CRNN, FROZEN AST, AND DOMAIN-STYLE FUSION FOR CROSS-DOMAIN MOSQUITO SPECIES CLASSIFICATION

Anant Gupta, Gourav Gupta, and Aditya Gupta

Arogya Pandit, India
anant.gupta@arogyapandit.com

ABSTRACT

This report describes four systems submitted to BioDCASE 2026 Task 5, Cross-Domain Mosquito Species Classification. The common base is a convolutional recurrent neural network trained on 8 kHz audio represented as 64-bin log-mel spectrograms. It uses convolutional feature extraction, a bidirectional GRU, attention pooling, and a gradient-reversal domain head. The submitted variants include a frozen public Audio Spectrogram Transformer auxiliary classifier, a fixed species-9 calibration hedge, and a domain-style CRNN trained to preserve species evidence under synthetic recording-style changes. The final system fuses the domain-style and mild supervised-contrastive CRNN variants. No external labelled mosquito data was used, and evaluation audio was used only for inference.

Index Terms-- mosquito species classification, domain generalization, CRNN, gradient reversal, domain-style transfer, Audio Spectrogram Transformer

1. INTRODUCTION

BioDCASE 2026 Task 5 evaluates closed-set mosquito species recognition under domain shift. The official ranking emphasizes balanced accuracy on unseen-domain clips, BA_unseen, because a system that performs well on familiar recording conditions may still fail when microphone, location, or acoustic background changes. This work therefore treats BA_unseen as the primary development metric and uses seen-domain performance and DSG only as supporting diagnostics.

2. DATASET AND EVALUATION

The released development set contains 271,380 labelled clips covering nine mosquito species and five acquisition domains. The official development split contains 213,647 training IDs, 30,516 validation IDs, and 27,217 development-test IDs. Domain imbalance is severe, with D5 contributing the majority of clips; this makes naive optimization prone to learning recording-domain shortcuts rather than species cues.

The final evaluation set contains 15,573 randomized audio files with hidden labels. Participants submit only file identifiers and predicted species IDs. Evaluation audio was used only for final inference in all systems described here. No official final score or MSE value is claimed in this report.

3. BASE CRNN SYSTEM

The base model converts raw audio to 8 kHz, peak-normalizes each clip, and extracts a 64-bin log-mel spectrogram using a 512-point FFT and 10 ms hop. A three-block convolutional frontend extracts local harmonic patterns. The resulting time sequence is projected to 256 dimensions, processed by a bidirectional GRU with hidden size 128, and summarized by learned attention pooling.

The pooled representation feeds a species classifier for nine species. During training, the same representation also feeds a five-domain classifier through a gradient-reversal layer. The domain loss weight was 0.25 and the maximum GRL strength was 0.35. Utility weighting, danger weighting, and global KDE filtering were disabled because the leaderboard metric is unweighted species-balanced accuracy.

4. SUBMITTED SYSTEMS

Submitted systems

Label	System summary
Gupta_AP_task5_1	CRNN epoch 6 + shift TTA
Gupta_AP_task5_2	0.85 CRNN + 0.15 frozen AST auxiliary fusion
Gupta_AP_task5_3	Slot 2 + fixed species-9 logit boost +0.10
Gupta_AP_task5_4	0.70 domain-style e3 + 0.30 mild-SupCon e3

Gupta_AP_task5_1 is the recovered CRNN checkpoint from epoch 6 evaluated with deterministic dropout-off inference and temporal shift averaging over shifts -8, 0, and 8. Gupta_AP_task5_2 adds a frozen public Audio Spectrogram Transformer representation. A shallow auxiliary classifier is trained on AST embeddings, and final probabilities are fused as 0.85 CRNN plus 0.15 AST.

Gupta_AP_task5_3 uses the Slot 2 probabilities and applies a fixed species-9 logit boost of +0.10 selected on the development split. Gupta_AP_task5_4 uses domain-style transfer during training: frequency-wise statistics from a different recording domain are imposed on a content clip while retaining its species label. It fuses epoch-3 probability outputs from the domain-style CRNN (0.70) and a mild supervised-contrastive CRNN (0.30).

5. DEVELOPMENT RESULTS

Development results

Slot	Unseen	Seen	DSG
1	0.274307	0.645648	0.371341
2	0.281629	0.692455	0.410826
3	0.288093	0.692318	0.404225
4	0.319175	0.485682	0.166506

All results use the released official development split. BA_unseen is the primary model-selection metric; BA_seen and DSG are reported as diagnostics.

The clean CRNN shift-TTA system exceeded the official baseline unseen balanced accuracy on the development setting. Adding the AST auxiliary branch improved BA_unseen from 0.274307 to 0.281629. The fixed species-9 hedge reached 0.288093. The domain-style and mild-SupCon fusion reached 0.319175 on the official development split, with a substantially smaller domain-shift gap. These development values were used only to select fixed inference variants before producing evaluation predictions.

6. DISCUSSION

The main design choice was to prioritize domain-robust species recall rather than raw accuracy. Earlier ablations showed that gentle GRL helped relative to no-GRL, while several generic pretrained and calibration-only variants did not consistently improve BA_unseen. The final domain-style system instead targets the core failure mode: recordings from different devices and environments alter spectral statistics while the species identity remains unchanged.

The AST representation is retained only as a frozen auxiliary signal, so it can provide complementary acoustic evidence without destabilizing the CRNN anchor. The domain-style model preserves the core CRNN architecture but exposes it to synthetic cross-domain statistics. The SupCon term is kept mild to encourage same-species cohesion without collapsing nearby species representations. Evaluation audio never contributed to training, style donors, probability-weight selection, or calibration.

7. REFERENCES

[1] Y. Hou et al., BioDCASE 2026 Challenge Baseline for Cross-Domain Mosquito Species Classification, arXiv:2603.20118, 2026.

[2] Yuanbo2020, CD-MSC: BioDCASE 2026 Challenge Baseline for Cross-Domain Mosquito Species Classification, https://github.com/Yuanbo2020/CD-MSC.

[3] Y. Hou et al., Learning Domain-Robust Bioacoustic Representations for Mosquito Species Classification with Contrastive Learning and Distribution Alignment, ICASSP 2026.

[4] Y. Gong et al., AST: Audio Spectrogram Transformer, Interspeech, 2021.

[5] K. Zhou et al., Domain Generalization with MixStyle, ICLR, 2021.