Frozen Bird Embeddings for Heart Murmur Detection

Method: Frozen Perch embeddings (1280-dim, trained on bird sounds) + Logistic Regression

Experiment 1: What is Perch Learning?

Setup: Binary classification (Present vs Absent/Unknown) to test if bird embeddings can distinguish heart murmurs.

Results: Recording-level AUROC=0.863 (mean), 0.865 (max), AUPRC=0.754 | Window-level AUROC=0.850, AUPRC=0.732

Confusion Matrix (Recording-level, threshold=0.5):

```
TN=456  FP=85   Specificity: 84.3%
FN=32   TP=88   Sensitivity: 73.3%
                Precision:   50.9%
```

Key Finding: The model successfully detects murmurs and is not simply predicting all-negative (balanced confusion matrix with 73% sensitivity, 84% specificity).

Experiment 2: How Does Perch Compare to Baselines?

Setup: Binary classification with fair comparison against statistical audio features.

Results (AUROC):

Method	Recording	Window
Perch	**0.863**	**0.850**
VGGish	0.818	0.806
MFCC+Spectral	0.769	0.759
MFCC	0.765	0.756
Random	0.481	0.485

Key Finding: Frozen bird embeddings outperform both audio-specific embeddings (VGGish, +4.5%) and traditional signal processing features.

Experiment 3: Competition Comparison

Setup: 3-class classification (Present/Unknown/Absent) using exact PhysioNet 2022 competition metrics.

Perch Performance: AUROC=0.793, AUPRC=0.611, Weighted Accuracy=0.759 → **Rank 6/40**

Leaderboard Context:

Rank	Team	AUROC	AUPRC	Weighted Acc
1	HearHeart	0.884	0.716	0.780
4	PathToMyHeart	0.880	0.684	0.771
6	**Perch (Tuned)**	**0.793**	**0.611**	**0.759**
6	Care4MyHeart	0.891	0.717	0.757
9	ISIBrno-AIMT	0.897	0.746	0.755

Key Finding: Using frozen bird embeddings with logistic regression ranks 6th among 40 teams, competitive with approaches specifically trained on heart sound data.

Takeaways

1. **Frozen embeddings work on heart audio:** 86% AUROC, balanced performance without cardiovascular training.
2. **Beats domain-specific models:** Outperforms VGGish (+4.5%), MFCC, spectral features.
3. **Competitive in real competition:** Ranks 6/40 without task-specific training.

Bird embeddings transfer effectively to medical audio, a data-scarce domain.