# Called Strike Probability Model

Madden Gleason, Ashlynn Simmons, Dina Taing, Addeson Zappa, Madison Ziegler

Department of Mathematics, Louisiana State University

Advisors: Dr. Nadejda Drenska & Dr. Zachary Jermain

## Introduction

Softball umpires do not always call the rulebook strike zone. Instead, each umpire appears to use a personal zone that depends on pitch location, count, batter, pitcher, and game situation.

Our project builds a **called strike probability model** for LSU Softball. The model estimates the probability that a taken pitch (no swing) is called a strike, based on pitch location, pitcher handedness, and batter handedness.

Because LSU pitch–tracking data were not available at the beginning of the semester, we first trained and tested our model using MLB Statcast data (2020–2024). Once LSU Softball TrackMan data became available late in the term, we adapted and fine–tuned the model for collegiate softball.

## Objectives

Goals

- Quantify how location, pitch type, count, and handedness influence called strike probability.
- Visualize effective strike zones for different situations using probability heatmaps.
- Transfer a model trained on MLB data to LSU Softball using machine–learning techniques.

Timeline

- **Midterm:** clean MLB data, build baseline strike probability model, and generate MLB heatmaps.
- **Final:** receive LSU TrackMan data, adapt the model, incorporate handedness, and generate LSU–specific heatmaps.

## LSU Softball Data

LSU Trackman Data (2025)

- Received near the end of the semester; significantly smaller than MLB.
- Includes pitch location (PlateLocSide, PlateLocHeight), pitch type, count (Balls, Strikes), pitcher/batter handedness, and pitch outcome.
- We removed rows with missing locations and kept only the variables needed for our strike zone model.

## Initial Results: MLB Baseline

MLB Statcast (2020–2024)

- MLB pitch–by–pitch tracking data from Baseball Savant (baseballsavant.mlb.com).
- Each recorded pitch includes: horizontal and vertical location at the plate, pitch type, velocity, batter/pitcher handedness, count, and pitch result.
- We filtered to taken pitches and defined $Y = 1$ for a called strike and $Y = 0$ otherwise.

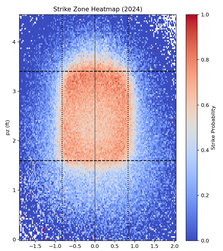


Figure 1: Called strike probability heatmap from the initial MLB model.

## Activation Functions

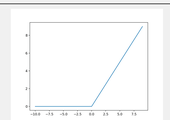

Figure 2a: ReLU

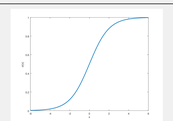

Figure 2b: Sigmoid

## Neural Network Strike Zone Model

We use a Multilayer Perceptron (MLP) architecture to build the model and generate the strike zone predictions:

$$\text{sum} = \left(\sum_{i=1}^{m} w_i x_i\right) + b$$

where $\mathbf{x}$ is the input feature vector, $w_i$ are the learned weights, $b$ is the bias term, and the expression computes the neuron's pre-activation value $z$.

Final Architecture (LSU Model)

- **Input layer (5 neurons):** horizontal location ($x$), vertical location ($y$), swing indicator, pitcher handedness, batter handedness.
- **Hidden layer 1:** 32 neurons, ReLU activation
- **Hidden layer 2:** 64 neurons, Sigmoid activation
- **Output layer:** 1 neuron returning the called–strike probability

$$W_1 = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,5} \\ \vdots & \vdots & \ddots & \vdots \\ w_{32,1} & w_{32,2} & \cdots & w_{32,5} \end{bmatrix}, \quad W_2 = \begin{bmatrix} w_{1,1} & \cdots & w_{1,32} \\ \vdots & \ddots & \vdots \\ w_{64,1} & \cdots & w_{64,32} \end{bmatrix}, \quad W_3 = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,64} \end{bmatrix}.$$

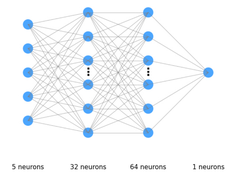


Figure 3: Neural network architecture used for strike probability.

## Training and Transfer Learning

Stage 1: Small Initial Dataset

We began with a limited dataset and a small model that offered very little flexibility in feature engineering. The model could only learn from a narrow set of inputs, restricting its predictive capacity.

Stage 2: Larger Dataset and Model Expansion

With a more detailed and much larger dataset, we added new input variables without discarding what the model had already learned. Using **weight surgery**, we expanded the input layer by copying the original weight matrix into a larger one and initializing the new feature columns to zero. This preserved the model's previous behavior while allowing it to learn the new variables safely.

Stage 3: Fine–Tuning on the Enhanced Dataset

The surgically expanded model was then fine–tuned on the more comprehensive dataset. Using the previously learned weights provided stability, improved convergence, and reduced overfitting during training.

## Strike Probability Model Outputs

Our neural network converts pitch location and handedness information into a smooth strike-probability surface. These maps show how likely a pitch is to be called a strike at each point in the zone.

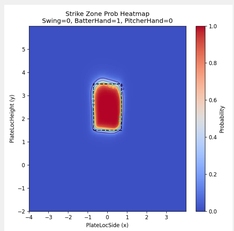


Figure 4a

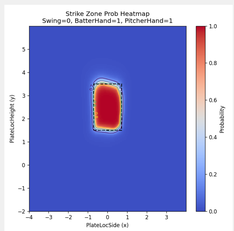


Figure 4b

Model Interpretation:

- A high-probability center where called strikes are most consistent.
- A fading boundary reflects borderline call regions that vary by umpire and context.
- Subtle shifts in the zone depending on pitcher and hitter handedness.

## LSU Softball Heatmaps: Pitch Type, Count, and Handedness

Using the cleaned LSU TrackMan data, we generated called–strike heatmaps for each pitch type and count, separated by pitcher handedness. These visualizations help reveal strike zone tendencies in real LSU softball settings.

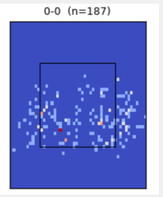


Figure 5a: RHP curveball, 0–0 count (called strikes).

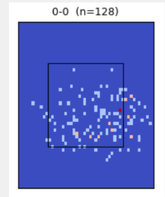


Figure 5b: LHP curveball, 0–0 count (called strikes).

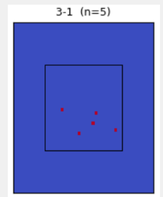


Figure 6a: RHP fastball, 3–1 count (called strikes).

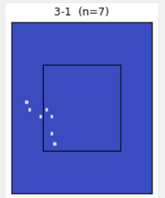


Figure 6b: LHP fastball, 3–1 count (called strikes).

Strike Zone Patterns:

- Certain pitch types cluster tightly in specific corners of the zone.
- Called strike zones shift with the count: pitchers get more of the edges in pitchers' counts and less in hitters' counts.
- Right– and left–handed pitchers show different patterns, especially on breaking pitches (screwballs and curveballs).

## Limitations

- **Small sample size:** Some pitch type / count / handedness combinations have very few pitches to produce stable heatmaps.
- **MLB-to-NCAA Softball Differences:** Differences in strike zone geometry, pitch movement, and umpire behavior do not allow MLB patterns to transfer perfectly.
- **Scope:** This model uses only pitch location and handedness. It does not yet include framing, umpire identity, or swing / take decisions.

## Conclusions and Next Steps

This project demonstrates that a location-based strike probability model can be adapted from MLB to NCAA softball, revealing consistent differences in pitch behavior across types, counts, and pitcher handedness. Although the current dataset is limited, the trends are clear, and the model provides a practical foundation for training applications.

Future work will involve collecting more LSU softball seasons, incorporating additional features (framing, swing behavior, and umpire tendencies), and building individualized strike zone visualizations for players and umpires. These additions will improve accuracy and strengthen the model's usefulness for preparation, development, and in-game decision support.

## Acknowledgements

We thank Dr. Nadia Drenska and Dr. Wolenski for their instruction, guidance, and support in completing this project. We also thank Dr. Zachary Jermain and the LSU Softball program for their expertise and contributions that shaped this model. Additional thanks go to Fernando Heidercheidt, Maganizo Kapita, and the extended course staff for their help and feedback throughout the semester.

## References

1. Major League Baseball Advanced Media. *Baseball Savant: Statcast Search and Player Visualizations*. Online, 2025. Retrieved from https://baseballsavant.mlb.com. Accessed: Oct. 21, 2025.
2. Franke, K. *Strike Probability Model and Catcher Framing Using Random Forest*. Medium, 2021. Accessed: Oct. 21, 2025.
3. TrackMan Baseball. *Radar Measurement Glossary of Terms (V3)*. Online, 2025. Retrieved from https://support.trackmanbaseball.com. Accessed: Oct. 21, 2025.
4. Nestico, T. *Classifying MLB Pitch Zones and Predicting MiLB Zones*. Medium, 2021. Retrieved from https://medium.com/@thomasjamesnestico/classifying-mlb-pitch-zones-and-predicting-milb-zones-7e95cf308254. Accessed: Dec. 1, 2025.
5. NCAA. *2024–2025 Softball Rules Book*. National Collegiate Athletic Association, Indianapolis, IN, 2024. Accessed: Nov. 11, 2025