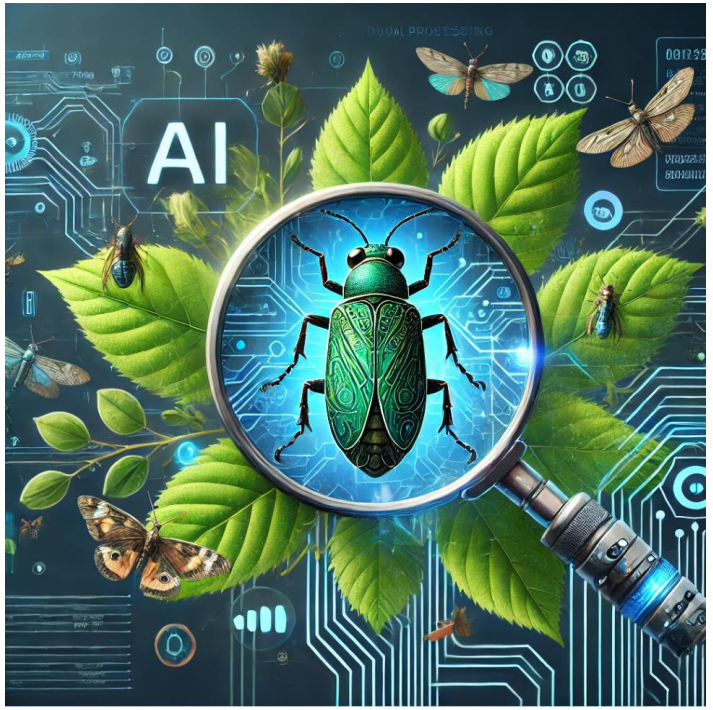
AI-driven advances in Species Identification and Biomass Analysis

Hossein Shirali

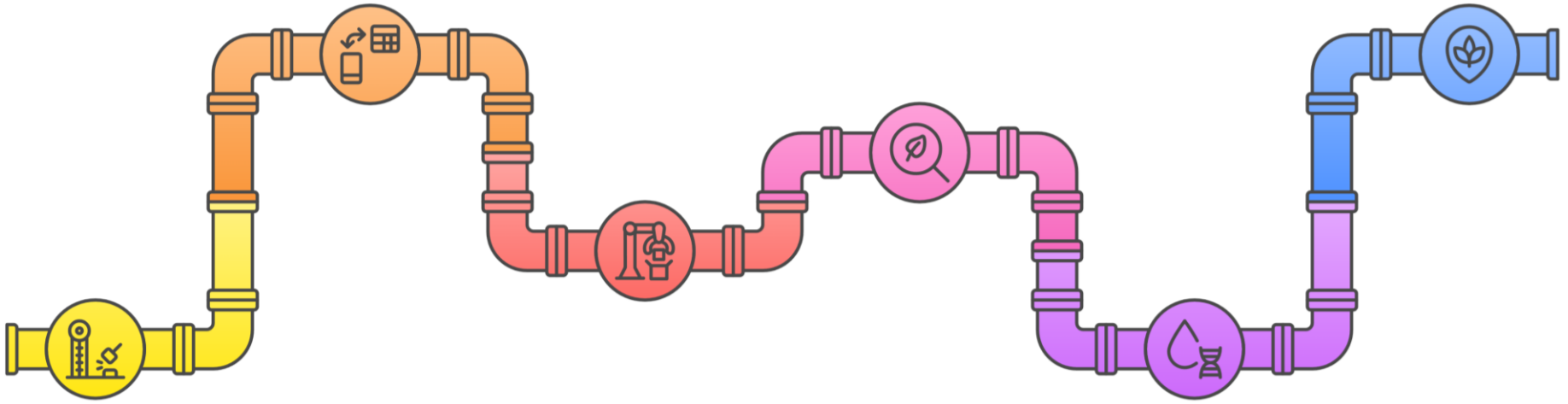
Karlsruhe Institute of Technology

Introduction: AI for Streamlining Entomological Research

- **Entomological Research Challenges:**
  - Scale and Diversity
  - Labor & Time
- **AI as a Solution:**
  - Efficiency & Accuracy
    - Focus:
      - Species Identification
      - Biomass Analysis



Integrated Pipeline for AI-Driven Entomological Analysis



01
Sampling
Collecting specimens using traps and nets

02
Sorting
Pre-sorting and separating samples

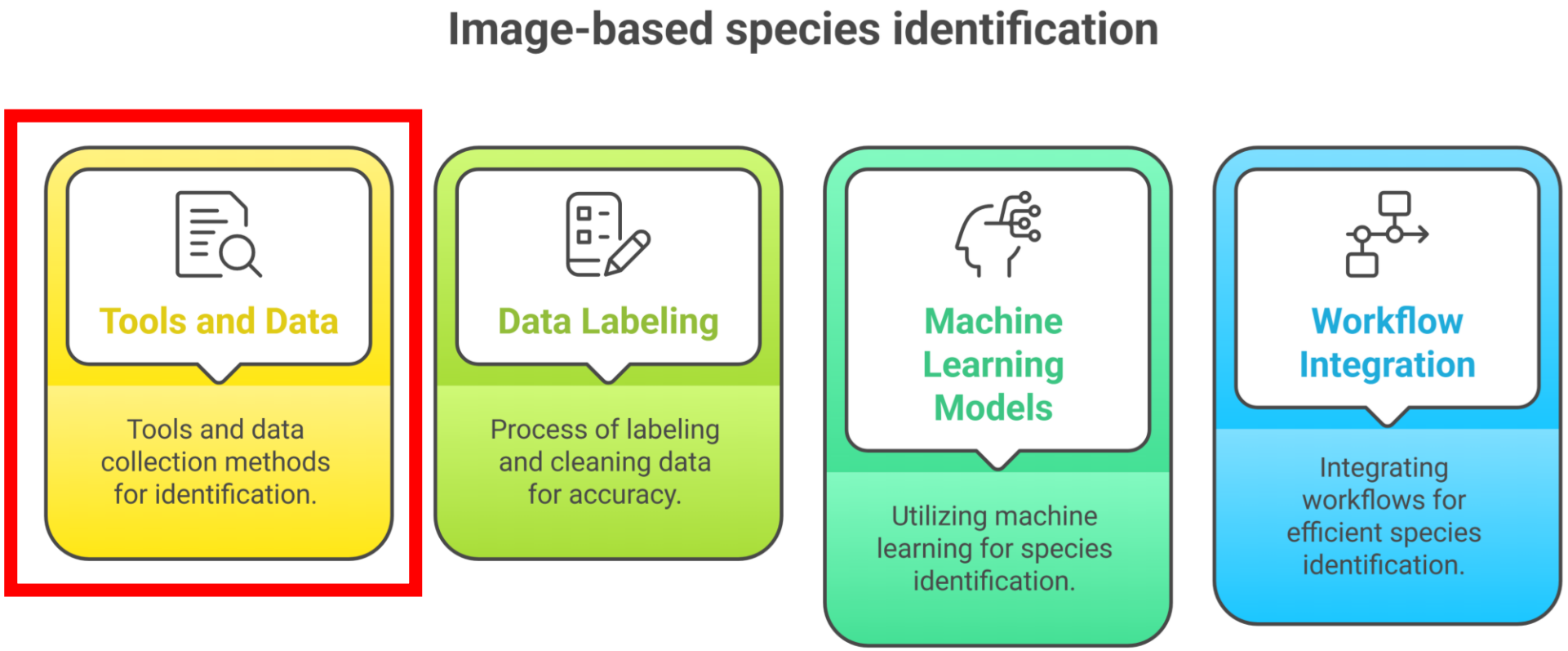
03
Imaging
Capturing automated images of specimens

04
Classification
Identifying species using AI image analysis

05
Sequencing
Validating findings through DNA barcoding

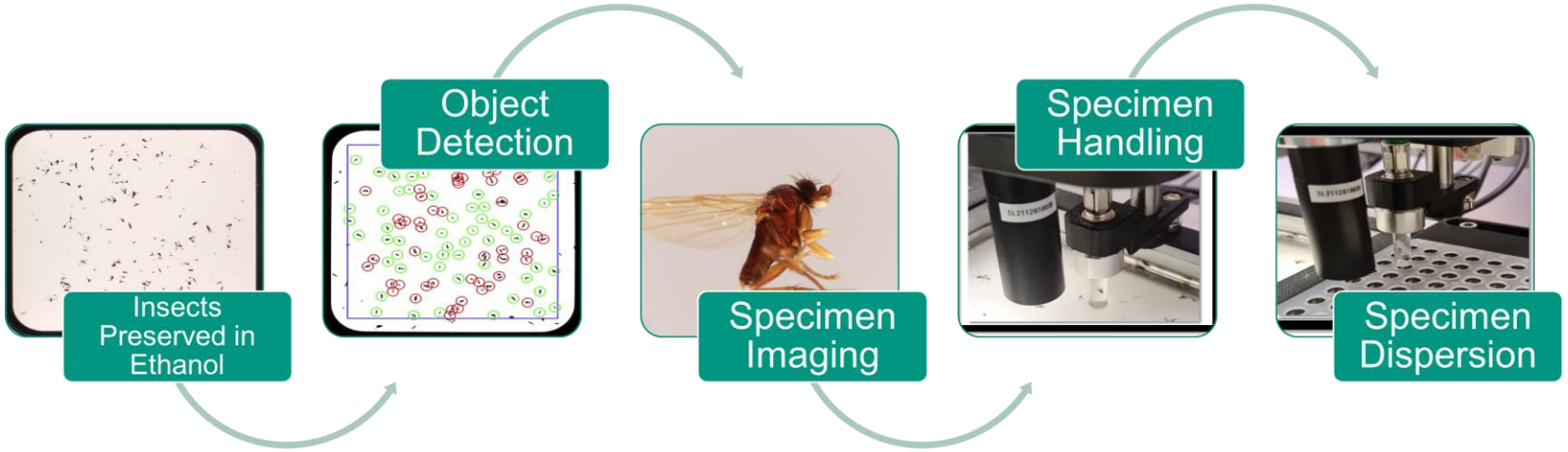
06
Morphometric Analysis
Estimating biomass with AI algorithms

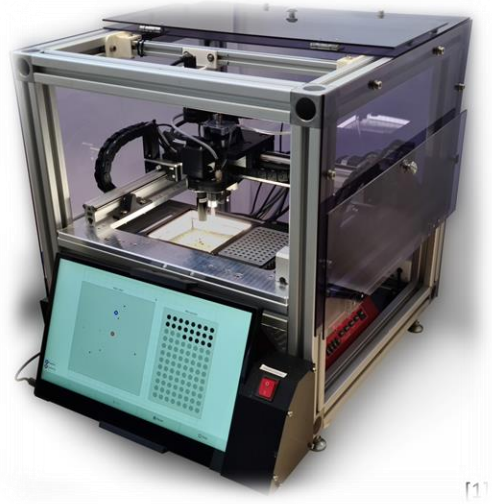
AI in Entomological Research: Image-based Species Identification



Tools and Data Collection: DiversityScanner & Entomoscope

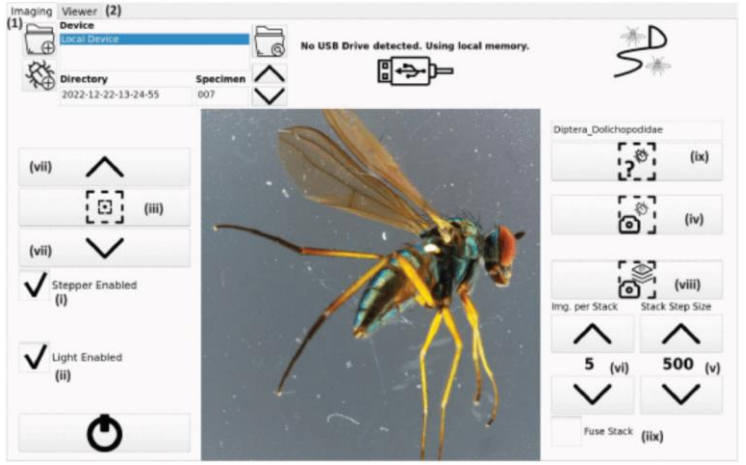
- **DiversityScanner** [3]:
  - Automatically Capture high-resolution images and sort specimens.





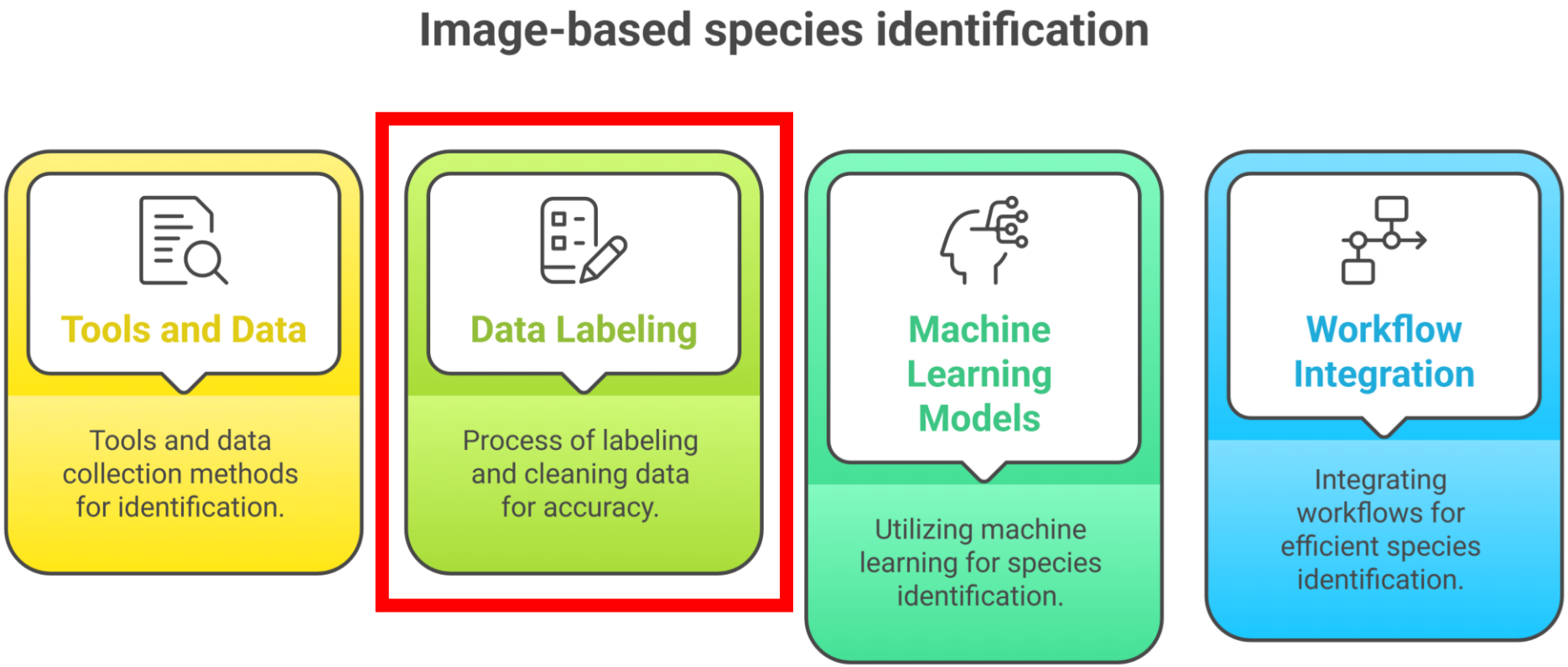
Tools and Data Collection

- **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.

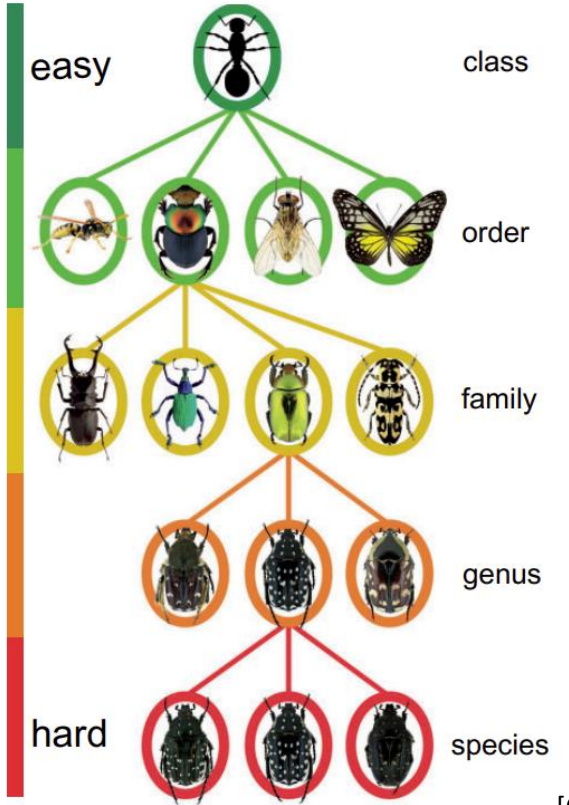
AI in Entomological Research: Image-based Species Identification



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.

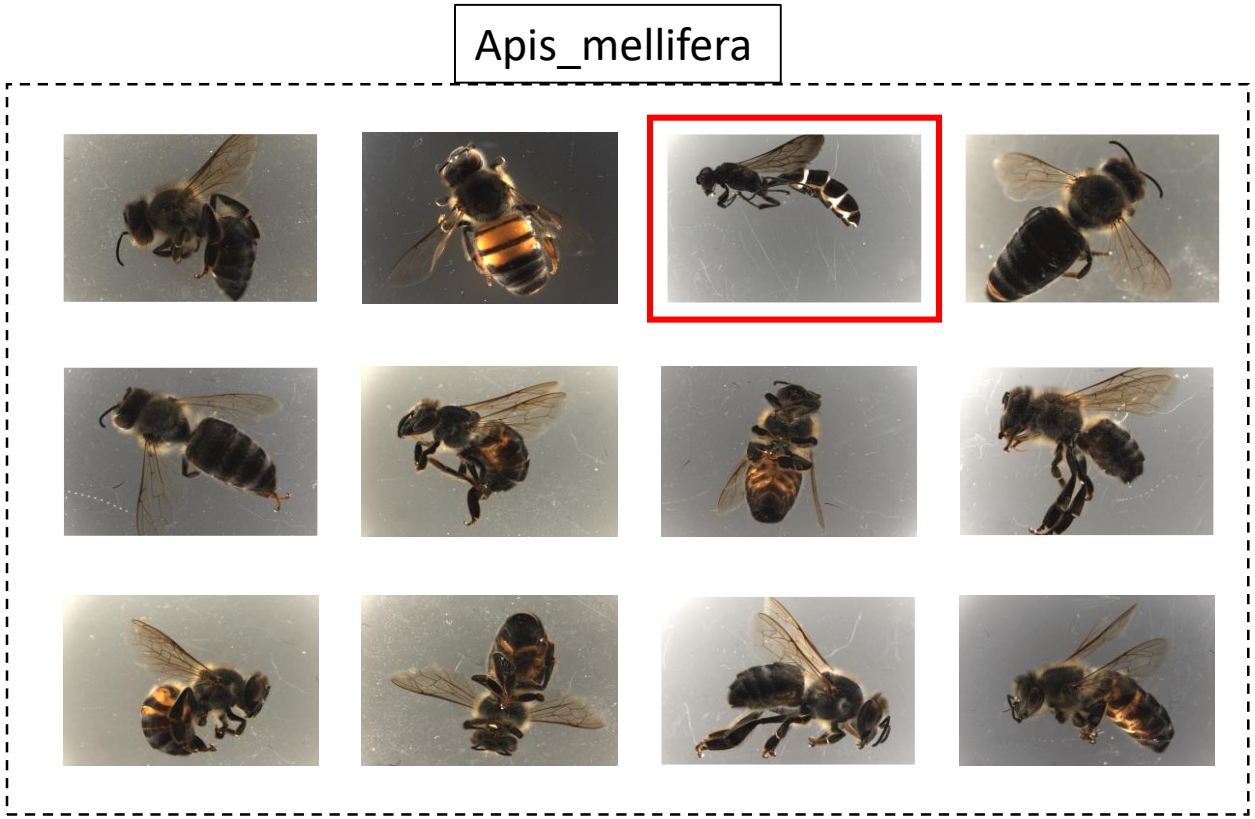


[6]

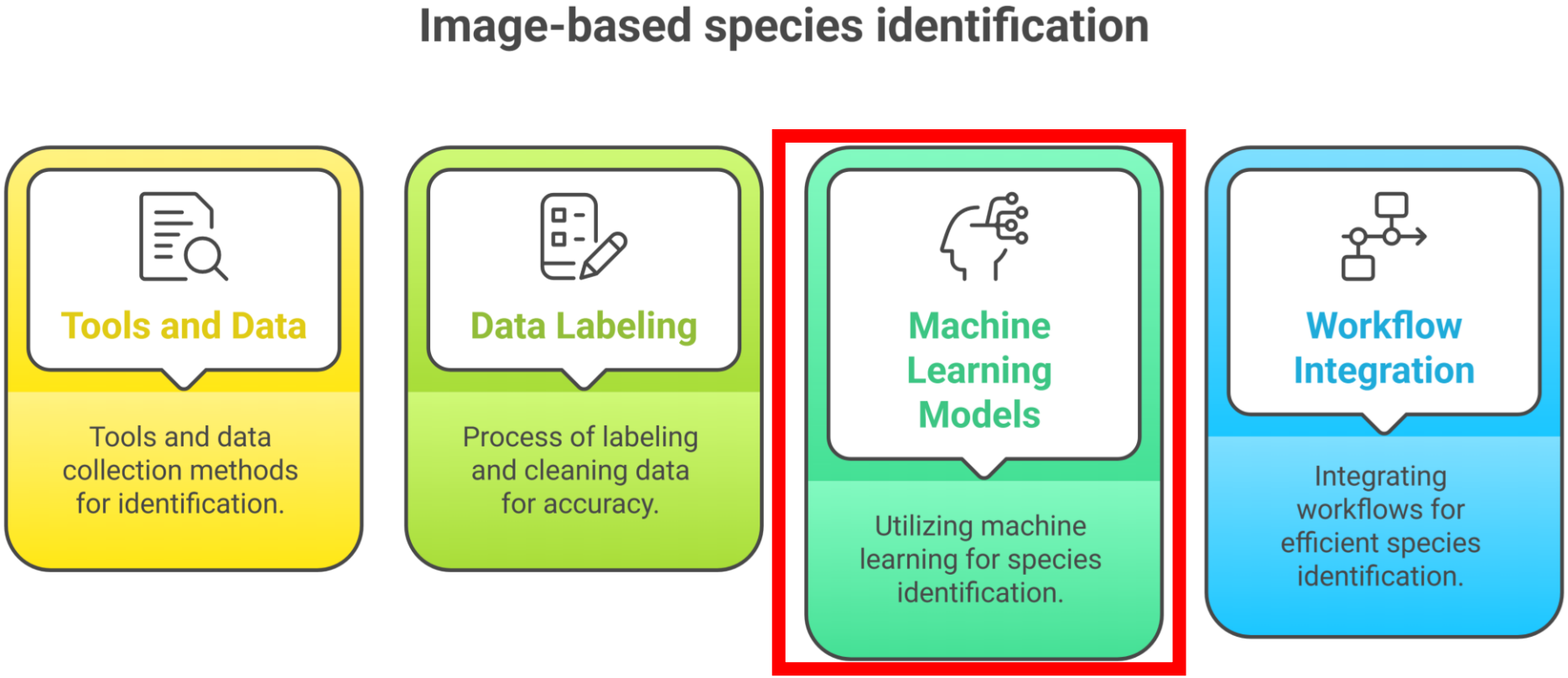
Data Labeling and Cleaning

Outlier Detection Model:

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



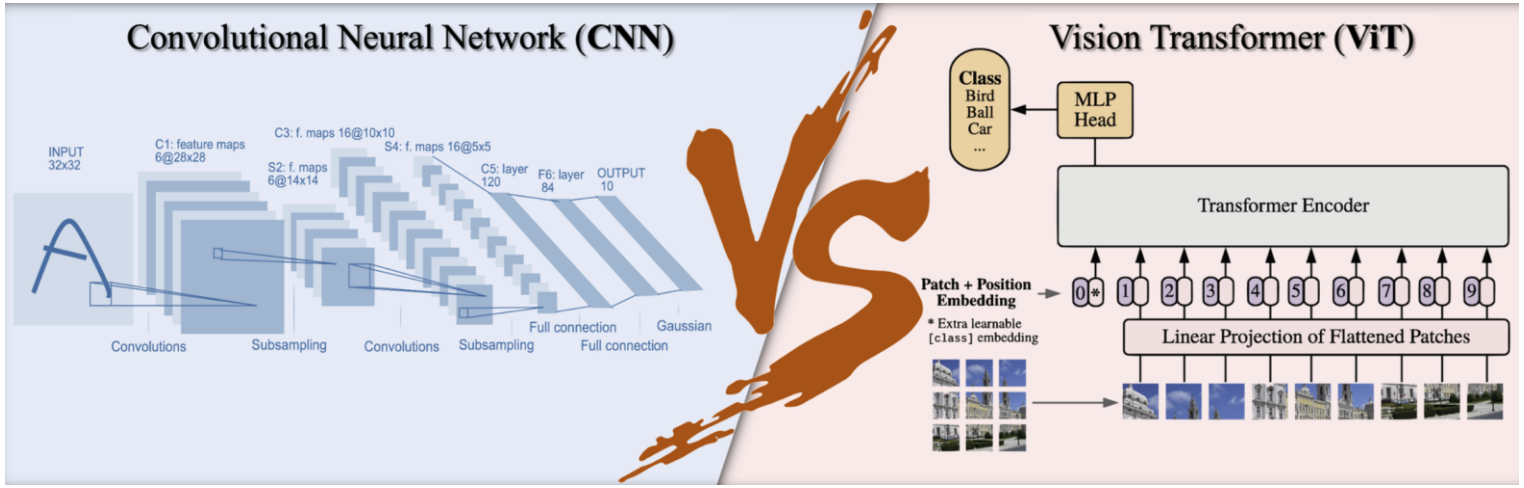
AI in Entomological Research: Image-based Species Identification



Machine Learning Models for Species Identification

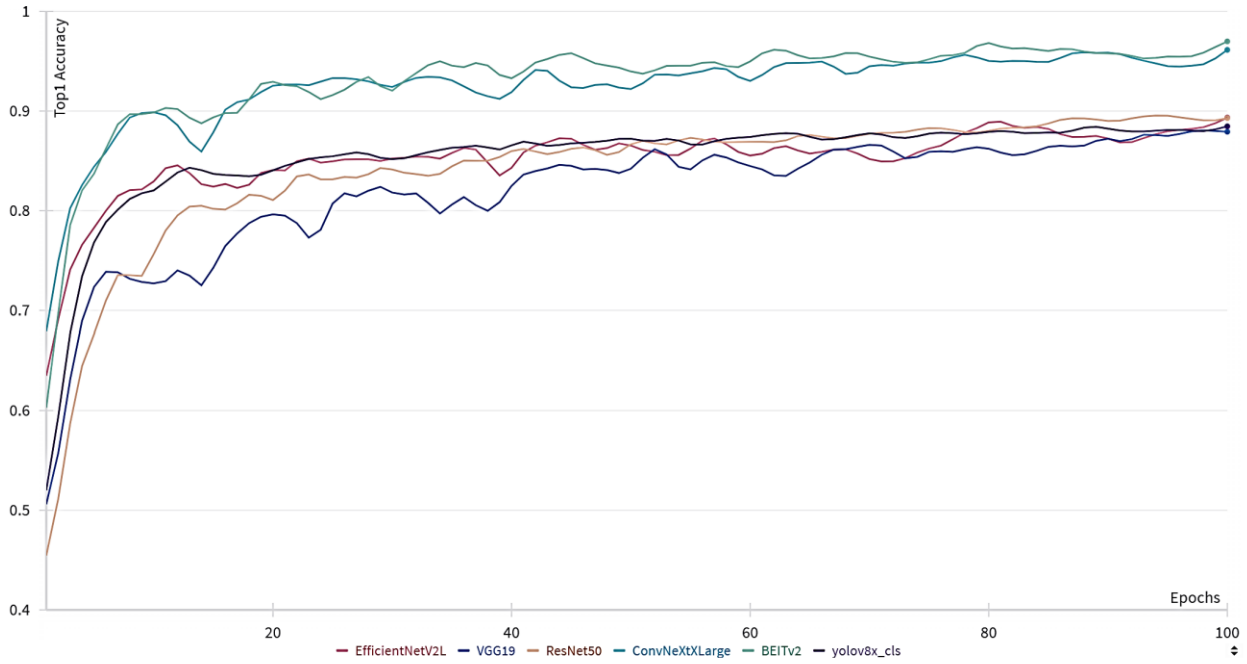
Core Models:

- **Convolutional Neural Networks (CNNs)**
- **Vision Transformers (ViTs)**



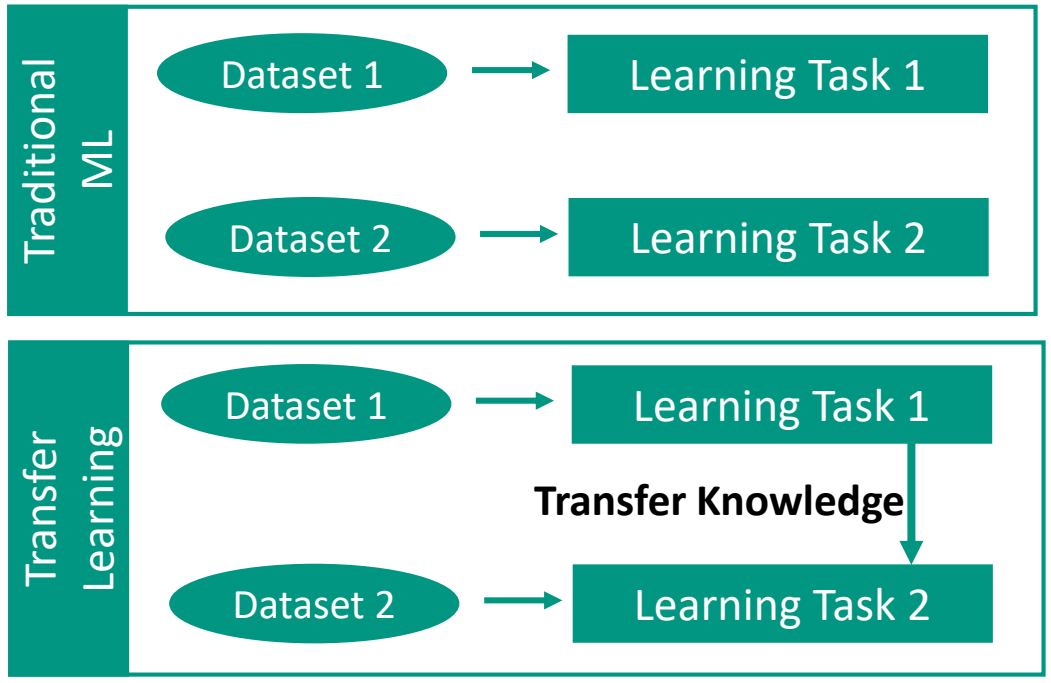
Model Selection and Evaluation

- **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

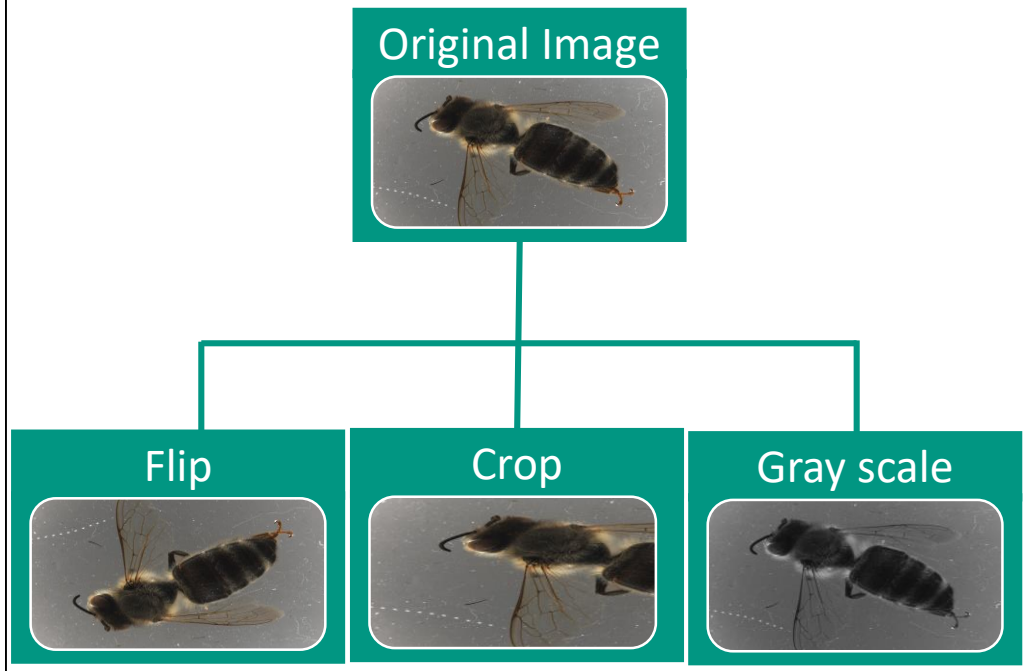


Advanced Techniques: Enhancing Model Performance

Transfer Learning

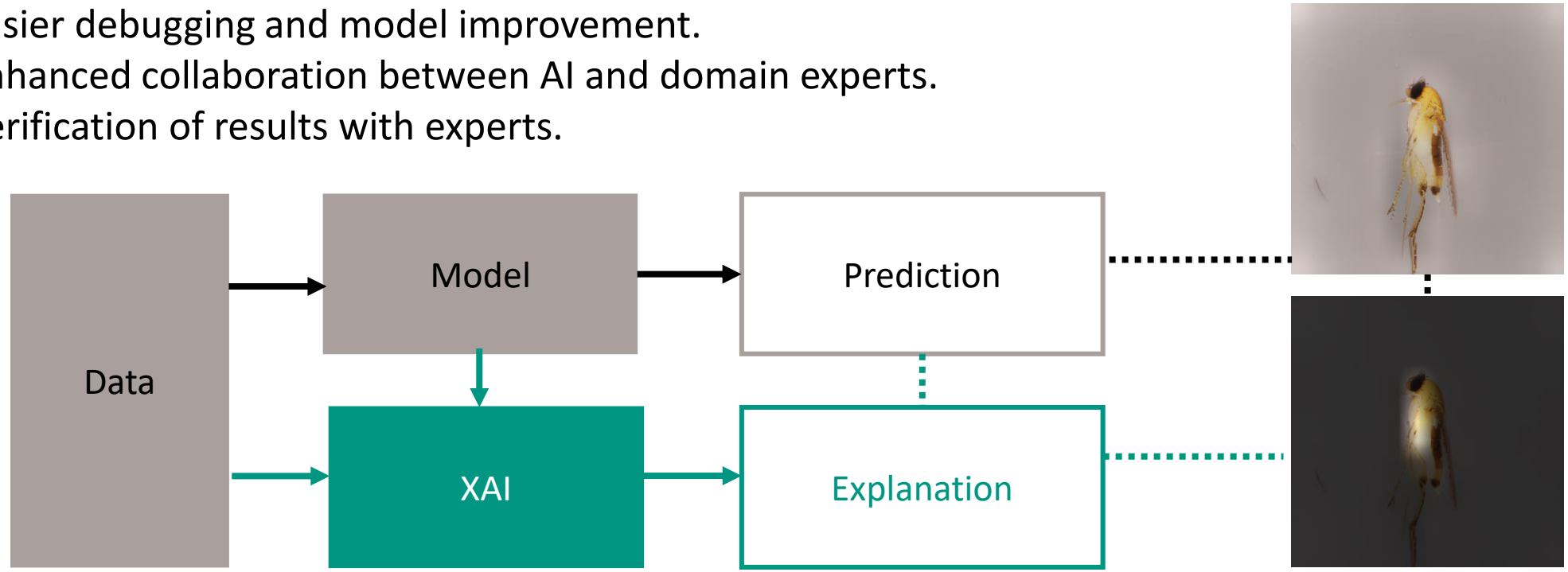


Data Augmentation



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.

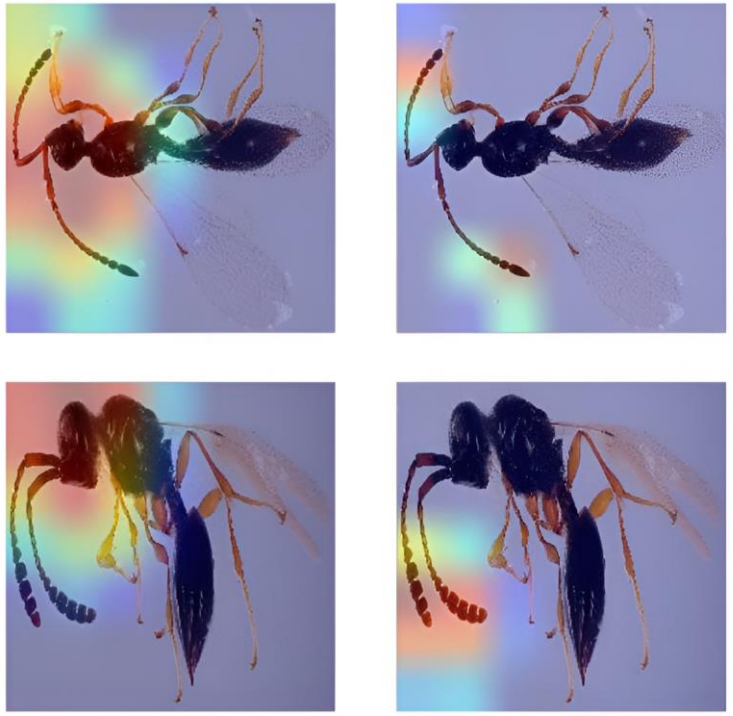


Case Study: Image-based recognition of parasitoid wasps using advanced neural networks [5]

- 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’.

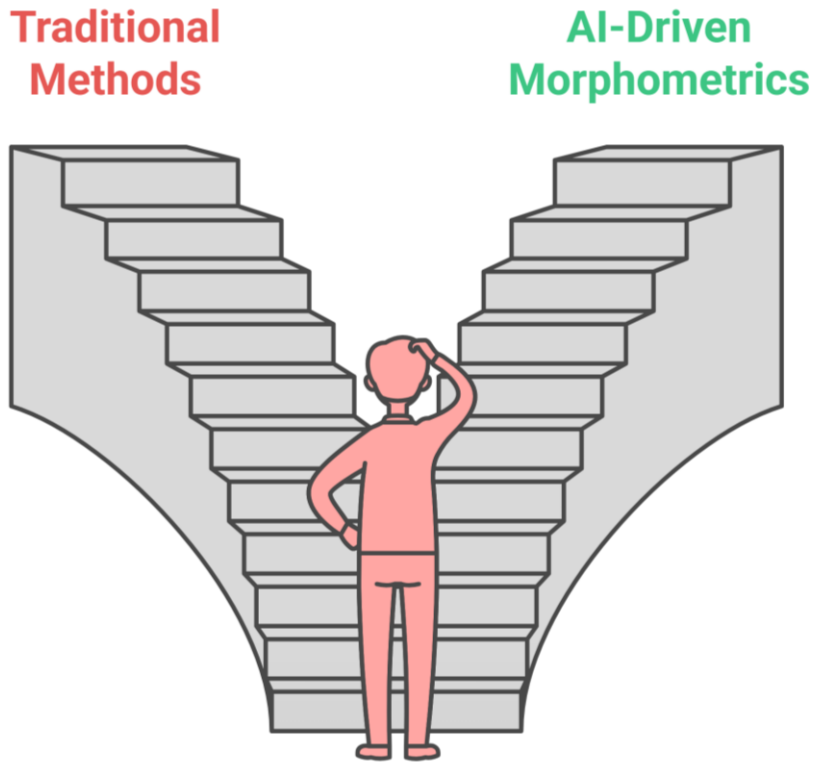
Image-based recognition of parasitoid wasps using advanced neural networks [5]

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



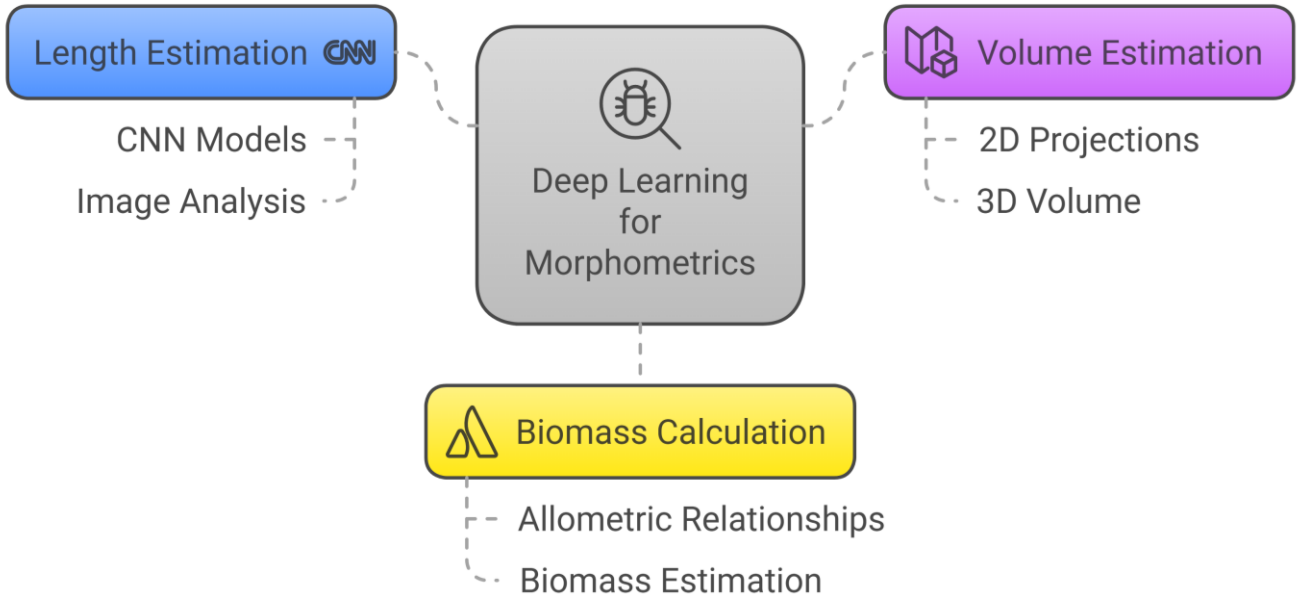
AI-Driven Biomass Analysis: Towards Automated Morphometrics

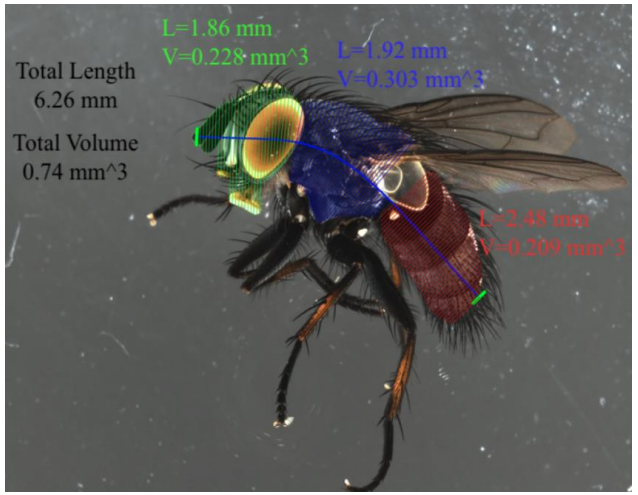
- **Biomass Analysis**
  - Ecological Significance
  - Traditional Methods
  - Need for Automation
- **Advantages of AI-Driven Morphometrics**
  - Non-destructive
  - Automated & Efficient
  - Scalable
  - Accurate



Automated Insect Morphometrics: Deep Learning for Biomass Estimation

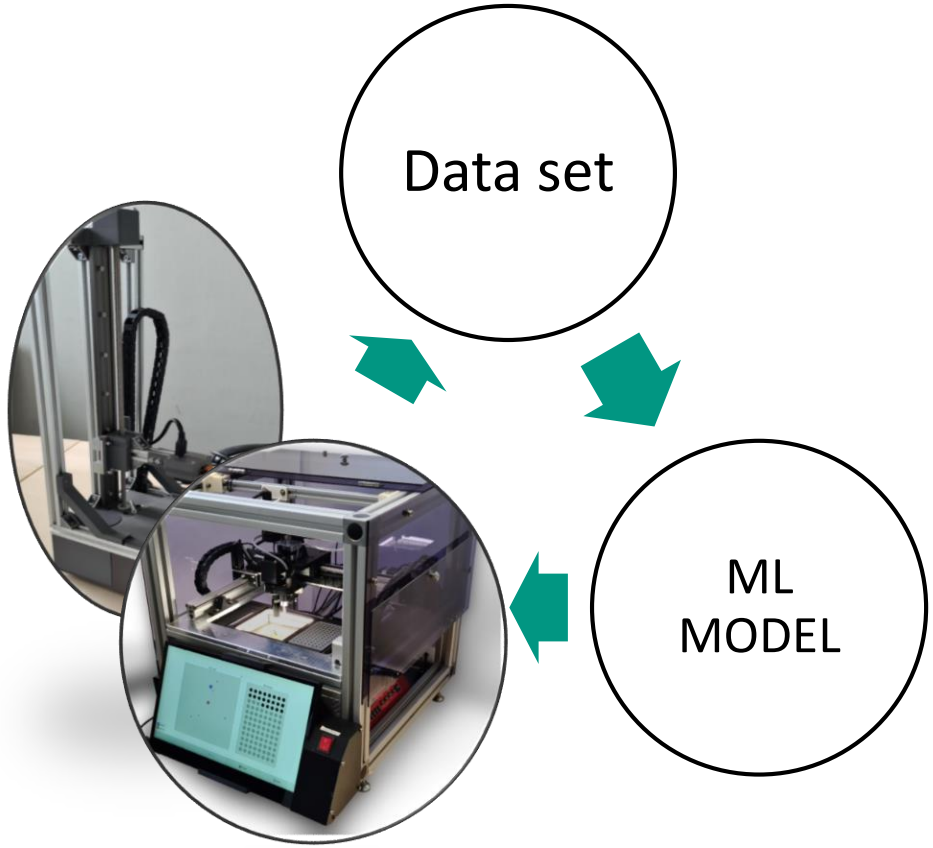
- **Introducing Automated Insect Morphometrics:**
  - **Our Approach:** Deep Learning-based methods for length, volume, and biomass estimation from 2D images.





Automation and Synergistic Insights: Integrated AI for Enhanced Entomological Research

- **Integrated Automation for Species ID and Biomass Analysis:**
  - End-to-End Efficiency
  - High-Throughput Analysis
  - Scalable Ecological Monitoring
  - Reduced Manual Labor



Future Directions and Impact

- **Scaling Up Classification and Biomass Estimation:**
  - Enhancing model capabilities and training with diverse datasets (across species, environments).
  - Developing robust and generalizable models applicable to wider insect diversity.

Thank you for your attention!

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