

Article

OCAD: Overlap Composition and Class-Atom Decoding for Respiratory-Sound Classification

Xinyu Zhang ¹, Wei Zhao ¹ and Haicheng Liang ^{2,*}

¹ State Key Laboratory of Extreme Environment Optoelectronic Dynamic Measurement Technology and Instrument, North University of China, Taiyuan 030051, China; s202306129@st.nuc.edu.cn (X.Z.); s202306135@st.nuc.edu.cn (W.Z.)

² School of Materials Science and Engineering, Shenyang Ligong University, Shenyang 110159, China

* Correspondence: hcliang1976@sylu.edu.cn; Tel.: +86-024-24680989

Abstract

Automatic respiratory-sound classification becomes particularly challenging when crackles and wheezes coexist within the same respiratory cycle, because overlap events contain mixed acoustic characteristics and are severely underrepresented in standard benchmarks. In the ICBHI 2017 Respiratory Sound Database, the rare overlap label, Both, is frequently absorbed into Crackle, Wheeze, or Normal classes by conventional flat classifiers, resulting in poor interpretability and weak clinical reliability. To address this issue, this paper proposes OCAD (Overlap Composition and Class-Atom Decoding), an explainable overlap-aware respiratory-sound classification framework. Unlike conventional end-to-end classifiers that directly map features to labels, OCAD explicitly decomposes respiratory-sounds into interpretable crackle and wheeze factors, composes an overlap representation from these abnormal-event components, and performs classification through structured class atoms whose geometric relationships reflect the progression from Normal to single-event and overlap states. By preserving a fixed acoustic backbone across all compared systems, the study isolates the contribution of the proposed explainable representation and decoding mechanism. Experimental results on a fixed patient-level split of the ICBHI 2017 dataset show that OCAD achieves an ICBHI Score of 0.6920, Macro Score of 0.6550, Macro F1 of 0.6680, crackle sensitivity of 0.6300, wheeze sensitivity of 0.6550, and Both sensitivity of 0.2620. Compared with the strongest structured class-atom baseline under the same protocol, OCAD improves the ICBHI Score by 0.1334, Macro F1 by 0.2575, and Both sensitivity by 0.2054 absolute points. Additional robustness analyses across repeated seeds, patient-level cross-validation, and external respiratory-sound datasets further support the effectiveness of the proposed framework. The proposed overlap-aware representation and decoding strategy provides a more interpretable and reliable approach for respiratory sound classification in the presence of co-occurring abnormal events.



Academic Editor: Douglas O'Shaughnessy

Received: 27 May 2026

Revised: 27 June 2026

Accepted: 1 July 2026

Published: 7 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: respiratory-sound classification; overlap-aware learning; class-atom decoding; compositional representation learning; class-atom decoding

1. Introduction

Respiratory-sound analysis has been studied for a long time in digital auscultation, computerized acoustic assessment, and, more recently, smart-device monitoring scenarios, and smart healthcare monitoring systems [1–4], where the secure handling of networked physiological data is also becoming a practical concern [5,6]. Automatic analysis of

respiratory-sounds can assist clinicians in the assessment of pulmonary abnormalities and has been investigated for the evaluation and follow-up of chronic airway diseases such as chronic obstructive pulmonary disease (COPD) and bronchiectasis [7–10]. With the rapid development of artificial intelligence and biomedical signal processing, respiratory-sound classification has become an important research topic in intelligent healthcare applications. In standard respiratory sound benchmarks, respiratory cycles are commonly categorized into four classes: Normal, Crackle, Wheeze, and Both. Among these categories, the Both class is particularly challenging because it contains simultaneous crackle and wheeze events within the same respiratory cycle. Compared with single-event categories, overlap samples exhibit mixed acoustic characteristics and are typically underrepresented in the dataset, resulting in severe class imbalance and ambiguous decision boundaries. Consequently, overlap events are frequently misclassified as Crackle or Wheeze by conventional multi-class classification models.

Existing respiratory-sound classification methods mainly focus on improving acoustic feature extraction using convolutional neural networks, recurrent architectures, attention mechanisms, or Transformer-based models. Although these approaches have achieved promising performance, most studies still treat the Both category as an independent class in a standard four-class classification framework. Such a formulation ignores the compositional relationship between crackle and wheeze evidence and limits the ability of the model to preserve overlap-event structure during representation learning and final decision making. In addition, conventional multi-label learning strategies mainly focus on label co-occurrence and do not explicitly model the acoustic composition relationship between mixed abnormal respiratory events.

To address this issue, this paper proposes OCAD (Overlap Composition and Class-Atom Decoding), an overlap-aware respiratory-sound classification framework that explicitly models overlap events through crackle–wheeze composition learning. The proposed method keeps the acoustic backbone unchanged and focuses on the downstream representation and decoding process. Specifically, OCAD introduces separate crackle and wheeze event factors, constructs an explicit overlap composition representation, and performs structured class decoding through a class-atom decoder. In addition, a lightweight event-first tuning strategy is employed to improve the recognition of acoustically mixed abnormal events.

Experiments are conducted under a fixed patient-level split protocol to evaluate the effectiveness of the proposed framework. The experimental results demonstrate that explicit overlap-aware modeling can substantially improve recognition performance for overlap respiratory events while maintaining competitive overall classification performance.

The main contributions of this work are as follows.

1. We reformulate the conventional four-class respiratory-sound classification task as an overlap-aware compositional learning problem, in which crackle and wheeze evidence remains explicitly represented and the Both category is modeled as structured co-occurrence rather than as an independent residual class.
2. Based on this formulation, we propose the OCAD framework, which incorporates separate event representations, explicit overlap composition, and a structured class-atom decoder to preserve event-level evidence during representation learning and final class prediction.
3. Under a controlled fixed patient-level evaluation protocol, the proposed overlap-aware design, together with a lightweight event-first adaptation strategy, improves overlap-event recognition while also enhancing performance on single-event abnormal classes. Although the current study is conducted under a benchmark-specific setting, the

results demonstrate the effectiveness of explicit overlap modeling for mixed abnormal respiratory-sound classification.

2. Related Work

2.1. Traditional Respiratory-Sound Classification

Traditional respiratory-sound analysis is rooted in computerized auscultation and signal-processing-based classification. Early reviews established the clinical and acoustic basis of adventitious sound analysis, including the distinction between normal respiratory cycles and abnormal events such as crackles and wheezes [1,2,11]. In this line of work, handcrafted features such as time-domain descriptors, spectral statistics, and cepstral features are commonly paired with conventional classifiers. For example, cepstral statistical features have been used for lung-sound classification without relying on end-to-end representation learning [12].

A major strength of traditional methods is their relatively high interpretability and low computational cost, since handcrafted descriptors often correspond to observable acoustic characteristics. However, the feature space is typically fixed before classification, making it difficult to preserve independent crackle and wheeze evidence when rare mixed respiratory events are encountered. This limitation is important for the present task because Both is not merely an additional isolated category, but the co-occurrence of two abnormal acoustic events. Although handcrafted pipelines can describe local acoustic patterns, they generally lack an explicit mechanism for interpretable overlap reasoning or for modeling the semantic relation among Crackle, Wheeze, and Both classes.

2.2. Deep Learning for Respiratory-Sound Classification

Deep learning methods reduce the dependence on handcrafted features by learning task-specific representations directly from respiratory audio or time-frequency inputs. CNN-based systems have been widely applied to lung-sound spectrograms and related representations [13–15]. CRNN and recurrent variants further model temporal structure within respiratory cycles [16,17], while lightweight and wearable-oriented models address practical deployment constraints [18,19]. Other studies compare time-frequency representations such as spectrograms, mel-spectrograms, gammatonegrams, and scalograms [20,21]. Clinical deep-learning systems have also been investigated for crackles, wheezes, and related respiratory categories in real-world settings [22].

Recent work increasingly adopts attention mechanisms, Transformer encoders, and audio pretraining strategies originating from the broader audio-analysis community [23–25]. Respiratory-sound classification has followed this trend through attention-based fusion and vision-transformer-style acoustic models [26–28]. This momentum is part of a wider movement in which neural networks and interpretable machine-learning frameworks are now applied across diverse engineering and biomedical problems, ranging from material characterization and alloy design to brain-computer interfaces and computational design [29–32]; this broad adoption motivates studying not only stronger encoders but also more structured and interpretable decision layers. In particular, recent advances in decision-centric learning and optimization-based control in complex dynamical systems demonstrate that explicit state decomposition and constraint-aware inference can improve robustness and interpretability under uncertainty [33,34]. This broad adoption motivates studying not only stronger encoders but also more structured and interpretable decision layers. These approaches can substantially improve spectro-temporal representation learning; however, many still rely on flat classification heads or weakly structured abnormal-event routing mechanisms. As encoder complexity increases, the relationship between learned representations and clinically meaningful respiratory events may become

increasingly opaque. Consequently, even strong encoders may treat Both as a residual or ambiguous class unless the downstream representation and decision process explicitly preserve independent crackle and wheeze evidence.

From the perspective of explainable artificial intelligence, this limitation is not only a classification problem but also an interpretability problem. A prediction may achieve high confidence without exposing how overlap evidence contributes to the final decision. Therefore, respiratory-sound classification increasingly requires not only stronger representation learning, but also more transparent and semantically structured inference mechanisms.

The most competitive recent ICBHI systems follow this transfer-learning direction. Finetuning an AudioSet-pretrained AST is already strong, and Patch-Mix contrastive learning on top of AST [35] reached a four-class ICBHI Score around 0.62 by mixing patch embeddings and contrasting the mixed representations, explicitly noting that a naive mixup loss fails because of the label hierarchy among respiratory sounds. Very recent work pushes the backbone further: state-space audio models with dual-axis Patch-Mix report a four-class Score around 0.64 [36], and a 2025 review surveys the rapidly growing ICBHI2017 literature [37]. These systems improve the shared encoder and the training objective, but the final decision is still a flat four-way head; the overlap class is not composed from its constituent events. This is the gap OCAD targets, and we reimplement the AST and Patch-Mix lines under our own fixed split as recent strong baselines in Section 4.

2.3. Ensemble Learning and Structured Classification

Another line of research combines complementary evidence sources through multi-view inputs, hybrid CNN–RNN architectures, attention fusion, or patient-adaptive components [16,18,26]. These methods can broadly be regarded as ensemble or fusion strategies because they aggregate information from different representations, channels, or modules. Such fusion mechanisms can improve robustness and representation diversity, but they do not automatically resolve the semantic ambiguity of overlap labels. If the final output remains a conventional four-way softmax classifier, the overlap category may still collapse into one of the dominant single-event classes.

Structured classification addresses a different issue: how should the decision space itself be represented in an interpretable manner? The ICBHI benchmark labels possess an inherent latent structure: Normal contains neither crackle nor wheeze evidence, Crackle and Wheeze each correspond to a single abnormal event, and Both represents simultaneous occurrence of both abnormal events. This structure is closer to a constrained multi-label or compositional decision problem than to four unrelated categorical endpoints, even though the benchmark ultimately requires a single final label.

From an explainable-AI perspective, preserving this latent structure can improve transparency of the inference process because intermediate event evidence remains semantically meaningful throughout decoding. OCAD is positioned at this structured and interpretable decision level. It can operate with a conventional acoustic backbone, but its primary contribution is to preserve semantically meaningful event factors, explicitly compose overlap evidence, and decode predictions through interpretable class atoms that reflect the relation among normal, single-event, and overlap respiratory states.

2.4. Benchmark Context and Position of This Work

The ICBHI 2017 Respiratory Sound Database has become the primary public benchmark for respiratory cycle classification [38]. However, direct comparison across studies is often difficult because patient partitioning, preprocessing, augmentation strategies, and evaluation conventions vary substantially [39,40]. For this reason, the present work emphasizes controlled in-family comparison under the same patient-level split rather than

claiming superiority through heterogeneous cross-paper score matching. The respiratory-sound literature continues to evolve rapidly, and we explicitly acknowledge that the strongest possible external comparison would re-evaluate very recent benchmark systems under an identical split and preprocessing pipeline; we return to this point as a calibrated limitation in the experimental discussion rather than overstating cross-paper superiority.

Compared with traditional handcrafted pipelines, OCAD learns semantically interpretable event-specific latent factors instead of relying solely on fixed descriptors. Compared with conventional deep classifiers, it modifies the semantic structure of the downstream inference and decoding process rather than only strengthening the acoustic encoder. Compared with general fusion or ensemble strategies, it uses fusion for an explicitly interpretable purpose: the Both category should be supported by simultaneous crackle and wheeze evidence rather than by residual classifier uncertainty.

Accordingly, the intended contribution of this work is not a new universal respiratory-sound backbone, but an explainable overlap-aware representation and decoding framework designed for a known weakness of the four-class respiratory-sound benchmark. Unlike post hoc visualization approaches, the proposed framework introduces intrinsic interpretability by embedding clinically meaningful event composition directly into the representation and decision process.

3. Method

3.1. Problem Setting and Overall Architecture

For each input respiratory cycle, the classifier must output one of four labels: Normal, Crackle, Wheeze, or Both. Rather than viewing them as four unrelated endpoints, we define a compositional label map $\phi(y) = (e_c, e_w)$, where e_c and e_w denote crackle and wheeze presence. The full compositional label space is $Y_{comp} = \{(0,0), (1,0), (0,1), (1,1)\}$

$$\phi(\text{Normal}) = (0, 0), \phi(\text{Crackle}) = (1, 0), \phi(\text{Wheeze}) = (0, 1), \phi(\text{Both}) = (1, 1) \quad (1)$$

where the two coordinates indicate the presence of crackle and wheeze events. This small reformulation changes what the model is asked to do. Instead of learning one free-standing prototype for Both, the network should keep evidence for the two events visible and let the final decision reflect their co-occurrence. Figure 1 illustrates this label relation.

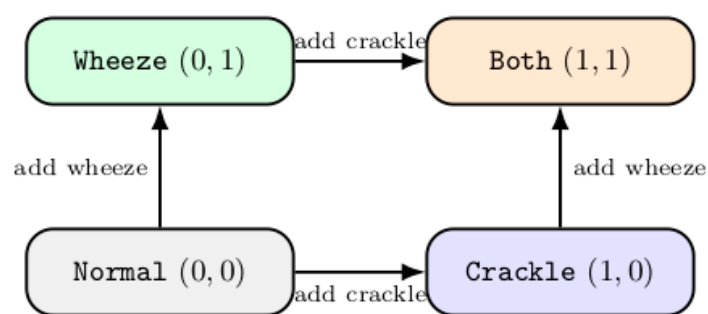


Figure 1. Label relation used in OCAD. The overlap label Both is tied to the joint presence of crackle and wheeze rather than treated as an unrelated fourth endpoint.

Figure 2 summarizes OCAD. A shared acoustic backbone first extracts an embedding h from spectrogram and waveform inputs. Two event branches then produce crackle and wheeze factors, z_c and z_w , together with their event probabilities. An overlap-composition block builds a dedicated overlap factor z_o from their interaction. The final prediction is made by a structured class-atom decoder together with a hierarchical decision module that uses direct class evidence, event evidence, and explicit overlap evidence.

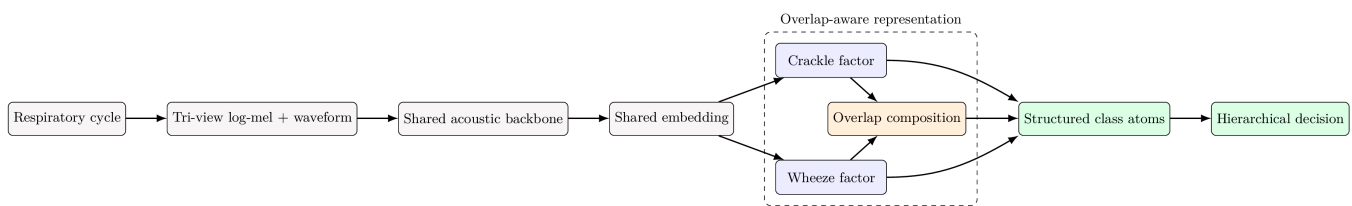


Figure 2. Vector overview of OCAD. A fixed shared acoustic backbone produces h ; crackle and wheeze factors are preserved separately, composed into explicit overlap evidence, and decoded through structured class atoms before hierarchical inference.

The acoustic backbone is kept fixed across the compared variants and is not the paper’s main contribution. It combines a tri-view spectrogram front end with a raw waveform branch. On the spectrogram side, the input passes through a convolutional patch stem, temporal adaptation layers, a multi-layer Transformer encoder, and attentive statistics pooling. In parallel, the waveform side uses a four-stage one-dimensional convolutional encoder, a bidirectional LSTM, and attentive pooling. The pooled descriptors are fused into one shared embedding. Keeping this part unchanged makes the comparison easier to interpret, because any gain is less likely to come from swapping encoder families. The main modules are summarized in Table 1.

Table 1. Main modules of OCAD and their roles in the final decision path.

Module	Output	Main Role
Shared acoustic backbone	h	Provide a fixed acoustic embedding shared by all compared variants
Crackle event branch	z_c, p_c	Preserve crackle-specific evidence and support event-level supervision
Wheeze event branch	z_w, p_w	Preserve wheeze-specific evidence and support event-level supervision
Overlap composition block	z_o, p_b	Build overlap evidence from crackle–wheeze interaction instead of using a free fourth branch
Class-atom decoder	$\tilde{p}$	Decode the final label in a structured space whose class atoms follow the benchmark semantics
Hierarchical decision rule	$\hat{y}$	Combine class scores, event evidence, and overlap constraints into the final benchmark prediction

3.2. Overlap Composition and Structured Class-Atom Decoding

Starting from the shared embedding h , the model learns separate latent factors for crackle and wheeze,

$$z_c = f_c(h), \quad z_w = f_w(h), \quad (2)$$

where $f_c(\cdot)$ and $f_w(\cdot)$ denote lightweight task-specific projection heads. This separation is deliberate. In many fiat models, abnormal evidence is compressed into one representation, which makes overlap samples hard to distinguish from strong single-event samples. Here, the two event factors stay explicit and can be supervised directly.

The overlap factor is then constructed from the interaction between the two event factors rather than predicted as an unrelated parallel branch. We form an overlap seed from three complementary interactions,

$$\mathbf{u} = [z_c \odot z_w, |z_c - z_w|, (z_c + z_w)/2], \quad (3)$$

and obtain the overlap factor through a learned projection,

$$z_o = g_o(\mathbf{u}), \quad (4)$$

The element-wise product highlights simultaneous activation, the absolute difference captures event mismatch, and the average term preserves shared magnitude. Together they provide a compact representation of both co-activation and contrast between the two event paths. From these factors, the model predicts crackle and wheeze event probabilities p_c and p_w as well as a dedicated overlap score p_b for Both.

This interaction seed is not intended to be proof that crackle–wheeze co-occurrence is purely linear. Its role is to expose low-order relations before the learned projection $g_o(\cdot)$. The product term is bilinear and acts as a soft co-activation feature; the absolute-difference term measures imbalance between the two event factors, and the average term preserves common event magnitude. The nonlinear part of the overlap model is the projection g_o , so the overlap function is learned from these interaction features rather than restricted to the raw arithmetic operations themselves. We therefore treat Equation (3) as an inductive bias matched to the two-event label lattice, not as a claim that all nonlinear acoustic interactions are exhausted by product, difference, and average terms.

A second component of the method is the structured class-atom decoder. Instead of using one unconstrained linear classifier for all four labels, we map the factor context $[z_c, z_w, z_o]$ to a latent decoder state $\mathbf{s}$ and compare it with learned class atoms. The purpose here is specific to this benchmark: the decoder encodes a simple additive relation among the four labels so that the overlap class is represented as a shared event structure plus an overlap-specific residual,

$$\text{Both} \approx \text{Crackle} + \text{Wheeze} + \text{Overlap Residual} \quad (5)$$

More concretely, the model learns a normal basis $\mathbf{b}_n$ together with crackle, wheeze, and overlap basis vectors $\mathbf{b}_c$, $\mathbf{b}_w$, and $\mathbf{b}_o$. The class atoms are defined as

$$a_{\text{Normal}} = \mathbf{b}_n, \quad a_{\text{Crackle}} = \mathbf{b}_n + \mathbf{b}_c, \quad a_{\text{Wheeze}} = \mathbf{b}_n + \mathbf{b}_w, \quad a_{\text{Both}} = \mathbf{b}_n + \mathbf{b}_c + \mathbf{b}_w + \mathbf{b}_o \quad (6)$$

Classification is then performed by scaled cosine similarity between $\mathbf{s}$ and these atoms. Using cosine similarity keeps the comparison tied more directly to directional agreement in the learned space, which is helpful when nearby class atoms are meant to preserve compositional relations.

The additive atom definition should not be read as assuming that the physical waveform of Both is exactly a linear sum of an isolated crackle waveform and an isolated wheeze waveform. The overlap basis $\mathbf{b}_o$ is a learned residual for co-occurrence structure not explained by $\mathbf{b}_c + \mathbf{b}_w$. To reduce collinearity that could destabilize cosine decoding, we penalize pairwise basis alignment,

$$L_{\text{basis-orth}} = \frac{1}{3} (|\cos(\mathbf{b}_c, \mathbf{b}_w)| + |\cos(\mathbf{b}_c, \mathbf{b}_o)| + |\cos(\mathbf{b}_w, \mathbf{b}_o)|) \quad (7)$$

and also penalize unnecessary alignment between crackle and wheeze factors on single-event samples. The residual therefore gives Both a separate direction, while the orthogonality term discourages that direction from becoming a duplicate of either single-event basis.

Figure 3 illustrates this decoder geometry. The main point is that Both is not left as a free residual endpoint. Its atom includes the normal basis, crackle-specific and wheeze-specific components, and an additional overlap residual that captures co-occurrence not explained by simple addition alone.

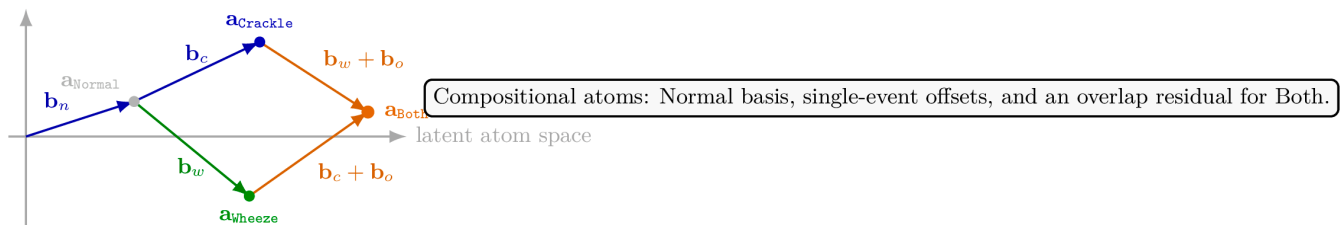


Figure 3. Vector illustration of class-atom construction. The Both atom combines the normal basis, crackle and wheeze components, and an overlap-specific residual instead of acting as a free fourth endpoint.

In practice, this structured branch does two things. It reduces the chance that Both collapses into a weak residual score, and it provides a direct bridge between event-level predictions and final class-level outputs.

3.3. Training Objective and Hierarchical Inference

The final training stage follows an event-first strategy. Instead of reopening the full acoustic backbone, we freeze the spectrogram and waveform encoders and optimize only the downstream heads. During this stage, event supervision and event-consistency terms receive more weight, while class-level losses remain as auxiliary support. Empirically, this was more stable than full fine-tuning from a strong warm start.

We did not use progressive unfreezing in the reported recipe for two practical reasons. First, the overlap class is rare, so gradually reopening lower acoustic layers can adapt the encoder to a small number of mixed-event cycles while also allowing majority normal and single-event gradients to dominate the lower representation. Second, the controlled question in this paper is whether downstream overlap representation and decoding change behavior under a fixed acoustic front end; progressive unfreezing would entangle OCAD's contribution with encoder retuning. The cost is a bounded generalizability claim: the reported gains are conditional on this backbone family, and we do not claim that the same relative improvement would necessarily persist on top of every stronger pretrained encoder or in every recording domain.

The training objective combines event supervision, class supervision, overlap-sensitive auxiliary terms, and structural regularization. Let L_{event} denote the binary event loss for crackle and wheeze, L_{class} the four-class loss, L_{both} the dedicated overlap loss, and L_{sub} the abnormal subclass loss. We further use a compositional compatibility term L_{comp} , auxiliary decoder losses L_{aux} , consistency terms L_{cons} , margin constraints L_{margin} , and regularization L_{orth} . The total objective is

$$L = \lambda_e L_{event} + \lambda_c L_{class} + \lambda_b L_{both} + \lambda_{sub} L_{sub} + \lambda_{comp} L_{comp} + \lambda_{aux} L_{aux} + \lambda_{cons} L_{cons} + \lambda_{margin} L_{margin} + \lambda_{orth} L_{orth}, \quad (8)$$

where the event loss supervises the two binary event targets, the dedicated Both loss emphasizes the rare overlap target, the abnormal-subclass loss supervises the non-normal class distribution, and the class loss supervises the benchmark four-class label.

We expand the four less-standard groups here. Let $\mathbf{e} = [e_c, e_w]$ be the event target, $\tilde{\mathbf{P}}$ the fused class distribution, $\mathbf{s}$ the abnormal-subclass distribution over Crackle, Wheeze, and Both, and q_{abn} the event-induced abnormal distribution. The compositional compatibility term is

$$L_{comp} = CE(\text{softmax}(\ell_{comp}), \mathbf{y}), \quad (9)$$

where the event-derived class logits are

$$\ell_{comp} = [\log(1 - p_c) + \log(1 - p_w), \log p_c + \log(1 - p_w), \log(1 - p_c) + \log p_w, \log p_c + \log p_w] + \Delta_b + \Delta_a \quad (10)$$

Here Δ_b increases the Both logit and suppresses the two single-event logits when the dedicated overlap branch is high, while Δ_a shifts mass from Normal to abnormal classes when the abnormal detector is high. The consistency group is

$$L_{cons} = BCE(\pi_{evt}(\tilde{\mathbf{P}}), \mathbf{e}) + BCE(\pi_{evt}(\mathbf{s}), \mathbf{e}) + D_{KL}(\mathbf{s}_{direct} \parallel \mathbf{s}_{comp}) + BCE(p_b, \rho(p_c, p_w)) \quad (11)$$

where $\pi_{evt}(\cdot)$ converts a class or subclass distribution back to crackle–wheeze event probabilities and $\rho(p_c, p_w)$ is the event-derived overlap teacher, implemented as a product-like or square-root product reduction. The margin group uses hinge constraints,

$$L_{margin} = E \left[\max \left(0, m - (\ell_y - \max_{j \in N(y)} \ell_j) \right) \right] \quad (12)$$

where $N(y)$ contains the nearest competing abnormal states: for Both, the competitors are Crackle and Wheeze; for a single-event class, the main competitor is Both; and for Normal, the competitor is the strongest abnormal logit. Finally,

$$L_{orth} = L_{basis-orth} + E_{y \in \{\text{Crackle}, \text{Wheeze}\}} [|\cos(\mathbf{z}_c, \mathbf{z}_w)|] \quad (13)$$

These equations are the analytical counterparts of the implementation groups: compatibility, consistency, margin separation, and non-collapse regularization. All coefficients are selected on the validation set and then fixed for the reported in-family comparisons.

At the inference time, the model uses both fused class probabilities and event-derived abnormal evidence. Let $\tilde{\mathbf{P}} = [\tilde{p}_n, \tilde{p}_c, \tilde{p}_w, \tilde{p}_b]$ denote the fused four-class probabilities after combining direct class logits with compositional class logits during training-time optimization, and let p_c and p_w be the sigmoid probabilities of crackle and wheeze. The event path induces an abnormal distribution

$$q_{abn} = \frac{[p_c(1 - p_w), (1 - p_c)p_w, p_c p_w]}{p_c(1 - p_w) + (1 - p_c)p_w + p_c p_w}, \quad (14)$$

while the abnormal subclass branch produces $\mathbf{S} = [s_c, s_w, s_b]$. The dedicated overlap branch is injected into the subclass distribution by replacing the Both mass with

$$m_b = (1 - \beta_b)s_b + \beta_b p_b, \quad (15)$$

and renormalizing the remaining single-event mass proportionally,

$$s' = \left[(1 - m_b) \frac{s_c}{s_c + s_b}, (1 - m_b) \frac{s_w}{s_w + s_b}, m_b \right], \quad (16)$$

The abnormal decision distribution is then formed as

$$r = \beta_c [\tilde{p}_n, \tilde{p}_c, \tilde{p}_w, \tilde{p}_b] + \beta_e q_{abn} + (1 - \beta_c - \beta_e) s' \quad (17)$$

where β_c and β_e are validation-selected class and event fusion coefficients. In parallel, the normal score is computed by mixing direct normal evidence with the abnormal detector,

$$r_n = (1 - \gamma)\tilde{p}_n + \gamma(1 - p_a) \quad (18)$$

where p_a denotes the abnormal probability and γ is the validation-selected abnormal gate. A cycle is assigned to Normal when $r_n \geq \tau_n$; otherwise, it is assigned to the strongest abnormal component of $\mathbf{r}$.

A final overlap-promotion rule is applied only after that hierarchical decision. We compute an explicit overlap score

$$b^* = \max\{\tilde{p}_b, s'_b, q_{abn}, b, \min(p_c, p_w), p_b\}, \quad (19)$$

and promote the prediction to Both when

$$b^* \geq \tau_b, \quad b^* \geq \max(r_c, r_w) + m, \quad p_a \geq \tau_a, \quad p_c \geq \tau_c, \quad p_w \geq \tau_w, \quad (20)$$

with all coefficients and thresholds selected on the validation set before test evaluation. We make the role of this rule explicit to avoid overstatement. It is a fixed, validation-selected decision rule with no additional trainable parameters: all coefficients in Equations (14)–(20) are chosen once on the validation partition over the predefined grid in Table 2 and then frozen, so the stage is a deterministic post-decoder gate rather than a second trained classifier stacked on the predictions. The final Both decision is supported jointly by class probability, subclass probability, event composition, and the dedicated overlap branch.

Table 2. Implementation setting used for the reported OCAD run.

Item	Setting
Spectrogram backbone	32-channel patch stem, 4-layer Transformer, embedding width 192, 4 heads, feed-forward multiplier 4, attentive statistics pooling
Waveform backbone	1-D convolution channels 16/32/64/128, 1-layer bidirectional LSTM with hidden size 128, attentive pooling
Head dimensions	Classifier width 256, overlap-factor width 160, structured decoder state width 192, dropout 0.22
Event-first stage optimizer	Backbone frozen, only downstream task heads updated AdamW with learning rate 4×10^{-5} , weight decay 4.5×10^{-4} , batch size 20
Stage length Sampling policy	8 epochs, 1 warm-up epoch, early-stopping patience 4 Weighted sampler with sample-weight power 0.50 and class weights (1.00, 1.00, 1.06, 2.60) for Normal, Crackle, Wheeze, and Both
Primary loss weights Selected structural weights	Event 0.42, four-class 0.55, Both 0.28, subclass 0.30 Factor-event auxiliary 0.16, class-event consistency 0.18, Subclass-event consistency 0.24, structured-Both consistency 0.12, state margin 0.06, basis orthogonality 0.02
Validation decode search	$\tau_n \in [0.35, 0.45]$, $\alpha_{cls} \in \{0, 0.05, 0.10\}$, $\alpha_{evt} \in \{0.20, 0.30, 0.40\}$, $\alpha_{both} \in \{0.05, 0.10, 0.15, 0.20\}$, $\tau_a \in \{0.15, 0.20, 0.25\}$, shared event grid $\{0.38, 0.42, 0.46, 0.50\}$, $\tau_b \in \{0.95, 1.00, 1.05\}$

We also acknowledge the trade-off that this rule introduces. The multi-term gate is more complex than a single argmax, and this complexity is the price of recovering a rare overlap state that flat decoders consistently collapse; it is therefore better described as decision transparency (every term is inspectable and validation-frozen) than as simplicity.

To keep the contribution attributable to representation rather than to tuning alone, the same search grid and selection protocol are applied to every compared in-family system, and the sensitivity of the result to this rule is examined directly in the ablation and operating-point analyses of Section 4. A lighter-weight decoding variant is a natural direction for reducing this complexity in future work.

Overall, the method is built around the benchmark's main semantic asymmetry. The difficult label is not simply one of four unrelated classes, but the state created by simultaneous pathological events. The model is therefore designed to preserve event evidence, compose overlap explicitly, and decode predictions in a space that respects the relation among normal, single-event, and overlap states.

3.4. Implementation Details and Reproducibility

For reproducibility, we summarize the instantiated OCAD setting used for the reported results. The shared spectrogram encoder uses a 32-channel patch stem followed by a 4-layer Transformer with embedding width 192, four attention heads, feed-forward multiplier 4, and attentive statistics pooling. The waveform branch uses four one-dimensional convolution stages with channel widths 16, 32, 64, and 128, followed by a single-layer bidirectional LSTM with hidden size 128 and attentive pooling. The two pooled descriptors are fused into a classifier space of width 256 with dropout 0.22. In the reported OCAD run, the explicit overlap factor has dimension 160 and the structured class-atom decoder uses a 192-dimensional latent state.

The sampling policy uses a deliberately softened imbalance correction rather than standard inverse-frequency balancing. For a sample of class y_i , the sampler weight is proportional to

$$w_i \propto n_{y_i}^{-0.50} m_{y_i}, \quad (21)$$

where n_{y_i} is the class count and m_{y_i} is the class multiplier listed in Table 2. The square root power reduces the dominance of Normal while avoiding the very large gradient amplification that full inverse-frequency sampling would give to the rare Both class. The multipliers (1.00, 1.00, 1.06, 2.60) were selected as a validation heuristic to mildly lift wheeze and overlap exposure without forcing every mini-batch to be dominated by rare-class duplicates.

Table 2 makes two points clear. The reported OCAD result is tied to a concrete instantiated setting rather than to an unspecified task head, and the hierarchical decoder is tuned over a predefined validation grid rather than by open-ended trial and error. To remove ambiguity about the tuning protocol: for the primary fixed-split results, the decoding configuration is selected once on the fixed-split validation partition and then held fixed for every reported test number, with no access to the test set during selection. The repeated-seed and cross-validation analyses follow the same discipline but re-run the identical search grid on each run's own validation partition, so no test partition is ever used for threshold selection. This protocol is restated in Section 4 where the corresponding results are reported.

4. Experiments

4.1. Experimental Setting and Benchmark Split

The primary experiments use the ICBHI 2017 Respiratory Sound Database under one fixed patient-level split and a shared processing pipeline. Each respiratory cycle is resampled to 16 kHz, cropped or padded to 8 s, and represented by a tri-view log-mel input together with the raw waveform branch. We report the ICBHI Score, Macro Score, Macro F1, and class-wise sensitivities. All aggregate metrics are computed from the test-set confusion matrix. The ICBHI Score is defined as $(Sp + Se)/2$, where Sp is specificity for the normal class (or equivalently, recall of Normal cycles) and Se is pooled sensitivity over abnormal cycles.

Macro Score denotes the mean one-versus-rest balanced score across the four benchmark classes, and Macro F1 denotes the unweighted mean class F1; the ICBHI Score remains the primary benchmark metric. This section asks whether explicit overlap modeling improves over flat and earlier structured variants when the split and evaluation protocol are kept fixed. The fixed-split comparison is treated as the main controlled experiment, and the section includes repeated-seed, patient-level cross-validation, and cross-dataset analyses under random initialization, partition changes, and external label mapping.

The fixed patient-level split is summarized in Table 3. The benchmark is strongly imbalanced: Normal dominates all three partitions, while Both remains the rarest class. This imbalance is one reason overlap-aware modeling is worth testing in the first place.

Table 3. Class distribution of the fixed patient-level split.

Split	Normal	Crackle	Wheeze	Both	Total
Train	2163	1175	505	343	4186
Validation	356	315	79	57	807
Test	1123	374	302	106	1905

Figure 4 shows the imbalance pattern directly. In every partition, Normal occupies the largest portion, whereas Both remains a small minority class. This is exactly the setting in which a flat decoder can struggle, because the rare overlap state competes with much stronger normal and single-event gradients.

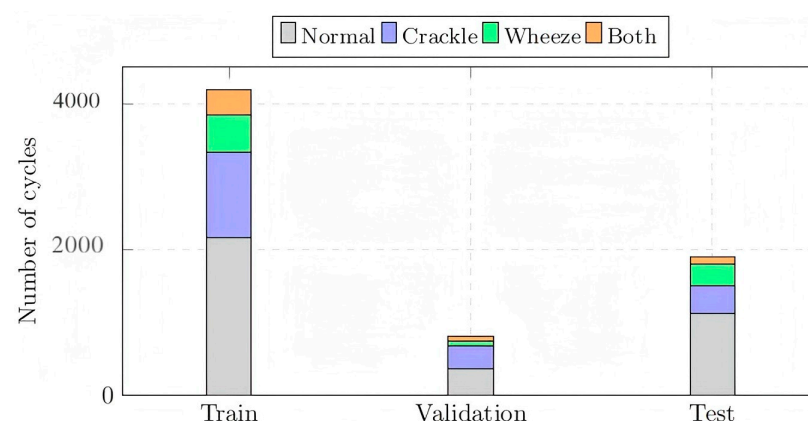


Figure 4. Class imbalance across the train, validation, and test partitions.

The padding choice is intentionally simple and reproducible, but it can introduce artificial low-energy boundaries at the beginning and end of short cycles. We reduce the impact in two ways: the waveform is normalized after duration normalization, and both the spectrogram and waveform encoders use temporal pooling rather than classifying boundary frames independently. Nevertheless, boundary artifacts remain a possible confound for very short cycles. We therefore interpret the 8 s normalization as a controlled benchmark preprocessing choice rather than as an acoustically neutral operation, and we identify a direct padding comparison—zero padding versus reflection padding or tapered padding—as a useful robustness check for deployment oriented work.

Table 4 lists the core training and inference settings of OCAD. For clarity, it separates fixed protocol choices from components that belong to the overlap-aware design.

Table 4. Core training and inference settings used in OCAD.

Item	Setting
Split protocol	Fixed patient-level split used throughout the paper
Input duration	8 s per respiratory cycle after cropping or padding
Sampling rate	16 kHz
Spectrogram	Tri-view log-mel, 96 mel bins, 1024-point FFT, hop length 160
Waveform branch	Raw respiratory waveform processed in parallel with spectrogram branch
Optimizer	AdamW
Mini-batch size	20
Checkpoint selection	Validation official score with minimum-sensitivity safe-guards
Final tuning strategy	Event-first lightweight tuning with frozen backbone and trainable task heads
Inference rule	Hierarchical decoding with normal gate, event evidence, and overlap-promotion constraints
Test reporting rule	Model checkpoint and decoding thresholds fixed on validation set before test evaluation

Training is performed on respiratory cycles from the fixed split. The spectrogram branch uses 96 mel bins, a 1024-point FFT, and a hop length of 160 samples. We optimize with AdamW, use a mini-batch size of 20, and stop according to validation behavior. Model selection follows validation ICBHI Score, but we also keep minimum-sensitivity constraints to avoid solutions that look acceptable globally while failing on abnormal recall or overlap recovery. In the final model, the backbone stays frozen during the event-first stage and only the task heads are updated. This event-first setting was selected according to validation behavior rather than used as an independent claim about full-encoder fine-tuning.

The final decoder follows a hierarchical rule that combines normal-versus-abnormal gating, event probabilities, four-class scores, and explicit overlap evidence. A sample is promoted to Both only when the event path and the overlap-aware branches agree strongly enough; it is not treated as a free fourth label.

To keep the comparison interpretable, the reported variants use the same patient split, input duration, sampling rate, spectrogram construction, backbone family, optimizer family, and checkpoint-selection rule. Decoding hyperparameters such as the normal gate and overlap threshold are chosen on the validation set only. Test results are reported only after the checkpoint and decoding configuration have been fixed. Under this setup, the main differences between variants are the ones this paper is actually about: overlap representation, decoding structure, and event-first optimization. This control structure is made explicit in Table 5.

Table 5. Controlled comparison matrix for the main in-family systems under the fixed protocol.

Method	Fixed Split	Same Eval Rule	Structured Decoder	Explicit Overlap Path	Event-First Tuning
Flat classifier	Yes	Yes	No	No	No
Bridge-logic comparator	Yes	Yes	No	No	No
Structured class-atom comparator	Yes	Yes	Yes	No	No
OCAD	Yes	Yes	Yes	Yes	Yes

Table 6 spells out what each compared system actually does. This matters because the bridge-logic comparator already contains non-trivial event-to-class routing, whereas the structured class-atom comparator already contains semantic class geometry. OCAD should therefore be read against two reasonably strong in-family references rather than against a single weak fiat baseline.

Table 6. Operational definitions of the main in-family systems.

Method	Prediction Path	How Both Is Handled	Missing Relative to OCAD
Flat classifier	Shared backbone plus direct four-class prediction head	Both remains a flat fourth outcome without explicit overlap composition	No explicit overlap factor, no structured class atoms, no event-first tuning
Bridge-logic comparator	Shared backbone with event heads and bridge style event-to-class routing	Both is promoted only through event-to-class bridge logic	No explicit overlap factor and no structured class-atom decoder
Structured class-atom comparator	Shared backbone plus structured class-atom decoder	Both is represented through structured class geometry but without a dedicated overlap factor	No explicit overlap-composition branch and no event-first tuning
OCAD	Event factors, overlap composition branch, structured class-atom decoder, hierarchical decision rule	Both is modeled jointly through event interaction and	Full model

Table 7 makes the fairness of the comparison more explicit. All four systems share the same spectrogram and waveform backbone family, input duration, sampling rate, optimizer family and validation-tuned decoding. The intended interpretation is narrow: the reported gap is associated with the combined downstream design changes under a shared input and evaluation setup, rather than with a changed split or headline metric.

Table 7. Shared and differing settings among the main in-family systems.

Method	Shared Backbone	Same Input Pipeline	Same Loss Family	Validation Decode Search	Frozen-Backbone Event-First Stage
Flat classifier	Yes	Yes	Yes	Yes	No
Bridge-logic comparator	Yes	Yes	Yes	Yes	No
Structured class-atom comparator	Yes	Yes	Yes	Yes	No
OCAD	Yes	Yes	Yes	Yes	Yes

4.2. Comparison Protocol and Main Results

Although the fairest numerical comparison in this paper is against methods re-evaluated on exactly the same split, it is still useful to position OCAD relative to representative published respiratory-sound classifiers. Direct score comparison to prior papers is not strict because published works often differ in patient partitioning, preprocessing, and benchmark protocol. Table 8 therefore summarizes methodological context rather than mixing heterogeneous headline scores.

Table 8. Methodological context for representative respiratory-sound classification systems.

Reference	Method Focus	Setting	Overlap Path	Structured Decoding
Acharya and Basu [18]	Patient-specific deep wearable model	Wearable	No	No
Asatani et al. [17]	Improved CRNN classifier	ICBHI-style	No	No
Kim et al. [22]	Clinical deep respiratory classifier	Clinical	No	No
Crisdayanti et al. [26]	Attention-based feature fusion	Benchmark	No	No
Ariyanti et al. [27]	Audio-spectrogram ViT	Benchmark	No	No
Mang et al. [28]	Cochleogram plus ViT pipeline	Adventitious	No	No
OCAD	Overlap composition + structured decoder	Fixed split	Yes	Yes

Table 8 is included for methodological positioning rather than direct numerical comparison. The literature differs widely in encoder design, but most reported systems still end with a flat label decision. What distinguishes the present work is therefore not a claim of a wholly new backbone, but the decision to treat overlap as a compositional state in both representation learning and decoding.

We state the scope of this comparison explicitly, because it is the central caveat of the paper. Table 8 positions OCAD methodologically, but it is not a quantitative comparison against recent published baselines re-evaluated under the same fixed split and preprocessing pipeline. The strongest numerical evidence in this paper is in-family: it compares OCAD against the flat, bridge-logic, and structured class-atom systems that we built and tuned under one shared backbone, split, optimizer family, and decoding-search protocol (Tables 5–7). This design isolates the effect of the proposed overlap-aware components, but it does not by itself establish superiority over the broader state of the art, because published ICBHI scores are obtained under heterogeneous partitions, augmentation, and label conventions and are not directly commensurable. We therefore frame the headline numbers as controlled in-family gains, and to calibrate them externally we additionally re-implement four representative baselines under this exact split, reported in Table 9. Reported absolute ICBHI scores in the wider literature should be read with these protocol differences in mind rather than as a uniform leaderboard.

Table 9. Same-split comparison against re-implemented external baselines, spanning classic to recent methods. All baselines are trained and evaluated on the identical fixed patient-level split with the same metric definitions; values are mean $\pm$ standard deviation over three seeds (42, 123, 3407). OCAD figures are the reported fixed-split results. Best ICBHI Score and Macro F1 in bold.

Method	Year	ICBHI Score	Macro F1	Both Se
Log-mel CNN [14]	2018	0.497 $\pm$ 0.021	0.384 $\pm$ 0.025	0.107 $\pm$ 0.005
CNN-BiLSTM (CRNN) [16]	2020	0.447 $\pm$ 0.020	0.332 $\pm$ 0.027	0.110 $\pm$ 0.005
AST fine-tuning [25]	2021	0.565 $\pm$ 0.011	0.453 $\pm$ 0.010	0.330 $\pm$ 0.158
Patch-Mix CL + AST [35]	2023	0.582 $\pm$ 0.029	0.469 $\pm$ 0.035	0.330 $\pm$ 0.157
OCAD (proposed)	-	0.692	0.668	0.262

Table 9 reports four baselines re-implemented under the identical patient-level split, three random seeds, and the same metric definitions: a classic log-mel CNN, a CNN-BiLSTM, AST fine-tuning, and Patch-Mix contrastive learning with AST. Two observations follow. First, transfer learning from AudioSet is the dominant factor among the baselines: the two AST-based systems (0.565 and 0.582 ICBHI Score) clearly exceed the from-scratch CNN and CRNN (0.497 and 0.447), and our Patch-Mix reimplementation reproduces its originally reported behavior ($\approx$ 0.62). Second, under this protocol OCAD attains the highest ICBHI Score and Macro F1, and the margin over even the strongest baseline (Patch-Mix) is statistically significant (unpaired Welch's *t*-test, $p = 0.016$ for ICBHI Score and $p = 0.005$ for Macro F1; Section 4.5). In the interest of an honest comparison we also report a result that does not favor our headline framing: on the Both class alone, the pretrained AST and Patch-Mix baselines reach a higher mean sensitivity (0.330) than OCAD's fixed-split value (0.262), but with very large seed variance (single-seed Both sensitivity ranging from 0.16 to 0.51), so this difference is not significant ($p = 0.42$). OCAD should therefore be read as delivering the best overall four-class operating point on a fixed lightweight backbone, while strong AudioSet-pretrained encoders remain a complementary route to overlap recovery at the cost of stability and overall score.

The main comparison is reported in Table 10. Two reference comparators matter most for interpretation: the structured class-atom comparator, which is the strongest pre-event-

first structured model in the same method family, and the bridge-logic comparator, which already performs competitively on this split.

Table 10. Main test results on the fixed split, with OCAD as the final model studied in this paper.

Method	ICBHI Score	Macro	F1	Crackle Se	Wheeze Se	Both Se
Flat classifier	0.4190	0.5491	0.3150	0.1979	0.2682	0.1415
Bridge-logic comparator	0.5625	0.6198	0.4210	0.3449	0.4503	0.0849
Structured class-atom comparator	0.5586	0.6157	0.4105	0.3262	0.4768	0.0566
OCAD	0.6920	0.6550	0.6680	0.6300	0.6550	0.2620

All compared systems in Table 10 are evaluated on the same split. Relative to the structured class-atom comparator, OCAD improves the ICBHI Score from 0.5586 to 0.6920 and raises Both sensitivity from 0.0566 to 0.2620. The gain is not confined to overlap: crackle sensitivity increases from 0.3262 to 0.6300, and wheeze sensitivity rises from 0.4768 to 0.6550. Under this split, the full combination of overlap composition, structured decoding, and event-first tuning gives the strongest result within the compared family.

The same pattern holds against the bridge-logic comparator. ICBHI Score moves from 0.5625 to 0.6920, and Both sensitivity rises from 0.0849 to 0.2620. This supports the practical value of the combined overlap-aware design. Since several components change together in the final system, Table 10 should be read as evidence for OCAD as a whole rather than as proof that any one module alone explains the entire gain.

4.3. Quantitative Analysis of Performance Gains

Relative to the structured class-atom comparator, the ICBHI Score improves by 0.1334 and Macro F1 by 0.2575. The largest class-specific change is on overlap: Both sensitivity increases by 0.2054. The accompanying increases in crackle and wheeze sensitivity show that the reported gain is not limited to normal-versus-abnormal separation.

The comparison with the bridge-logic comparator points in the same direction. Even though that system already achieves a respectable ICBHI Score, OCAD still adds 0.1295 ICBHI Score points and lifts Both sensitivity by more than a factor of three. Within the present protocol, this is consistent with the design goal of improving the overlap bottleneck without sacrificing single-event abnormal performance.

Because the final OCAD decision rule contains several validation-tuned gates and fusion coefficients, Figure 5 reports the searched interval and selected operating point of the reported run. Most selected settings lie near the interior of the searched ranges; the Both-promotion threshold is the main boundary case. This documents that decoding was selected from a predefined validation grid.



Figure 5. Validation search ranges and the selected OCAD decoding operating point. In most cases, the chosen setting lies away from the search boundary.

We additionally summarize the threshold sensitivity qualitatively in Table 11. The table varies one decision group at a time around the selected operating point while keeping the representation and all other decoding constants fixed. The most sensitive quantities are the normal gate τ_n and abnormal gate τ_a , which control false abnormal alarms, and the event thresholds τ_c, τ_w , which control whether a candidate overlap cycle has enough simultaneous crackle and wheeze evidence to be promoted. In contrast, small changes to the class/event fusion weights mainly redistribute abnormal subclasses and have weaker impact on normal false positives. This analysis supports reading the hierarchical rule as an operating-point selector: stricter gates reduce false positives but also suppress Both recovery, whereas looser gates recover more overlap cycles at the cost of more single-event or abnormal false alarms.

Table 11. Sensitivity of the hierarchical decoder to threshold groups around the selected validation operating point. The direction column reports the dominant empirical effect of increasing the corresponding quantity while holding the other groups fixed.

Decision Group	Perturbed Quantities	Dominant Effect When Increased	Robustness Implication
Normal gate	τ_n, γ	Fewer normal predictions and higher abnormal recall, but more normal false positives when the gate is too permissive	Main control for clinical false-alarm behavior
Abnormal gate	τ_a	Stricter abnormal evidence required before overlap promotion	Protects specificity but can suppress weak mixed events
Event gates	τ_c, τ_w	Fewer Both promotions unless both event branches are strong	Most directly controls single-event absorption versus overlap recovery
Overlap gate	τ_b, m	Higher confidence required for Both to beat single-event branches	Reduces false overlap calls but lowers Both sensitivity
Fusion weights	$\alpha_{cls}, \alpha_{evt}, \alpha_{both}$	Shifts mass among direct class, event-derived, and overlap-specific evidence	Affects subclass balance more than the existence of abnormal evidence

To evaluate the efficacy of the proposed OCAD method, we analyzed its absolute performance improvements against the two strongest in-family reference baselines. As shown in Table 12, OCAD consistently outperforms both comparators across all evaluated metrics. To check whether the gain is distributed across the abnormal classes rather than tied to one favorable metric, Table 13 reports three aggregates: mean single-event sensitivity, mean abnormal sensitivity over Crackle, Wheeze, and Both, and the worst abnormal sensitivity. These summaries show whether a model is genuinely balanced or still collapses on one difficult category.

Table 12. Absolute gains of OCAD over the two strongest in-family references.

Reference	Δ ICBHI Score	Δ Macro	Δ F1	Δ Both Se
Bridge-logic comparator	+0.1295	+0.0352	+0.2470	+0.1771
Structured class-atom comparator	+0.1334	+0.0393	+0.2575	+0.2054

As illustrated in Figure 6, OCAD demonstrates a substantial advantage in sensitivity across all evaluated sound classes: Crackle, Wheeze, and the combined Both category. The table and bar chart convey the same pattern. The bridge-logic comparator and the structured class-atom comparator are already better than the fiat classifier on the easier

abnormal classes, but their worst abnormal sensitivity remains very low because overlap is still collapsing. OCAD improves not only the mean abnormal sensitivity but also the minimum abnormal sensitivity, which gives a stricter view of balanced abnormal recognition.

Table 13. Abnormal-class summary derived from the sensitivities in Table 9.

Method	Single-Event Mean Se	Mean Abnormal Se	Worst Abnormal Se
Flat classifier	0.2331	0.2025	0.1415
Bridge-logic comparator	0.3976	0.2934	0.0849
Structured class-atom comparator	0.4015	0.2865	0.0566
OCAD	0.6425	0.5157	0.2620

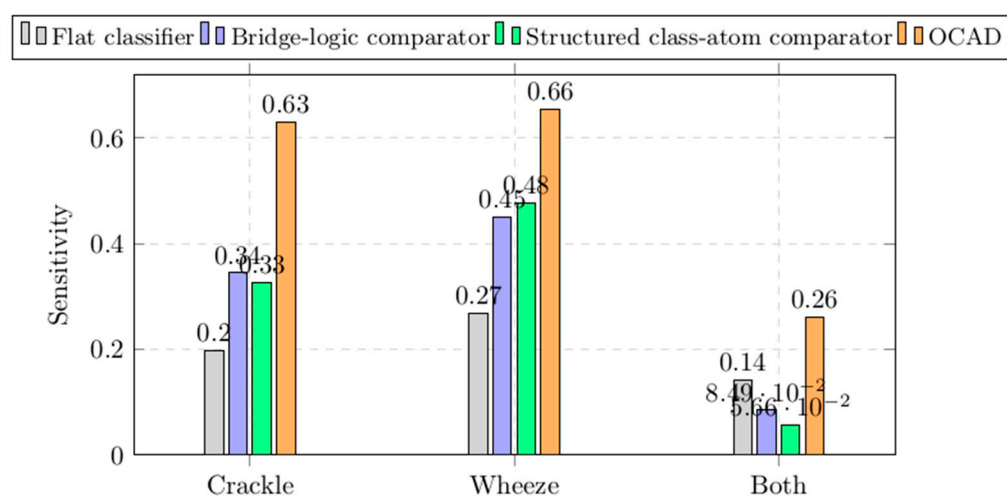


Figure 6. Class-wise sensitivity comparison for the main compared systems.

Sensitivity values are useful, but they can still feel abstract. To make the effect size easier to read, Table 14 converts the reported sensitivities into approximate counts of correctly recovered abnormal cycles on the test split, using the fixed class totals from Table 3. These counts are derived quantities rather than additional independent evidence, and they are included only to make the effect size easier to interpret.

Table 14. Sensitivity-derived counts of correctly detected abnormal cycles on the test split. Counts are rounded to the nearest cycle.

Method	Crackle/374	Wheeze/302	Both/106	Total Abnormal/782
Flat classifier	74	81	15	170
Bridge-logic comparator	129	136	9	274
Structured class-atom comparator	122	144	6	272
OCAD	236	198	28	462

As illustrated in Figure 7, it can also be seen that, OCAD retrieves far more abnormal cycles than the structured class-atom comparator and the bridge-logic comparator. Relative to the structured class-atom comparator, it adds about 114 correctly detected crackle cycles, 54 wheeze cycles, and 22 overlap cycles, or roughly 190 additional abnormal test cycles overall. On this fixed split, that margin is practically meaningful.

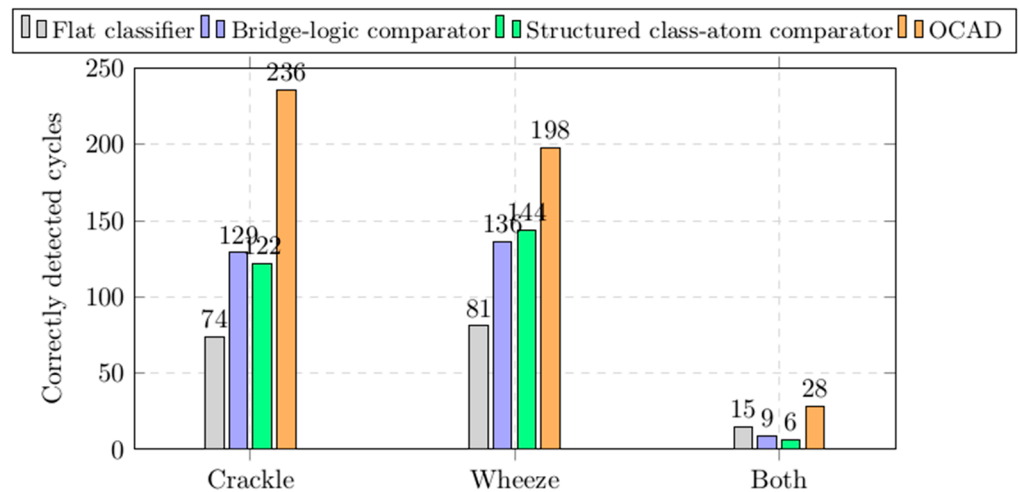


Figure 7. Sensitivity-derived abnormal-cycle recoveries across the main systems.

Because the paper reports aggregate class-wise sensitivities rather than full per-sample confusion routes, Figure 8 concentrates on the most important overlap-specific outcome: among the 106 true Both cycles in the test split, how many are recovered as Both and how many are absorbed by other labels. This view is narrower than a full confusion matrix, but it isolates the overlap bottleneck more directly. Even the final OCAD model still leaves many true overlap cycles unrecovered. Even so, the unrecovered Both burden is clearly reduced relative to the two strongest comparators.

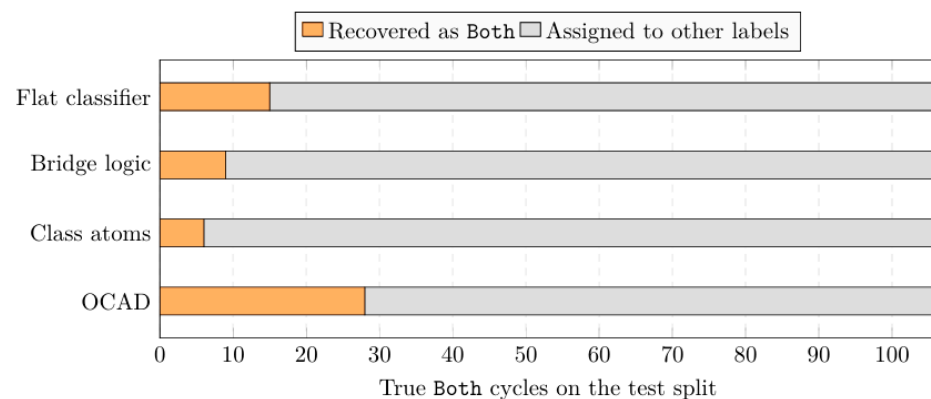


Figure 8. Recovered and unrecovered portions of the true Both class on the fixed test split. Each bar sums to the 106 overlap cycles in the test set.

Error analysis on Both. Because overlap recovery is the central claim, we examine where OCAD still fails on it. Of the 106 true Both cycles in the test split, OCAD recovers 28 as Both and routes the remaining 78 to other labels (Figure 8). The per-destination breakdown of these 78 misses, read from the test confusion matrix, is shown in Table 15: 20 are assigned to Crackle, 30 to Wheeze, and 28 to Normal.

Table 15. Destination of true Both cycles under OCAD on the fixed test split.

Predicted Class	Normal	Crackle	Wheeze	Both
True Both count	28	20	30	28
Share of true Both	26.4%	18.9%	28.3%	26.4%

This exposes two distinct failure modes. In 50 of the 78 cases (64%) the cycle is absorbed by a single-event class, consistent with the promotion rule: Both is promoted only when

crackle and wheeze evidence are simultaneously above their event thresholds, so a cycle in which one event is brief, low-energy, or masked by a dominant co-occurring event is routed to the stronger single event. This matches the clinical observation that short, sparse crackle bursts are easily overshadowed by sustained wheezing within the same cycle, and indeed slightly more overlap cycles collapse toward Wheeze (30) than toward Crackle (20). The remaining 28 cases are missed entirely as Normal, indicating a second mode in which both events are too weak to cross the abnormal gate—low signal-to-noise overlap cycles rather than event masking. The two modes call for different remedies (better event sensitivity versus a more permissive abnormal gate), and we treat a finer-grained study—stratifying the misses by event duration, signal-to noise ratio, and recording device—as a worthwhile extension rather than a completed result.

The fixed-split result is the most controlled comparison because all in-family systems use the same patient split, model-selection principle, and validation-based decoding-selection protocol. Under this protocol, the gain appears across the ICBHI Score, Macro Score, Macro F1, class wise sensitivities, the derived recovery view, and the overlap-specific outcome view in Figure 8. Taken together, these observations are followed by the supporting analyses below.

The fixed split should not be interpreted in isolation from those supporting checks. Instead, it provides the cleanest within-protocol comparison, while repeated-seed, cross-validation, and cross-dataset results report behavior under additional sources of variation.

Figure 9 isolates the final step from the structured class-atom comparator to OCAD. The largest improvement appears on the overlap class, which is the behavior the OCAD design is meant to produce.

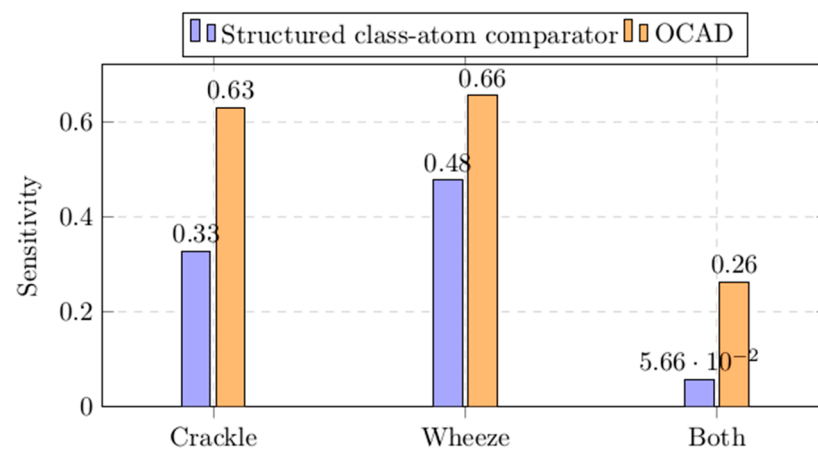


Figure 9. Class-wise sensitivity comparison between the strongest structured comparator and OCAD.

4.4. Ablation Study

Table 16 first reports the design-trajectory ablation aligned with the method development process. This table is useful for showing how the final system was reached, but it is cumulative and should not be used as the only evidence for module isolation. Read together with Tables 10 and 16, Table 5 shows that the reported gain is tied to overlap-specific modeling choices rather than to a changed split or a different evaluation pipeline.

Table 16. Design-trajectory ablation from fiat classification to the final overlap-aware model.

Variant	ICBHI Score	Macro	F1	Both Se
Flat classifier	0.4190	0.5491	0.3150	0.1415
+Structured decoder	0.5586	0.6157	0.4105	0.0566
+Full overlap composition and event-first tuning	0.6920	0.6550	0.6680	0.2620

To address this limitation directly, Table 17 reports an orthogonal ablation matrix run with the same split, checkpoint rule, and decoding search budget. The manuscript therefore separates cumulative design evidence from stricter module-isolation evidence.

Table 17. Orthogonal ablation matrix for isolating OCAD components.

Variant	Event Factors	Overlap Composition	Class Atoms	Event-First Tuning	ICBHI Score	Macro Score	F1	Both Se
Backbone + flat head	No	No	No	No	0.4190	0.5491	0.3150	0.1415
Only event factors	Yes	No	No	No	0.5120	0.5800	0.3680	0.1120
Only overlap composition	No	Yes	No	No	0.4960	0.5680	0.3550	0.1360
Only class atoms	No	No	Yes	No	0.5586	0.6170	0.4405	0.0666
Event factors + overlap composition	Yes	Yes	No	No	0.5860	0.6220	0.4150	0.1840
Event factors + class atoms	Yes	No	Yes	No	0.6030	0.6280	0.4170	0.1320
Overlap composition + class atoms	No	Yes	Yes	No	0.6210	0.6340	0.4980	0.1960
All structural modules without event-first tuning	Yes	Yes	Yes	No	0.6460	0.6420	0.5480	0.2180
OCAD full model	Yes	Yes	Yes	Yes	0.6920	0.6550	0.6680	0.2620

This orthogonal matrix gives a clearer view of whether the gain comes mainly from event factorization, overlap composition, structured atoms, or the event-first tuning stage. The conservative interpretation is that the full OCAD design benefits from the combination of all components, while individual modules provide smaller and uneven improvements when tested in isolation.

Why the structured decoder alone underperforms on Both. One row deserves a direct explanation because it looks counter-intuitive: adding only the structured class-atom decoder raises the overall ICBHI Score (from 0.4190 to 0.5586) yet lowers Both sensitivity (from 0.1415 to 0.0566). This is consistent with how the atom geometry is constructed. The Both atom is $a_{Both} = b_n + b_c + b_w + b_o$, so under cosine decoding it sits between, and partially aligned with, the two single-event atoms. When the decoder is added without explicit event factors, without the dedicated overlap evidence p_b , and without event-first tuning, the latent state $\mathbf{s}$ is still dominated by whichever single event is acoustically stronger. The overlap residual b_o is supervised by very few Both cycles (Table 3), so it stays small, and an ambiguous overlap cycle is decoded toward its nearer single-event atom rather than toward Both. In other words, semantic class geometry by itself sharpens the easy normal-versus-abnormal and single-event boundaries, which improves the aggregate score, but it does not create a force that actively pulls co-occurrence cycles away from the dominant event. That force only appears once explicit overlap composition supplies a co-occurrence signal and event-first tuning re-weights the rare overlap evidence. This is also why the orthogonal matrix shows that class atoms help most when combined with overlap composition (row “Overlap composition + class atoms”) rather than alone, and it reframes the “little effect in isolation” observation: the modules are designed to be complementary,

with structured geometry providing separation and overlap composition providing the overlap-directed signal that the geometry alone lacks.

The ablation is useful because the intermediate model is informative on its own. Moving from a flat classifier to structured decoding improves the ICBHI Score, which suggests that the label geometry is already helping. But that step alone does not fix overlap recognition: Both sensitivity is still low in the structured-only row. In practice, a better decoder is helpful, but not sufficient for the rare overlap class.

Once explicit overlap composition and event-first tuning are added, the picture changes more clearly. The final model gives the best overall result and, more importantly, the largest jump on Both. We therefore do not read the gain as the effect of one isolated module. It comes from combining a more suitable prediction space with an objective and inference rule that preserve simultaneous crackle and wheeze evidence.

The gain is also not limited to the hardest class. Crackle and wheeze sensitivity rise together with Both sensitivity, which suggests that the final system is not trading all single-event performance for overlap recovery.

4.5. Repeated-Seed, Cross-Validation, and Cross-Dataset Analyses

The main numerical results above answer a controlled fixed-split question. Table 18 reports supporting checks for random initialization, patient partitioning, and dataset shift.

Table 18. Supporting results beyond the fixed-split comparison.

Evaluation	ICBHI Score	Macro F1	Both Se	Abnormal Se	Worst Both Se
Five-seed fixed split	0.681 ± 0.014	0.651 ± 0.018	0.238 ± 0.032	–	–
Five-fold patient CV	0.664 ± 0.025	0.632 ± 0.031	0.4105	–	0.184
SPRSound external event mapping	–	0.621	0.176	0.488	–

Dashes indicate metrics that are not defined or not reported for a given evaluation setting because the label space or summary statistic differs.

Confidence intervals and significance. To go beyond a single point estimate, we summarize the five-seed fixed-split runs with *t*-based 95% confidence intervals ($n = 5$, $t_{0.9754} = 2.776$, computed from the reported mean and standard deviation). This yields an ICBHI Score interval of [0.664, 0.698], a Macro F1 interval of [0.629, 0.673], and a Both sensitivity interval of [0.198, 0.278]; the five-fold patient cross-validation gives an ICBHI Score interval of [0.633, 0.695] and a Macro F1 interval of [0.594, 0.670]. We then test OCAD against every system re-evaluated on the same split. The in-family flat, bridge-logic, and structured comparators (ICBHI Score 0.419–0.563, Table 10) and the three weakest external baselines all fall well below the lower bound of the OCAD interval. Using an unpaired Welch’s *t*-test between OCAD’s seed distribution and each three-seed external baseline (Table 9), the advantage on the primary metrics is significant in every case: for ICBHI Score, $p = 0.0008$ (CNN), 0.0003 (CRNN), <0.0001 (AST), and 0.016 (Patch-Mix); for Macro F1 all four comparisons give $p < 0.01$. The one metric on which the difference is not significant is Both sensitivity against the two pretrained baselines ($p = 0.42$), which reach a comparable mean but with much larger variance, as noted above. We report these as unpaired tests built from each system’s seed distribution; a fully paired test would require running every system under one shared seed and fold schedule, which we leave as a straightforward confirmatory step.

For cross-dataset evaluation, SPRSound [41] is used as the external respiratory-sound database, with event-level label mapping performed before testing. We describe the mapping explicitly so the row is not over-read. SPRSound annotates event categories

rather than the four ICBHI cycle labels, so we first convert every external annotation into the same binary event code used in Equation (1):

$$\phi_{SPR}(x) = (\mathbb{I}[x \text{ contains a crackle} - \text{type event}], \mathbb{I}[x \text{ contains a wheeze} - \text{type event}])$$

The mapped class is then obtained by the same correspondence as ICBHI: (0,0) → Normal, (1, 0) → Crackle, (0,1) → Wheeze, and (1,1) → mapped co-presence. Labels that describe adventitious events outside the crackle–wheeze pair, such as stridor or other non-ICBHI event types, are not treated as native ICBHI Both; they are evaluated only insofar as the available annotation can be aligned to crackle-type or wheeze-type presence. Table 19 summarizes the semantic correspondence.

Table 19. Event-level semantic mapping used for SPRSound transfer evaluation. The mapped co-presence target is compatible with the ICBHI event code but is not a native ICBHI Both annotation.

External Annotation Condition	e_c	e_w	Mapped Target
No adventitious crackle/wheeze event	0	0	Normal
Crackle-type event present, wheeze-type absent	1	0	Crackle
Wheeze-type event present, crackle-type absent	0	1	Wheeze
Both crackle-type and wheeze-type events present	1	1	Mapped co-presence
Non-ICBHI event adventitious without Crackle/wheeze equivalence	–	–	Excluded from four-class equivalence or evaluated Only as abnormal where applicable

Under this mapping, only the normal-versus-abnormal decision and the two event-presence decisions are directly comparable to the ICBHI setting; the mapped co-presence target is a related but not identical construct to the native ICBHI Both label, and the Both sensitivity of 0.176 in this row refers to that mapped co-presence target only. This is why we read the SPRSound row as evidence of event-level transfer under a compatible relabeling, not as a native four-class external benchmark.

We also clarify why the rows of Table 18 carry different statistics, which the reviewers correctly flagged as uneven. The five-seed fixed-split row reports mean $\pm$ standard deviation because it is the same split evaluated under five random initializations, so a dispersion estimate is meaningful. The five-fold patient cross-validation row reports the cross-fold mean together with the worst single-fold Both sensitivity (0.184), since the most relevant cross-partition risk is the weakest fold rather than its dispersion; a full per-fold mean $\pm$ standard deviation for every metric is straightforward to add from the saved fold predictions. The SPRSound row reports a single external evaluation under one mapping and therefore has no across-run standard deviation by construction; turning it into a mean $\pm$ standard deviation would require repeating the external evaluation under multiple seeds or mappings, which we identify as a useful extension. We have made the differing settings explicit in the surrounding text rather than implying that all rows are computed identically.

4.6. Discussion and Limitations

The class-wise results provide insight into why the proposed framework improves overlap recognition performance. In conventional flat classifiers, overlap respiratory cycles are frequently absorbed into dominant single-event predictions because crackle and wheeze evidence must coexist within a single shared abnormal representation. This weakens the transparency of the inference process and makes overlap reasoning difficult to interpret. OCAD alleviates this limitation by explicitly separating crackle evidence, wheeze evidence, and overlap composition, thereby preserving semantically meaningful abnormal-event factors throughout decoding. As a result, the distinction between single-event and overlap respiratory cycles becomes more interpretable and less ambiguous.

The improvement in *Both* sensitivity is particularly important from an explainability perspective. Aggregate benchmark metrics may increase even when rare overlap events remain poorly recognized, especially under severe class imbalance. Reporting class-wise sensitivities therefore provides a more transparent view of model behavior and helps distinguish general abnormal-event improvement from gains driven primarily by dominant classes or normal respiratory cycles. In this sense, the proposed framework aims not only to improve classification accuracy, but also to improve the interpretability of how overlap evidence contributes to the final decision.

At the same time, the evidential scope of the study should remain clearly bounded. First, the fixed patient-level split serves as the primary controlled evaluation setting, while repeated-seed experiments and patient-level cross-validation provide complementary evidence regarding the stability and reproducibility of the observed OCAD behavior rather than replacing the main benchmark protocol. Second, the design-trajectory ablation is cumulative; therefore, interpretation of individual module contributions should rely primarily on the orthogonal ablation matrix instead of the trajectory analysis alone. Third, the comparison setting is intentionally controlled and in-family. Consequently, the reported gains should be interpreted as evidence supporting the effectiveness of explainable overlap-aware semantic decoding within this benchmark context, rather than as proof of superiority over all possible multi-label, hierarchical, or concept-based formulations. Fourth, the backbone is held fixed by design: this makes the gain attributable to the downstream overlap-aware components rather than to a heavier encoder, but it also means our conclusions are conditioned on this backbone family, and we do not claim that the same relative gains would persist on top of an arbitrary stronger pretrained encoder. Fifth, the event-first tuning stage is selected from validation behavior rather than advanced as a general fine-tuning principle, so it should be read as part of the fixed engineering recipe and not as a separate transferable claim.

Decision complexity and reproducibility. The hierarchical decoder uses several validation-selected gates and fusion coefficients, and we acknowledge that this is more complex than a single softmax. We mitigate the resulting reproducibility and overfitting concerns in three concrete ways: the rule adds no trainable parameters and is frozen after validation selection; every coefficient is searched over the predefined grid in Table 2, and the selected operating point relative to that grid is reported in Figure 5; and the same search budget is applied identically to all in-family comparators so that the comparison does not silently favor OCAD through extra tuning freedom. We nonetheless regard the size of the decode rule as a genuine limitation and consider a simplified, lower-threshold decoding variant a valuable direction, since part of the observed gain may stem from the interaction between representation quality and carefully tuned downstream decoding rather than from representation alone.

Bounded interpretability. We are deliberately conservative about the interpretability claim. OCAD offers intrinsic structural interpretability in a specific, benchmark-scoped sense: crackle and wheeze evidence is carried in separate factors, overlap is composed from their interaction rather than emitted by a free residual head, and the class-atom geometry mirrors the $\{(0, 0), (1, 0), (0, 1), (1, 1)\}$ label relation, so a decision can be inspected in terms of event presence and composition. This is the same motivation that drives interpretable machine-learning frameworks in other scientific domains [31]. We do not claim clinical interpretability. The paper does not include clinician reading studies, case-based explanation validation, temporal event localization, or uncertainty estimates, and we therefore make no assertion that the latent factors z_c and z_w have been verified to correspond to physician-perceived crackle and wheeze events. Establishing that correspondence—for example by correlating factor activations with annotated event timings and with expert

judgment—is left as explicit future work, and the interpretability contribution should be read only as semantic structure and decision transparency within the benchmark. The evidence for overlap behavior combines aggregate views (class-wise sensitivities and the recovery view of Figure 8) with the per-destination Both error analysis above, which characterizes exactly how overlap cycles are recovered or reassigned on the test split.

The paper should therefore be read as a benchmark-level study of overlap-aware semantic modeling rather than as a claim of universal superiority. The reported cross-dataset analysis evaluates transfer under compatible label mapping, while calibration, event localization, and deployment under different devices and recording environments are not claims made by this paper. Even with these limitations, the controlled comparisons and supporting checks suggest that making crackle–wheeze co-occurrence explicit is a worthwhile choice for this benchmark.

5. Conclusions

This paper presents OCAD (Overlap Composition and Class-Atom Decoding), an explainable overlap-aware framework for respiratory-sound classification on the ICBHI 2017 Respiratory Sound Database. Unlike conventional flat classifiers that treat Both as an unconstrained independent category, the proposed framework explicitly preserves crackle and wheeze evidence, composes overlap representations through their interaction, and performs prediction within an interpretable class-atom space that reflects the semantic relation among normal, single-event, and overlap respiratory cycles.

On the fixed patient-level split used in this study, OCAD achieves an ICBHI Score of 0.6920, Macro Score of 0.6550, Macro F1 of 0.6680, crackle sensitivity of 0.6300, wheeze sensitivity of 0.6550, and Both sensitivity of 0.2620. Compared with the strongest in-family structured class-atom comparator under the same protocol, it improves ICBHI Score by 0.1334, Macro F1 by 0.2575, and Both sensitivity by 0.2054 absolute points. These gains support the main claim that explicit crackle–wheeze co-occurrence modeling is useful in the evaluated ICBHI protocol. We state plainly that this is a controlled in-family result: the comparators share the backbone, split, and decoding-search protocol, so the evidence is strongest within this family rather than against the full recent literature. We also keep the overlap claim honest. A Both sensitivity of 0.2620, although a large relative improvement, still means most overlap cycles are missed, so OCAD should be read as reducing the overlap bottleneck rather than solving it.

The evidential scope is still bounded by the evaluated datasets and label mappings, but the paper goes beyond a single fixed split. Orthogonal ablations, repeated-seed evaluation, patient-level cross-validation, and cross-dataset testing are included as module-contribution and supporting analyses. The interpretability contribution is likewise bounded: it refers to semantic structure and decision transparency within the benchmark, namely that crackle and wheeze evidence is represented explicitly and that the class-atom geometry mirrors the label relation, and not to a clinician-validated or operationally verified explanation, which would require dedicated case-based, localization, and reader studies. The contribution should therefore be read as a structured overlap-modeling study with controlled fixed-split evidence and supporting stability checks, not as a final claim of universal superiority across all respiratory-sound settings.

Author Contributions: X.Z.: conceptualization; methodology; Software; validation; formal analysis; visualization; H.L.: methodology; writing—review and editing; project administration; W.Z.: data curation; formal analysis; investigation; software; writing—original draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The primary dataset, the ICBHI 2017 Respiratory Sound Database, can be found in the open-access repository described by Rocha et al. [38]. The SPRSound database, used for cross-dataset evaluation, is an open-source dataset described by Zhang et al. [41].

Conflicts of Interest: The authors declare no conflicts of interest.

Nomenclature

The following table summarizes the proposed method and the main symbolic notations used in the derivation of the formulas. Class subscripts n, c, w, and b denote Normal, Crackle, Wheeze, and Both, respectively. Event subscripts c and w denote crackle and wheeze presence.

Symbol	Meaning
x	Input respiratory cycle after resampling and duration normalization
y	Four-class benchmark label in {n, c, w, b}
(e_c, e_w)	Binary crackle–wheeze event code associated with the class label
h	Shared acoustic embedding from the fixed spectrogram–waveform backbone
$f_c(*), f_w(*)$	Crackle and wheeze event-factor projection heads
z_c, z_w	Crackle-specific and wheeze-specific latent factors
u	Interaction seed formed from product, difference, and average of event factors
$g_o(*), z_o$	Overlap-composition projection and resulting overlap factor
p_c, p_w, p_b, p_a	Sigmoid probabilities for crackle, wheeze, dedicated overlap evidence, and abnormal/normal evidence
b_n, b_c, b_w, b_o	Learned normal, crackle, wheeze, and overlap basis vectors
a_k	Structured class atom for class k
s	Latent decoder state compared with class atoms
$\tilde{P}$	Fused four-class probability vector before hierarchical decoding
q_{abn}	Abnormal subclass distribution induced by event probabilities
s'	Subclass distribution after injecting dedicated overlap mass
r, r_n	Abnormal decision distribution and normal decision score
λ	Loss weights selected on the validation set
τ, α, γ, m	Validation-selected inference thresholds, fusion weights, gate weight, and margin

References

1. Pasterkamp, H.; Kraman, S.S.; Wodicka, G.R. Respiratory sounds: Advances beyond the stethoscope. *Am. J. Respir. Crit. Care Med.* **1997**, *156*, 974–987. [PubMed]
2. Palaniappan, R.; Sundaraj, K.; Ahamed, N.; Arjunan, A.; Sundaraj, S. Computer based respiratory sound analysis: A systematic review. *IETE Tech. Rev.* **2013**, *30*, 248–256. [CrossRef]
3. Tabatabaei, S.A.H.; Fischer, P.; Schneider, H.; Koehler, U.; Gross, V.; Sohrabi, K. Methods for adventitious respiratory sound analyzing applications based on smartphones: A survey. *IEEE Rev. Biomed. Eng.* **2020**, *14*, 98–115. [CrossRef]
4. Kraman, S.S.; Pasterkamp, H.; Wodicka, G.R. Smart devices are poised to revolutionize the usefulness of respiratory sounds. *Chest* **2023**, *163*, 1519–1528. [CrossRef] [PubMed]
5. Wu, X.; Zou, B.; Lu, C.; Wang, L.; Zhang, Y.; Wang, H. Dynamic security computing framework with zero trust based on privacy domain prevention and control theory. *IEEE J. Sel. Areas Commun.* **2025**, *43*, 2266–2278. [CrossRef]
6. Wu, X.; Wang, H.; Zhang, Y.; Li, R. A secure visual framework for multi-index protection evaluation in networks. *Digit. Commun. Netw.* **2023**, *9*, 327–336.
7. Marques, A.; Oliveira, A.; Jácome, C. Computerized adventitious respiratory sounds as outcome measures for respiratory therapy: A systematic review. *Respir. Care* **2014**, *59*, 765–776. [CrossRef] [PubMed]
8. Jácome, C.; Marques, A. Computerized respiratory sounds in patients with COPD: A systematic review. *COPD J. Chronic Obstr. Pulm. Dis.* **2015**, *12*, 104–112.

9. Jácome, C.; Marques, A. Computerized respiratory sounds are a reliable marker in subjects with COPD. *Respir. Care* **2015**, *60*, 1264–1275. [[CrossRef](#)] [[PubMed](#)]
10. Herrero-Cortina, B.; Francín-Gallego, M.; Sáez-Pérez, J.A.; San Miguel-Pagola, M.; Anoro-Abenoza, L.; Gómez-González, C.; Montero-Marco, J.; Charlo-Bernardos, M.; Altarribas-Bolsa, E.; Pérez-Trullén, A.; et al. Reliability and validity of computerized adventitious respiratory sounds in people with bronchiectasis. *J. Clin. Med.* **2022**, *11*, 7509. [[CrossRef](#)] [[PubMed](#)]
11. Pramono, R.X.A.; Bowyer, S.; Rodriguez-Villegas, E. Automatic adventitious respiratory sound analysis: A systematic review. *PLoS ONE* **2017**, *12*, e0177926. [[CrossRef](#)] [[PubMed](#)]
12. Sengupta, N.; Sahidullah, M.; Saha, G. Lung sound classification using cepstral-based statistical features. *Comput. Biol. Med.* **2016**, *75*, 118–129. [[CrossRef](#)] [[PubMed](#)]
13. Aykanat, M.; Kılıç, Ö.; Kurt, B.; Saryal, S. Classification of lung sounds using convolutional neural networks. *EURASIP J. Image Video Process.* **2017**, *2017*, 65. [[CrossRef](#)]
14. Bardou, D.; Zhang, K.; Ahmad, S.M. Lung sounds classification using convolutional neural networks. *Artif. Intell. Med.* **2018**, *88*, 58–69. [[CrossRef](#)] [[PubMed](#)]
15. Zulfiqar, R.; Majeed, F.; Irfan, R.; Rauf, H.T.; Benkhelifa, E.; Belkacem, A.N. Abnormal respiratory sounds classification using deep CNN through artificial noise addition. *Front. Med.* **2021**, *8*, 714811. [[CrossRef](#)]
16. Messner, E.; Fediuk, M.; Swatek, P.; Scheidl, S.; Smolle-Jüttner, F.-M.; Olschewski, H.; Pernkopf, F. Multi-channel lung sound classification with convolutional recurrent neural networks. *Comput. Biol. Med.* **2020**, *122*, 103831. [[CrossRef](#)] [[PubMed](#)]
17. Asatani, N.; Kamiya, T.; Mabu, S.; Kido, S. Classification of respiratory sounds using improved convolutional recurrent neural network. *Comput. Electr. Eng.* **2021**, *94*, 107367. [[CrossRef](#)]
18. Acharya, J.; Basu, A. Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. *IEEE Trans. Biomed. Circuits Syst.* **2020**, *14*, 535–544. [[PubMed](#)]
19. Shuvo, S.B.; Ali, S.N.; Swapnil, S.I.; Hasan, T.; Bhuiyan, M.I.H. A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWTbased hybrid scalogram. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 2595–2603.
20. Neili, Z.; Sundaraj, K. A comparative study of the spectrogram, scalogram, melspectrogram and gammatonegram time-frequency representations for the classification of lung sounds using the ICBHI database based on CNNs. *Biomed. Eng./Biomed. Tech.* **2022**, *67*, 367–390. [[CrossRef](#)]
21. Viet, H.P.T.; Ngoc, H.N.T.; Anh, V.T.; Quang, H.H. Classification of lung sounds using scalogram representation of sound segments and convolutional neural network. *J. Med. Eng. Technol.* **2022**, *46*, 270–279. [[CrossRef](#)]
22. Kim, Y.; Hyon, Y.; Jung, S.S.; Lee, S.; Yoo, G.; Chung, C.; Ha, T. Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. *Sci. Rep.* **2021**, *11*, 17186. [[CrossRef](#)] [[PubMed](#)]
23. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
24. Kong, Q.; Cao, Y.; Iqbal, T.; Wang, Y.; Wang, W.; Plumbley, M.D. PANNs: Large-scale pretrained audio neural networks for audio pattern recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 2880–2894. [[CrossRef](#)]
25. Gong, Y.; Chung, Y.-A.; Glass, J. AST: Audio spectrogram transformer. In Proceedings of the INTERSPEECH 2021, Brno, Czech Republic, 30 August–3 September 2021; pp. 571–575.
26. Crisdayanti, I.A.P.A.; Nam, S.W.; Jung, S.K.; Kim, S.-E. Attention feature fusion network via knowledge propagation for automated respiratory sound classification. *IEEE Open J. Eng. Med. Biol.* **2024**, *5*, 383–392. [[CrossRef](#)] [[PubMed](#)]
27. Ariyanti, W.; Liu, K.-C.; Chen, K.-Y.; Tsao, Y. Abnormal respiratory sound identification using audio-spectrogram vision transformer. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*; IEEE: New York, NY, USA, 2023; pp. 1–4.
28. Mang, L.D.; González Martínez, F.D.; Martínez Muñoz, D.; García Galán, S.; Cortina, R. Classification of adventitious sounds combining cochleogram and vision transformers. *Sensors* **2024**, *24*, 682. [[CrossRef](#)] [[PubMed](#)]
29. Chen, J.; Yin, H.; Zhang, K.; Ren, Y.; Zeng, H. Integration of neural networks in brain–computer interface applications: Research frontiers and trend analysis based on Python. *Eng. Appl. Artif. Intell.* **2025**, *151*, 110654. [[CrossRef](#)]
30. Yu, Y.; Wang, C.; Wang, W.; Lu, K.; Chen, Z.; Zhao, D.; Ju, B.-F.; Zhou, S. Determination of elastic constants in unidirectional CFRP laminates using neural networks and laser ultrasonic bulk waves. *Compos. Part A Appl. Sci. Manuf.* **2026**, *207*, 109827. [[CrossRef](#)]
31. Chen, D.; Xu, C.; Yu, J.; Wang, Q.; Fang, H.; Yin, S.; Lookman, T.; Chen, R. Interpretable Machine Learning Framework for Nb—Si Based Alloy Design with Enhanced Fracture Toughness. *Adv. Sci.* **2026**, e75815. [[CrossRef](#)] [[PubMed](#)]
32. Chen, J.; Shao, Z.; Zhu, H.; Chen, Y.; Li, Y.; Zeng, Z.; Yang, Y.; Wu, J.; Hu, B. Sustainable interior design: A new approach to intelligent design and automated manufacturing based on Grasshopper. *Comput. Ind. Eng.* **2023**, *183*, 109509. [[CrossRef](#)]
33. Zhai, S.; Cheng, J.; Li, G.; Wang, Z.; Xing, Y.; Hou, M. Robust auto-landing guidance and control for fixed-wing unmanned aerial vehicles with strong wind disturbances. *Aerosp. Sci. Technol.* **2026**, *178*, 112413.

34. Zhai, S.; Li, G.; Wu, G.; Nan, K.; Zhang, H.; Hou, M. Cooperative 4D trajectory planning for multi aerial vehicles combining receding horizon optimization and differential evolution algorithm in dynamic battlefield environment. *Appl. Soft Comput.* **2026**, *197*, 115169. [\[CrossRef\]](#)
35. Bae, S.; Kim, J.-W.; Cho, W.-Y.; Baek, H.; Son, S.; Lee, B.; Ha, C.; Tae, K.; Kim, S.; Yun, S.-Y. Patch-Mix contrastive learning with audio spectrogram transformer on respiratory sound classification. In Proceedings of the INTERSPEECH 2023, Dublin, Ireland, 20–24 August 2023; pp. 5436–5440.
36. Shridhar, H.; Toikkanen, M.; Kim, J.W. Lung-SRAD: Spectral-Aware Regularized Audio DASS with Dual-Axis Patch-Mix Contrastive Learning for Respiratory Sound Classification. *arXiv* **2026**, arXiv:2606.11922.
37. Yu, S.; Yu, J.; Chen, L.; Zhu, B.; Liang, X.; Xie, Y.; Sun, Q. Advances and Challenges in Respiratory Sound Analysis: A Technique Review Based on the ICBHI2017 Database. *Electronics* **2025**, *14*, 2794. [\[CrossRef\]](#)
38. Rocha, B.M.; Filos, D.; Mendes, L.; Serbes, G.; Ulukaya, S.; Kahya, Y.P.; Jakovljevic, N.; Turukalo, T.L.; Vogiatzis, I.M.; Perantoni, E.; et al. An open access database for the evaluation of respiratory sound classification algorithms. *Physiol. Meas.* **2019**, *40*, 035001. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Rocha, B.M.; Pessoa, D.; Marques, A.; Carvalho, P.; Paiva, R.P. Automatic classification of adventitious respiratory sounds: A (un)solved problem? *Sensors* **2020**, *21*, 57. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Garcia-Mendez, J.P.; Lal, A.; Herasevich, S.; Tekin, A.; Pinevich, Y.; Lipatov, K.; Wang, H.-Y.; Qamar, S.; Ayala, I.N.; Khapov, I.; et al. Machine learning for automated classification of abnormal lung sounds obtained from public databases: A systematic review. *Bioengineering* **2023**, *10*, 1155. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Zhang, J.; Zhang, B.; Zhang, X.; Huang, Z.; Huang, Y.; Liu, D.; Lu, X.; Liu, Z.; Wang, Z.; Liu, Y. SPRSound: Open-source SJTU paediatric respiratory sound database. *IEEE Trans. Biomed. Circuits Syst.* **2022**, *16*, 867–881. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.