ICE 2024 KYOTO
New Discoveries through Consilience

Karlsruhe Institute of Technology

AI as a Catalyst in Entomological Research by Simplifying Species Identification

Hossein Shirali

Karlsruhe Institute of Technology

Introduction

Challenges in Species Identification[1] **:**

- High diversity and morphological similarity among species.
- Manual identification prone to human error.
- Time-consuming and labor-intensive processes.
- Requires specialized expertise and extensive training.
- Limited scalability for large-scale studies.

Introduction

Opportunities of AI[2]

- AI enhances the efficiency and accuracy of rapid species identification.
- Automated image recognition and classification.
- Machine learning algorithms tailored for entomological data.
- Enhanced data collection and analysis.
- Facilitates large-scale biodiversity studies.

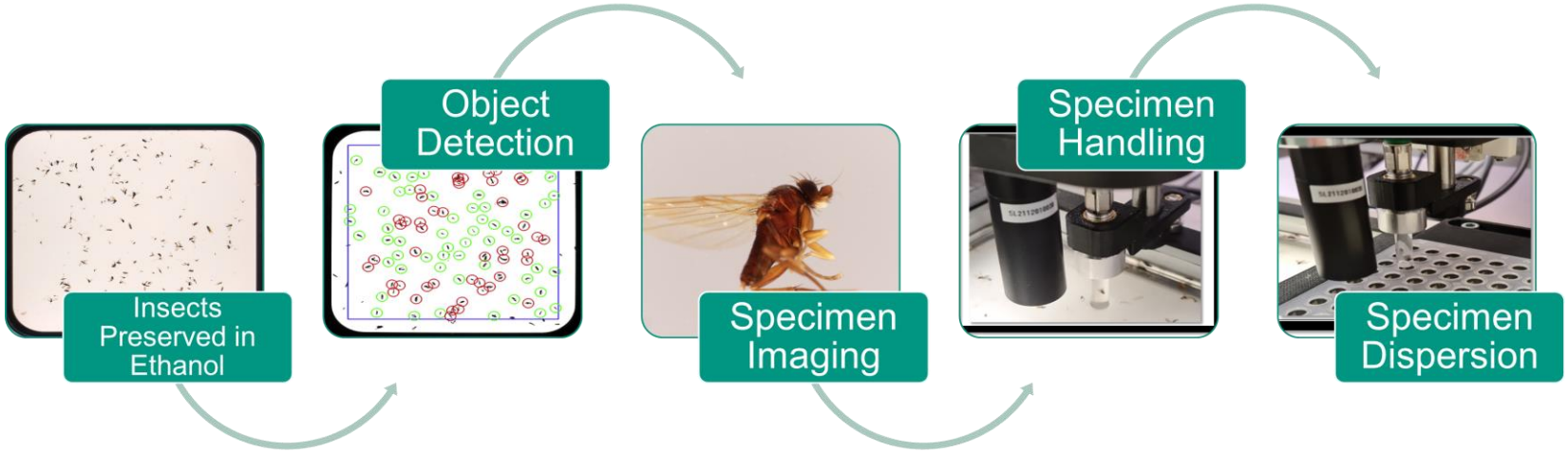
AI in Entomological Research

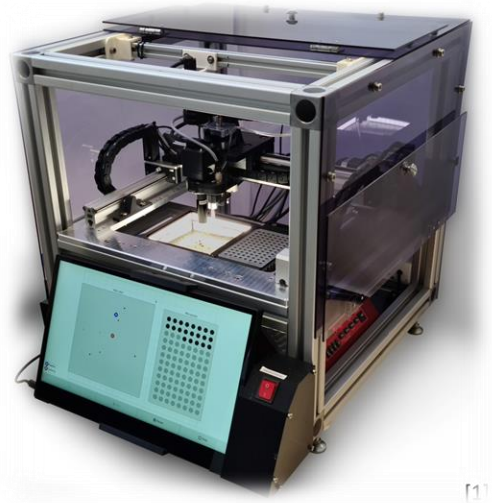
Image-based species identification

1. Tools and Data Collection
2. Data Labeling and Cleaning
3. Machine Learning Models for Species Identification
4. Automation in specimen classification and sorting

Tools and Data Collection

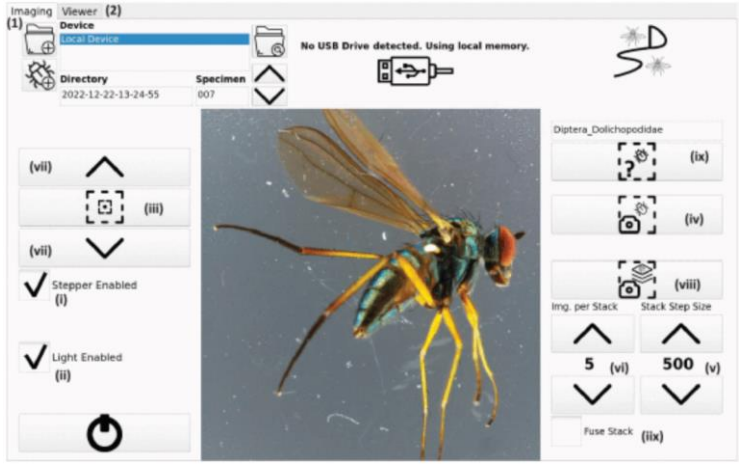
- **DiversityScanner** [3]:
  - Automatically Capture high-resolution images and sort specimens.





Tools and Data Collection

- **Entomoscope** [4]:
  - A DIY photomicroscope for imaging invertebrates preserved in ethanol.

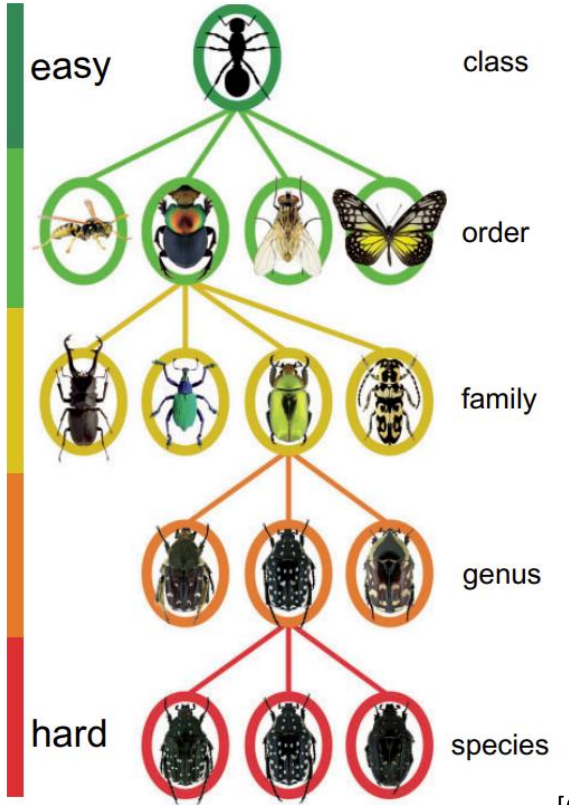


- **Workflow Integration**:
  - Streamlines data collection processes.
  - Significantly improves research efficiency.
  - Reduces time spent on data collection.

Data Labeling and Cleaning

- Expert Labeling & DNA Analysis:
  - Accurate labeling by experts.
  - Morphologically or using DNA analysis.
  - Depending on the project's goal, labeling can range from class to species level.

- Importance of Clean Data:
  - Clean data is crucial for training effective machine learning models.
  - Ensures accurate predictions and reliable insights.

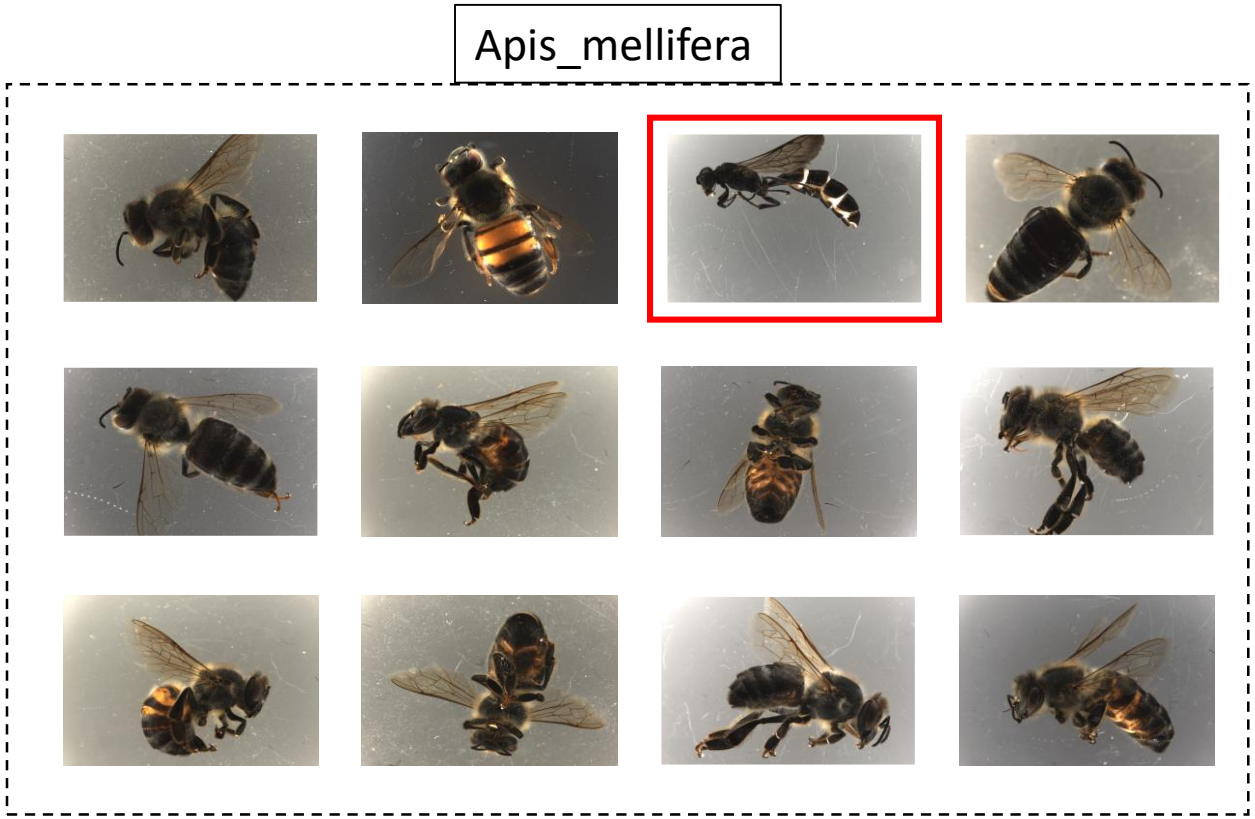


[6]

Data Labeling and Cleaning

Outlier Detection Model:

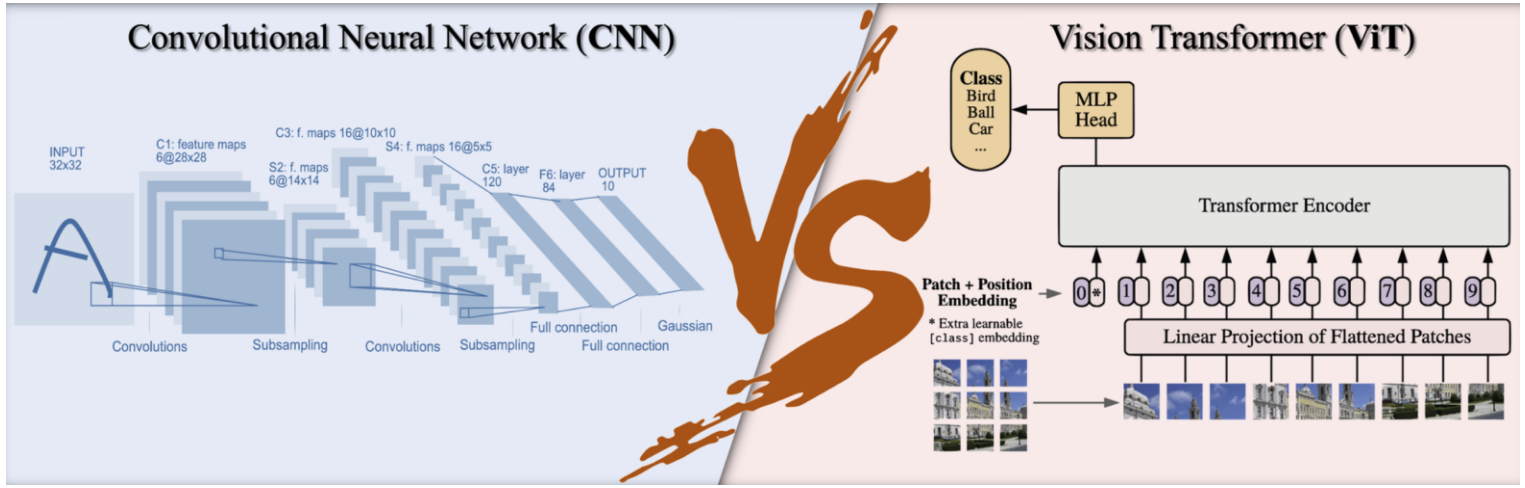
- Identify contaminations using unsupervised methods
- Flagging Potential Issues
- Enhancing Dataset Integrity
- Efficiency in Quality Control



Machine Learning Models for Species Identification

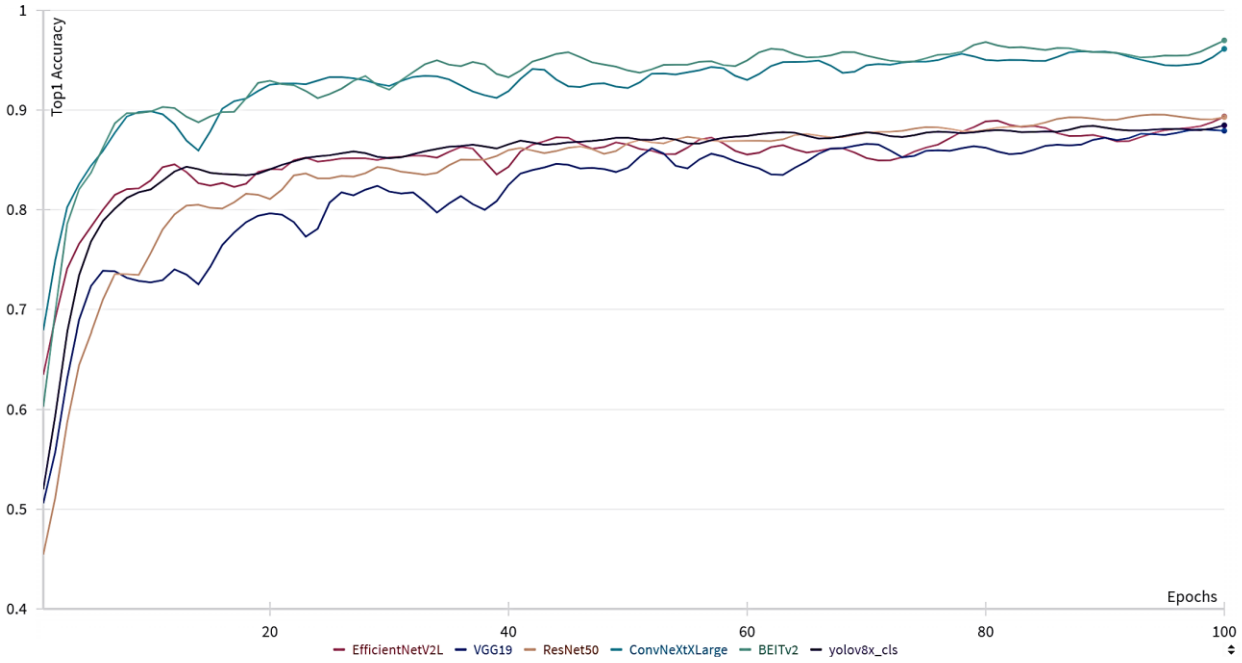
Core Models:

- Convolutional Neural Networks (CNNs): Highly effective for extracting and processing spatial features in image data.
- Vision Transformers (ViTs):Advanced architecture designed for capturing complex patterns and global relationships in images.



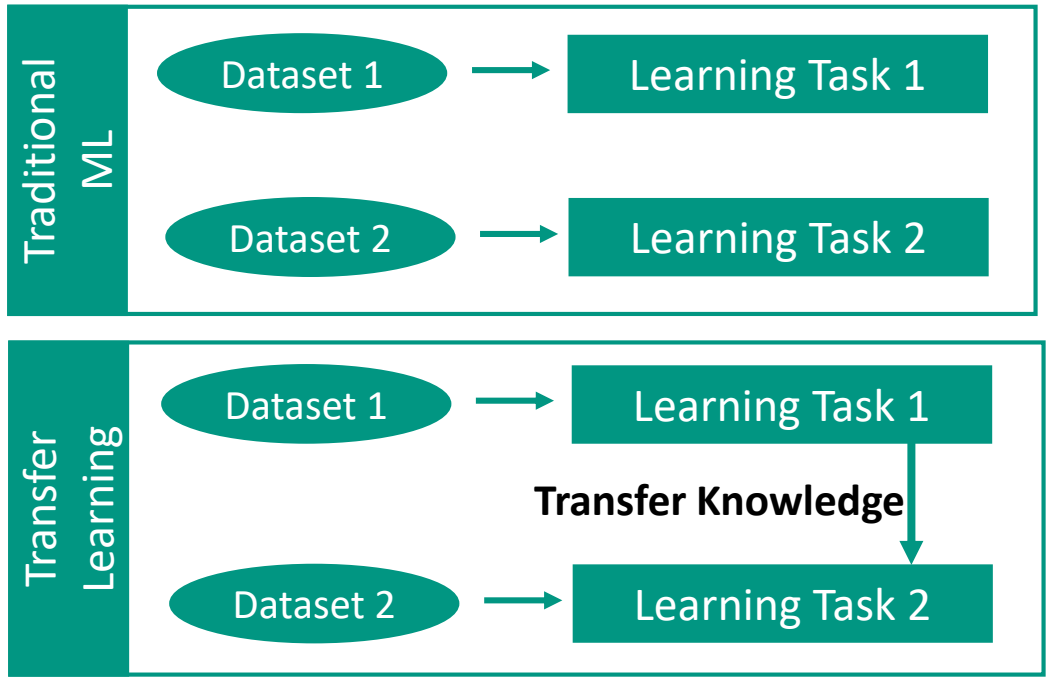
Model Selection and Evaluation

- Model Benchmarking:
  - Systematic evaluation of multiple models to identify the most suitable one for our dataset.
- Evaluation Metrics:
  - Using metrics such as accuracy, precision, recall, and F1-score and Confusion matrix to rigorously assess model performance and reliability

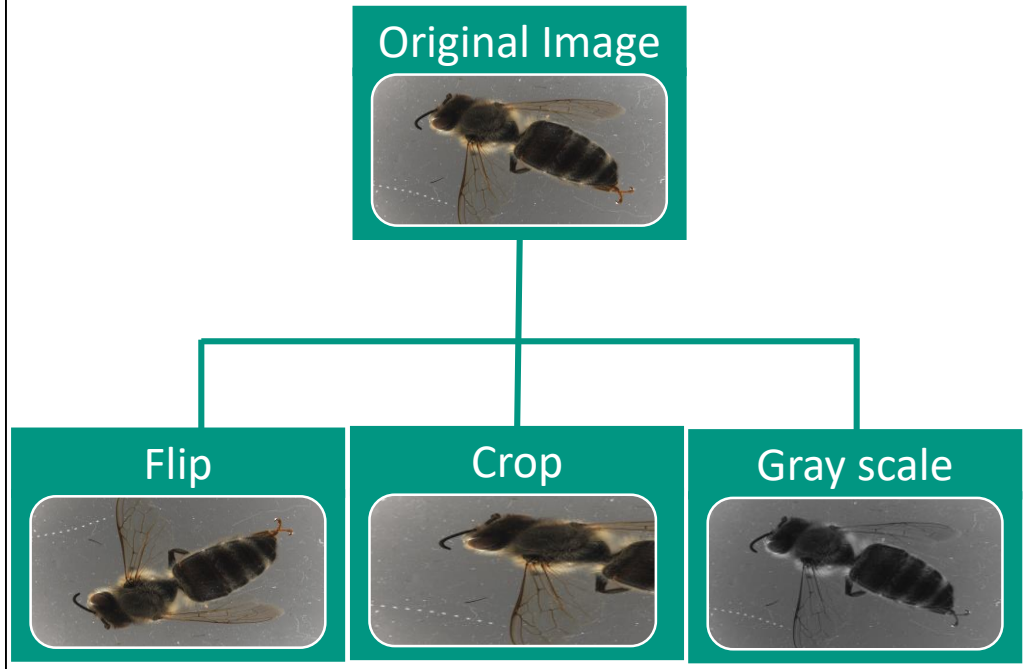


Advanced Techniques

Transfer Learning



Data Augmentation



Explainable AI (XAI)

- Increased trust in AI models.
- Easier debugging and model improvement.
- Enhanced collaboration between AI and domain experts.
- Verification of results with experts.

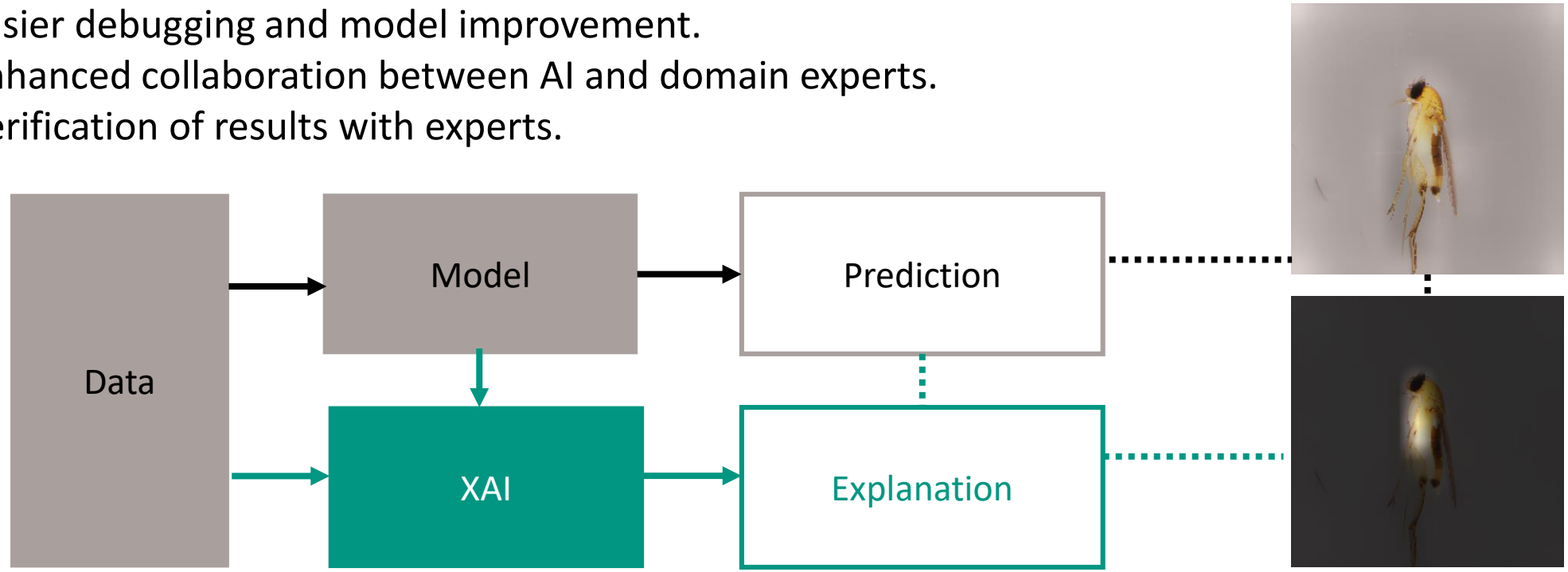
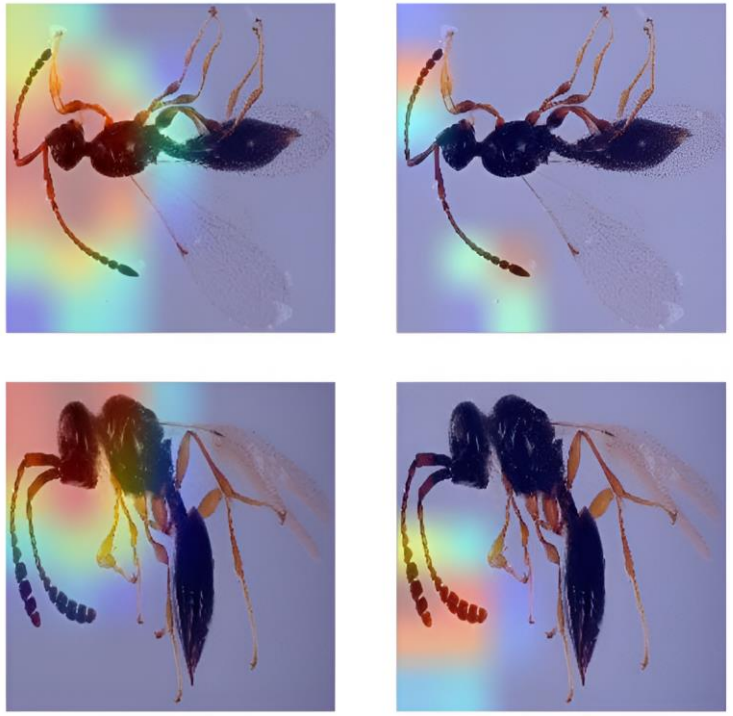


Image-based recognition of parasitoid wasps using advanced neural networks [5]

- Prepared a dataset of 2,257 images.
- Labeled using DNA barcoding and expert morphological analysis.
- Taxonomically identified to the genus and gender levels:
  - 11 diaprid genera.
  - An additional mixed group of mixed ‘other Hymenoptera’.

Image-based recognition of parasitoid wasps using advanced neural networks [5]

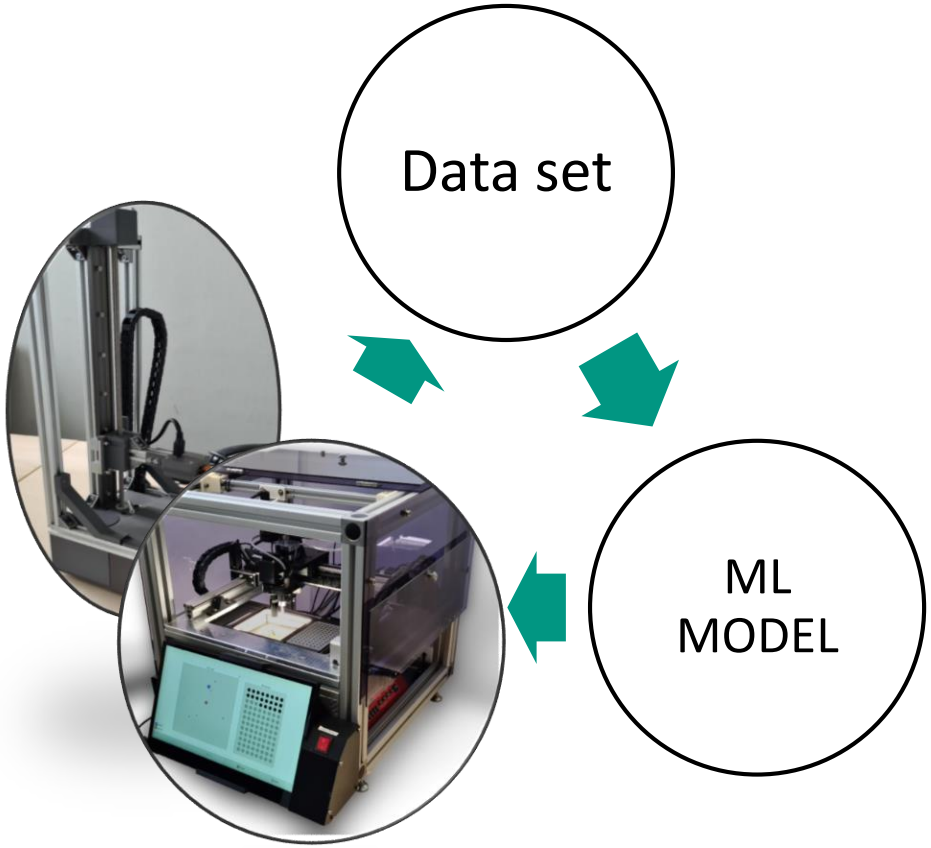
Architectures	Genus accuracy	Sex accuracy
BEITV2	0.96	0.97
ConvNeXt XLarge	0.94	0.95
YOLOv8	0.89	0.94



Automation in Specimen Classification and Sorting

- Integration of trained models with DiversityScanner and Entomoscope:

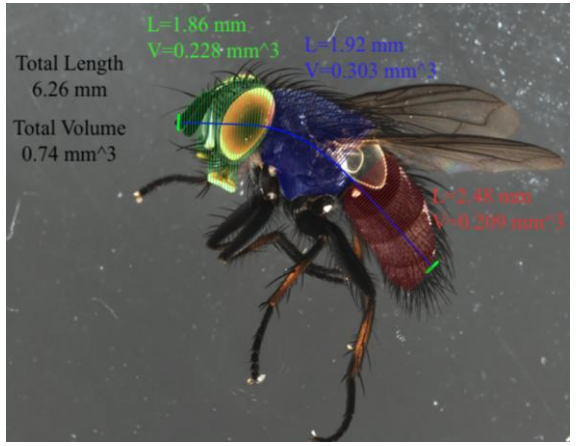
- Benefits:
  - Fast and high-accuracy identification of species.
  - Facilitates the identification of unknown species.
  - Enables scaling of models through active learning.
  - Enhances model performance with iterative expert feedback and retraining.



Future Directions and Impact

- Scaling Up Classification:
  - Enhancing model capabilities and training with diverse datasets.
  - Goal to identify the most common species captured in a Malaise trap.

- Biomass Estimation from 2D Images:
  - Developing new techniques to estimate biomass (length and volume) from 2D images.
  - Integrating image analysis with ecological data for comprehensive insights.



Thank you for your attention!

MUSEUM FÜR NATURKUNDE BERLIN

Let's connect !

References

[1] Guerrero, M. S., & Cayabyab, B. F. (2024). DNA barcoding of Locusta migratoria manilensis (Orthoptera: Acrididae) reveals insights into the species and subspecies differentiation. Journal of Entomological Science, 59, 125-132.
[2] Sharma, R. P., et al. (2023). Exploring the significance of insects in ecosystems: A comprehensive examination of entomological studies. International Journal of Environment and Climate Change. https://doi.org/...
[3] Wührl, L., Pylatiuk, C., Giersch, M., Lapp, F., Balke, M., Schmidt, S., Cerretti, P., & Meier, R. (2022). DiversityScanner: Robotic handling of small invertebrates with machine learning methods. Molecular Ecology Resources, 22(4), 1626-1638. https://doi.org/10.1111/1755-0998.13567

References

[4] Wührl, L., Rettenberger, L., Meier, R., Hartop, E., Graf, J., & Pylatiuk, C. (2024). Entomoscope: An open-source photomicroscope for biodiversity discovery. IEEE Access, 12, 11785-11794. https://doi.org/10.1109/ACCESS.2024.3355272

[5] Shirali, H., Hübner, J., Both, R., Raupach, M., Reischl, M., Schmidt, S., & Pylatiuk, C. (2024). Image-based recognition of parasitoid wasps using advanced neural networks. Invertebrate Systematics, 38, IS24011.

[6] Miroslav Valan, Karoly Makonyi, Atsuto Maki, Dominik Vondráček, Fredrik Ronquist, Automated Taxonomic Identification of Insects with Expert-Level Accuracy Using Effective Feature Transfer from Convolutional Networks, *Systematic Biology*, Volume 68, Issue 6, November 2019, Pages 876–895, https://doi.org/10.1093/sysbio/syz014