Interpretable Machine Learning to Understand Variation in Manakin Display Types

FSU | FLORIDA STATE UNIVERSITY

Brandon Bonamarte, Kaya Simmons, Elle Daily, Jo Melissa Fernandez-Jenkins, Emily H. DuVal

FSU UNDERGRADUATE RESEARCH OPPORTUNITY PROGRAM
CENTER FOR UNDERGRADUATE RESEARCH & ACADEMIC ENGAGEMENT

Background

Complex Movement and Machine Learning

Quantifying complex movement is a challenge across biological systems, from animal communication to human health. These behaviors generate high dimensional time-series data that is difficult to analyze with traditional approaches. Machine learning enables classification of dynamic movement patterns, and interpretable models allow us to identify which features drive biologically meaningful variation.

The Lance-tailed Manakin

The lance-tailed manakin, *Chiroxiphia lanceolata*, is a small tropical bird that performs highly coordinated, cooperative courtship displays in which male pairs execute repeated, structured movement sequences. This system provides an ideal opportunity to test how interpretable machine learning can quantify behavioral variation and link display structure to reproductive success.

Research Question:
How does the structure of the male courtship display vary among male pairs, and which features of variation relate to copulation success?

Study System: *Chiroxiphia lanceolata*

Figure 1: An image showing a male lance-tailed manakin resting on a perch. [1]

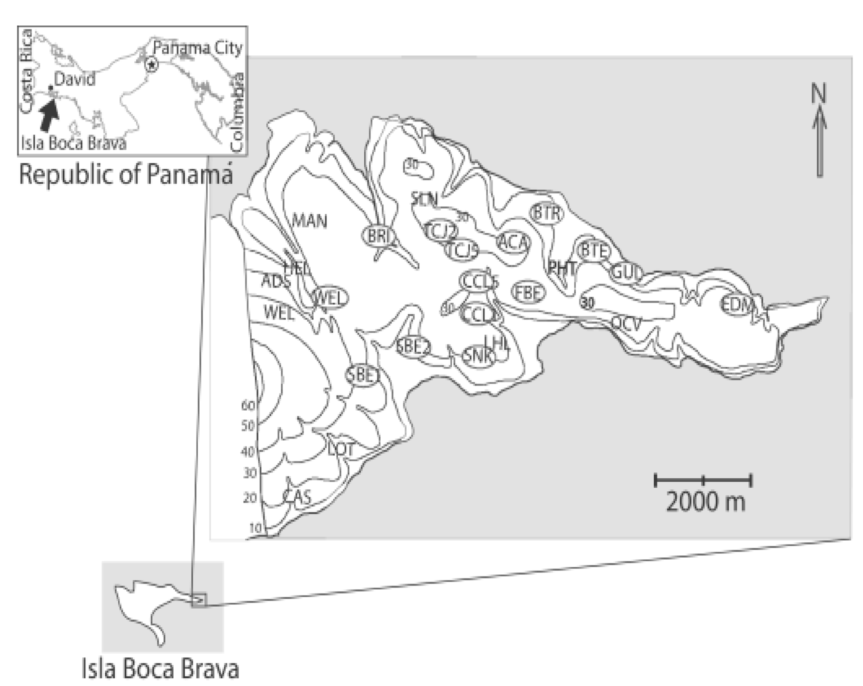


Figure 2: A map illustrating the study site, and location of individual displays across the island, Boca Brava. [2]

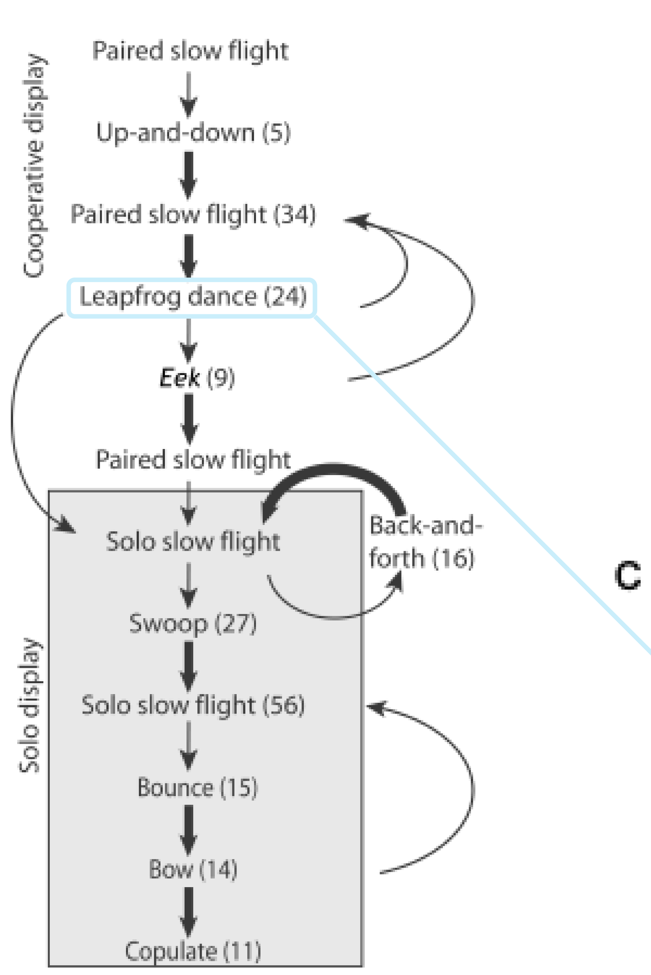


Figure 3: A diagram illustrating the full display of the lance-tailed manakin. [2]

The lance-tailed manakin, *Chiroxiphia lanceolata*, is a tropical bird in which male pairs form long-term cooperative alliances to perform elaborate courtship displays. These displays consist of repeated, highly coordinated movement sequences that are believed to be tied to mating success. This study focuses on the "leapfrog dance," a rapid, cyclical display in which two males revolve around one another. Its stereotyped structure and repetition make it ideal for quantifying fine-scale variation in movement.

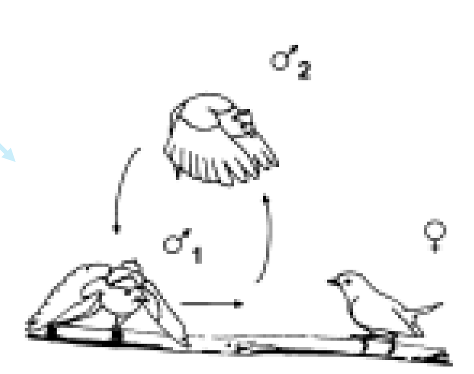


Figure 4: A graphic illustrating the leapfrog element of the display [2]

Figure 5: A QR code that leads to a video of the leapfrog element of the display

Part 1: Initial Data Processing

In studies of mate choice, researchers often must make subjective decisions about what traits to measure. To ensure data extraction remains objective and free from human assumptions, we utilize DeepLabCut (DLC), a machine-learning-based pose estimation tool [3]. By tracking every body part directly from raw video, we minimize subjective interpretation. This iterative process is currently ongoing, with model accuracy continually improving as we integrate additional training data.

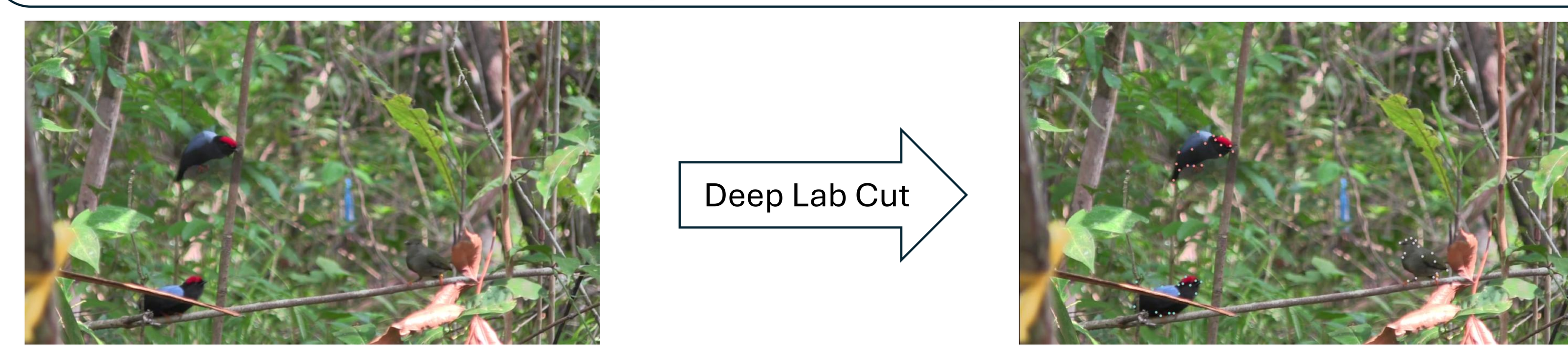


Figure 6: A diagram illustrating the Deep Lab Cut program's input and output. On the left is the raw image that DLC receives, and the right is the labelled image it outputs, with 27 possibly body part points labeled per bird.

Part 2: Data Cleaning

We addressed two main challenges in using the pose-estimation data and preparing it for machine learning. (1) Normalization. Different cameras recorded displays at different zooms. Therefore, the post DLC output, pixel position over time, is meaningless when compared across videos. To fix this, we find the unique size standardization (centimeters/pixel) and compute scalar multiplication across the entire dataset. This normalized data allows for just comparison across videos. (2) Gaps in data. Even on relatively well-labelled videos from DLC, small to large gaps in position data is possible. To accommodate this, we filter out the body parts that have too large gaps, and compute spline interpolation on those with smaller gaps.

Data Challenge #1: Videos are recorded at many different magnitudes.	Data Solution #1: Normalize post-DLC files through unique scalar multiplication of $\frac{cm}{px}$.

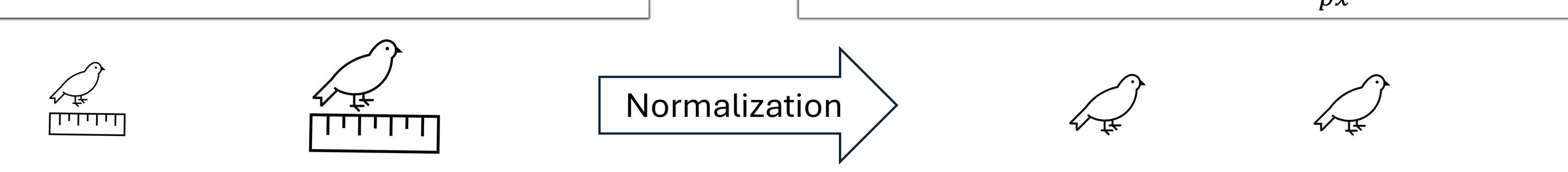


Figure 7: A diagram illustrating two birds the same size, at different magnitudes. An arrow pointing to the left represents the process of normalization. On the left, is the two birds, now the same size.

Data Challenge #2: DLC is an imperfect program, sometimes leaving large gaps of unlabeled data.	Data Solution #2: Exclude data with too large gaps and employ cubic spline interpolation on smaller gaps.

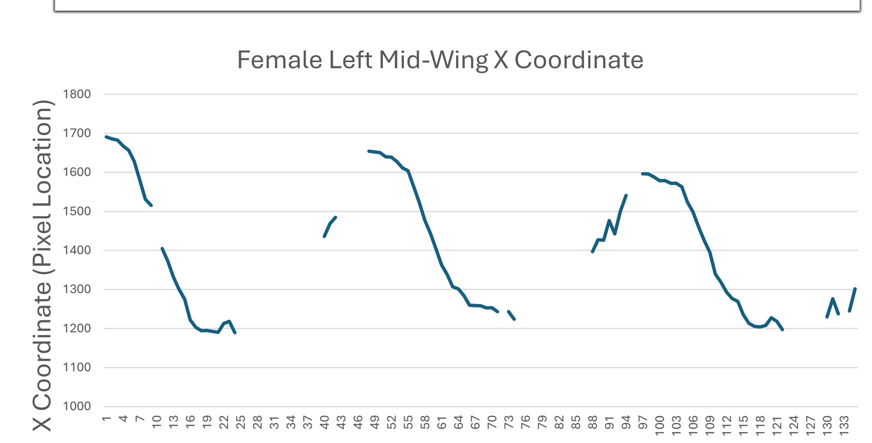


Figure 8: A graph showing the raw X Coordinate of a females left mid-wing (Pixel Location in Video) over frames.

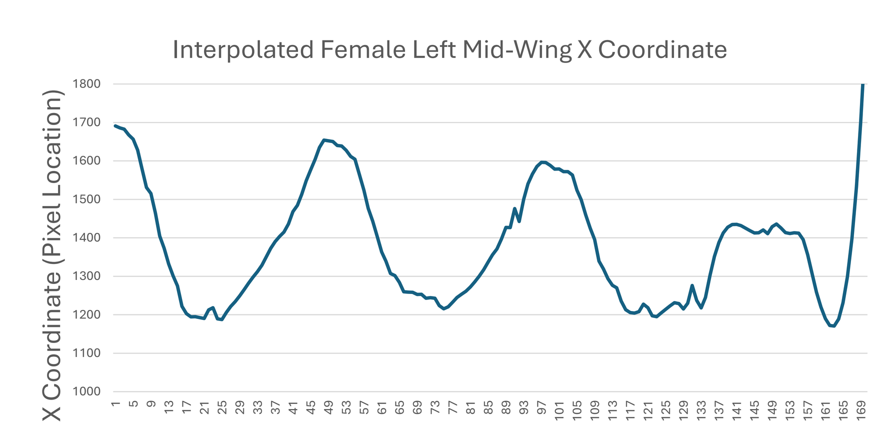
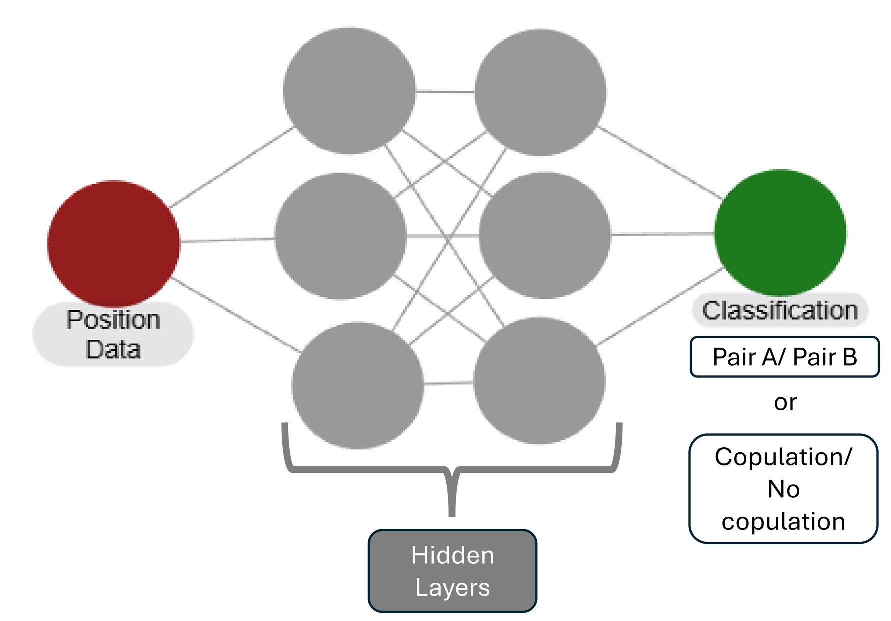


Figure 9: A graph showing the same data as figure 8, but after undergoing spline interpolation.

Part 3: Machine Learning Models

The true essence of this research lies in our proposed machine learning approach, utilizing two distinct models to interpret the data. The first model classifies by male pair to predict performer identity, allowing us to determine if and how dances vary by individual regardless of the outcome. The second model classifies by successful copulation to determine if specific leapfrog components indicate success later in the courtship sequence, pinpointing the timing and nature of that significance. While both models will utilize a neural network approach, the specific architectures and novel methods remain to be determined.

Model 1: Classify time-series position data by performing male pair



Model 2: Classify time-series position data by successful copulation

Figure 10: A diagram illustrating a neural network. The red node represents the position data, the gray nodes represent the hidden layers of the network, and the green node represents the classification made by the network.

Part 4: Interpretation Methods

Performance Metrics: Evaluate the quality of classifications	+	Post-hoc Methods: Evaluate how classifications are made

After training the machine learning models, we evaluate the existence of variation between male pairs and its impact on copulation success using two complementary approaches. We utilize performance metrics derived from confusion matrices, defined as nxn tables comparing true classifications against the models' predictions, to determine if variation exists and its broad relationship to success. These metrics, such as accuracy, precision, and recall, validate whether the models are correctly predicting both performer identity and copulation outcomes. Simultaneously, post-hoc methods like gradient tracking, permutation, and decomposition reveal the decision-making logic of each model, pinpointing exactly how and when this variation emerges and affects the respective predictions of performer and copulation success.

Next Steps

- ❑ DeepLabCut successfully labels videos
- ❑ Models designed and trained
- ❑ Interpretation of models

References and Acknowledgements

Figure 11: A QR code the leads to the references and acknowledgements for the poster.

This work was supported in part by grant DEB LTREB # 2243423 to E.H.D. We thank members of the DuVal Lab for useful discussion and feedback on this work.