prevent overfitting.
- **Early Stopping & AUC** – Stops training if validation AUC plateaus.

Model Performance



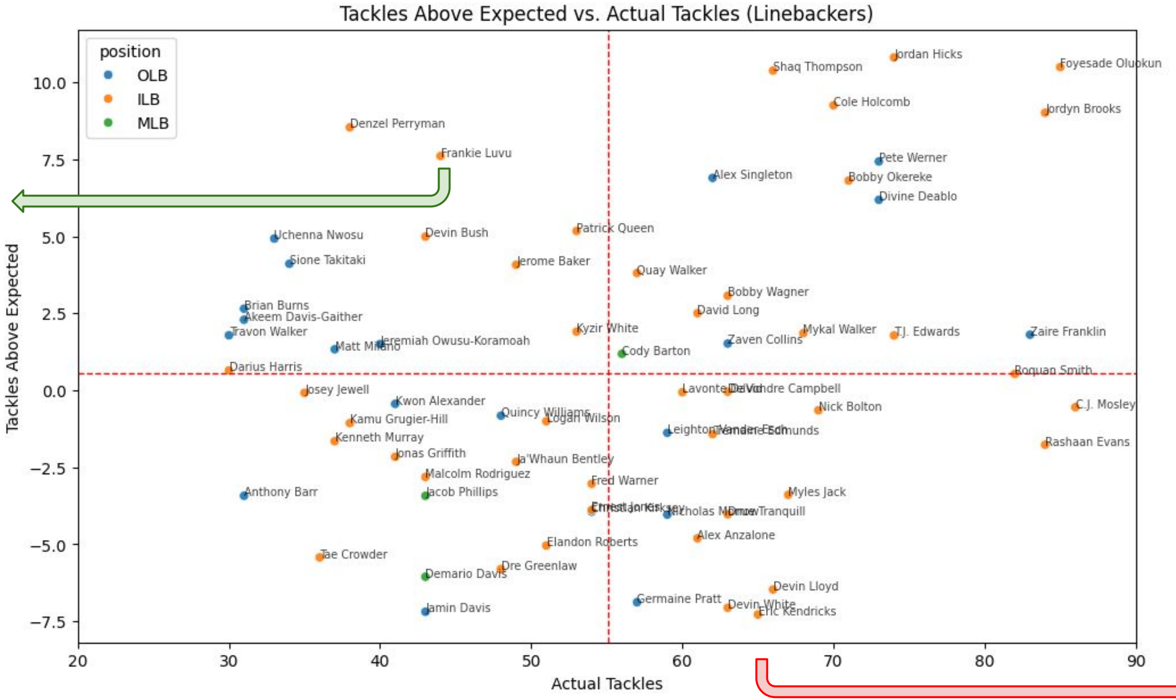
- **Accuracy (83.33%)**: The overall percentage of correct predictions.
- **Precision (84.00%)**: When the model predicts a tackle, it is correct **84.00%** of the time.
- **Recall (90.05%)**: The model correctly identifies **90.05%** of all actual tackles.
- **F1-Score (87.12%)**: A balance between **precision and recall**, indicating strong overall model performance.

Key Takeaway:

- The model performs **consistently well across all metrics**, showing a good balance between **precision and recall**.

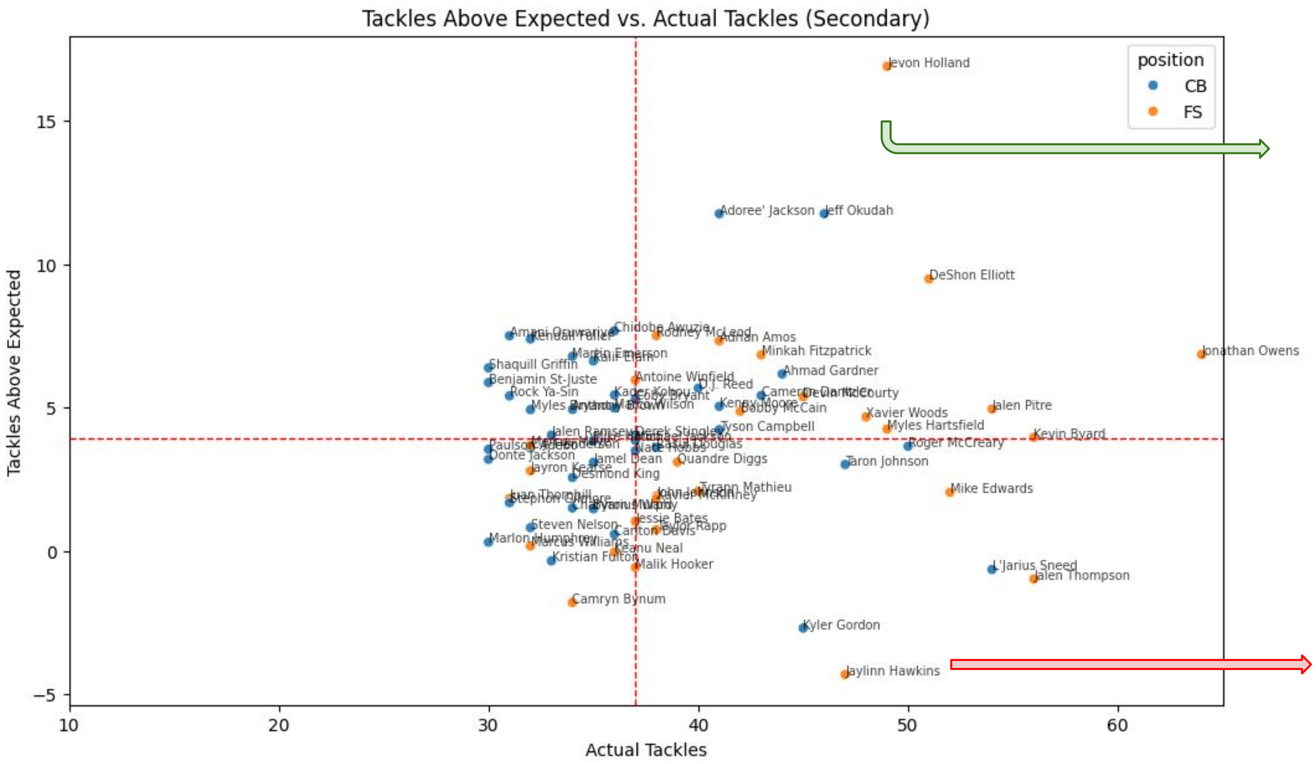
Model Insights (Linebackers)

Frankie Luvu



Eric Kendricks

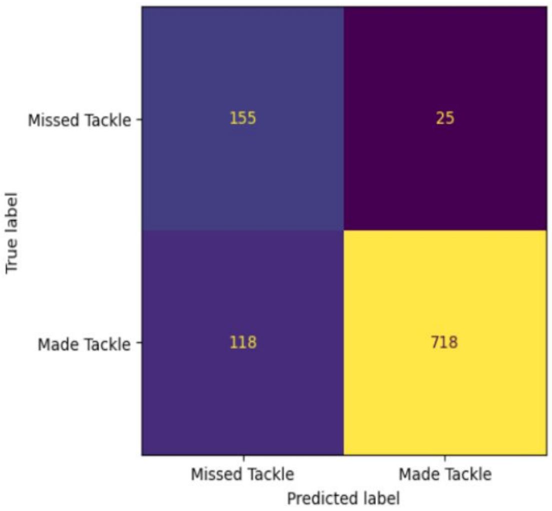
Model Insights (Secondary)



Jevon Holland

Jaylinn Hawkins

Other Models Comparisons (Uncovering Tackle Opportunities and Missed Opportunities)



Class	Precision	Recall	F1-Score
Made Tackle	0.97	0.86	0.91
Missed Tackle	0.57	0.86	0.68

https://www.kaggle.com/code/matthewpchang/uncovering-missed-tackle-opportunities

"We generated the training dataset with 8000 made tackles and 1583 missed tackles from weeks 1-8 tracking data…The test dataset includes 836 made tackles and 180 missed tackles exclusively from week 9 to prevent leakage."

1. **Frame-Level Data Captures More Context** – Tracks player movement, positioning, and decision-making across time, unlike their play-based approach.
2. **Improved Temporal Dynamics** – Models real-time reactions and adjustments, providing deeper insights into tackle success vs. failure.
3. **More Data, Better Generalization** – 200,000 frames improve learning, while their limited sample (~9,819 plays) risks overfitting.

Balanced Dataset Advantage

- Prevents **bias toward tackles** (their 5:1 ratio skews predictions).
- Improves **missed tackle detection**, crucial for defensive evaluation.
- Enhances **model stability**, avoiding trivial predictions.

Why My Model Is Unique

✅ **Frame-Level Analysis** – Evaluates every **moment leading up to the tackle**, not just final outcomes.

✅ **Beyond Proximity** – Identifies **why** some defenders **miss tackles even when close**, using movement, angles, and ball carrier characteristics.

✅ **Balanced Dataset** – Prevents bias by focusing on relevant moments, ensuring **missed tackles are properly learned**, unlike models that overpredict tackles.

✅ **Predictive, Not Reactive** – Unlike models that **only detect tackles happening**, this model **forecasts tackle success** before impact.

Model Applications & Next Steps

- **Player Evaluation & Scouting** – Identifies defenders who consistently **maximize tackle opportunities** and those who struggle, helping teams with **talent evaluation**.
- **Coaching & Player Development** – Pinpoints **why tackles are missed** (e.g., bad angles), allowing for **targeted training** to improve tackling efficiency.
- **Game Strategy & Defensive Schemes** – Helps coaches understand **which defenders perform best in specific scenarios**, optimizing defensive assignments and formations.