HYU SUBMISSION TO DCASE 2026 TASK 7: MULTI-BRANCH FUSION AND CONSERVATIVE ROUTING FOR DOMAIN-AGNOSTIC INCREMENTAL AUDIO CLASSIFICATION

Technical Report

Gihun Son[1,*], *Pil Moo Byun*[2,*], *Joon-Hyuk Chang*[1,2,†]

[1] Department of Artificial Intelligence Application, Hanyang University, Seoul, Republic of Korea
[2] Department of Artificial Intelligence, Hanyang University, Seoul, Republic of Korea
{gihunson, fordream0309, jchang}@hanyang.ac.kr

ABSTRACT

This paper presents the Hanyang University team's submission to DCASE 2026 Task 7 on domain-agnostic incremental audio classification. The task requires adapting an audio classifier to sequentially revealed acoustic domains while preserving performance on previously learned domains and performing inference without domain labels. To address this challenge, we explore four systems based on stage-specific branch adaptation and frozen inference-time routing. System 1 uses stage-local acceptors with a fixed D3 override rule, while Systems 2 and 3 perform multi-branch fusion over a retention-oriented balanced path and a more plastic specialist path with different fusion rules. System 4 adopts a conservative low-fire disagreement router to reduce harmful D3 overrides. On the released development report split, System 3 achieves the best D2/D3 average among our submissions, while System 2 provides stronger D2 retention and System 1 achieves the highest D3 accuracy. The submitted systems offer complementary operating points along the retention–plasticity trade-off in domain-agnostic incremental learning.

***Index Terms*—** Domain-agnostic incremental learning, continual learning, audio classification, knowledge distillation

1. INTRODUCTION

DCASE 2026 Task 7 addresses domain-agnostic incremental learning for audio classification, where a system must learn sequentially revealed acoustic domains while retaining performance on previously learned domains [1]. Unlike standard supervised learning, all domains are not available jointly, and later-stage training cannot rely on replaying previous-domain raw samples. At evaluation time, the true domain identity is unavailable, so the system must perform class prediction in a domain-agnostic manner. This setting requires controlling the stability–plasticity trade-off between preserving previous-domain knowledge and adapting to newly observed domains.

The official baseline uses a CNN-based classifier with domain-specific normalization paths [1, 2]. However, selecting the correct path is difficult when domain labels are not given at inference time. To address this issue, our submission explores branch-based incremental adaptation with frozen sample-wise routing. We use stage-specific branches and adapters near the upper part of the network, while previous-stage behavior is preserved using frozen teachers and regularization terms inspired by knowledge distillation (KD) [3], learning without forgetting (LwF) [4], L2 starting point regularization (L2-SP) [5], and elastic weight consolidation (EWC) [6].

We submit four complementary systems that represent different operating points between previous-domain retention and D3 adaptation. System 1 uses stage-local acceptors with frozen D2/D3 acceptor thresholds and a fixed no-training D3 override rule. Systems 2 and 3 perform multi-branch fusion over four frozen branch outputs built from a retention-oriented balanced path and a more plastic specialist path. System 2 applies confidence-gap-entropy soft fusion over the four D2/D3 branch outputs, whereas System 3 uses entropy-weighted fusion within D2 and D3 branch pairs followed by a fixed D2/D3 mixture. In contrast, System 4 avoids broad multi-branch fusion and uses a conservative two-expert low-fire disagreement router that switches from the frozen D2-stage expert to the adapted D3 expert only under strong sample-local evidence. All submitted systems use only the official D1 checkpoint and the provided task data, without external data, external pretrained audio models, pseudo-labels, evaluation-set statistics, adaptive batch normalization, test-time training, or cross-sample decision rules.

2. METHOD

2.1. Common Setup

All submitted systems start from the official D1 checkpoint and sequentially adapt to D2 and D3 using only the data available at each stage, without replaying previous-domain raw audio. The systems are based on an MCnn14-style classifier that converts the input waveform into log-mel features and processes them with a convolutional backbone. Systems 1–3 use stage-specific upper branches on a shared convolutional backbone, whereas System 4 uses a frozen D2-stage expert and a separately adapted D3 expert with a conservative fixed router.

Several systems use simple branch-level reliability statistics. For a branch posterior, confidence denotes the maximum posterior probability, gap denotes the difference between the largest and second-largest posterior probabilities, and entropy measures posterior uncertainty. A confidence-gap-entropy score refers to a hand-designed reliability score that increases with confidence and gap and decreases with entropy. These statistics are used only for branch comparison or routing; they are not separately trained classifiers.

*Equal contribution
†Corresponding author

2.2. System 1: Stage-Local Acceptor

System 1 uses stage-local branch adaptation with acceptor-based routing. Starting from the official D1 checkpoint, we first train a D2 branch using only D2 data. This branch adds task-specific batch normalization, copied upper convolutional blocks, a residual adapter, and a classifier head near the top of the shared backbone. After D2 training, a D2 acceptor is fitted to decide whether the D2 branch should replace the frozen D1 path for each input sample.

The acceptor is a sample-wise binary gate rather than a class classifier. It compares the current branch and the older path using reliability and agreement features, including confidence, gap, entropy, score differences, cross-branch probabilities, KL-divergence terms, prediction agreement, and prototype-similarity features. Positive switching examples correspond to samples where the current branch is correct and the older path is wrong, whereas harmful switching examples correspond to the opposite case. The acceptor threshold is selected on an internal split of the available training data to improve macro accuracy while penalizing harmful switches.

The model is then continued to D3. The D3 branch is initialized from the D2 top modules and trained using only D3 data, while frozen D1 and D2 checkpoints are used as teachers on D3 inputs. The training objective combines supervised classification with retention-preserving regularization based on KD [3], LwF [4], L2-SP [5], and EWC [6]. After D3 training, a D3 acceptor is fitted using D3 samples.

At inference time, System 1 first applies the frozen D2 acceptor to choose between the D1 and D2 paths, and then applies the frozen D3 acceptor to decide whether to switch to the D3 branch. The submitted system also includes a fixed no-training D3 override rule. This rule allows an additional D3 switch only when the D3 accept probability is at least the acceptor threshold minus a fixed probability margin and the D3 branch score exceeds both the selected old path and the D2 path by predefined margins.

2.3. Systems 2 and 3: Multi-Branch Fusion

Systems 2 and 3 perform fusion over four frozen branch outputs. The four branches are built from two sequential training paths: a retention-oriented balanced path and a more plastic specialist path. Both paths are initialized from the official D1 checkpoint, trained on D2 data only, and saved as D2 checkpoints. Each D2 checkpoint is then independently continued using D3 data only, producing balanced-D3 and specialist-D3 checkpoints. The official D1 model is used for initialization and as a frozen teacher, but it is not one of the four final branch outputs.

The balanced path is designed to preserve previous-domain behavior more strongly. It uses stronger retention constraints, including KD, L2-SP, EWC, embedding/prototype-level distillation, center regularization, and branch-rank regularization, and initializes the D3 head from the D2 head. The specialist path is designed to adapt more strongly to the current training domain at each stage. It uses weaker retention constraints and a larger residual adapter, and its D3 head is not initialized from the D2 head. In both paths, the trainable scope is concentrated in the current task-specific batch-normalization parameters, upper convolutional blocks, residual adapter, and classifier head, while the lower convolutional backbone is kept frozen.

At inference time, Systems 2 and 3 use the same four outputs: balanced-D2, specialist-D2, balanced-D3, and specialist-D3. They differ only in their fixed fusion rule. System 2 applies confidence-gap-entropy soft routing over all four outputs: each branch receives a softmax mixing weight computed from its confidence-gap-entropy score with a fixed branch bias, and the final posterior is the weighted sum of the four branch posteriors. System 3 uses a simpler entropy-weighted fusion: it first fuses the balanced and specialist branches within each stage by giving larger weight to the lower-entropy branch, and then combines the D2 and D3 mixtures using fixed stage weights of 0.45 and 0.55, respectively.

2.4. System 4: Low-Fire Disagreement Router

While Systems 2 and 3 perform multi-branch fusion for broader coverage, System 4 takes the opposite design choice. It avoids broad branch fusion and instead uses a conservative two-expert router that switches to the D3 expert only under strong sample-local evidence. Therefore, System 4 is intended as a low-fire, previous-domain-protective variant.

System 4 uses the frozen D2-stage model as the previous-stage expert and a separately adapted D3 model as the D3 expert. Unlike System 1, it does not apply a full D1–D2–D3 acceptor cascade. Instead, it focuses on the final transition from the frozen D2-stage expert to the adapted D3 expert. The D3 expert is initialized from the D2 checkpoint and trained using only D3 internal-fit data. Its trainable scope is restricted to the final convolutional block, task-specific batch-normalization parameters, and classifier head. No D1 or D2 audio samples are used during this D3 adaptation stage.

At inference time, System 4 applies a fixed low-fire disagreement router. The router is called low-fire because it is intentionally reluctant to switch from the previous-stage expert to the D3 expert, and disagreement-based because it mainly evaluates D3 overrides when the two experts predict different classes. If the two experts agree, the agreed prediction is kept. If they disagree, the D3 expert is allowed to override only when several frozen sample-local conditions support the switch.

The override conditions include posterior confidence, entropy, posterior margin, support-space evidence, and a previous-stage expert veto. Support-space evidence is computed from frozen D3 internal-fit embeddings using k-nearest-neighbor agreement, prototype-similarity margin, and prototype gain. The router also compares the uncertainty of the D3 expert and the previous-stage expert. The previous-stage expert veto rejects a switch when the previous-stage expert is already highly confident and the D3 confidence gain is insufficient. All router thresholds, prototype statistics, and k-nearest-neighbor support statistics are fixed before evaluation, and each test sample is classified independently. Therefore, System 4 permits D3 corrections only when both posterior evidence and support-space evidence strongly favor the D3 expert.

3. EXPERIMENTS AND RESULTS

3.1. Experimental Setup

We evaluate the submitted systems on the official development report split of DCASE 2026 Task 7. Since the released development report split contains D2 and D3 but not D1 audio, we report class-wise macro accuracy on D2 and D3 and use their arithmetic mean as the main development summary.

All systems start from the official D1 checkpoint. D2-stage training uses D2 training samples only, and D3-stage training uses D3 training samples only. No external audio data, external pretrained audio model, pseudo-label, evaluation-domain label, adaptive batch normalization, or test-time adaptation is used. All final inference rules are fixed before evaluation and are applied independently to each audio clip.

Table 1: Summary of the four submitted systems.

System	Submission folder	Main idea	Final inference rule
1	`Chang_HYU_task7_1`	Stage-local acceptor with fixed D3 override	Frozen D2/D3 acceptors plus a fixed D3 override rule
2	`Chang_HYU_task7_2`	Multi-branch fusion	Confidence-gap-entropy multi-branch fusion
3	`Chang_HYU_task7_3`	Multi-branch fusion	Entropy-weighted D2/D3 mixture with 0.45/0.55 weights
4	`Chang_HYU_task7_4`	Low-fire memory-protection router	D3 override only under disagreement and strong D3 evidence

Table 2: D2-stage and final D3-stage development results.

System	After D2 learning	After D3 learning		
	D2 macro	D2 macro	D3 macro	Avg.
System 1	75.9254	59.6866	65.5349	62.6108
System 2	74.8503	70.6078	61.2889	65.9484
System 3	74.8503	69.6477	62.7225	66.1851
System 4	69.1303	69.3365	56.4194	62.8780

3.2. Results

Table 1 summarizes the four submitted systems. System 1 uses stage-local acceptors with a fixed D3 override rule. Systems 2 and 3 use the same four branch outputs but apply different frozen multi-branch fusion rules. System 4 uses a conservative low-fire disagreement router.

Table 2 reports the development results after D2 learning and after the final D3-stage inference rule. The D2-stage column shows D2 macro accuracy before D3 adaptation, while the final columns report the packaged system performance on D2 and D3. System 3 obtains the highest final D2/D3 average among the submitted systems, while System 2 gives the highest final D2 score and System 1 gives the highest final D3 score. This indicates that the submitted systems provide different operating points along the retention–plasticity trade-off.

The D2-stage results show that Systems 1–3 obtain higher D2 accuracy before D3 adaptation than after the final D3-stage inference rule is applied. Systems 2 and 3 preserve D2 performance better than System 1 after D3 learning, while System 1 provides the strongest D3 adaptation. System 4 maintains a conservative D2-oriented behavior, but its lower D3 activation leads to the lowest final D3 score among the submitted systems.

4. CONCLUSION

We presented four submitted systems for DCASE 2026 Task 7 based on sequential adaptation from the official D1 checkpoint and frozen inference-time routing. All systems use only the provided task data at each stage and operate without domain labels or test-time adaptation. On the development report split, System 3 obtains the best D2/D3 average by combining D2 and D3 branch mixtures with entropy-weighted fusion, while System 2 provides slightly stronger D2 retention and System 1 provides the highest D3 accuracy. These results show that the submitted systems cover different operating points in the retention–plasticity trade-off under domain-agnostic incremental audio classification.

5. REFERENCES

[1] R. Casciotti, M. Mulimani, M. Harju, J. R. Jensen, and A. Mesaros, "Domain-agnostic incremental learning for sound classification. a dcase 2026 challenge task," *arXiv preprint arXiv:2606.02173*, 2026.

[2] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "PANNs: Large-scale pretrained audio neural networks for audio pattern recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2880–2894, 2020.

[3] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[4] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2935–2947, 2017.

[5] L. Xuhong, Y. Grandvalet, and F. Davoine, "Explicit inductive bias for transfer learning with convolutional networks," in *International conference on machine learning*, 2018, pp. 2825–2834.

[6] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.