ABCD: Air-Bone Cough Contrastive Detector for Dual-Path Edge Audio Health Monitoring

Langchu Huang[1,2], Feiya Xiang[1], Joseph de la Viesca[1], Mayur Sanap[1], Edgar Lobaton[1*]

[1] *Department of Electrical and Computer Engineering, North Carolina State University*

[2] *Department of Computer Science and Engineering, Southern University of Science and Technology*

INTRODUCTION

❖ **Cough is a key biomarker**

An early symptom of respiratory conditions such as asthma and infections.

Objective, continuous monitoring improves diagnosis and treatment tracking.

❖ **Multimodal approaches enhance robustness**

Current studies indicate that fusing AC with motion or physiological signals improves cough detection under real-world noise and movement.

However, **synchronized air–bone conduction (AC–BC) fusion remains largely unexplored for joint cough–speech detection.**

❖ **Bone-conduction signals carry complementary cues**

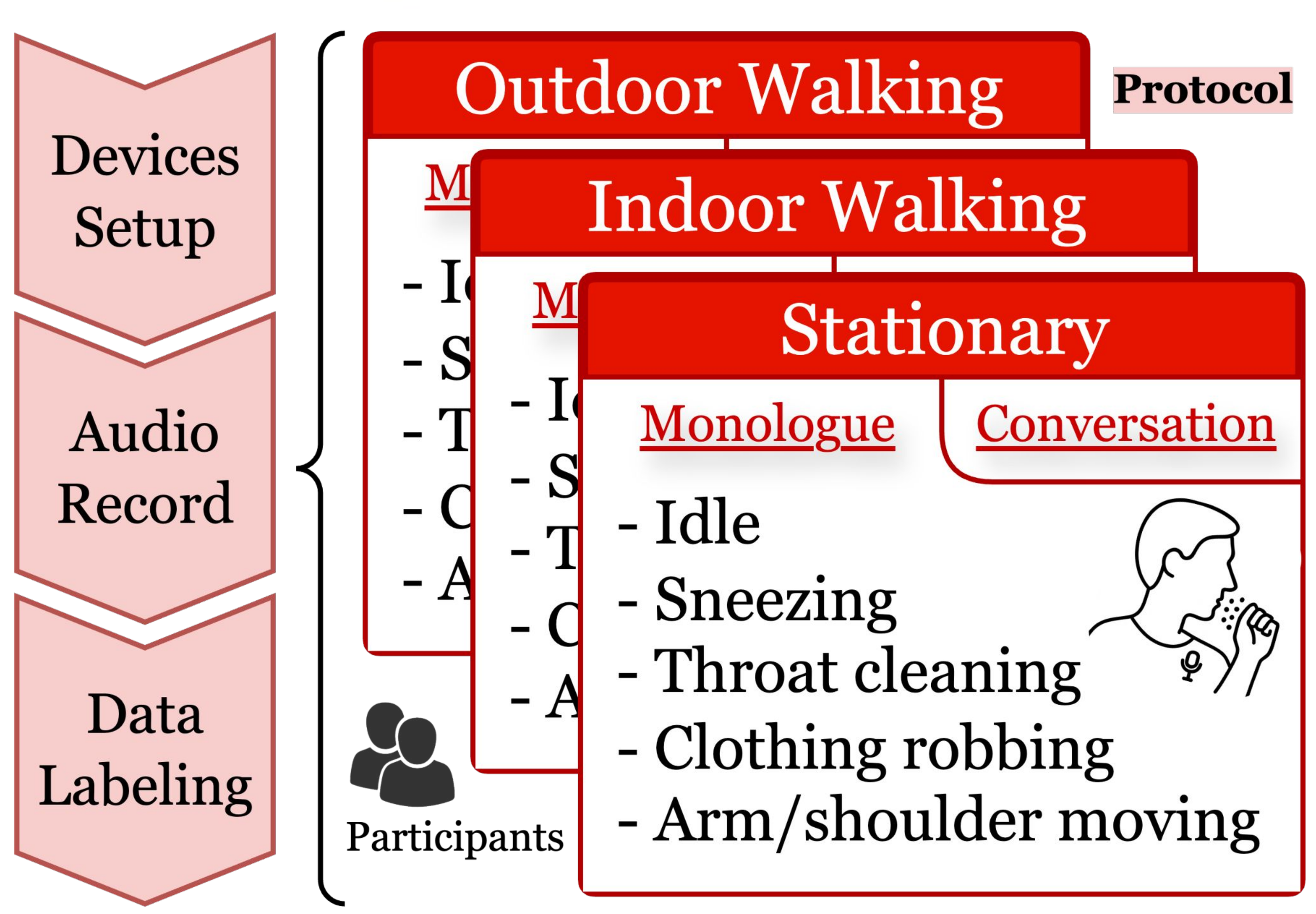
BC audio suppresses ambient noise and speech intelligibility, offering robustness and privacy.

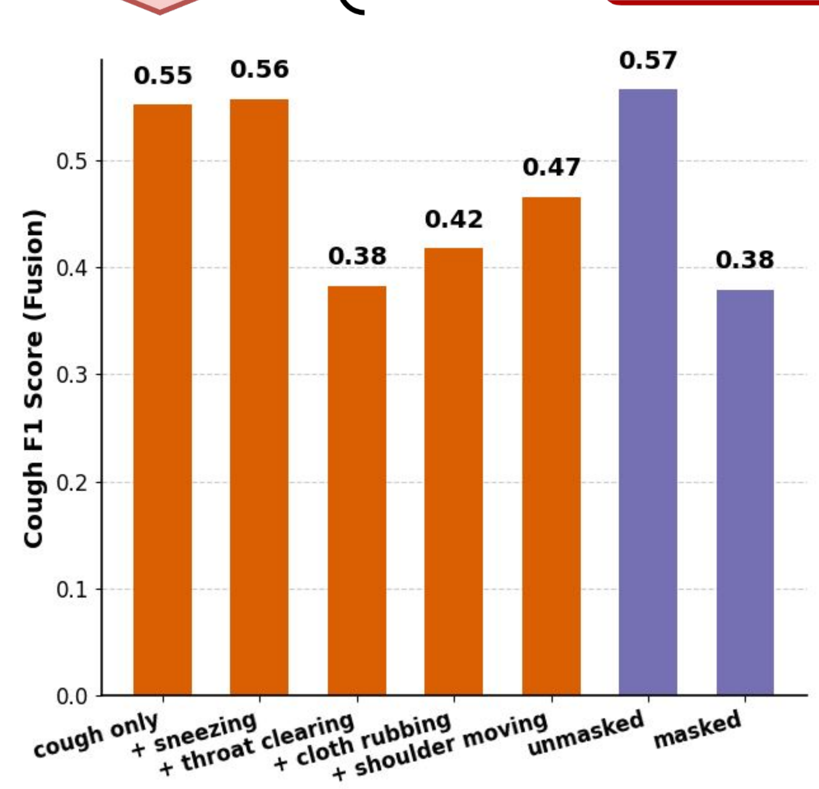
Yet due to their limited bandwidth and muffled sound quality, **effective BC feature utilization requires specialized fusion strategies.**

DATA COLLECTION

Air Conduction Microphone (AC)

Bone Conduction Microphone (BC)





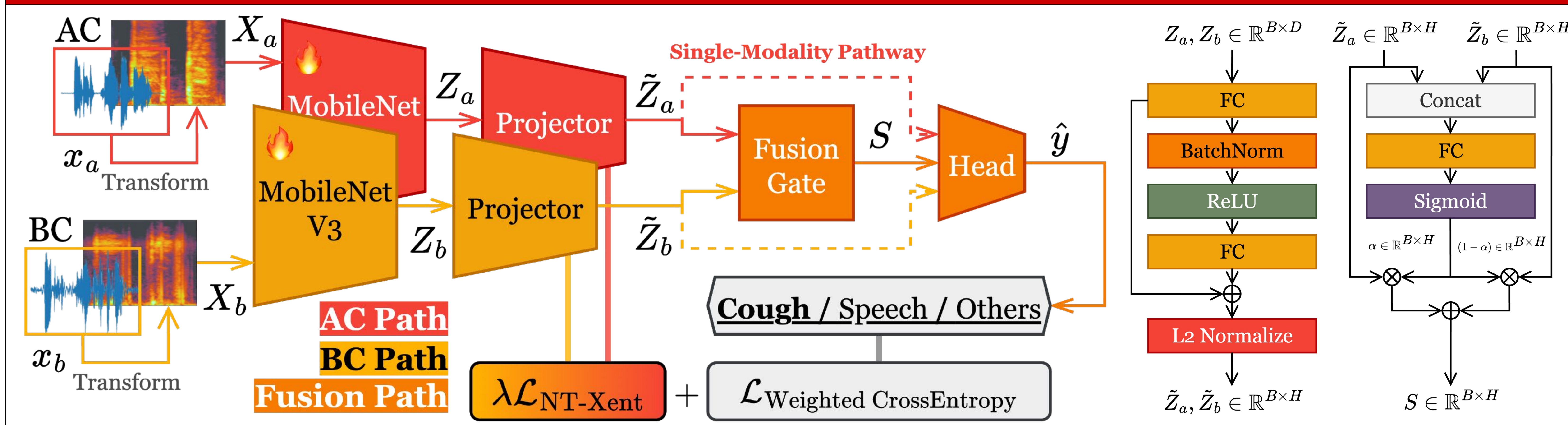
The model exhibits varying degrees of confusion under different interferences:

throat clearing > masked > cloth rubbing > shoulder moving > sneezing ≈ idle

The model performs worse under the masked condition compared to unmasked.

The results reveal how various interferences affect the efficiencies of cough detection.

METHODOLOGY

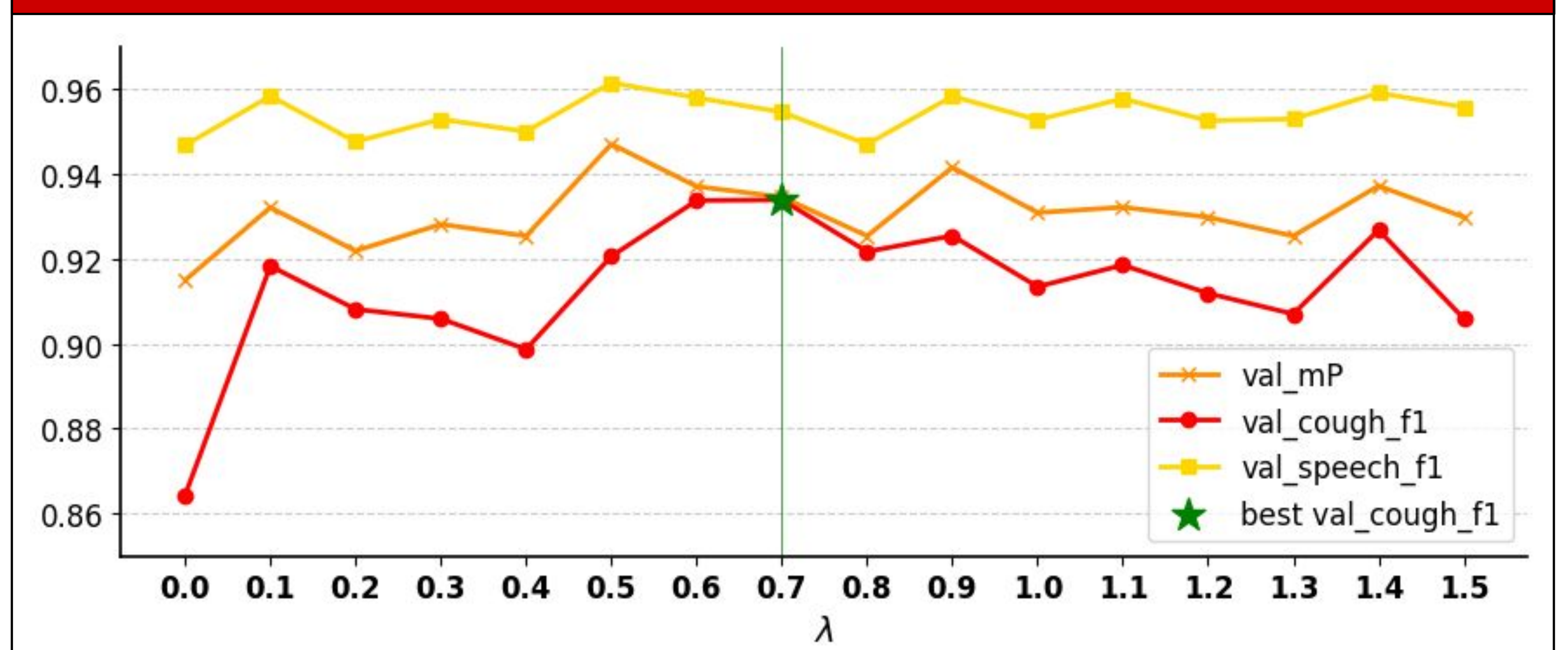


(a) Overview of the proposed ABCD framework for joint cough and speech detection using synchronized air-conduction (AC) and bone-conduction (BC) audio signals. The model processes log-Mel spectrograms through dual MobileNet encoders and projects them into a shared latent space via MLP-based projectors. A gated fusion module adaptively combines the two modalities before classification. The architecture supports both modality-specific and fused pathways, and is trained with a hybrid loss combining weighted cross-entropy and contrastive alignment.

(b) Projector: 2-layer MLP with BatchNorm and residual path, followed by L2 normalization.

(c) Fusion Gate: Learns modality weights via sigmoid gating over concatenated features.

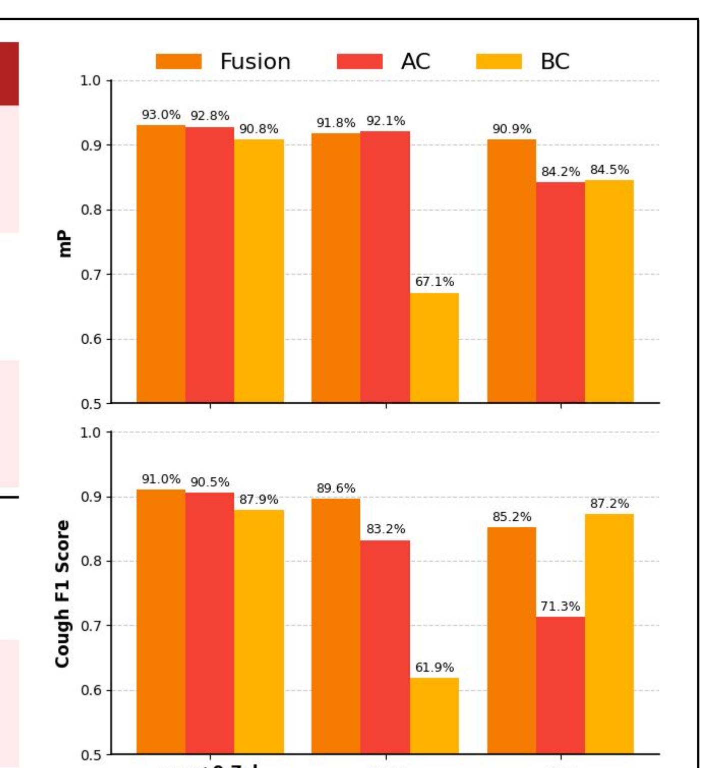
EXPERIMENTS



(d) Validation performance vs. contrastive loss weight λ. The figure shows how validation mP, cough F1, and speech F1 change as the weight of the contrastive loss λ increases. The optimal cough F1 score is achieved at λ = 0.7 (highlighted by a green star), suggesting that a moderate contrastive loss improves performance by effectively balancing classification and representation learning.

Method	Fusion	AC	BC
$\text{ABCD}_{\text{wce+0.7cl}}$	mP: <u>0.930</u> cough_f1: <u>0.910</u>	mP: 0.928 cough_f1: 0.905	mP: 0.908 cough_f1: 0.879
ABCD_{wce}	mP: 0.918 cough_f1: 0.896	mP: 0.921 cough_f1: 0.832	mP: 0.671 cough_f1: 0.619
ABCD_{ce}	mP: 0.909 cough_f1: 0.852	mP: 0.842 cough_f1: 0.713	mP: 0.845 cough_f1: 0.872
$\text{ABCD}_{\text{wce\_ac}}$	—	mP: 0.910 cough_f1: 0.884	mP: 0.023 cough_f1: 0.127
$\text{ABCD}_{\text{wce\_bc}}$	—	mP: 0.132 cough_f1: 0.0	mP: 0.885 cough_f1: 0.848

(e) Ablation Evaluation (mP and cough F1) across different training strategies and modalities. The upper section compares three loss configurations—(1) contrastive + weighted cross-entropy, (2) weighted cross-entropy, and (3) standard cross-entropy—all trained jointly on AC and BC data, and evaluated on three test modalities. The lower section presents results from models trained on a single modality only: (4) one using AC data and (5) the other using BC data, both employing weighted cross-entropy loss.



(f) Bar plots visualization of the upper section of Table (e), showing the mP (top) and cough F1 (bottom) scores for models trained with wCE+CL, wCE, and CE on AC+BC data.

CONCLUSION

(d) The hyperparameter tuning on contrastive loss weight λ confirms that the best validation cough F1 score is achieved at **λ = 0.7**, significantly outperforming the model without contrastive loss.

(e) The model trained with the combined loss (wCE + CL) after tuning achieves the highest overall performance (**mP: 93.0%, cough F1: 91.0%**) and remains strong under single-modality inference.

The upper section of Table (e) shows that both weighted cross-entropy and contrastive loss contribute positively to performance improvement.

The bottom section further indicates that single-modality trained models not only underperform compared to the fusion model, but also fail to generalize across modalities.

(f) These results suggest that contrastive learning not only boosts performance but also aligns the latent spaces of AC and BC representations, enabling the shared classification head to better adapt to both individual modalities and their linear fusion. This alignment significantly enhances single-modality inference, approaching the performance of full fusion.

FUTURE WORK

- ❖ **Pretraining**: Leverage public AC–BC datasets to pretrain the encoders for improved extraction and generalization.
- ❖ **Benchmark Evaluation**: Evaluate on established cough detection benchmarks and aim to match or surpass state-of-the-art results.
- ❖ **Multi-modal Extension**: Incorporate IMU data to explore a tri-modal fusion architecture (AC + BC + IMU).
- ❖ **Modality-aware Augmentation**: Design augmentation strategies tailored to AC–BC characteristics to enhance cross-modal robustness.
- ❖ **Parameter Tuning**: Further refine model configurations, including audio transformations and hyperparameters.

REFERENCE