Using Machine Learning to Quantify Complex Behavior in a Tropical Bird

Kaya Simmons, Lillian Gilfry, Emily DuVal

Florida State University, Department of Biological Science

Introduction

Behavioral repertoire in animal species can be complicated and difficult to characterize objectively. Researchers often quantify behavior manually through watching and annotating video files, which can be limited by user bias, requires expert knowledge, and can be very time-consuming (Luxem et al. 2023). Recent technological developments allow for more efficient analysis of complex behaviors while minimizing human error. Machine learning uses computer algorithms to identify patterns in data with minimal human guidance. This process involves training an algorithm to apply deep-learning methods to analyze videos (Nath et al. 2019). Several tools have become freely available to perform this type of analysis, creating exciting new possibilities for studying animal behavior.

We investigated the ability of machine learning software to accurately quantify complex behavior through:

1) identifying a best fit program for our model
2) training the program and analyzing results to assess accuracy

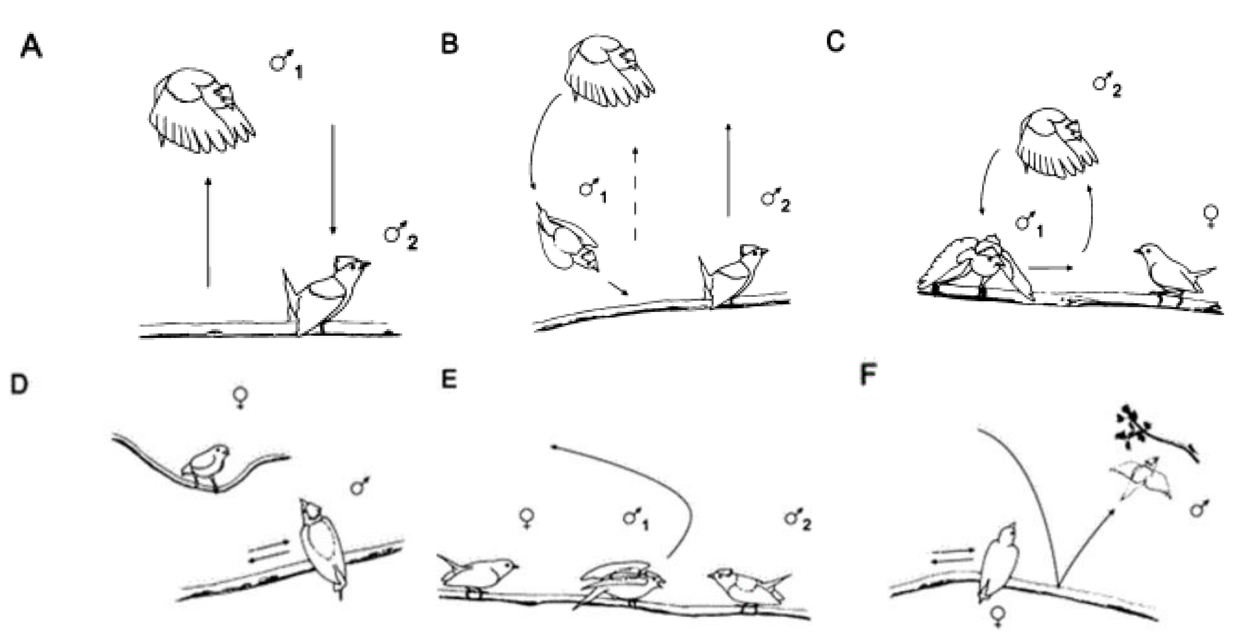

Figure 1. Example manakin display elements (DuVal, 2007)

Our System

Lance-Tailed Manakins (*Chiroxiphia lanceolata)* are small, tropical passerine birds that participate in complex cooperative courtship displays. These birds have a lek mating system, in which males remain in one location to perform displays while females explore options and freely choose who to copulate with. These displays involve multiple males performing a fast-moving, coordinated dance. They can involve up to 11 unique display elements and can last a long time (DuVal 2007). These qualities make observation and quantification of courtship displays difficult for humans to manage. Machine learning holds great potential to provide new insight.

Aim 1: Program Comparison

Broad literature review of ideal qualities for 3 potential programs →

	Multi-Animal Displays	Unsupervised Learning*	Individual Identification	Easy to use for beginners	Able to run
DLC	✓	✓	✓	✓	✓
LEAP	✗	✓	✗	✓	✓
SIPEC	✓	✓	✓	✗	✗

Figure 2. Chart displaying criteria used for determining best fit program

*Unsupervised learning is a technique in which the algorithm analyzes and clusters unlabeled data without the need for human intervention (Todd et al., 2017).

Aim 2: Data Accuracy

Methods:

- We extracted 150 automated frames from a manakin display involving a female and multiple males.
- We identified and labeled 24 body parts across frames and used these frames to created a training dataset that the algorithm used to predict points.
- After training, we allowed the program to evaluate its accuracy. This generated data on confidence for each predicted point (Figures 3 and 4), as well as frames to visualize performance (Figure 5).

Results:

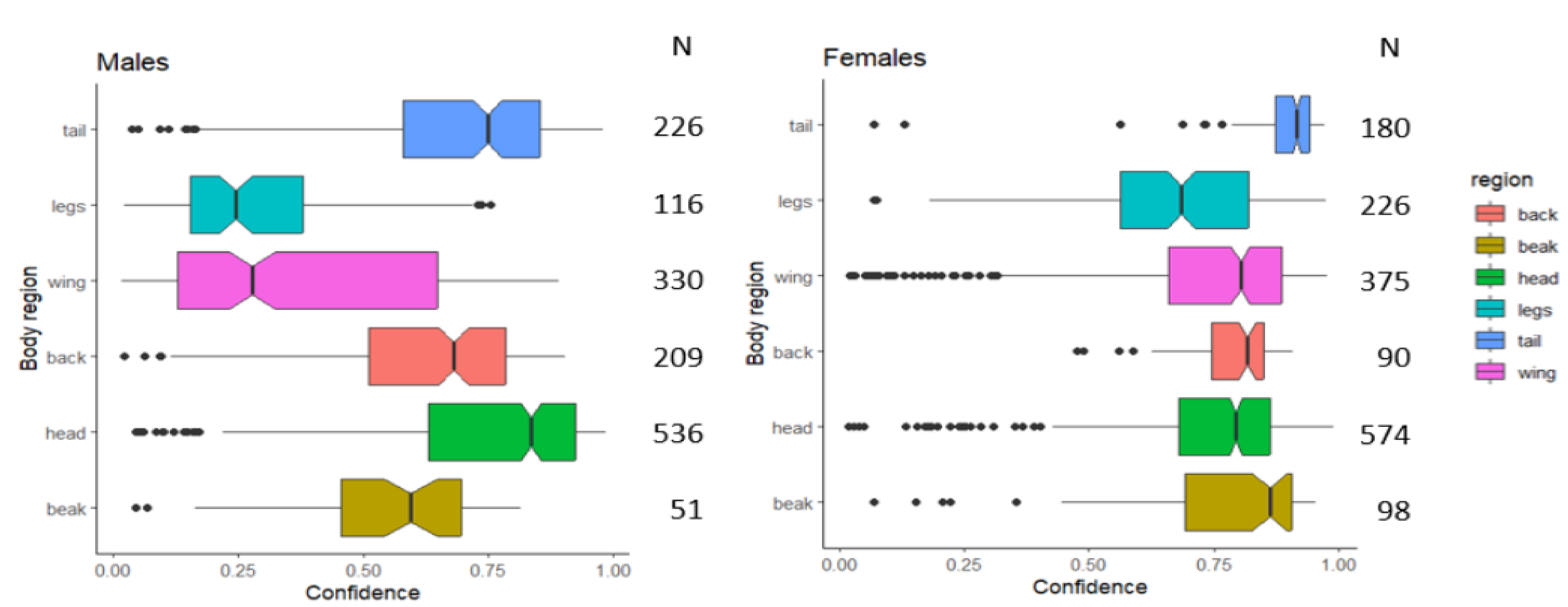

Figure 3. Box plots comparing program's confidence in identifying body part clusters by sex. N is the number of measures compared. Notches represent confidence interval around the median.

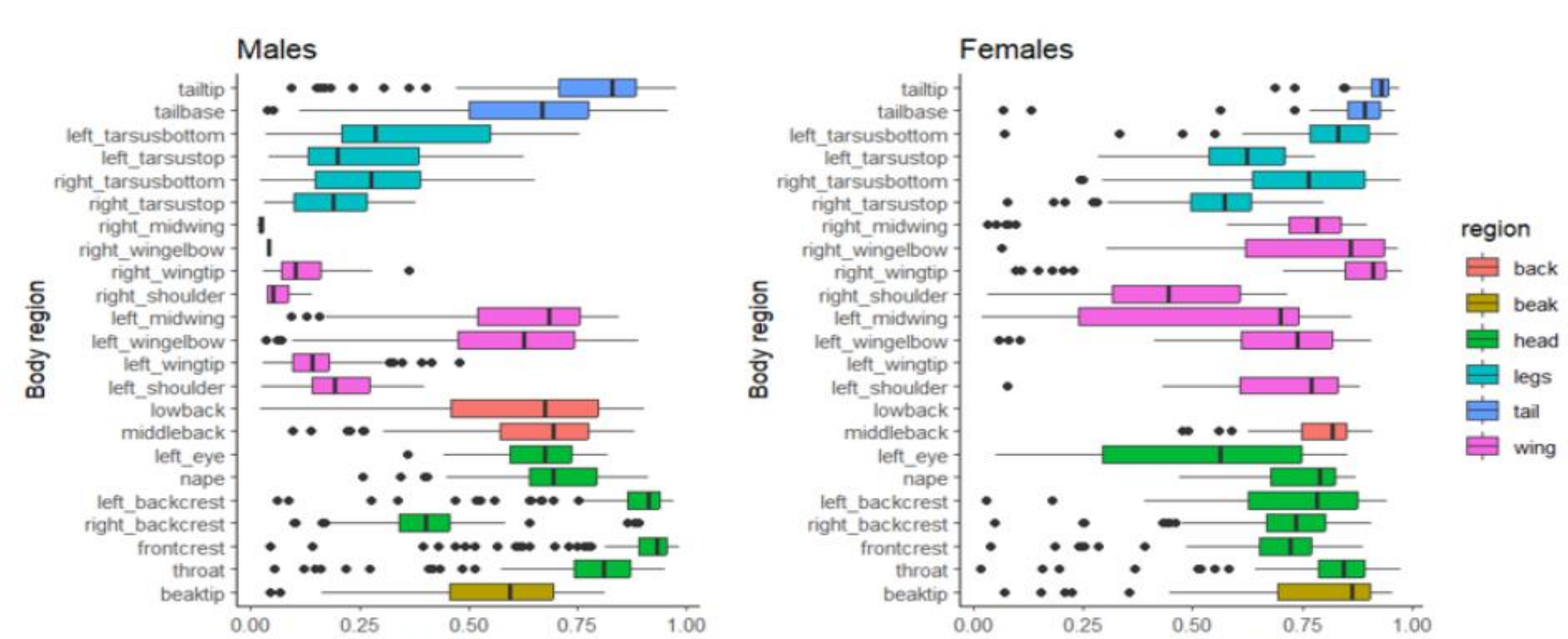

Figure 4. Box plots comparing program's confidence in identifying all labeled body parts by sex.

- Body parts of females were on average more confidently assigned than males (Kruskal-Wallis X^2 (1)= 252.8, p<0.0001).
- Body regions displayed significantly different levels of confidence within males (Kruskal-Wallis $X^2(5)$= 518.67, p < 0.0001) and within females (Kruskal-Wallis X2(5)= 212.96, p < 0.0001).

Conclusion

- There was large variation in the confidence levels of different points, both between and within sexes.
- Overall, confidence was significantly higher in females compared to males. This is likely due to the behavioral differences in the sexes. Males participate in quick, acrobatic displays, leaving body parts or the entire bird blurry in frames. Females tend to sit relatively still, making them easier to assign points to.
- Within a sex, variation in confidence could be in part caused by the variation in movement of certain body parts over others. For example, despite having more than 100 additional points of data, points on wings were assigned with significantly less confidence than points on the tail, likely because of variation in position and clarity.
- Our next steps will be to assess the extent to which labeling additional frames improves accuracy and apply this protocol to novel videos.

Figure 5. Example frames depicting program's performance. Plus signs (+) are human-labeled points, dots are predictions with a likelihood >.6 P-cutoff, and crosses (X) are predictions with a likelihood <.6.

Acknowledgements

We would like to thank all the members of the DuVal Lab for their continued support, feedback and encouragement throughout this project and beyond.

References

DuVal E.H. (2007) Cooperative display and lekking behavior of the lance-tailed manakin (Chiroxiphia lanceolata). *The Auk* 124 (4): 1168-1185.

Luxem, K., Sun, J. J., Bradley, S. P., Krishnan, K., Yttri, E., Zimmermann, J., ... & Laubach, M. (2023). Open-source tools for behavioral video analysis: Setup, methods, and best practices. *Elife*, 12, e79305.

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature protocols*, 14(7), 2152-2176.

Todd, J. G., Kain, J. S. & de Bivort, B. L. (2017) Systematic exploration of unsupervised methods for mapping behavior. *Phys. Biol.* 14, 015002