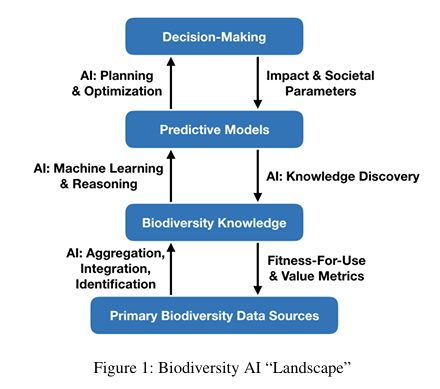
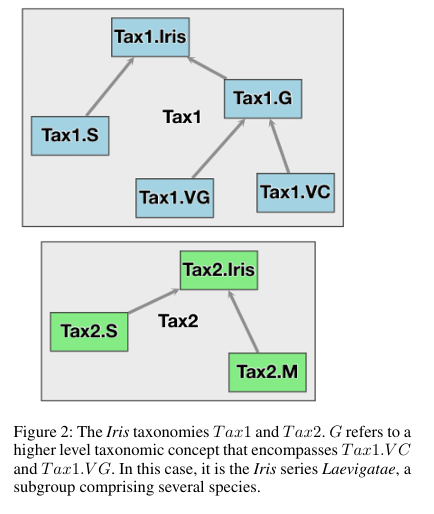
Revolutionizing Biodiversity Research with AI: Machine Learning Meets Automated Reasoning

**Introduction**

Managing biodiversity data is a complex challenge due to its massive volume, multi-modal nature, and conflicting taxonomies. Traditional methods rely on static taxonomies, which fail to account for evolving classifications or inconsistencies.

In their paper “***Combining Machine Learning & Reasoning for Biodiversity Data Intelligence***” [Sen et al., 2021], researchers propose an innovative AI-driven framework. This framework combines **Machine Learning (ML)** and **automated reasoning** to align conflicting taxonomies and integrate biodiversity data, enabling scalable and dynamic solutions for ecological research.





**The Challenges in Biodiversity Data Integration**

**1. Conflicting Taxonomies**

Taxonomic conflicts arise when species classifications differ between regions or evolve over time.

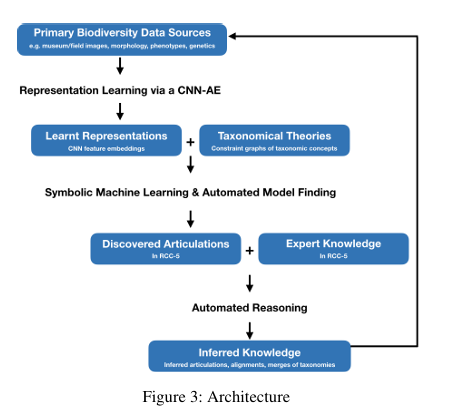
Example (Figure 2):

In the genus *Iris,* species like **Iris versicolor (VC)** and **Iris virginica (VG)** in Taxonomy 1 (Tax1) might merge into a single species **M** in Taxonomy 2 (Tax2). These conflicts prevent effective integration of biodiversity data.

**2. Data Complexity and Volume**

Biodiversity data includes images, text, genetic sequences, and geospatial data, often with missing or incomplete records. As shown in **Figure 1**, the complexity of aligning such fragmented data sources poses significant challenges.

**Key Problem**: Traditional tools assume taxonomies are static and fail to address real-world inconsistencies.



**Proposed Solution: Taxonomic Intelligence**

The authors present a two-part framework:

**1. Machine Learning for Species Identification (Figure 3)**

The ML component uses a **Convolutional Neural Network Autoencoder (CNN-AE)** to classify species from images.

**Key Features:**

**- Architecture**: 8 convolutional layers for encoding and decoding, with a dense embedding layer (10 features).

**- Optimizer**: Adam

**- Loss Function**: Mean Squared Error

**- Training**: 10 epochs on an NVIDIA K80 GPU

The CNN-AE extracts meaningful features to classify species, addressing intra-species variation and inter-species similarity.

**Formula: Mapping Features to Morphological Traits**

The relationship between learned features () and morphological traits () is defined as:

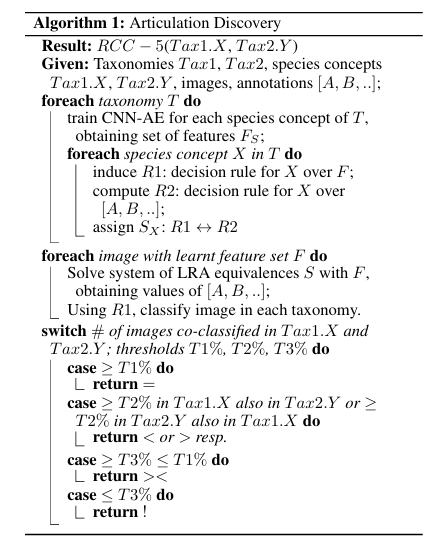
This formula establishes how the model links visual features to species-defining characteristics.

**2. Automated Reasoning for Taxonomy Alignment**

The reasoning component uses the **Microsoft Z3 Solver** to align taxonomies through logical relationships defined by the **RCC-5 calculus**.

**Algorithm 1: Articulation Discovery**

The following steps align taxonomies:



**RCC-5 Relationships Explained** (Figure 3):

**- Equality (=):** Taxa match completely.

**- Proper inclusion (<):** One taxon is fully contained within another.

**- Overlap (><):** Taxa partially overlap.

**- Disjointness (!):** Taxa have no overlap.

**Example:**

For *Iris* species:

- Tax1.VC < Tax2.M (Proper inclusion)

- Tax1.S = Tax2.S (Equality)

**Case Study: Herbarium Plant Dataset**

**Dataset**

The framework was tested on the **Herbarium Challenge Dataset**, containing:

- **46,469 images** from 683 species across 63 genera of the Melastomataceae family.

**Challenges Addressed**

1. **High Intra-Species Variation**: Differences within the same species caused by environmental or curatorial factors.

2. **Low Inter-Species Variation**: Similar-looking species.

**Results**

- **Classification Accuracy**: 89.8% on known species.

- The reasoning component successfully aligned taxonomies dynamically, showing how taxonomic relationships like inclusion and overlap were resolved.

**Applications of Taxonomic Intelligence**

**1. Wildlife Trade Monitoring**

Illegal wildlife trade often uses vernacular species names. The system aligns these names with scientific taxonomies, aiding enforcement of treaties like CITES.

**2. Climate Change Research**

By aggregating biodiversity data, the framework helps monitor habitat changes and extinction risks under climate change scenarios.

**3. Conservation Policy**

The framework supports consistent data integration for treaties like the Convention on Biological Diversity (CBD).

**Limitations and Future Directions**

While promising, the framework has some limitations:

**1. Uncertainty Handling**

The framework relies on logical assumptions that may not always hold universally. Improving uncertainty management is a key area for future work.

**2. Scalability**

Incorporating additional data types, such as genetic sequences and geospatial data, will enhance the framework's applicability to a broader range of biodiversity studies.

**3. Real-Time Updates**

Taxonomies evolve, and the system needs mechanisms to dynamically incorporate new findings and updates.

**Conclusion**

The proposed AI framework offers a dynamic solution to one of the most significant challenges in biodiversity research. By combining **machine learning** and **automated reasoning**, it achieves:

- **Dynamic taxonomy alignment** for real-world biodiversity data.

- Applications in **wildlife monitoring**, **climate research**, and **conservation policy**.

The system is a step forward in leveraging technology for ecological preservation, enabling researchers and policymakers to make data-driven decisions.

What’s your perspective on AI’s role in biodiversity research? Let’s discuss in the comments!

**References**

1. Atriya Sen et al. *“Combining Machine Learning & Reasoning for Biodiversity Data Intelligence.”* AAAI Conference on Artificial Intelligence, 2021.

2. Dataset Reference: Herbarium Challenge Dataset, 2019.