澳門理工大學
Universidade Politécnica de Macau
Macao Polytechnic University

HYPERSEC
ICML International Conference On Machine Learning
2026

Spherical Procrustes Alignment for Reliable Medical Audio Diagnosis

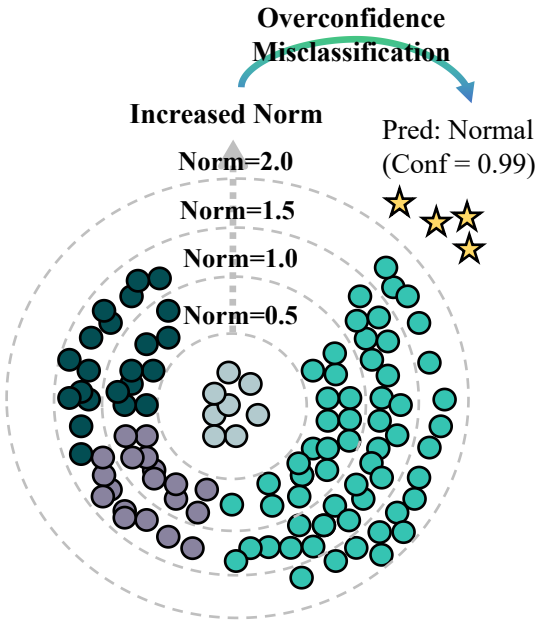
Ying Wang[1], Guoheng Huang[2], Chan-Tong Lam[1], Xiaochen Yuan[1]

[1]Macao Polytechnic University, xcyuan@mpu.edu.mo

Code: github.com/wangying1586/SPA

Motivation – The Overconfidence Problem

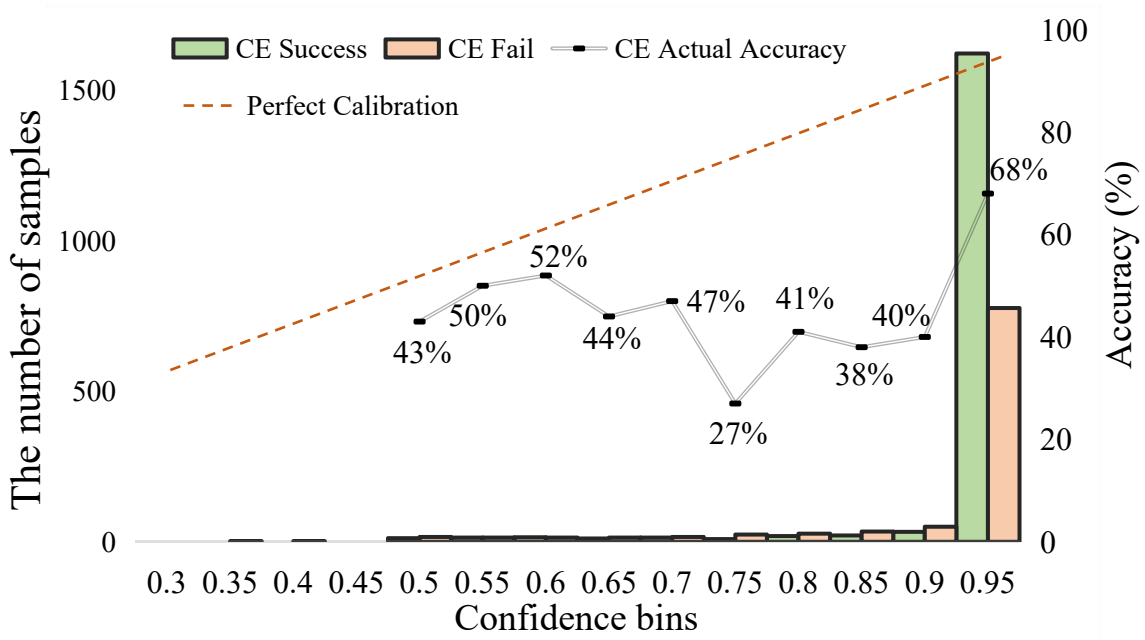
norm-biased radial space
→ high-norm noise triggers 0.99 confidence



(a) **CE**

Reliability diagram (CE baseline)
→ far from diagonal (Perfect Calibration)

CE's true accuracy is around 60%, but most of the confidence distribution is at 0.95.

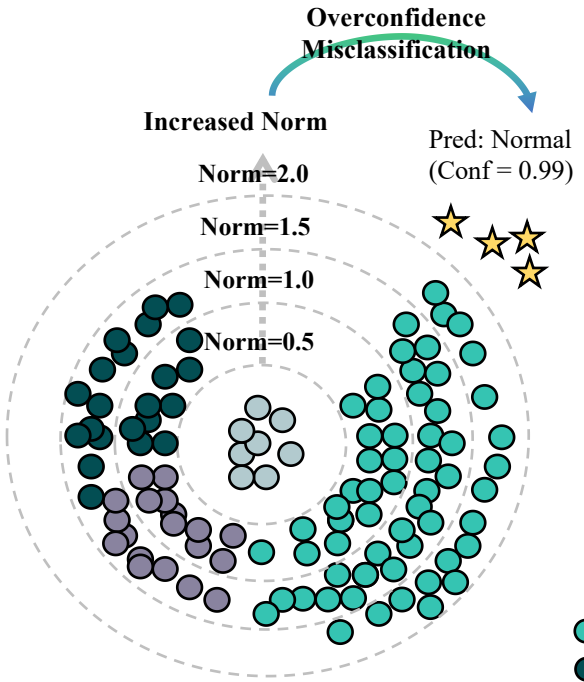


(c) **CE vs. SPA**

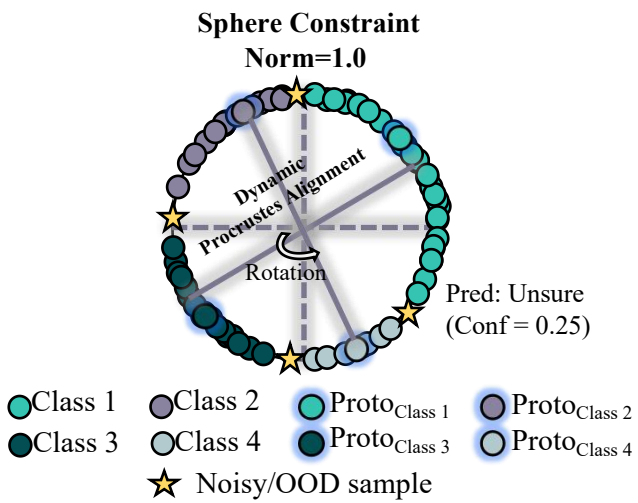
Clinical diagnosis needs both accuracy and well-calibrated confidence.

Root Cause & Our Insight

Root cause: Norm bias – logit magnitude dominates, not semantic alignment
Why existing methods fail: Loss-level or post-hoc; ignore geometric pathology

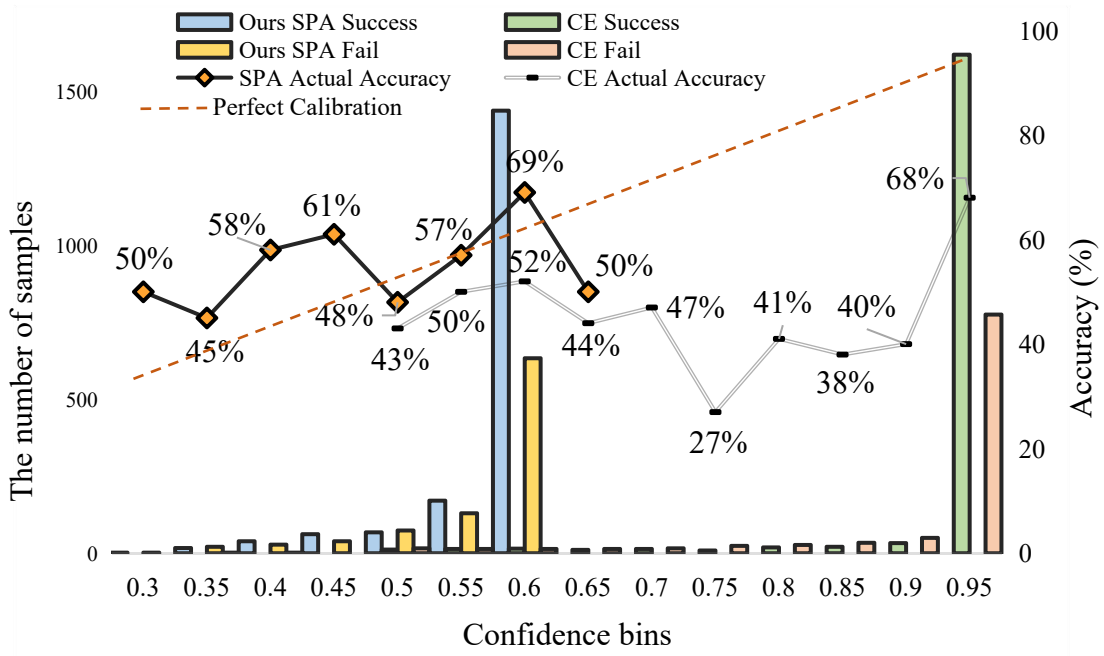


(a) **CE**



(b) **SPA**

Ours SPA is closer to the diagonal (Perfect Calibration).



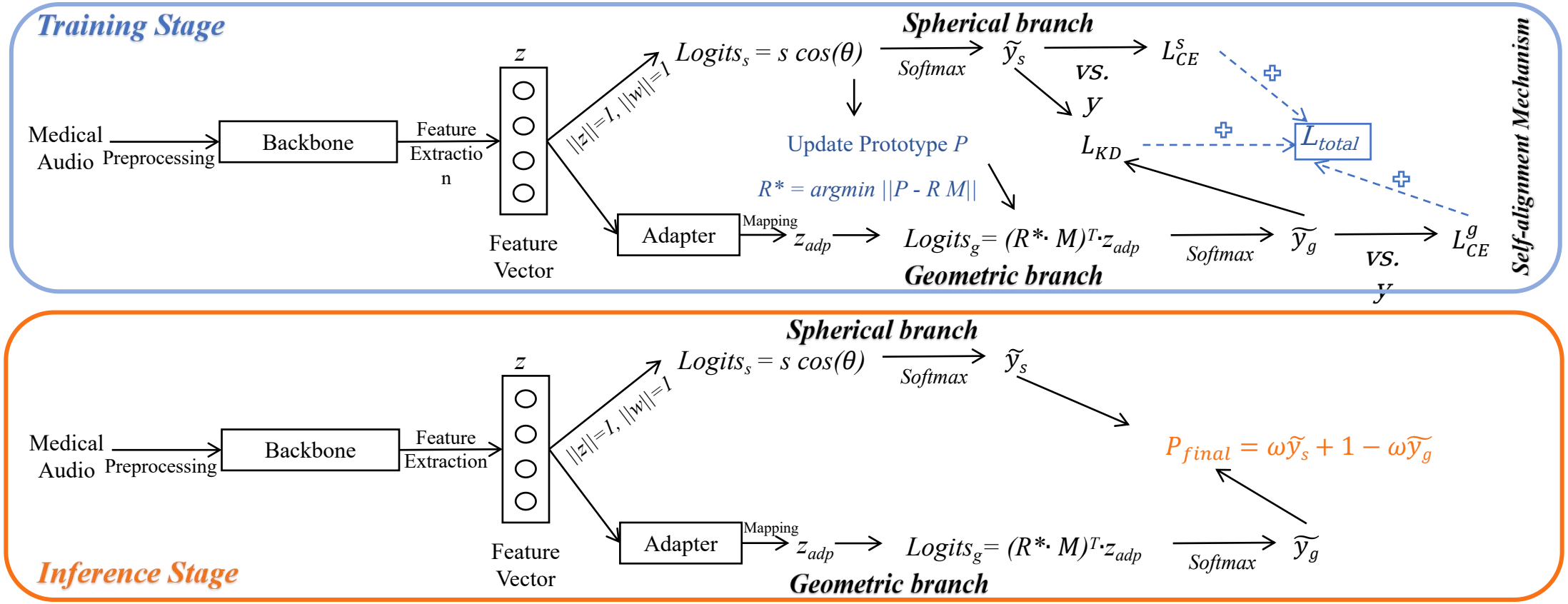
(c) **CE vs. SPA**

Our insight - Enforce spherical constraint plus dynamic geometric alignment with ETF

Spherical Procrustes Alignment Overview

Training stage:

- **Spherical branch:** unit hypersphere, eliminate norm bias
- **Geometric branch:** fixed ETF plus dynamic Procrustes alignment
- **Self-alignment Mechanism:** fuse via KL divergence



Inference stage: weighted fusion

Spherical Procrustes Alignment – Dual branch architecture

Spherical Branch – Decouple Norm

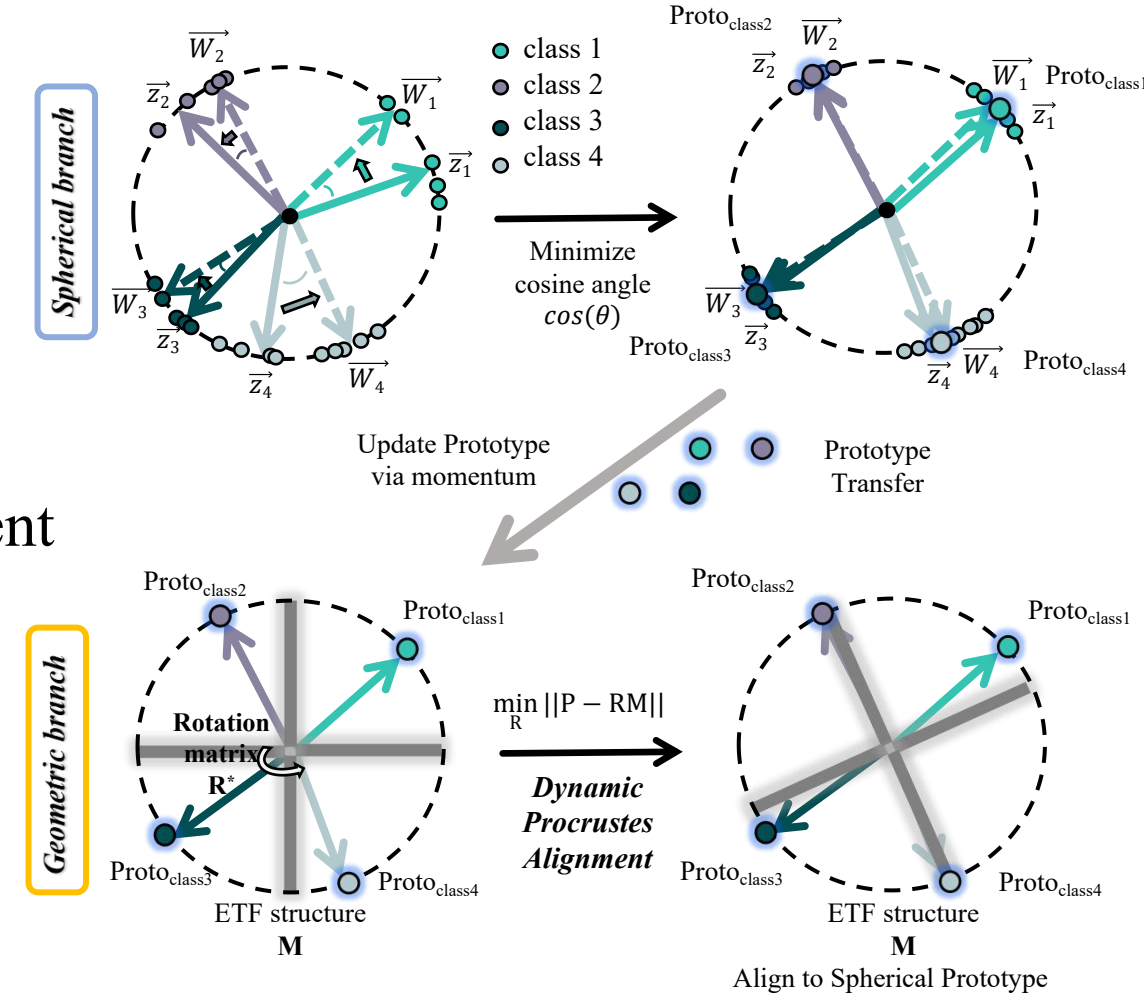
- Normalize features $\tilde{z}$ and weights $\widetilde{W}$
- Logit = $s \cdot \cos\theta \rightarrow$ confidence purely angular

Momentum prototype update for stability

Geometric Branch – Dynamic Procrustes Alignment

- Fixed Simplex ETF M (maximally separated)
- Solve Orthogonal Procrustes:
- $min_R \parallel P - RM \parallel_F \rightarrow R^* = UV^\top$ via SVD

Avoid gradient jitter
→ stable geometry even on small/noisy data



Spherical Procrustes Alignment – Training & Inference

Joint loss: $\mathcal{L}_{total} = \gamma\left(\mathcal{L}_{CE}^{s} + \mathcal{L}_{CE}^{g}\right) + \lambda\mathcal{L}_{KD}$

Inference: $P_{final} = \omega \cdot \sigma(logits_s) + (1 - \omega) \cdot \sigma\left(logits_g\right)$

No extra inference cost

Main Results – ICBHI & CirCor & Calibration Methods

ICBHI (BEATs): AS ↑ 65.27%, ECE ↓ from 23.36% to **6.05%**

Backbone	Method	Classification Performance			Reliability
		Se (%) ↑	Sp (%) ↑	AS (%) ↑	ECE (%) ↓
CNN6	+ CE Loss (Kong et al., 2020b)	$35.56_{\pm 4.99}$	$81.75_{\pm 4.33}$	$58.66_{\pm 0.54}$	$23.53_{\pm 5.01}$
	+ Focal Loss (Lin et al., 2017)	$35.11_{\pm 5.36}$	$79.18_{\pm 5.01}$	$57.14_{\pm 0.58}$	$17.51_{\pm 2.24}$
	+ LDAM-DRW (Cao et al., 2019)	$33.37_{\pm 3.96}$	$80.49_{\pm 4.13}$	$56.93_{\pm 0.85}$	$17.47_{\pm 3.85}$
	+ SPA (Ours)	$35.24_{\pm 2.92}$	$82.96_{\pm 4.82}$	**59.10**$_{\pm 1.24}$	**7.51**$_{\pm 0.72}$
AST	+ CE Loss (Gong et al., 2021)	$42.58_{\pm 2.71}$	$79.52_{\pm 2.07}$	$61.05_{\pm 0.54}$	$21.59_{\pm 6.99}$
	FBS (Fraihi et al., 2025)	$43.22_{\pm 2.44}$	$84.19_{\pm 3.08}$	$64.28_{\pm 1.09}$	$19.82_{\pm 5.13}$
	FBS + SPA (Ours)	$44.81_{\pm 2.65}$	$84.92_{\pm 3.19}$	**65.01**$_{\pm 2.77}$	**5.49**$_{\pm 1.13}$
BEATs	+ CE Loss (Chen et al., 2023)	$50.60_{\pm 1.79}$	$77.20_{\pm 3.22}$	$63.90_{\pm 1.15}$	$28.51_{\pm 4.91}$
	+ SPA (Ours)	$49.77_{\pm 2.17}$	$79.06_{\pm 2.81}$	**64.42**$_{\pm 0.78}$	**4.44**$_{\pm 1.02}$
	PAFA (Jeong & Kim, 2025)	$48.72_{\pm 3.75}$	$80.19_{\pm 4.07}$	$64.45_{\pm 0.52}$	$23.36_{\pm 7.83}$
	PAFA + SPA (Ours)	$49.34_{\pm 4.12}$	$81.21_{\pm 4.48}$	**65.27**$_{\pm 0.57}$	**6.05**$_{\pm 3.63}$

CirCor (M2D): Wacc ↑ 84.23%, ECE ↓ to **4.63%**

Backbone	Method	Classification		Reliability
		W_{acc} (%) ↑	UAR (%) ↑	ECE (%) ↓
Previous Studies				
Wav2vec	Panah et al. (Panah et al., 2023)	80.0	70.0	-
HSMM	CUED Acoustics (McDonald et al., 2022)	80.0	68.0	-
Pre-trained extractors				
CNN14	+ CE Loss (Kong et al., 2020a)	$57.47_{\pm 3.25}$	$53.63_{\pm 2.89}$	$11.24_{\pm 2.14}$
	+ SPA (Ours)	**58.32**$_{\pm 2.47}$	**54.47**$_{\pm 2.18}$	**6.08**$_{\pm 1.24}$
BYOL-A	+ CE Loss (Niizumi et al., 2021)	$54.34_{\pm 3.71}$	$54.81_{\pm 2.67}$	$12.13_{\pm 2.53}$
	+ SPA (Ours)	**55.73**$_{\pm 2.85}$	**55.52**$_{\pm 2.36}$	**6.42**$_{\pm 1.47}$
AST	+ CE Loss (Gong et al., 2021)	$64.12_{\pm 2.38}$	$66.38_{\pm 2.95}$	$10.35_{\pm 1.87}$
	+ SPA (Ours)	**65.24**$_{\pm 1.96}$	**66.91**$_{\pm 2.07}$	**5.58**$_{\pm 1.12}$
M2D	+ CE Loss (Niizumi et al., 2024)	$83.51_{\pm 1.92}$	$72.14_{\pm 2.14}$	$9.17_{\pm 1.65}$
	+ SPA (Ours)	**84.23**$_{\pm 1.53}$	**72.96**$_{\pm 1.78}$	**4.63**$_{\pm 0.89}$

Method	AS (%) ↑	ECE (%) ↓	Inf. Time (ms)
CE Loss	$63.90_{\pm 1.15}$	$28.51_{\pm 4.91}$	4.1
Label Smoothing	$62.58_{\pm 0.59}$	$33.91_{\pm 4.41}$	4.1
Mixup	$62.16_{\pm 0.66}$	$8.67_{\pm 6.29}$	4.1
Post-hoc & Sampling Methods			
Temperature Scaling	$63.90_{\pm 1.15}$	$21.99_{\pm 6.97}$	4.1
MC Dropout (5 passes)	$63.82_{\pm 0.89}$	$25.90_{\pm 4.18}$	19.1
Deep Ensembles (5 models)	$64.40_{\pm 0.93}$	$22.29_{\pm 4.89}$	20.5
SPA (Ours)	**64.42**$_{\pm 0.78}$	**4.44**$_{\pm 1.02}$	**4.2**

Outperforms Label Smoothing, Mixup, Temperature Scaling, MC Dropout, and Deep Ensembles

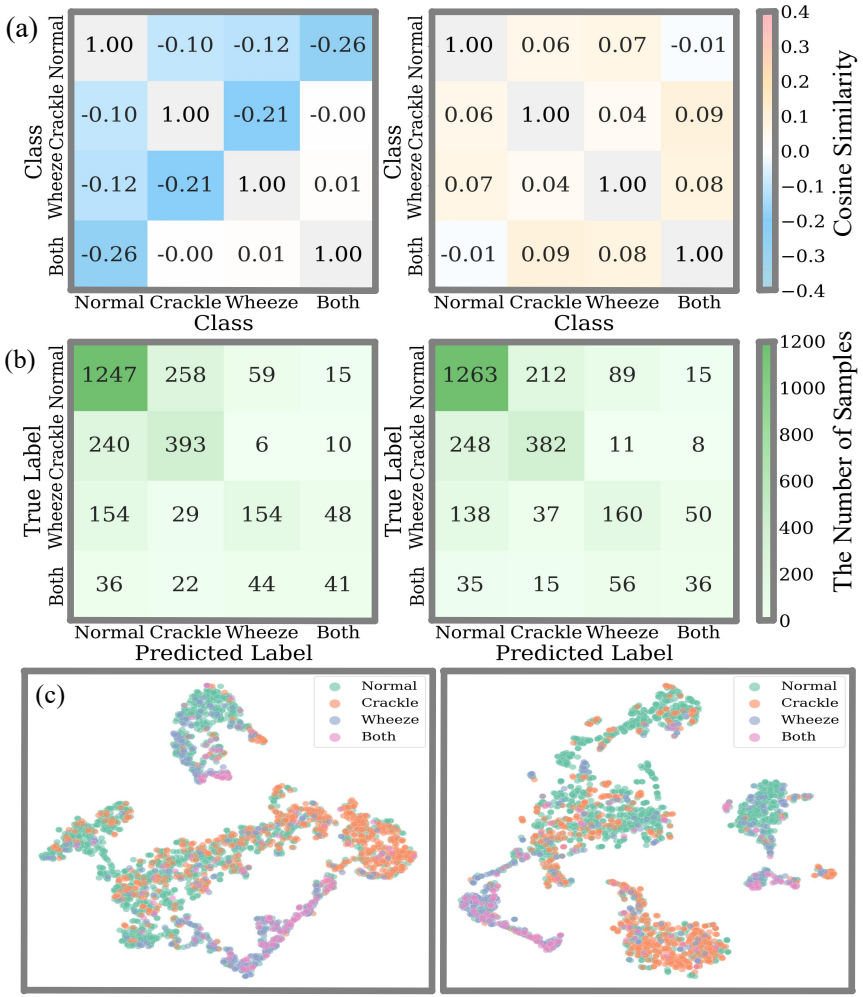
without consuming extra inference

澳門理工大學
Universidade Politécnica de Macau
Macao Polytechnic University

HYPERSEC
ICML International Conference On Machine Learning
2026

Visualization

Comparison on ICBHI



(a) Cosine similarity heatmap

SPA *enforces near-zero inter-class correlation*

(b) Confusion matrix

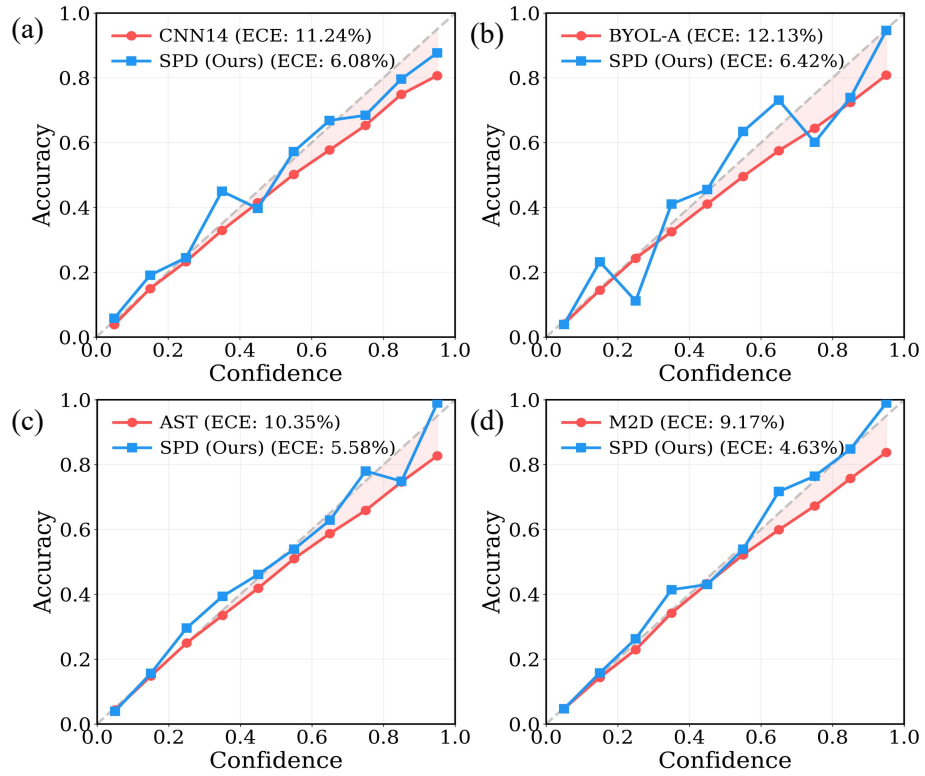
SPA *improves minority recall without hurting majority*

(c) t-SNE projection

SPA *yields compact, angularly separated clusters*

Reliability diagram on CirCor

SPA (blue) aligns near diagonal vs. baseline (red)



Conclusion

Norm bias is the geometric root of overconfidence in medical audio

SPA: spherical constraints plus dynamic Procrustes alignment

SOTA accuracy and calibration, no extra inference cost

Future Work

Robust prototype estimation, OOD detection

This work was supported by
the Macao Polytechnic University under grant RP/FCA-12/2025 and the Science and Technology Development Fund of Macau SAR under Grant 0053/2025/RIB2.