MANTIS: Multi-granularity Adaptive Network for Taxonomy-Informed Species Classification

Junyoung Park
Dept. of Life Science
Incheon National University
Incheon, South Korea
jy2bdoc@inu.ac.kr

Hyung Wook Kwon
Dept. of Life Science
Incheon National University
Incheon, South Korea
hwkwon@inu.ac.kr

*Sungtek Kahng
Dept. of Information & Telecommunication Engineering
Incheon National University
Incheon, South Korea
s-kahng@inu.ac.kr

***Abstract*—Fine-grained species classification is critical for pest monitoring and biodiversity assessment, yet existing classifiers treat all misclassifications equally regardless of taxonomic distance. When a classifier incorrectly predicts a species, it provides no useful information—even though knowing the correct genus or family could still enable appropriate action.**

We propose MANTIS (Multi-granularity Adaptive Network for Taxonomy-Informed Species classification), a hierarchical classification framework that enables *graceful degradation*: when species-level confidence is low, the model automatically falls back to higher taxonomic levels (genus, family) where predictions are more reliable.

Our approach uses a Graph Neural Network to model taxonomic relationships, producing multi-granularity outputs from a single forward pass. Experiments on a pest image dataset (301 species) demonstrate that MANTIS achieves 93.5% species accuracy (+0.67%) and 95.7% genus accuracy. More importantly, for truly unknown inputs, MANTIS shows significantly better graceful degradation at higher taxonomic levels: +2.4% for family-level fallback and +4.1% for order-level fallback. This demonstrates that hierarchical GNN learning produces more discriminative embeddings at coarser taxonomic levels. MANTIS is particularly suitable for real-world open-world deployment where reliability at *some* taxonomic level is preferable to confident but wrong species predictions.

***Index Terms*—Hierarchical Classification, Graph Neural Networks, Open-Set Recognition, Insect Classification, Taxonomy**

I. Introduction

Automated monitoring of disease vectors such as mosquitoes is critical for public health surveillance and outbreak prevention. Deploying classification systems in resource-limited settings—remote areas or developing regions—requires models that provide reliable predictions even when encountering unfamiliar species. However, the high visual similarity between related species makes fine-grained classification challenging, and state-of-the-art classifiers inevitably make mistakes.

The critical issue is *how* these mistakes occur. Consider a classifier that misidentifies *Aedes albopictus* (Asian tiger mosquito) as *Culex pipiens* (common house mosquito). These species belong to different genera with distinct disease transmission profiles and control strategies. Such a "distant mistake" renders the prediction useless for pest management.

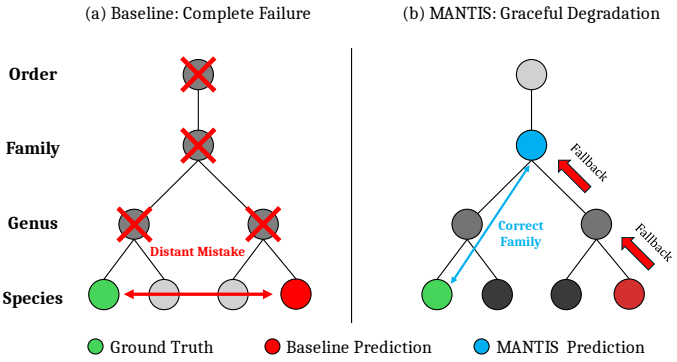

Fig. 1. Graceful degradation via hierarchical fallback. When species-level confidence is low, MANTIS falls back to higher taxonomic levels (genus, family, order) where predictions are more reliable. GraphSAGE-based embeddings capture taxonomic relationships, enabling meaningful fallback predictions.

Existing approaches to hierarchical classification [1], [2] focus on making "better mistakes"—ensuring that when wrong, predictions fall within the same genus or family. While valuable, these methods still output species-level predictions, which may be confidently wrong.

We propose a fundamentally different approach: **graceful degradation**. Instead of always predicting at the species level, our model can *choose* to output predictions at coarser taxonomic levels when uncertain. If species confidence is low but genus confidence is high, the model outputs a genus-level prediction. This mirrors how human experts operate—when uncertain about exact species, they report the genus or family instead of guessing (Fig. 1).

Our contributions are:

1) A hierarchical classification framework (MANTIS) that produces multi-granularity outputs (species, genus, family, order) from a single model.
2) A confidence-based fallback mechanism that automatically selects the appropriate taxonomic level based on prediction certainty.
3) Experimental validation on 301 species showing

This work was supported by the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (202002010006). This work is also supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-RS-2025-00259061121821038201O1) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

+2.4%/+4.1% improvement in open-set fallback accuracy at family/order levels for unknown genera/families.

II. Related Works

A. Hierarchical Classification

Traditional approaches embed class hierarchies into the loss function. Hierarchical Cross-Entropy (HXE) [1] weights mistakes by taxonomic distance, encouraging predictions within the correct genus even when species is wrong. Conditional Risk Minimization (CRM) [2] performs inference-time reranking using hierarchy information without retraining. Both methods improve mistake severity but still commit to species-level predictions. Our approach differs by explicitly *changing the prediction level* itself—falling back to genus or family when confidence is low.

B. Open-Set Recognition

Open-set recognition (OSR) aims to reject unknown classes during inference [3]. Approaches include threshold-based rejection [4], prototype learning [5], energy-based methods [6], and virtual outlier synthesis [18]. Open-world detection [17] extends this to object detection with contrastive clustering. While OSR focuses on binary known/unknown decisions, MANTIS provides *partial* answers by falling back to higher taxonomic levels.

C. Selective Prediction

The idea of abstaining when uncertain has been explored in reject option classifiers [7]. We extend this concept to hierarchical settings—rather than binary accept/reject, we fall back through taxonomy levels, providing partial but reliable information.

D. GNNs for Classification

Graph Neural Networks [14] have been applied to zero-shot learning by modeling class relationships [8], [9]. Prototypical networks [15] use distance-based classification with learned embeddings. We extend these ideas to hierarchical classification, using GraphSAGE [11] to propagate taxonomy-aware information.

E. Contrastive Learning

Supervised contrastive learning [16] improves upon self-supervised methods [13] by incorporating label information. Metric learning approaches [19] learn embeddings where similar samples are close. We adopt hierarchy-aware margins in our contrastive loss, encoding taxonomic distances into the embedding space.

III. Method

A. Problem Formulation

For an input image $\mathbf{x}$, we seek to produce predictions across labels at multiple taxonomy levels $\ell \in \mathcal{L}$: species y_{species}, genus y_{genus}, family y_{family}, and order y_{order}. These labels follow a strict hierarchy: each species belongs to exactly one genus, each genus to one family, and each family to one order.

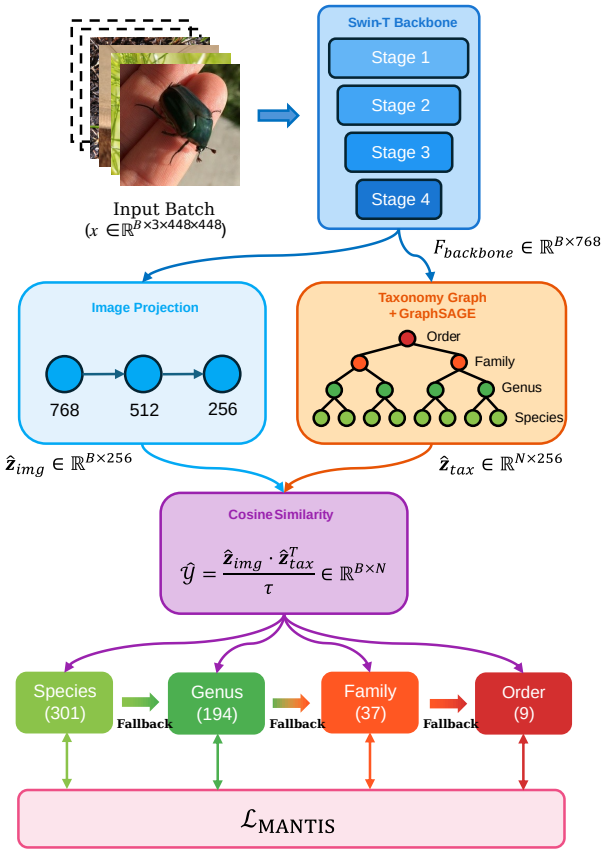

Fig. 2. MANTIS architecture. An input image is processed by Swin-T backbone ($\mathbf{F}_{\text{backbone}} \in \mathbb{R}^{768}$) and projected to 256-d embedding space ($\hat{\mathbf{z}}_{\text{img}}$). The taxonomy graph with $N = 541$ nodes produces embeddings via GraphSAGE ($\hat{\mathbf{Z}}_{\text{tax}}$). Classification is performed via cosine similarity at each taxonomy level, enabling graceful degradation from species to order.

Crucially, during inference, we may encounter samples from *unknown* classes at any level:

- **Unknown-Species (U1)**: Species not in training set, but genus is known.
- **Unknown-Genus (U2)**: Genus not in training set, but family is known.
- **Unknown-Family (U3)**: Family not in training set, but order is known.

For such cases, we aim to provide the most specific correct prediction possible.

B. Overview

MANTIS consists of three components: (1) a visual backbone that extracts image features, (2) a taxonomy-aware GNN that models hierarchical relationships, and (3) a multi-level classification head with confidence-based fallback. Fig. 2 illustrates the overall architecture.

C. Architecture

Let $\mathcal{L} = \{\text{species}, \text{genus}, \text{family}, \text{order}\}$ denote the set of taxonomy levels. Given an input image $\mathbf{x} \in \mathbb{R}^{3\times448\times448}$, we extract features using a Swin Transformer [10] backbone. We choose Swin-T over vanilla ViT because its shifted window

mechanism provides effective local feature aggregation, which is crucial for fine-grained insect classification where subtle morphological details (e.g., wing venation, antenna shape) distinguish closely related species.

$$\mathcal{F}_{\text{backbone}} = \text{SwinT}(\mathbf{x}) \in \mathbb{R}^{768} \tag{1}$$

1) Image Projection: The backbone features are projected to a 256-dimensional embedding space via a two-layer MLP:

$$\hat{\mathbf{z}}_{\text{img}} = \text{MLP}(\mathcal{F}_{\text{backbone}}) \in \mathbb{R}^{256} \tag{2}$$

where MLP : $768 \to 512 \to 256$ with LayerNorm, GELU activation, and dropout.

2) Taxonomy Graph: We encode the taxonomy as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $N = |\mathcal{V}| = 541$ nodes (301 species + 194 genera + 37 families + 9 orders), with edges $\mathcal{E}$ connecting each taxon to its parent. All nodes are initialized with learnable embeddings and updated through GraphSAGE [11]:

$$\mathbf{h}_v^{(k+1)} = \sigma\left(\mathbf{W}^{(k)} \cdot \text{CONCAT}\left(\mathbf{h}_v^{(k)}, \text{AGG}(\{\mathbf{h}_u^{(k)}\}_{u \in \mathcal{N}(v)})\right)\right) \tag{3}$$

After $K = 2$ layers of message passing, we obtain taxonomy embeddings $\hat{\mathbf{Z}}_{\text{tax}} \in \mathbb{R}^{N \times 256}$.

3) Cosine Similarity Classification: Classification logits at each taxonomy level $\ell \in \mathcal{L}$ are computed via scaled cosine similarity:

$$\hat{\mathbf{y}}_\ell = \frac{\hat{\mathbf{z}}_{\text{img}} \cdot \hat{\mathbf{Z}}_\ell^\top}{\tau}, \quad p_\ell = \text{softmax}(\hat{\mathbf{y}}_\ell) \tag{4}$$

where $\hat{\mathbf{Z}}_\ell \subset \hat{\mathbf{Z}}_{\text{tax}}$ are the embeddings for level ℓ, and $\tau = 0.1$ is the temperature parameter following contrastive learning practices [13].

4) Loss Function: We train with the MANTIS loss, combining cross-entropy supervision at each taxonomy level with auxiliary regularization terms:

$$\mathcal{L}_{\text{MANTIS}} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{aux}} \tag{5}$$

where the cross-entropy term aggregates losses across all taxonomy levels:

$$\mathcal{L}_{\text{CE}} = \sum_{\ell \in \mathcal{L}} \lambda_\ell \cdot \text{CE}(\hat{\mathbf{y}}_\ell, y_\ell) \tag{6}$$

with weights $\lambda_\ell = \alpha^{D - d_\ell}$ where D is the maximum depth, d_ℓ is the depth of level ℓ, and $\alpha = 0.5$. This exponential decay assigns higher importance to finer-grained levels while still supervising coarser levels.

The auxiliary loss $\mathcal{L}_{\text{aux}}$ includes three terms:

$$\mathcal{L}_{\text{aux}} = \lambda_1 \mathcal{L}_{\text{hier}} + \lambda_2 \mathcal{L}_{\text{cont}} + \lambda_3 \mathcal{L}_{\text{reg}} \tag{7}$$

5) Hierarchical consistency loss: $\mathcal{L}_{\text{hier}}$ supervises coarser levels using aggregated logits:

$$\mathcal{L}_{\text{hier}} = \text{CE}(\hat{\mathbf{y}}_{\text{genus}}, y_{\text{genus}}) + 0.5 \cdot \text{CE}(\hat{\mathbf{y}}_{\text{family}}, y_{\text{family}}) \tag{8}$$

6) Contrastive loss: $\mathcal{L}_{\text{cont}}$ enforces margin-based separation with hierarchy-aware margins:

$$\mathcal{L}_{\text{cont}} = \sum_{i,j} \max(0, \text{sim}(\hat{\mathbf{z}}_i, \hat{\mathbf{z}}_j) - (1 - m_{ij})) \tag{9}$$

where $m_{ij} \in \{0.2, 0.5, 1.0\}$ for same-genus, same-family, and different-family pairs respectively.

7) Regularization loss: $\mathcal{L}_{\text{reg}}$ encourages intra-genus cohesion in label embeddings:

$$\mathcal{L}_{\text{reg}} = 1 - \frac{1}{|G|} \sum_{i,j \in G} \text{sim}(\hat{\mathbf{z}}_i^{\text{tax}}, \hat{\mathbf{z}}_j^{\text{tax}}) \tag{10}$$

where G denotes species pairs within the same genus.

D. Graceful Degradation at Inference

The key innovation is the fallback mechanism. Given confidence threshold θ, we select the finest taxonomic level with sufficient confidence:

$$\text{output} = \begin{cases} \text{species} & \text{if } \max(p_{\text{species}}) \geq \theta \\ \text{genus} & \text{elif } \max(p_{\text{genus}}) \geq \theta \\ \text{family} & \text{elif } \max(p_{\text{family}}) \geq \theta \\ \text{order} & \text{otherwise} \end{cases} \tag{11}$$

This enables graceful degradation: uncertain species predictions are replaced with more reliable genus/family predictions, providing actionable information rather than confident mistakes.

IV. Experiments

A. Dataset and Setup

We evaluate on **iNat-Pests**, a curated subset of iNaturalist [12] focusing on pest-relevant insect species (Fig. 3). The dataset contains approximately 91,000 images spanning a 4-level taxonomy: 9 Orders $\to$ 37 Families $\to$ 194 Genera $\to$ 301 Species (known classes). The selected orders include agricultural pests (Lepidoptera, Coleoptera, Orthoptera), disease vectors (Diptera, Hemiptera), and structural pests (Blattodea, Hymenoptera). For open-set evaluation, we additionally hold out 9 species (U1), 9 genera (U2), and 3 families (U3) from training (Fig. 4).

We use Swin-T pretrained on ImageNet [21] as the visual backbone. Images are resized to 448×448. The GNN has 2 layers with hidden dimension 256. We train for 50 epochs using AdamW [24] optimizer with learning rate 5×10^{-5} and cosine decay.

B. Main Results

Table I compares MANTIS with the baseline (standard Swin-T with species-only classification). For fair comparison, we derive baseline's genus/family accuracy by looking up the predicted species' taxonomy.

MANTIS outperforms the baseline at all taxonomic levels. At 448×448 resolution, it achieves 93.49% species accuracy (+0.67%) and 95.68% genus accuracy.

Fig. 3. Representative images from each taxonomic order in the iNat-Pests dataset.

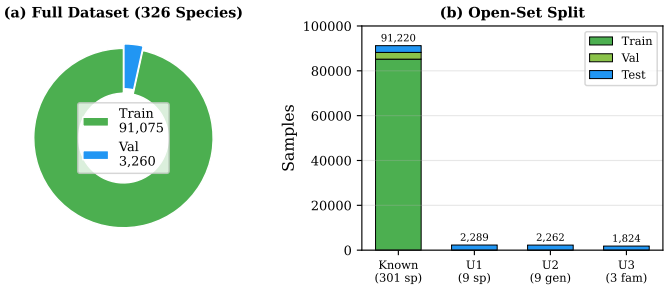

Fig. 4. Open-set split: 301 known species for training with three unknown test sets (U1: unknown species, U2: unknown genera, U3: unknown families).

TABLE I
CLASSIFICATION ACCURACY ON 301 KNOWN SPECIES

Method	Species	Genus	Family	Order
Baseline	92.82%	95.15%	97.31%	98.14%
MANTIS	**93.49%**	**95.68%**	**97.57%**	**98.37%**
Δ	+0.67%	+0.53%	+0.26%	+0.23%

C. Error Quality Analysis

Beyond raw accuracy, we analyze the *quality* of errors. Table II shows error distribution by taxonomic distance:

TABLE II
ERROR QUALITY ANALYSIS: TAXONOMIC DISTANCE OF MISCLASSIFICATIONS.

Error Type	Baseline	MANTIS
Same Genus (good)	38.89%	**41.94%**
Same Family	23.93%	25.81%
Same Order	11.54%	8.76%
Different Order (bad)	25.64%	**23.50%**

TABLE III
OPEN-SET GRACEFUL DEGRADATION: ACCURACY AT HIGHER TAXONOMIC LEVELS FOR TRULY UNKNOWN INPUTS (MODELS TRAINED ON 301 SPECIES).

Fallback Task	Baseline	MANTIS	Δ
U1: Unknown Sp. → Genus	54.65%	54.39%	-0.26%
U2: Unknown Gen. → Family	66.18%	**68.61%**	**+2.43%**
U3: Unknown Fam. → Order	31.20%	**35.31%**	**+4.11%**

TABLE IV
CALIBRATION COMPARISON ON SPECIES-LEVEL PREDICTIONS.

Metric	Baseline	MANTIS
ECE (↓ better)	0.0440	**0.0342**
Mean Confidence	0.9715	0.9318
Conf. on Unknowns	0.9961	**0.9367**

MANTIS makes taxonomically closer errors: 41.94% of mistakes fall within the same genus (vs. 38.89% for baseline). This means when MANTIS is wrong, the prediction is still more useful—belonging to a closely related species rather than a distant one.

D. Open-Set Graceful Degradation

The key result is how models handle *unknown* inputs. Table III shows accuracy when predicting at higher taxonomic levels for unknown classes:

MANTIS shows significantly better graceful degradation for higher-level fallbacks (Fig. 5). When encountering unknown genera (U2), MANTIS achieves 68.61% family-level accuracy (+2.43%). For unknown families (U3), the improvement is even larger: 35.31% order-level accuracy (+4.11%). This demonstrates that hierarchical GNN learning produces more discriminative embeddings at higher taxonomic levels, enabling more reliable fallback predictions.

Interestingly, for unknown species within known genera (U1), both models perform similarly (∼54%). This is expected since both models have learned genus-level representations from training data containing other species of the same genus. However, for higher-level fallbacks (U2, U3), where models must generalize to entirely unseen genera or families, the hierarchical GNN structure provides substantial benefit.

E. Calibration and Confidence Analysis

Table IV compares calibration quality [23] between models:

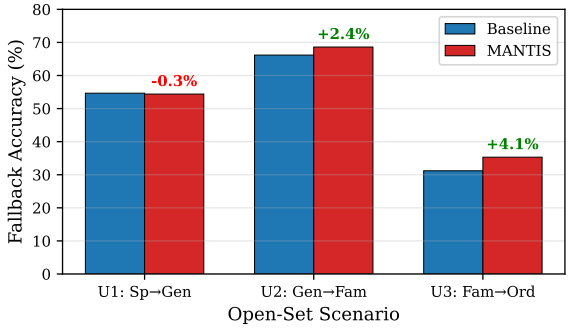

Fig. 5. Open-set graceful degradation: MANTIS shows increasing advantage at higher taxonomic levels (+2.4% for U2, +4.1% for U3), demonstrating better hierarchical representations.

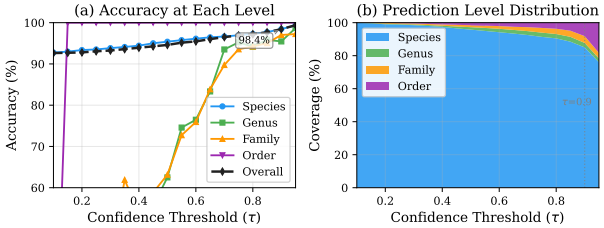

Fig. 6. Fallback curve: accuracy vs. confidence threshold. Higher thresholds yield higher accuracy on confident predictions while triggering more fallbacks to coarser taxonomic levels.

MANTIS shows better calibration (ECE: 0.034 vs. 0.044) and, critically, **lower confidence on unknown samples** (93.7% vs. 99.6%). This realistic uncertainty enables threshold-based filtering: at $\tau = 0.9$, MANTIS flags 4.81% of samples as uncertain (vs. only 0.74% for baseline), with both models achieving high accuracy on their respective confident predictions.

The baseline's overconfidence (99.6% on unknowns) makes it unsuitable for threshold-based deployment—unknown samples pass through as "confident" predictions.

Fig. 6 shows the trade-off between prediction confidence threshold and accuracy. As the threshold increases, fewer samples are predicted at the species level, but accuracy on those predictions improves. At $\theta = 0.9$, both models achieve near-perfect accuracy on their confident predictions, but MANTIS identifies 6.5× more uncertain samples for fallback (4.81% vs. 0.74%).

F. OOD Detection with Energy Score

We further enhance MANTIS with Virtual Outlier Synthesis (VOS) [18] fine-tuning to improve out-of-distribution detection. Table V compares OOD detection performance using energy score [6] (AUROC, higher is better):

VOS fine-tuning substantially improves energy-based OOD detection, particularly for U1 (+4.19%) and U2 (+4.26%) unknown classes. This enables reliable rejection of unknown samples before fallback, complementing the graceful degradation mechanism.

TABLE V
OOD DETECTION PERFORMANCE (AUROC) USING ENERGY SCORE.

Method	U1	U2	U3
Baseline	83.58%	84.55%	94.04%
MANTIS + VOS	**87.77%**	**88.81%**	**94.19%**
Δ	+4.19%	+4.26%	+0.15%

TABLE VI
DIRECT FALLBACK VS. TAXONOMY LOOKUP ($\tau = 0.7$)

	Baseline		MANTIS + VOS	
Task	Direct	Lookup	Direct	Lookup
U1→Genus	25.94%	30.69%	**36.16%**	35.21%
U2→Family	27.24%	40.00%	**46.90%**	45.19%
U3→Order	34.02%	26.00%	**35.26%**	26.94%

TABLE VII
ABLATION STUDY ON KEY COMPONENTS.

Metric	Base	+GNN	+Loss	Full
Known Classes				
Species	92.82	93.19	93.36	**93.49**
Genus	95.15	95.41	95.55	**95.68**
Family	97.31	97.44	97.51	**97.57**
Order	98.14	98.24	98.30	**98.37**
Open-set Fallback				
U1 → Genus	54.65	54.48	54.52	54.39
U2 → Family	66.18	67.52	68.14	**68.61**
U3 → Order	31.20	33.55	34.43	**35.31**

G. Direct Hierarchical Fallback

Beyond taxonomy lookup from predicted species, we evaluate *direct* fallback: when confidence $< \tau$, predict at the parent level using its dedicated logits. Table VI shows results at $\tau = 0.7$:

Critically, for baseline, direct fallback *degrades* performance vs. lookup (e.g., U2: 27.2% vs. 40.0%), meaning its genus/family embeddings are not discriminative enough for direct prediction. In contrast, MANTIS consistently *improves* with direct fallback, demonstrating that hierarchical GNN learning produces meaningful embeddings at every taxonomic level. The MANTIS + VOS model achieves 46.90% family accuracy on U2 unknowns via direct fallback, outperforming baseline's lookup by +6.9%.

H. Ablation Study

Table VII presents the incremental impact of each architectural component:

Each component contributes incrementally: GNN embeddings improve taxonomic consistency, multi-level loss enables explicit supervision at each level, and $\mathcal{L}_{\text{aux}}$ further encourages hierarchically coherent representations.

Fig. 7 shows the training dynamics. MANTIS converges smoothly and maintains consistent improvement across all taxonomy levels throughout training.

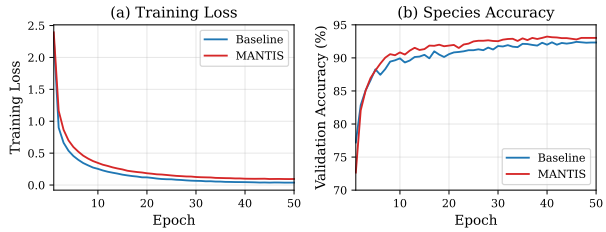

Fig. 7. Learning curves during training. MANTIS shows stable convergence with consistent accuracy improvements at species, genus, and family levels.

TABLE VIII
COMPUTATIONAL OVERHEAD COMPARISON.

Model	Parameters	Latency (ms)
Baseline (Swin-T)	27.75M	5.78 ± 0.11
MANTIS	28.56M	5.91 ± 0.08
Overhead	+2.9%	+2.2%

I. Overhead Analysis

We benchmark inference latency on an NVIDIA RTX 3090 GPU with Intel i9-13900K CPU, using 448×448 input images and batch size 1. Results are averaged over 100 runs (Table VIII). The GNN component adds minimal overhead: +2.9% parameters (27.75M → 28.56M) and +2.2% latency (5.78ms → 5.91ms), making it practical for real-time deployment.

V. CONCLUSION

We presented MANTIS, a hierarchical classification framework with graceful degradation for species identification. Unlike existing methods that always predict at species level, MANTIS can fall back to genus or family when uncertain, providing reliable predictions at *some* taxonomic level.

Key results show that: (1) MANTIS achieves 93.5% species accuracy with taxonomically closer errors (41.9% intra-genus vs. 38.9%); (2) for unknown genera/families, MANTIS shows +2.4%/+4.1% improvement at family/order-level fallback; (3) better calibration (ECE: 0.034 vs. 0.044) enables MANTIS to identify 6.5× more uncertain samples for fallback.

Current threshold selection requires validation data with known/unknown class distribution. The hierarchical fallback assumes a well-defined taxonomy, which may not exist for all domains. Real-world deployment faces additional challenges including inter-genera domain shift from field-captured specimens with varying conditions [26]. Future work will explore adaptive threshold selection, alternative backbone architectures (e.g., ConvNeXt, larger Swin variants), visual attention analysis (e.g., Grad-CAM) to verify that learned features align with expert-identified morphological keys [26], extension to other hierarchical domains beyond biological taxonomy, and integration with active learning for continuous model improvement.

REFERENCES

[1] L. Bertinetto, R. Mueller, K. Tertikas, S. Samangooei, and N. A. Lord, "Making better mistakes: Leveraging class hierarchies with deep networks," in *Proc. IEEE/CVF CVPR*, 2020, pp. 12506–12515.
[2] S. Karthik, J. Mena, and Z. Akata, "No cost likelihood manipulation at test time for making better mistakes in deep networks," in *Proc. ICLR*, 2021.
[3] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1757–1772, 2012.
[4] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proc. IEEE/CVF CVPR*, 2016, pp. 1563–1572.
[5] G. Chen, P. Peng, X. Wang, and Y. Tian, "Adversarial reciprocal points learning for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8065–8081, 2021.
[6] W. Liu, X. Wang, J. Owens, and Y. Li, "Energy-based out-of-distribution detection," in *Proc. NeurIPS*, 2020, pp. 21464–21475.
[7] C. Chow, "On optimum recognition error and reject tradeoff," *IEEE Trans. Inf. Theory*, vol. 16, no. 1, pp. 41–46, 1970.
[8] X. Wang, Y. Ye, and A. Gupta, "Zero-shot recognition via semantic embeddings and knowledge graphs," in *Proc. IEEE/CVF CVPR*, 2018, pp. 6857–6866.
[9] M. Kampffmeyer, Y. Chen, X. Liang, H. Wang, Y. Zhang, and E. P. Xing, "Rethinking knowledge graph propagation for zero-shot learning," in *Proc. IEEE/CVF CVPR*, 2019, pp. 11487–11496.
[10] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF ICCV*, 2021, pp. 10012–10022.
[11] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. NeurIPS*, 2017, pp. 1024–1034.
[12] G. Van Horn, O. Mac Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, and S. Belongie, "The iNaturalist species classification and detection dataset," in *Proc. IEEE/CVF CVPR*, 2018, pp. 8769–8778.
[13] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. ICML*, 2020, pp. 1597–1607.
[14] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. ICLR*, 2017.
[15] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. NeurIPS*, 2017, pp. 4077–4087.
[16] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Proc. NeurIPS*, 2020, pp. 18661–18673.
[17] K. J. Joseph, S. Khan, F. S. Khan, and V. N. Balasubramanian, "Towards open world object detection," in *Proc. IEEE/CVF CVPR*, 2021, pp. 5830–5840.
[18] X. Du, Z. Wang, M. Cai, and Y. Li, "VOS: Learning what you don't know by virtual outlier synthesis," in *Proc. ICLR*, 2022.
[19] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE/CVF CVPR*, 2015, pp. 815–823.
[20] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *Proc. ICLR*, 2017.
[21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE/CVF CVPR*, 2009, pp. 248–255.
[22] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," in *Proc. ICML*, 2021, pp. 10347–10357.
[23] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. ICML*, 2017, pp. 1321–1330.
[24] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. ICLR*, 2019.
[25] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.
[26] J. Park, D. I. Kim, B. Choi, W. Kang, and H. W. Kwon, "Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020.