

From Genetic Variation To Visual Representation: Image Predictions In Insect Databases Using SNP

Chi-Hsien Chang, Szu-Ping Chen, Christopher Childers, and Monica Poelchau
USDA, Agricultural Research Service, National Agricultural Library
June 2024

Outline

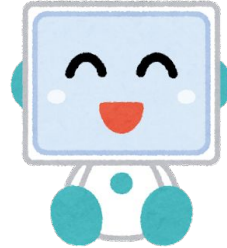
- **Motivation**
- **Method**
- **Experiment**
- **Conclusion & Future Work**

Motivation

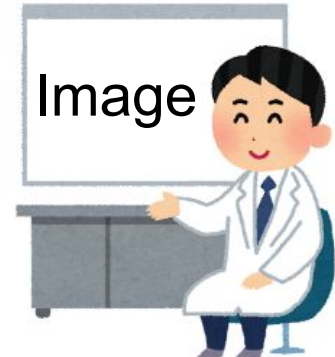
- Physical Attributes: Shape and Size



- Machine Learning (ML)

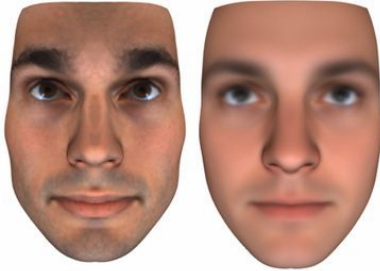


- The Potential of Medical Applications



Real (Left) vs. Predicted (Right)

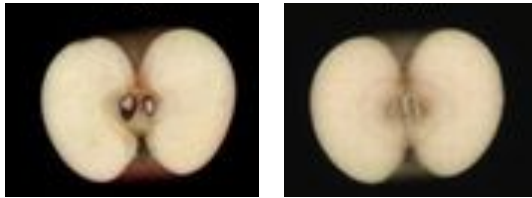
- Human



Lippert, C., Sabatini, R., Maher, M. C., Kang, E. Y., Lee, S., Arikan, O., ... & Venter, J. C. (2017). **Identification of individuals by trait prediction using whole-genome sequencing data.** Proceedings of the National Academy of Sciences.

- **Main Method: Ridge Regression and Classification**

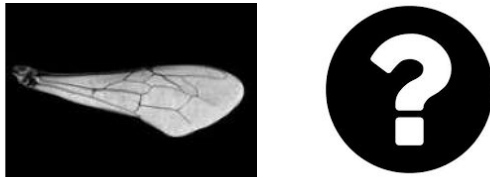
- Fruit



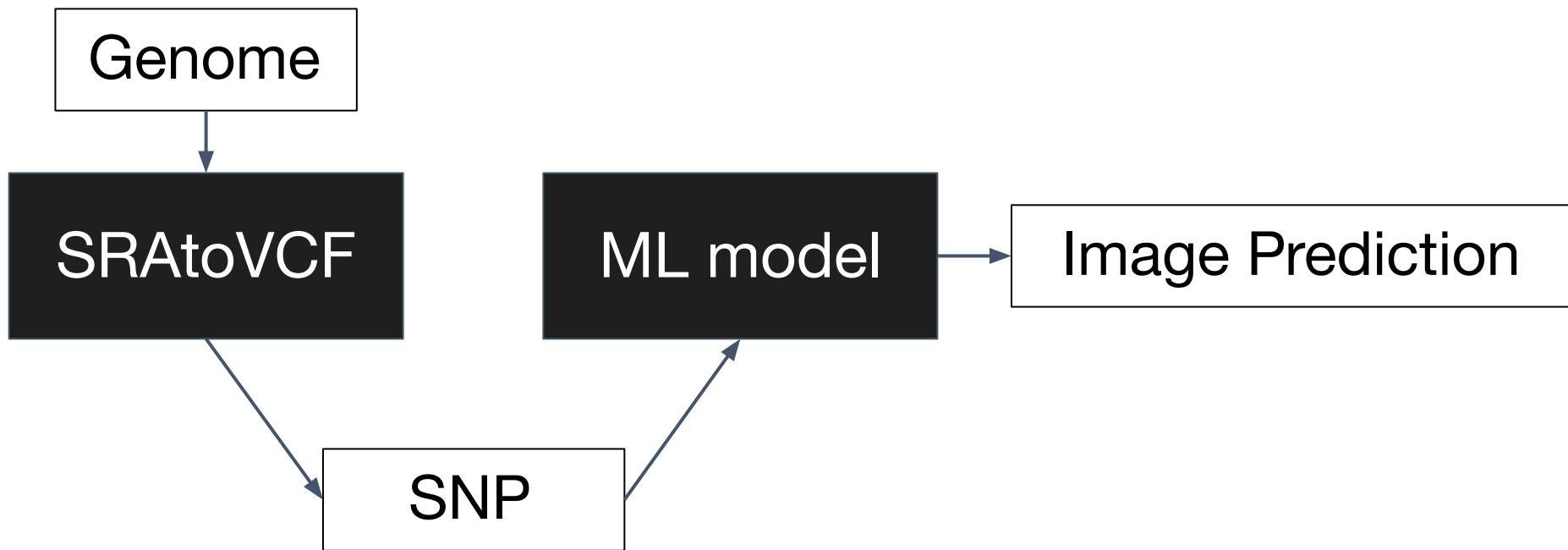
Jurado-Ruiz, F., Rousseau, D., Botía, J. A., & Aranzana, M. J. (2023). **GenoDrawing: An autoencoder framework for image prediction from SNP markers.** Plant Phenomics.

- **Main Method: Autoencoder and Embedding Predictor**
- **The approach consisted of training 2 models**

- Insect



Method: Predict Image from SNP



The source codes are available at GitHub:

- SRAToVCF: <https://github.com/NAL-i5K/SRAToVCF.git>
- ML model: https://github.com/Chi-HsienChang/Predict_Image_from_SNP.git

Experiment

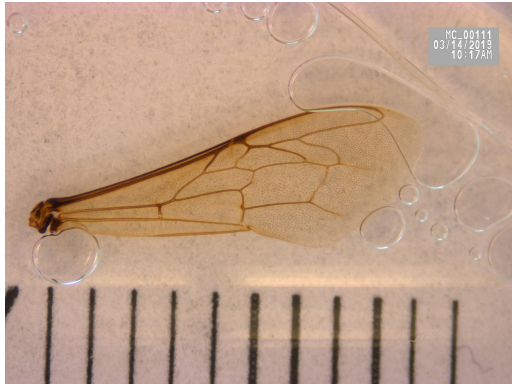
- **Dataset**



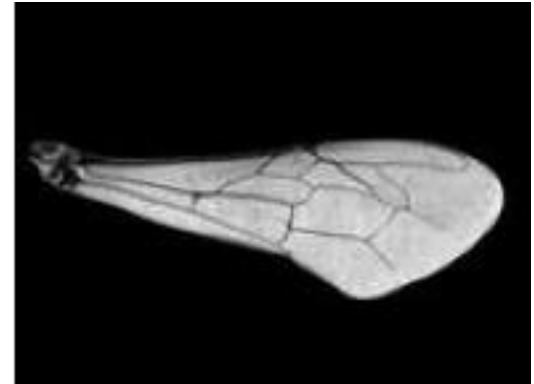
Calfee, Erin et al. (2020). **Apis mellifera wing images** (Africanized honey bees) [Dataset]. Dryad. <https://doi.org/10.25338/B8T032>

Train	Test
10	4

Original

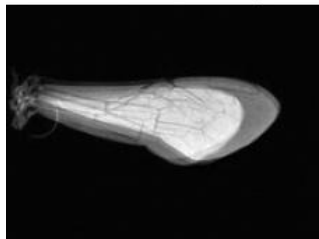


- Matting
- Black & White
- Length (Small vs. Large)
- Angle

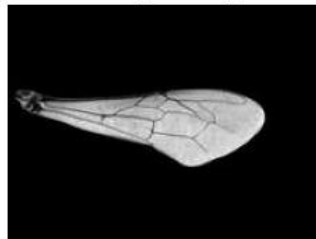


Experiment

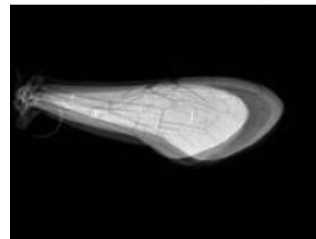
Predict 1



Target 1



Predict 2

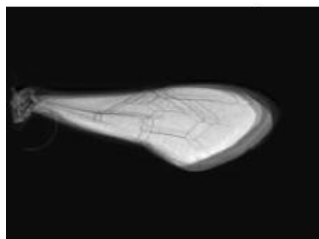


Target 2

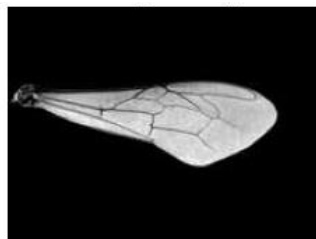


Small

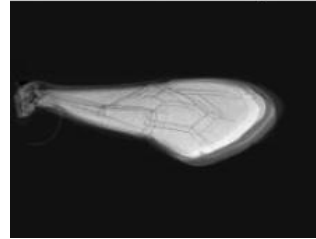
Predict 3



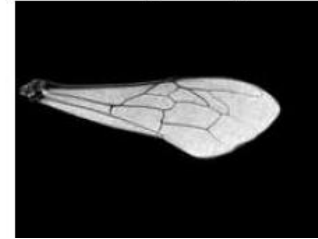
Target 3



Predict 4

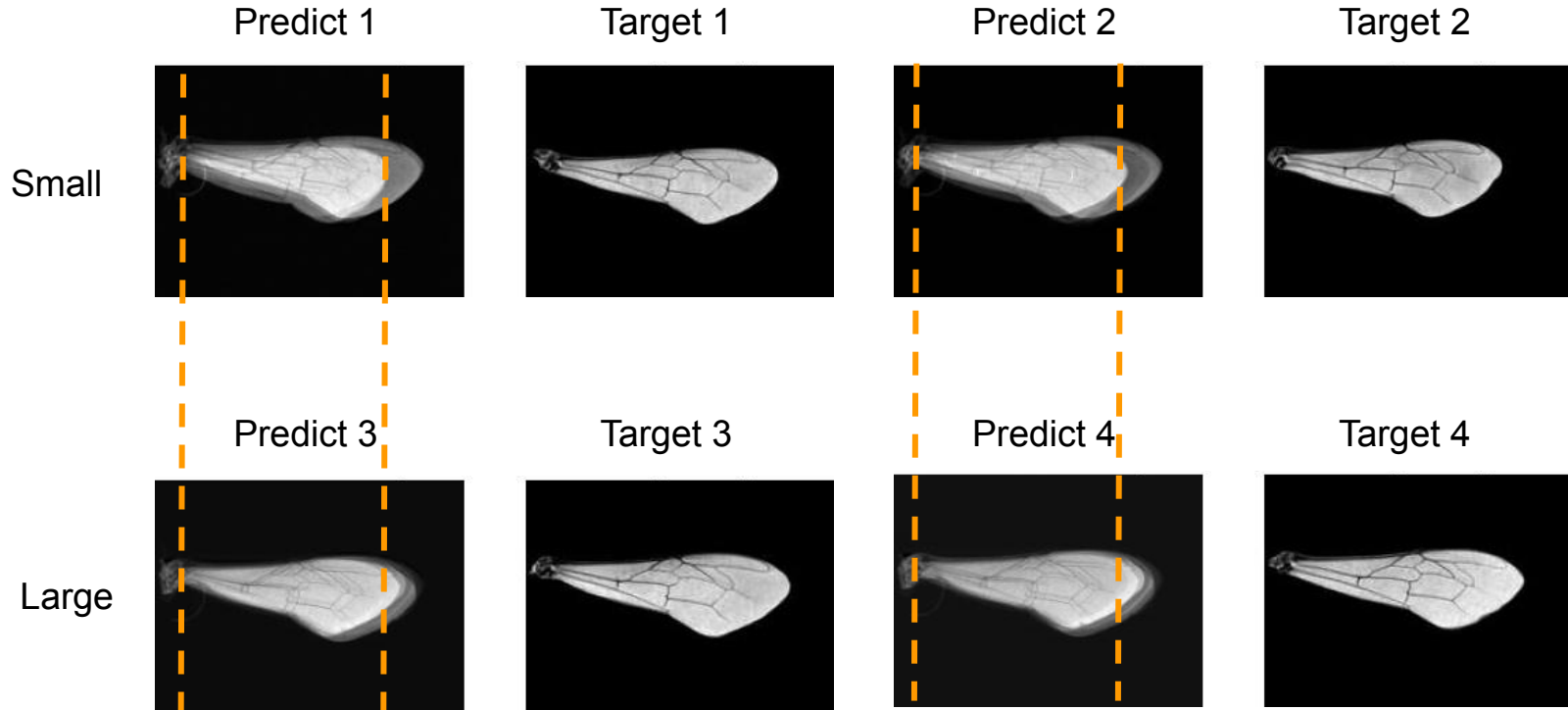


Target 4



Large

Results



Conclusion & Future Work

- Integrating ML with SNP data shows promise for medical imaging applications, despite limited dataset trials.
- Expand the dataset and advance research to enhance the ML model's accuracy and reliability for effective SNP-based image predictions.

Thank you.

Q & A