Statistical Validation of Current Approach in Model Construction

Logan Kelsch (Chick Counting Group)
COSC 426 (Software Engineering II)

October 2, 2025

Summary

In the "Chick Counting" group project with Perdue Farms in COSC 426 (Soft. Eng. II), the task before us is to construct a machine learning model capable of counting chicks in sets of 100 with maximal accuracy and minimal variance. The current performance standard we are looking to challenge is 97.5 % accuracy—more specifically, a mean detection rate of 100 per 100 chicks with a standard deviation $\lesssim \pm 3.133$[1]. Team member Anye Forti developed our strongest RGB detection and counting model (YOLO), capable of detecting essentially every chick across multiple environments. The drawback with this model is the inability to differentiate between one chick and a group of chicks, which leads to a consistent undercount. After analysis and visualization of performance over time, it became apparent that this inefficiency is nearly perfectly consistent. Building on that consistency, we assess the legitimacy of a simple coefficient meta-model as a first step toward a more robust meta-model approach.

[1]If absolute errors are modeled as half-normal with scale σ, then $\mathbb{E}[|Z|] = \sigma\sqrt{2/\pi} \approx 0.798\,\sigma$. A target mean absolute error of 2.5 implies $\sigma \approx 2.5/0.798 \approx 3.133$.

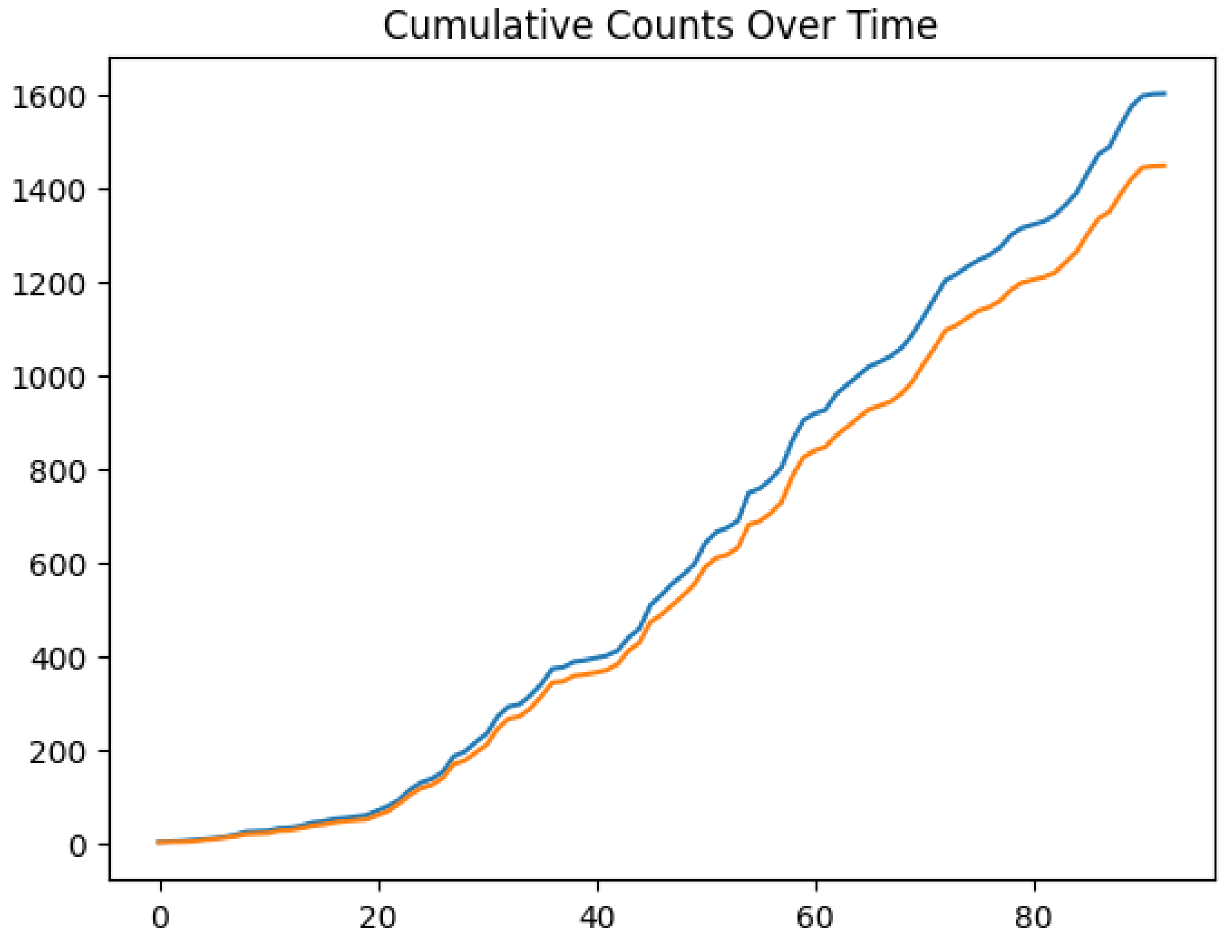

Figure 1: Cumulative True (blue) and predicted (orange) counts per capture over the validation sample.

Validation

This analysis uses Anye's hand-counted data compared against his model predictions over a validation sample of about 1,600 chicks. Data is loaded into a 2D NumPy array; true and predicted counts are calculated per capture.

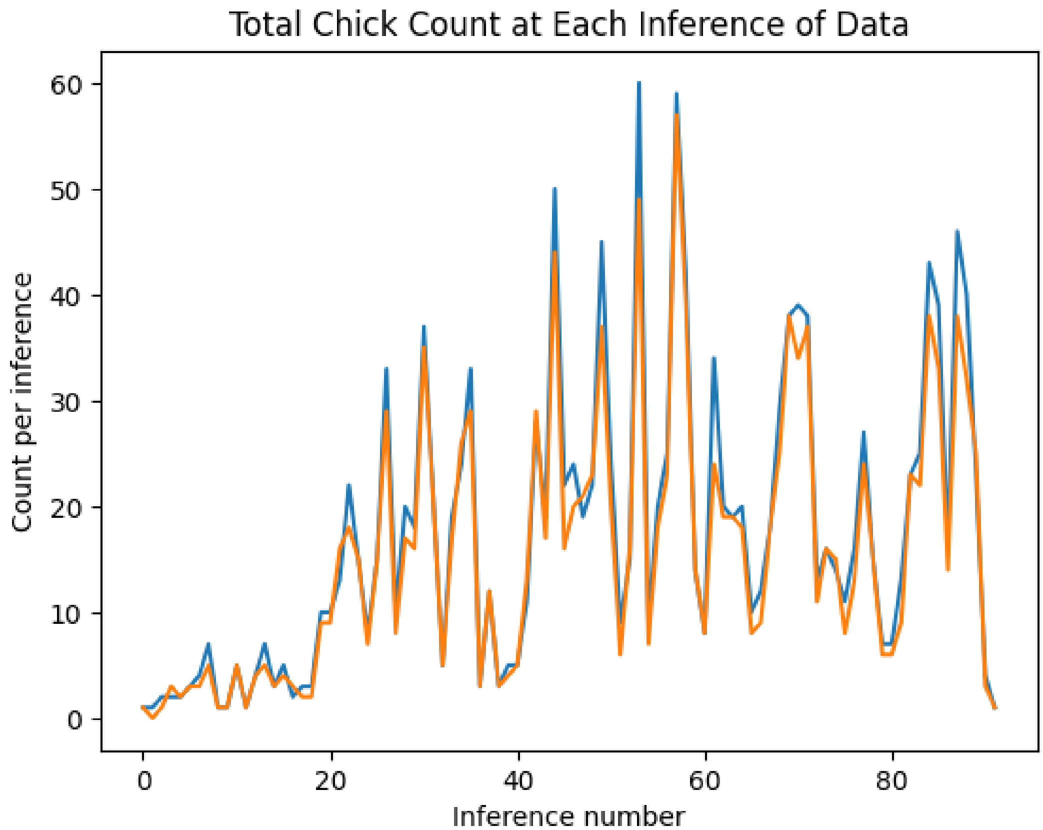

Figure 2: True (blue) and predicted (orange) counts for each capture over the validation sample. (Rough derivative)

Coefficient Meta-Model

Given that the current YOLO model has a mean prediction count of 90.4 per 100 chicks (90.4 %) with a fair variance of about 6 %, solving for a simple in-sample coefficient (assuming the approach is valid) quickly yields a mean meta-prediction of 100 per 100 chicks with a fair variance of about 6.3 %. Because the error is one-sided (undercount), a single coefficient has the potential to resolve most of it.

We define the one-parameter meta-model on the YOLO model prediction output $\hat{y}$ as

$$\hat{y}^{(\text{meta})} \;=\; c \cdot \hat{y},$$

where c is estimated on a training subset and then applied out-of-sample to produce $\hat{y}^{(\text{meta})}$ on the validation subset.

Monte Carlo Permutation Test (MCPT) [1,2]

We generate thousands of simulations on the same data. Several thousand random and shuffled 50/50 splits segment the data into training and validation sets. For each split i, the training subset is used to fit the coefficient c_i, which is then applied to the validation subset to obtain a validation error; the distribution of those validation errors is then plotted. Ideally, the variance of error is minimal and the mean centers near zero.

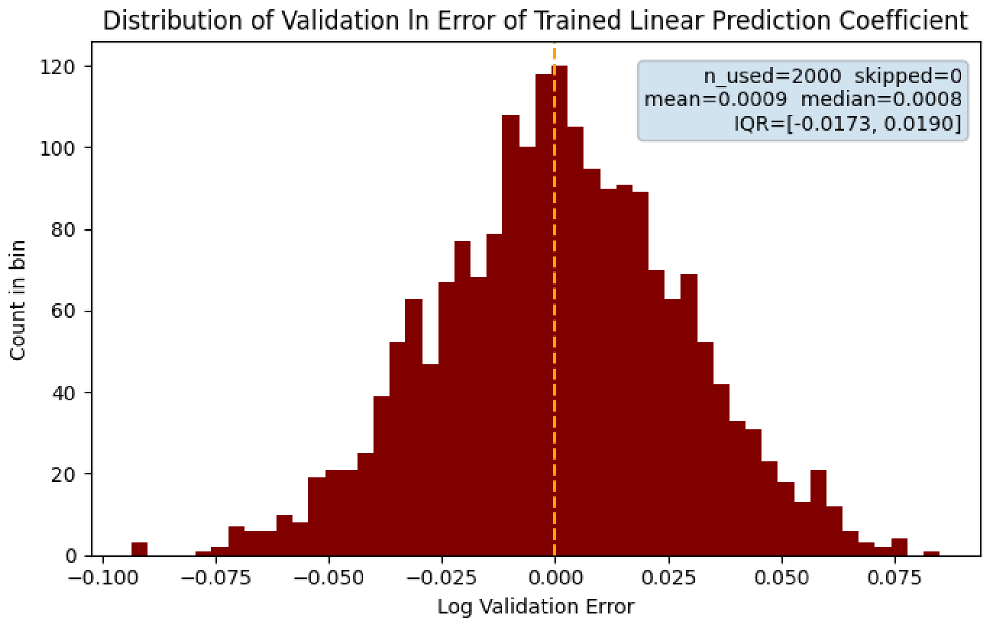

Figure 3: Histogram of validation errors across MCPT splits; mean near zero with narrow IQR.

Split-Threshold Robustness

Because the train/validation split threshold is the only subjective parameter, we repeat the same simulation across multiple thresholds $\in [0.1, 0.9]$. Kernel density estimates (KDEs) for each threshold show consistent variance and means near zero, indicating that the threshold is a robust parameter for this analysis. Using 5,000 simulations per threshold improves smoothness; visually, thresholds from 0.3 to about 0.7 yield essentially the same distribution.

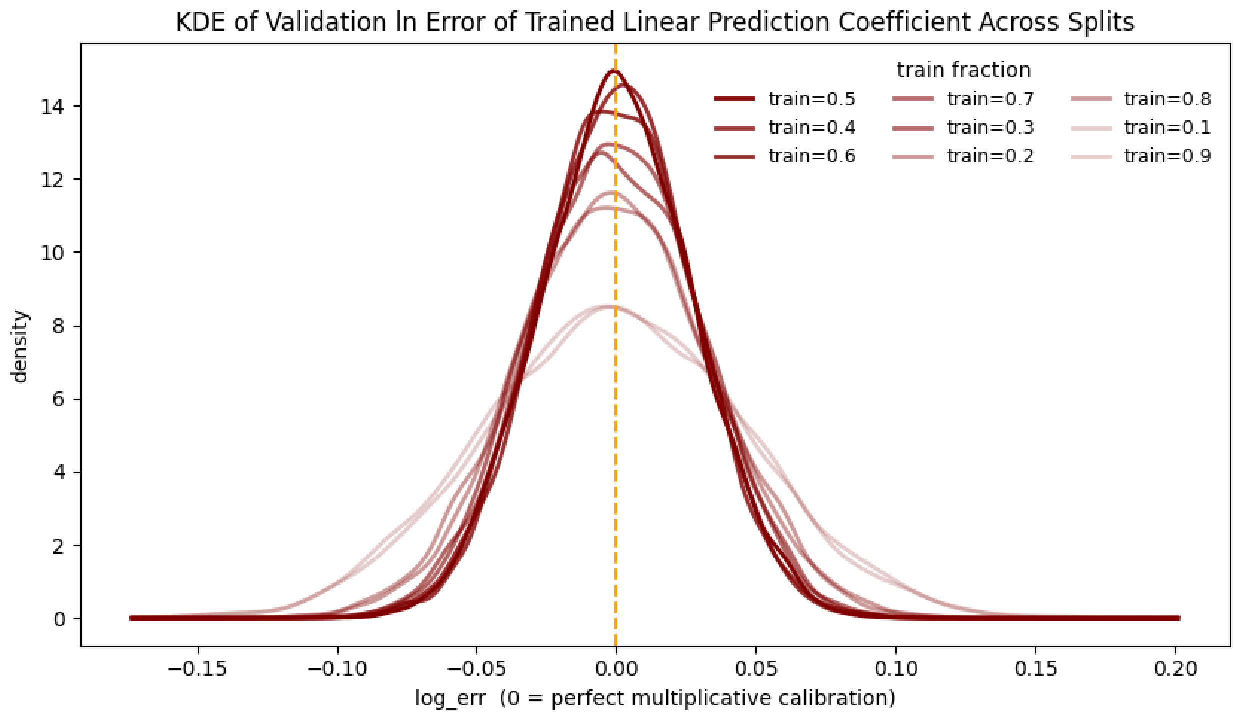

Figure 4: KDEs of validation-error distributions across split thresholds: consistent means near zero and similar spread $\in [0.3, 0.7]$.

Conclusion

The approach is logically sound: the mean lands essentially at zero with a narrow IQR (about $\pm 1.82\%$), and the train/validation split threshold behaves robustly. Given the extremely low validation error spread relative to our goal, moving forward with meta-model development is the strongest direction, with the obvious expectation that a learned meta-model will exceed the coefficient's performance and robustness given extensive data complexity to build from.

Please contact Logan Kelsch for any questions on this document.
lkelsch1@gulls.salisbury.edu

References

[1] Dwass, M. (1957). Modified Randomization Tests for Nonparametric Hypotheses. *Annals of Mathematical Statistics*, 28(1), 181–187.

[2] Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *Proceedings of NeurIPS 2017*, 4765–4774.