WIPSNet: Deep Learning for Paediatric Wheeze Detection from Overnight Impedance Pneumography

Anonymous Authors[1]

Abstract

Overnight impedance pneumography (IP) is used to monitor paediatric respiratory health. Its current clinical readout, the Expiratory Variability Index (EVI), compresses each IP recording into a single scalar and achieves an AUC of 0.63 for night-level wheeze classification. We introduce Wheeze Impedance Pneumography Scalogram Network (**WIPSNet**[2]), a 3D SE-ResNet operating on stacked continuous wavelet transform scalograms of overnight IP signals. On a 15-patient cohort (60 nights, 281 hours), WIPSNet achieves an AUC of 0.873 ± 0.019, outperforming EVI, a state-space model (Mamba), and two modern sleep-staging architectures. Performance increases with volumetric depth up to 32 minutes of temporal context, suggesting that multi-scale temporal aggregation is important for modelling nocturnal respiratory dynamics. Overall, these results indicate that structured time–frequency representations combined with 3D convolutional architectures provide an effective approach for learning from long, irregular physiological time series.

1. Introduction

Asthma is the most common chronic lung disease in childhood. Later asthma risk is associated with acute wheezing episodes in preschool-age children (Stern et al., 2020; Saglani et al., 2007). These wheezing episodes result from airway narrowing and bronchial obstruction, but are often difficult to recognise at home, delaying treatment and increasing morbidity. This motivates the need for remote paediatric wheeze monitoring tools (Global Initiative for Asthma, 2025; Makrinioti et al., 2024).

AI-powered digital stethoscopes can detect wheeze in controlled settings (Cheng et al., 2022; Habukawa et al., 2020), but require active cooperation from the child, making them unsuitable for continuous overnight monitoring. Impedance pneumography (IP) is a viable alternative that measures thoracic electrical impedance (Seppä et al., 2016). The Ventica® system is a commercially deployed device that summarises an overnight IP recording into a single scalar, Expiratory Variability Index (EVI). However, EVI's clinical utility is limited by high inter-individual variability (Seppä et al., 2020a; Burman et al., 2021; Seppä et al., 2020b). A recent systematic review (Hammod et al., 2026) concludes that the EVI is better suited for longitudinal monitoring of airway function than for wheeze diagnosis. This motivates the development of novel methods for wheeze diagnosis that exploit richer information contained in the IP waveform. In addition, overnight IP recordings are irregular due to motion artefacts and sensor displacement, resulting in variable-length and partially corrupted time series.

Deep learning can exploit the full IP waveform, but the extreme length ($\sim 10^7$ samples) and irregularity of overnight IP recordings challenge conventional architectures. Recurrent neural networks suffer from vanishing gradients (Pascanu et al., 2013), transformers scale quadratically (Vaswani et al., 2017), and state-space models such as Mamba (Gu & Dao, 2024) have limited long-context capabilities (Waleffe et al., 2024; Ben-Kish et al., 2025; Ye et al., 2025). In contrast, long physiological signals favour convolutional backbones operating on time-frequency inputs (Li et al., 2022; Ullah et al., 2020; Lekkas et al., 2025). These approaches can be viewed as learning structured representations that capture local spectral patterns and their evolution over longer time scales. Additionally, modern sleep-staging architectures, such as AttnSleep (Eldele et al., 2021) and SleePyCo (Lee et al., 2024), provide strong baselines for our task as they are designed for similar long, continuous physiological signals recorded during sleep.

We introduce **WIPSNet**, a deep learning pipeline for paediatric nocturnal wheeze classification from overnight IP recordings. Building on a 3D SE-ResNet over stacked continuous wavelet transform (CWT) scalograms, WIPSNet exploits the full IP waveform through a modular design that accommodates variable-length and partially corrupted

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

[2]Code: https://anonymous.4open.science/r/WIPSNet-SD4H/

recordings. We benchmark against EVI, AttnSleep, SleePyCo, and Mamba, achieving an AUC improvement over EVI ($+0.24$) and Mamba ($+0.148$). Our results suggest that explicitly structured time–frequency representations combined with convolutional inductive biases are particularly effective for modelling long, oscillatory physiological signals, and suggest that SSMs may struggle to fully exploit multi-scale structure in this setting.

2. Methods

2.1. Data and Labels

Overnight IP recordings were collected using the Ventica® system from 15 preschool children (mean age 3.3 ± 1.3 years; 60 nights; 281.3 hours; 250 Hz) in two London (UK) hospitals. Each night was labelled either wheezy or non-wheezy depending on whether WheezeScan®, a clinically validated acoustic device (Habukawa et al., 2020; OMRON Healthcare, 2025), detected wheeze or not, prior to sleep or upon waking. The cohort contains 21 wheezy and 39 non-wheezy nights.

2.2. WIPSNet

WIPSNet (Figure 1) takes a 32-minute segment of overnight IP signal at 250 Hz as input and outputs a wheeze probability for that segment. The pipeline processes raw IP signal through high-pass filters, transforms it into consecutive time-frequency scalograms, stacks them into 3D volumes, and processes them with a 3D CNN to predict wheeze. Night-level predictions are obtained by mean-pooling segment-level outputs, enabling variable-length recordings.

Each overnight recording is high-pass filtered (4th-order Butterworth, 0.2 Hz) to remove baseline drift, then split into non-overlapping 4-minute windows (60,000 samples) (Figure 1b–c). Each window is Z-score normalised and transformed via the continuous wavelet transform (CWT) into a 224×224 scalogram using a complex Morlet wavelet (Figure 1d). Eight consecutive scalograms are then stacked to form a 3D volume representing $8 \times 4 = 32$ minutes of continuous IP signal (Figure 1e). The volume depth, $D{=}8$, was selected via hyperparameter optimisation.

This design explicitly encodes a two-timescale structure in the data: short-term spectral content within scalograms and longer-term evolution across stacked windows. Within each scalogram, the 2D time-frequency representation captures breath-by-breath spectral content over a 4-minute window. The stacked scalograms encode how the spectral content evolves over 32 minutes, capturing the onset, persistence, and resolution of respiratory events. A 3D convolution over the 32-minute scalogram volume integrates these two timescales into a unified representation of respiratory dynamics.

Each scalogram volume is processed by a 3D SE-ResNet (Bottleneck3D, $[2, 2, 2, 2]$ layers, expansion 4) (He et al., 2016; Hu et al., 2018) (Figure 1f). Squeeze-and-Excitation (SE) modules selectively emphasise feature channels that contain relevant respiratory patterns at each resolution. This is particularly suited to scalograms, where wheezing episodes appear as narrow, localised frequency bands, allowing suppression of irrelevant background structure. A global-average-pooling layer followed by a linear head produces a wheeze logit per volume. All hyperparameters were tuned via Optuna (Akiba et al., 2019) (50 trials).

2.3. Baselines for benchmarking

State-space models: Mamba (Gu & Dao, 2024) was trained directly on raw 1D IP under the same Optuna HPO budget used for WIPSNet. The best configuration used 20-second segments (5,000 samples). To test context scaling explicitly, we trained *Mamba-Long* on 32-minute segments matching WIPSNet context length. To isolate whether the CWT representation alone suffices, *Mamba-Scalogram* feeds Mamba WIPSNet's scalogram volumes via ViT-style patch tokenisation (16×16 patches per frame, 1,152 tokens per volume).

Modern single-channel sleep-staging architectures: AttnSleep (Eldele et al., 2021) is a multi-resolution 1D CNN with adaptive feature recalibration and causal self-attention. SleePyCo (Lee et al., 2024) is a multi-scale feature pyramid and Transformer classifier. Both operate on raw single-channel 1D physiological signal at 100 Hz with epoch-based aggregation, extract multi-scale temporal features without EEG-specific priors (no band filters), and target the same regime of ultra-long overnight recordings with subject-wise cross-validation. We use published default configurations, replacing only the final layer with a binary classification head. Epoch probabilities are mean-pooled into a night-level prediction. Each 4-minute IP segment is resampled from 250 Hz to 100 Hz and split into eight 30-second sub-epochs. This matches the original architectures' expected input dimensions, ensuring that their default convolutional receptive fields accurately capture the intended physical time scales.

Non-deep-learning baselines: We considered EVI as the standard clinical baseline and XGBoost as a classical ML baseline. XGboost was trained on 44 handcrafted features extracted from the scalograms, including per-band mean energy, temporal variance, and spectral centroids.

Representation ablations: We performed representation and context ablations to validate WIPSNet's specific design choices. We replaced the CWT front-end with a mel spectrogram, short-time Fourier transform, and wavelet scattering, keeping the 3D SE-ResNet backbone identical. We examined the volumetric depth $D \in \{1, 2, 4, 8, 12, 16, 24, 32\}$ for the CWT variant to quantify the impact of temporal context (ranging from 4 to 128 minutes of continuous signal).

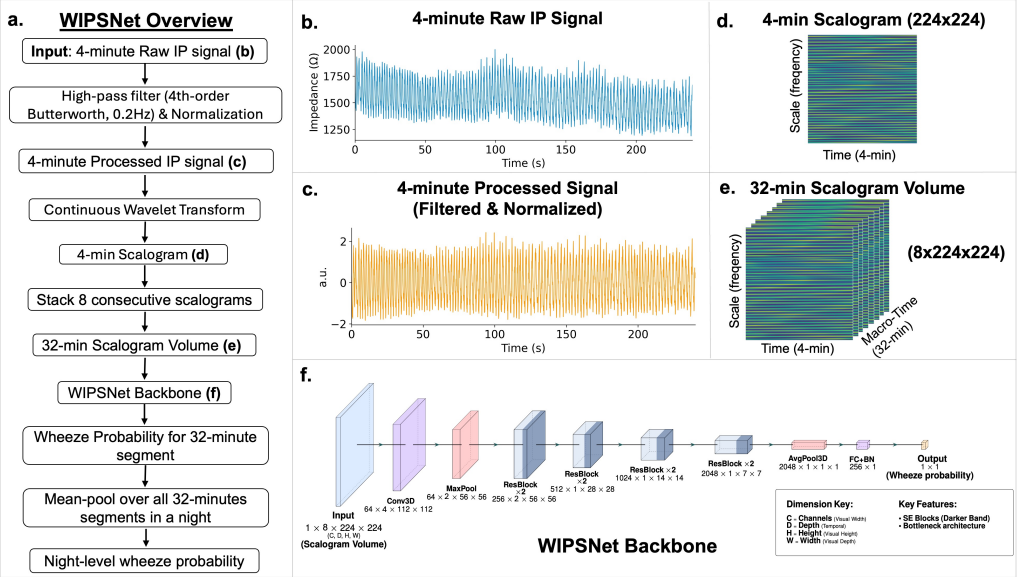

Figure 1. WIPSNet framework and architecture. (a) The IP signal is filtered, transformed into scalograms, grouped into 3D volumes, and passed through a 3D CNN; volume-level probabilities are aggregated into a night-level prediction. (b–d) Raw signal, processed signal, and 4-minute scalogram. (e) Volumetric input construction: D consecutive scalograms stacked along the depth axis. (f) The bottleneck 3D SE-ResNet backbone.

2.4. Evaluation

All models and baselines were evaluated using the patient-level leave-one-out cross-validation (LOOCV, 15 folds) across 5 random seeds, using night-level aggregation. While the cohort is small, LOOCV maximises training data and results are averaged across 5 seeds to ensure robustness. We report two threshold-independent metrics: Area Under the Receiver Operating Characteristic (AUC) and Area Under the Precision-Recall Curve (AUPRC). AUC captures global class separability, while AUPRC focuses on performance on the positive class, which is clinically meaningful since missing a wheezing episode is a missed opportunity to treat a patient. A random classifier achieves an AUPRC of 0.35 for our cohort prevalence of 35% positive nights (21/60).

3. Results

WIPSNet achieves an AUC of 0.873 ± 0.019 and an AUPRC of 0.758 ± 0.075 (Table 1). The performance improvement of WIPSNet over EVI ($+0.24$ AUC, $+0.325$ AUPRC) suggests that overnight IP recording contains discriminative information that the EVI discards. EVI was not computable on 3/60 nights because the sleep duration did not meet the algorithm's minimum signal-length requirement. XGBoost (0.595 AUC) indicates that handcrafted spectral features, without temporal context, are insufficient. In contrast, AUC of the three architectures specialised for continuous physiological signals (WIPSNet, AttnSleep and SleePyCo) cluster at 0.84–0.87, while that of Mamba trails at 0.725.

Table 1. Night-level wheeze classification results. Mean ± std over 5 seeds, 15-fold patient-level LOOCV. AUPRC random baseline at our class prevalence is 0.35. †Deterministic; 57/60 valid nights.

Model	Input & backbone	AUC	AUPRC
WIPSNet	CWT + 3D CNN	$\mathbf{.873}_{\pm.019}$	$\mathbf{.758}_{\pm.075}$
AttnSleep	Raw + MRCNN-attention	$.859_{\pm.025}$	$.713_{\pm.026}$
SleePyCo	Raw + pyramid-Transformer	$.842_{\pm.016}$	$.659_{\pm.013}$
Mamba	Raw + SSM	$.725_{\pm.014}$	$.560_{\pm.017}$
EVI†	Clinical scalar	.633	.433
XGBoost	Handcrafted features	$.595_{\pm.029}$	$.461_{\pm.032}$

WIPSNet's AUC advantage within this top-performing cluster is modest, but its AUPRC lead widens significantly. The AUPRC gaps between WIPSNet and the two raw-signal baselines (AttnSleep and SleePyCo) are wider than the corresponding AUC gaps. This suggests WIPSNet's superiority in handling the imbalanced class, providing more reliable predictions on minority wheeze-positive nights.

WIPSNet processes an overnight recording in ~ 0.34 s on a consumer GPU (NVIDIA RTX 4070), providing more than enough throughput for the cloud-based retrospective analysis of longitudinal wearable data.

3.1. Volumetric Depth as a Driver of Performance

Volumetric depth D controls the 3D CNN's temporal context, as D stacked 4-minute scalograms span $4D$ minutes. Sweeping $D \in \{1, 2, 4, 8, 12, 16, 24, 32\}$ (Figure 2), we find AUC climbs from 0.843 at D=1 (4 min) to peak at

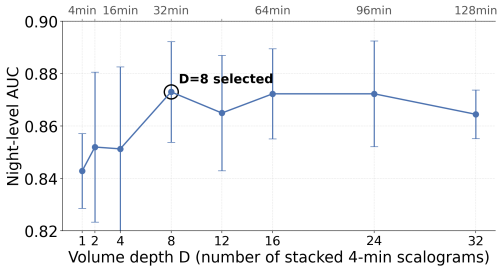

Figure 2. WIPSNet's depth-scaling behaviour. AUC improves with volumetric depth up to 32 minutes of temporal context and plateaus thereafter. Error bars: standard deviation over 5 seeds. Top axis: temporal context in minutes.

0.873 for D=8 (32 min), before plateauing. This scaling supports WIPSNet's two-timescale design: 2D scalograms capture breath-level spectra, while volumetric depth tracks their evolution over tens of minutes to mirror nocturnal oscillatory events. Fixed-sequence baselines (AttnSleep, SleePyCo) lack an explicit mechanism for such hierarchical temporal aggregation, which may limit their ability to model extended physiological dynamics. This volumetric advantage is robust to the front-end transform as STFT and mel spectrograms maintained strong performance alongside the 3D backbone (Appendix B). The volumetric architecture's macro-scale temporal modelling drives performance given adequate spectral information.

3.2. Ablating Context and Representation in Mamba

We conduct two controlled experiments (Appendix B) to test whether Mamba's underperformance is due to limited context or input representation. First, to test whether Mamba simply lacked sufficient context, *Mamba-Long* concatenates 8 consecutive 4-minute segments into a 32-minute training window to match WIPSNet's temporal field of view, but its AUC collapses to 0.452 ± 0.076. This observation is consistent with Mamba's documented bounded effective receptive field (Waleffe et al., 2024). Second, *Mamba-Scalogram* feeds Mamba the CWT volumes via patch tokenisation but only reaches AUC 0.620 ± 0.083. This converts each 2D time–frequency scalogram into a sequence of patches, removing local time–frequency structure that convolutional models exploit. Together, these results show that neither longer context nor improved representation alone closes the gap to convolutional baselines.

4. Discussion

4.1. Clinical Impact

WIPSNet is the first deep learning pipeline, to the best of our knowledge, for paediatric nocturnal wheeze classification from overnight IP recording. Its AUC improvement over EVI, a scalar metric with poor individual-level utility (Hammod et al., 2026), is large enough ($+0.24$) to potentially make WIPSNet a clinical decision-support tool. The sliding aggregation of 4-minute volumes makes WIPSNet independent of recording length, a practically important property since sleep duration varies night to night. Inference takes ~ 0.34 s per night recording, making WIPSNet feasible for real-world deployment. Reliable nocturnal wheeze detection, as an objective marker of airway obstruction, could enable earlier recognition of exacerbations, inform treatment decisions from longitudinal data rather than subjective parental report, and contribute to better management of acute wheeze exacerbations in early childhood.

4.2. ML Impact

Across architectures, three models (WIPSNet, AttnSleep, SleePyCo) cluster at AUC 0.84–0.87, while Mamba achieves 0.725 despite comparable tuning effort. Together with the ablation studies in Section 3.2, this is consistent with known challenges in state-space models for long-sequence generalization (Waleffe et al., 2024; Ben-Kish et al., 2025; Ye et al., 2025). Extending the input context (*Mamba-Long*) leads to a performance drop, while transforming inputs into structured time-frequency tokens (*Mamba-Scalogram*) does not recover performance, suggesting limitations in capturing multi-scale structure in this setting. These findings indicate that convolutional architectures remain a strong baseline for long, irregular, oscillatory physiological signals, as they preserve local patterns while aggregating them hierarchically across time. In contrast, raw-sequence state-space models appear to struggle to integrate such multi-scale structure over extended durations. WIPSNet's gains over AttnSleep and SleePyCo are smaller in AUC but more pronounced in AUPRC, and its consistent scaling with volumetric depth up to 32 minutes suggests that explicit hierarchical temporal aggregation is beneficial when relevant dynamics unfold over longer time scales.

4.3. Limitations and Future Work

While the dataset comprises over 280 hours of high-frequency recording (∼250 million samples), it remains small at the cohort level (15 children, 60 nights). External validation in larger and more diverse populations is required. Furthermore, labels derived from WheezeScan® at sleep boundaries may miss transient intra-night wheezing and are subject to measurement noise from the reference device. Addressing these limitations with larger cohorts and finer-grained labels would enable modelling of intra-night dynamics, including prediction of wheeze onset. Integrating these predictions with clinical data such as medication intake could further support personalised treatment strategies. Extending the framework to longitudinal multi-night modelling, together with improved wearable acquisition, would advance objective respiratory monitoring in home settings.

Impact Statement

This work develops a deep learning approach for nocturnal wheeze detection from overnight impedance pneumography, with potential to support earlier identification of respiratory exacerbations and improve home-based monitoring in young children. Such tools could contribute to more objective and timely management of wheezing disorders. However, the model is trained on a small cohort and may not generalise without further validation. Misclassifications could lead to missed events or unnecessary concern, so the system should be used as decision support rather than a replacement for clinical judgment. Continuous physiological monitoring also raises data privacy considerations, particularly in paediatric settings. Further validation, fairness assessment, and careful integration into clinical workflows are required before deployment.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631, 2019.

Ben-Kish, A., Zimerman, I., Abu-Hussein, S., Cohen, N., Globerson, A., Wolf, L., and Giryes, R. Decimamba: Exploring the length extrapolation potential of mamba, 2025. URL https://arxiv.org/abs/2406.14528.

Burman, J. et al. Observational study of inhaled corticosteroid treatment for improved expiratory variability index in steroid-naïve asthmatic children. *ERJ Open Research*, 8(1), 2021.

Cheng, Z. R., Zhang, H., Thomas, B., Tan, Y. H., Teoh, O. H., and Pugalenthi, A. Assessing the accuracy of artificial intelligence enabled acoustic analytic technology on breath sounds in children. *Journal of Medical Engineering & Technology*, 46(1):78–84, 2022.

Eldele, E., Chen, Z., Liu, C., Wu, M., Kwoh, C.-K., Li, X., and Guan, C. An attention-based deep learning approach for sleep stage classification with single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:809–818, 2021. doi: 10.1109/TNSRE.2021.3076234.

Global Initiative for Asthma. GINA strategy report 2025, 2025. URL https://ginasthma.org.

Gu, A. and Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2024.

Habukawa, C. et al. A wheeze recognition algorithm for practical implementation in children. *PLOS ONE*, 15(10): e0240048, 2020.

Hammod, S.-R., Kullberg, F., Hauerslev, M., Hansen, K. S., and Chawes, B. Impedance pneumography for diagnosing and monitoring asthma in young children: A systematic review. *Children*, 13(2), 2026. ISSN 2227-9067. doi: 10.3390/children13020193. URL https://www.mdpi.com/2227-9067/13/2/193.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. Squeeze-and-excitation networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.

Lee, S., Yu, Y., Back, S., Seo, H., and Lee, K. Sleepyco: Automatic sleep scoring with feature pyramid and contrastive learning. *Expert Systems with Applications*, 240:122551, April 2024. ISSN 0957-4174. doi: 10.1016/j.eswa.2023.122551. URL http://dx.doi.org/10.1016/j.eswa.2023.122551.

Lekkas, G., Vrochidou, E., and Papakostas, G. A. Time–frequency transformations for enhanced biomedical signal classification with convolutional neural networks. *BioMedInformatics*, 5(1), 2025. ISSN 2673-7426. doi: 10.3390/biomedinformatics5010007. URL https://www.mdpi.com/2673-7426/5/1/7.

Li, C., Qi, Y., Ding, X., Zhao, J., Sang, T., and Lee, M. A deep learning method approach for sleep stage classification with eeg spectrogram. *International Journal of Environmental Research and Public Health*, 19(10), 2022. ISSN 1660-4601. doi: 10.3390/ijerph19106322. URL https://www.mdpi.com/1660-4601/19/10/6322.

Makrinioti, H. et al. European Respiratory Society statement on preschool wheezing disorders. *European Respiratory Journal*, 64(3), 2024.

OMRON Healthcare. WheezeScan: The world's first wheeze detection device, 2025. URL https://www.omron-healthcare.co.uk.

Pascanu, R., Mikolov, T., and Bengio, Y. On the difficulty of training recurrent neural networks. *Proceedings of the 30th International Conference on Machine Learning*, pp. 1310–1318, 2013.

Saglani, S. et al. Early detection of airway wall remodeling and eosinophilic inflammation in preschool wheezers. *American Journal of Respiratory and Critical Care Medicine*, 176(9):858–864, 2007.

Seppä, V.-P., Pelkonen, A. S., Kotaniemi-Syrjänen, A., Viik, J., Mäkelä, M. J., and Malmberg, L. P. Tidal flow variability measured by impedance pneumography relates to childhood asthma risk. *European Respiratory Journal*, 47(6):1687–1696, 2016.

Seppä, V.-P. et al. Expiratory variability index (EVI) is associated with the severity of acute bronchial obstruction in small children. *Pediatric Allergy and Immunology*, 31 (6):636–642, 2020a.

Seppä, V.-P. et al. Reduced expiratory variability index (EVI) is associated with controller medication withdrawal and symptoms in wheezy children. *Pediatric Allergy and Immunology*, 31(5):489–495, 2020b.

Stern, J., Pier, J., and Litonjua, A. A. Asthma epidemiology and risk factors. *Seminars in Immunopathology*, 42(1): 5–15, 2020.

Ullah, A., Anwar, S. M., Bilal, M., and Mehmood, R. M. Classification of arrhythmia by using deep learning with 2-d ecg spectral image representation. *Remote Sensing*, 12(10), 2020. ISSN 2072-4292. doi: 10.3390/rs12101685. URL https://www.mdpi.com/2072-4292/12/10/1685.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Waleffe, R., Byeon, W., Riach, D., Norick, B., Korthikanti, V., Dao, T., Gu, A., Hatamizadeh, A., Singh, S., Narayanan, D., Kulshreshtha, G., Singh, V., Casper, J., Kautz, J., Shoeybi, M., and Catanzaro, B. An empirical study of mamba-based language models, 2024. URL https://arxiv.org/abs/2406.07887.

Ye, Z., Xia, K., Fu, Y., Dong, X., Hong, J., Yuan, X., Diao, S., Kautz, J., Molchanov, P., and Lin, Y. C. Longmamba: Enhancing mamba's long context capabilities via training-free receptive field enlargement, 2025. URL https://arxiv.org/abs/2504.16053.

A. Hyperparameter Search Spaces and Selected Values

WIPSNet and Mamba were both tuned via Optuna (50 trials each).

Table 2. WIPSNet hyperparameter search space and selected values. SE = Squeeze-and-Excitation.

Parameter	**Search Space**	**Selected Value**
Scalogram window length	2–16 min	60,000 (4 min)
Volume depth D	1–10 scalograms	8
Stride (scalogram)	1–4 scalograms	1
Backbone	ResNet / EfficientNet	ResNet
Batch size	{8, 16, 32, 64}	32
Learning rate	10^{-5}–10^{-2}	2.66×10^{-3}
Weight decay	10^{-6}–10^{-3}	8.20×10^{-6}
Dropout	0–0.5	0.049
SE attention	{True, False}	True
Mixup α	0–0.5	0.017
LR scheduler	{True, False}	False

Table 3. Mamba SSM hyperparameter search space and selected values.

Parameter	**Search Space**	**Selected Value**
Segment length	{5, 10, 15, 20, 30, 60, 240, 480}k samples	5,000 (20 s)
Stride (samples)	5k–240k samples	20,000
d_{model}	16–192	32
d_{state}	8–48	32
d_{conv}	2–4	3
Expand factor	1–4	3
Num. layers	2–8	3
Dropout	0–0.5	0.374
Batch size	{8, 16, 32, 64}	16
Learning rate	10^{-5}–10^{-2}	5.35×10^{-4}
Weight decay	10^{-6}–10^{-3}	2.54×10^{-4}
Scheduler	{True, False}	False

B. Full Results Table

Table 4. All models and ablations, sorted by AUC. Mean ± std over 5 seeds for DL models. AUC is the primary ranking metric; AUPRC (area under the precision–recall curve) is reported alongside to reflect performance on the minority (wheeze) class under class imbalance.

Model	Input & backbone	AUC	AUPRC
WIPSNet-CWT (D=8)	CWT + 3D CNN	.873 ± .019	.758 ± .075
WIPSNet-Mel (D=8)	Mel spectrogram + 3D CNN	.867 ± .022	.754 ± .041
WIPSNet-CWT (D=12)	CWT + 3D CNN	.865 ± .022	.777 ± .049
WIPSNet-STFT (D=8)	STFT + 3D CNN	.861 ± .039	.748 ± .043
AttnSleep	Raw 1D + MRCNN + attention	.859 ± .025	.713 ± .026
WIPSNet-CWT (D=2)	CWT + 3D CNN	.852 ± .029	.755 ± .063
WIPSNet-CWT (D=4)	CWT + 3D CNN	.851 ± .031	.752 ± .060
WIPSNet-CWT (D=1)	CWT + 3D CNN	.843 ± .014	.750 ± .014
SleePyCo	Raw 1D + pyramid + Transformer	.842 ± .016	.659 ± .013
WIPSNet-WST (D=8)	Wavelet scattering + 3D CNN	.830 ± .043	.685 ± .079
Mamba	Raw 1D + SSM	.725 ± .014	.560 ± .017
EVI	Clinical scalar	.633	.433
Mamba-Scalogram	CWT patches + SSM	.620 ± .083	.487 ± .069
XGBoost	Handcrafted features	.595 ± .029	.461 ± .032
Mamba-Long	Raw 1D 32 min + SSM	.452 ± .076	.404 ± .029