

RespiScan: Tuberculosis Triage Support from Cough Acoustics Using Continuous Wavelet Transform Scalograms and MobileNetV2

Gibran El Fahmi¹, Heri Pratikno^{2*}, Pauladie Susanto³, Weny Indah Kusumawati⁴

^{1,2*,3,4}Department of Information Systems, Universitas Dinamika, Surabaya, East Java, Indonesia

e-mail: 24410200007@dinamika.ac.id¹, heri@dinamika.ac.id^{2*}, pauladie@dinamika.ac.id³,

weny@dinamika.ac.id⁴

Article Information

Article History:

Received : 17 June 2026
Revised : 6 August 2026
Accepted : 1 October 2026
Published : 22 October 2026

*Correspondence:

heri@dinamika.ac.id

Keywords:

Clinical Pre-Screening, Continuous Wavelet Transform, Cough Acoustics, Focal Loss, MobileNetV2, Tuberculosis

Copyright © 2026 by Author.
Published by Universitas Dinamika.



This is an open access article under the CC BY-SA license.



10.37802/joti.v8i2.1467

Journal of Technology and Informatics (JoTI)

P-ISSN 2721-4842

E-ISSN 2686-6102

[https://e-](https://e-journals.dinamika.ac.id/index.php/joti)

[journals.dinamika.ac.id/index.php/joti](https://e-journals.dinamika.ac.id/index.php/joti)

Abstract:

Tuberculosis (TB) requires accessible pre-screening approaches for referral prioritization in resource-limited settings. Cough-acoustic studies have not always reported subject-level partitioning, patient-level evaluation, operating-point selection, and performance uncertainty consistently. This study presents RespiScan, a patient-level triage-support framework using complex Morlet Continuous Wavelet Transform scalograms and ImageNet-pretrained MobileNetV2 with corrected weighted focal loss. We curated a secondary cohort of 198 subjects (87 TB and 111 Non-TB) from TBscreen and Coswara. Audio was standardized to 16-kHz mono, one-second segments, capped at 10 clean recordings per subject, and partitioned strictly by subject into 137 training, 30 validation, and 31 internal-holdout subjects. Training-only augmentation used 10-dB environmental noise for both classes and speed perturbation for Non-TB. We selected the Non-TB: TB weight of 1.5:1.0 from three validation-only settings, then averaged patient scores before applying a validation-derived Youden threshold of 0.424084. On the internal holdout, RespiScan achieved an AUROC of 0.8866 (95% CI: 0.7479-0.9832), 77.42% accuracy, 71.43% sensitivity, and 82.35% specificity. Wide uncertainty, source-label confounding, and exploratory calibration preclude clinical-readiness claims; larger source-balanced prospective validation is required.

INTRODUCTION

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, remains a major infectious-disease burden. The World Health Organization reported that TB again became the leading infectious-disease killer in 2023 [1]. Confirmatory microbiological and radiological tests remain essential, but access can be limited by cost, equipment, sample handling, trained personnel,

and turnaround time in peripheral facilities. Low-complexity pre-screening tools may therefore help prioritize individuals for confirmatory testing without replacing clinical assessment. WHO target product profiles emphasize rapid, accessible technologies for TB detection at peripheral levels [2], [3]. Cough acoustics is attractive because cough can be captured using microphones in smartphones and low-cost edge devices. Pulmonary inflammation, mucus accumulation, airway obstruction, and parenchymal involvement may change cough time-frequency structure. Nevertheless, cough is non-specific; a model output should be interpreted as a triage score that may support referral prioritization, not as a diagnosis or an absolute probability of TB.

Prior studies have shown that cough recordings can support TB screening [4], [5], while multi-country resources and independent benchmarking have expanded the field [6], [7]. However, subject-level partitioning and patient-level reporting have not always been described consistently. If repeated recordings from one participant cross data partitions, performance may reflect participant-specific acoustics rather than disease patterns. Other concerns include unequal subject contribution, inconsistent preprocessing, source-domain confounding, segment-level inference, insufficiently justified thresholds, and limited uncertainty reporting.

RespiScan addresses this methodological gap through: (1) standardized complex Morlet CWT scalograms; (2) strict subject-level partitioning with zero overlap; (3) a maximum of 10 clean recordings per subject; (4) train-only augmentation with clean validation and holdout partitions; (5) validation-only class-weight comparison and a validation-derived operating threshold; (6) patient-level score averaging; and (7) uncertainty, multi-seed, source-domain, and exploratory calibration analyses. The study is positioned as a preliminary internal evaluation of a triage-support framework rather than a clinically deployable diagnostic system. Reporting of dataset composition, model development, validation, and performance followed the TRIPOD+AI reporting framework [13].

RELATED WORKS

Pahar et al. evaluated forced coughs from 51 symptomatic adults using participant-separated nested cross-validation and found that logistic regression on selected high-resolution MFCC features achieved strong discrimination [4]. Sharma et al. developed TBscreen using complex-Morlet scalograms and a ResNet18 classifier for passive coughs collected in a controlled cohort [5]. These studies support the feasibility of acoustic TB screening, but their cohort sizes, recording conditions, evaluation procedures, and operating-point definitions differ. Voluntary-cough classification studies also supported the broader feasibility of respiratory-disease screening from cough acoustics [18], while reviews highlighted challenges related to recording variability, labeling, dataset shift, and evaluation design [19].

Huddart et al. released a multi-country resource containing more than 700,000 solicited cough sounds from 2,143 individuals evaluated for TB [6]. The CODA TB DREAM Challenge used individual-level training and held-out test partitions to benchmark multiple cough-only and cough-plus-clinical models across seven countries [7]. Coswara provides a complementary respiratory-audio resource but was collected for remote respiratory screening rather than TB diagnosis [8]. These resources emphasize the value of subject-level separation, transparent test governance, and external validation.

For short non-stationary signals, CWT provides localized time-frequency analysis with variable resolution [9]. MobileNetV2 offers a lightweight architecture through inverted residuals and linear bottlenecks [10], while DEMAND provides real-world environmental recordings for controlled robustness exposure [11]. Focal loss can emphasize difficult samples

[12], but its focusing parameter and class weights must be reported transparently. Table 1 compares representative cough-based TB work with the present study. Recent JoTI studies also demonstrate the journal's applied machine-learning scope across cybersecurity, computer vision, text classification, and clinical-risk classification [20]–[24]. Earlier automatic cough-sound analysis by Botha et al. provided an early proof of concept for TB-related acoustic discrimination using conventional features and classifiers [25].

Table 1. Critical Comparison with Representative Cough-Based TB Studies

Study	Data and Model	Evaluation Design	Main Contribution	Main Limitation
Pahar et al. [4]	51 adults; 1,358 forced coughs; MFCC and logistic regression	Patient-separated nested cross-validation; inner-fold equal-error-rate threshold	Real-world clinical proof of concept	Small single-site cohort
Sharma et al. [5]	149 subjects; balanced T1 cohort of 90; CWT and ResNet18	Subject-independent fivefold cross-validation; threshold 0.50 and sensitivity at 70% specificity	Passive-cough scores associated with bacterial burden	Controlled cohort; limited passive-to-forced transfer
Huddart et al. [6]	2,143 individuals; over 700,000 cough sounds	Multi-country training and held-out test resource	Large microbiologically annotated dataset	Dataset descriptor, not a classifier study
RespiScan	198 subjects; CWT, MobileNetV2, and focal loss	Strict 137/30/31 subject split; patient-level evaluation; validation-derived threshold	Sample cap, validation-only selection, bootstrap CI, and source audit	Small internal holdout and source-label confounding

METHOD

Dataset Source, Ethical Considerations, and Preprocessing Pipeline

This study, illustrated in Figure 1, used a curated secondary cough-audio cohort from selected anonymized records in the TBscreen clinical cough database and the Coswara database. Only records meeting audio-quality, label-availability, and preprocessing requirements were retained. The initial curated cohort consisted of 225 subjects, including 149 from TBscreen and 76 from Coswara. Based on the original diagnostic metadata, the raw TBscreen subset contained 110 subjects from Cohort A and 39 from Cohort C. Cohort A was not treated as a purely TB-positive cohort; final class labels were assigned using the available ground-truth diagnostic metadata. In the raw TBscreen subset, Cohort A contained 103 TB-positive subjects and 7 Non-TB controls, while Cohort C contained 39 Non-TB controls. After quality control, the final standardized cohort consisted of 198 subjects, including 87 TB-positive subjects and 111 Non-TB subjects. All data were de-identified secondary recordings, and the authors conducted no direct patient recruitment or clinical intervention. Table 2 summarizes source-level subject

composition after quality control.

Table 2. Source-Level Subject Composition After Quality Control

Data source	Raw subjects	Excluded subjects	Active subjects	TB subjects	Non-TB subjects
TBscreen Cohort A	110	18	92	87	5
TBscreen Cohort C	39	9	30	0	30
Coswara LFS	76	0	76	0	76
Total	225	27	198	87	111

All recordings were standardized to mono audio, resampled to 16 kHz, and cropped or zero-padded into non-overlapping one-second segments. The 16-kHz rate retains acoustic content up to 8 kHz while reducing computational cost. Each valid segment was converted into a 224 x 224 CWT scalogram. A fixed subject manifest enforced strict separation so no participant appeared in more than one partition; post-split verification confirmed zero overlap across training, validation, and internal holdout. Across the three partitions, 1,267 clean recordings were retained. After augmentation of the training partition only, the final analysis set comprised 2,740 CWT scalograms: 2,312 training, 229 validation, and 199 internal-holdout samples as summarized in Table 3.

Table 3. Subject And CWT Scalogram Sample Distributions Across Final Dataset Splits

Split	Subjects	TB subjects	Non-TB subjects	Total samples	TB samples	Non-TB samples	Description
Train	137	60	77	2,312	1,044	1,268	Clean data plus train-only augmentation.
Validation	30	13	17	229	127	102	Clean-only validation for threshold calibration.
Test	31	14	17	199	133	66	Clean-only internal holdout.

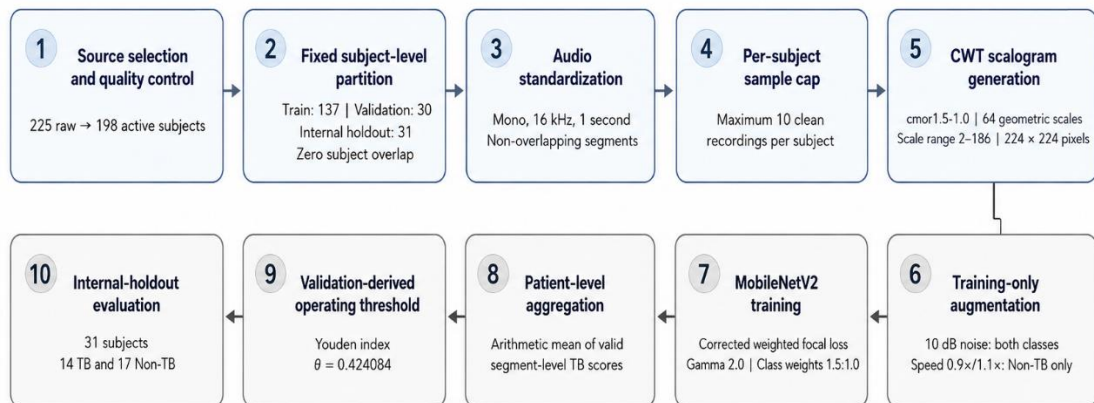


Figure 1. RespiScan patient-level triage workflow from source selection and quality control to final internal-holdout evaluation.

Continuous Wavelet Transform Scalogram Generation

The Continuous Wavelet Transform (CWT) was used to represent cough sounds as two-dimensional time-frequency images. For a signal $x(t)$, the CWT coefficient at scale a and translation b is defined as:

$$W(a, b) = \left[\frac{1}{\sqrt{|a|}} \right] \left[\int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \right] \quad (1)$$

where $\psi^*(.)$ denotes the complex conjugate of the mother wavelet, a controls frequency dilation, and b controls temporal translation. This study used the complex Morlet wavelet `cmor1.5-1.0`, with bandwidth parameter 1.5 and center frequency 1.0. The scale widths were defined using 64 geometric values:

$$a_i = 10^{\left[\log_{10}(a_{min}) + i \frac{(\log_{10}(a_{max}) - \log_{10}(a_{min}))}{(N-1)} \right]}, i = 0, 1, \dots, 63 \quad (2)$$

where $a_{min} = 2$, $a_{max} = 186$, and $N = 64$. The magnitude $|W(a,b)|$ was rendered as a CWT scalogram. This representation was selected because cough sounds are short, transient, and non-stationary, making variable-resolution time-frequency analysis more appropriate than a single fixed-window representation.

Environmental Noise Robustness and Class-Balancing Protocol

Environmental-noise mixing was applied only to the training partition. Noise segments from kitchen, park, and hallway environments were mixed with clean cough signals from both TB and Non-TB training samples at a controlled SNR of 10 dB:

$$x_{noisy}(t) = s(t) + \alpha n(t) \quad (3)$$

$$\alpha = \sqrt{\frac{\sum_t s(t)^2}{SNR_{linear} \cdot \sum_t n(t)^2}} \quad (4)$$

where $s(t)$ is the clean cough signal, $n(t)$ is the noise signal, and $SNR_{linear} = 10^{(SNR_{dB}/10)}$. For 10 dB, $SNR_{linear} = 10$. All 839 clean training recordings received one noise-mixed version. Speed perturbation at 0.9x and 1.1x was additionally applied to the 317 clean Non-TB training recordings, producing 634 samples. Validation and internal-holdout partitions remained clean and unaugmented. This augmentation protocol was predefined for training and was not established as globally optimal.

MobileNetV2 Architecture and Training

MobileNetV2 was retained because its depthwise separable convolutions, inverted residual blocks, linear bottlenecks, and small parameter footprint align with the study's edge-oriented triage objective [10]. The analysis uses one reproducible lightweight architecture and does not claim superiority over heavier models. The ImageNet-pretrained backbone was followed by a 512-unit fully connected layer, ReLU activation, dropout with $p = 0.5$, and a two-class output layer for Non-TB and TB. For kernel size $D_k \times D_k$, input channels M , and output channels N , the computational costs of standard and depthwise-separable convolution are:

$$Parameters = Dk^2M + MN \quad (5)$$

Training used the corrected weighted focal-loss formulation with $\gamma = 2.0$ [12].

The focal modulation term was calculated from the unweighted true-class probability $p_{t,i}$, while the class-specific factor α_c was applied separately. Gamma was fixed as a literature-informed setting rather than independently optimized. Three Non-TB:TB weight settings (1.0:1.0, 1.5:1.0, and 2.0:1.0) were compared using validation data only; 1.5:1.0 was selected among the evaluated settings and was not derived from inverse class frequency.

Optimization used Adam with a learning rate of $1e-5$ for the feature extractor and $1e-4$ for the classifier head, batch size 32, ReduceLROnPlateau scheduling, a maximum of 30 epochs, and early-stopping patience of seven. The checkpoint was selected by maximum validation patient-level AUROC.

Table 4. Final Hyperparameters and Selection Rationale

Component	Final setting	Selection basis and rationale
Audio and CWT	Mono 16 kHz, 1 s; cmor1.5-1.0; 64 scales; 2–186; 224×224	Predefined preprocessing configuration balancing time-frequency resolution and computational cost
Subject cap	Maximum 10 clean recordings	Limits dominance by highly represented participants
Augmentation	10-dB noise for both classes; $0.9\times/1.1\times$ speed for Non-TB	Predefined training-only strategy for environmental and class-specific variation
Architecture	ImageNet-pretrained MobileNetV2; dropout 0.5	Lightweight architecture aligned with edge-oriented use
Optimization	Adam; LR $10^{-5}/10^{-4}$; batch 32	Lower backbone learning rate and faster classifier adaptation
Training control	Maximum 30 epochs; patience 7; ReduceLROnPlateau	Validation-monitored checkpoint selection
Focal loss	$\gamma = 2.0$; Non-TB:TB = 1.5:1.0	Literature-informed γ ; class weight selected on validation data
Patient decision	Mean patient score; validation Youden threshold	Patient-level inference with validation-only operating-point selection

where $p_{t,i}$ denotes the model probability assigned to the true class of sample i , $\gamma = 2.0$ is the focusing parameter, and α_c denotes the class-specific weight.

Patient-Level Score Averaging and Threshold Selection

Clinical triage decisions are made at the patient level. Therefore, segment-level TB outputs from each participant were averaged with equal weight to obtain a patient-level model score:

$$P_j = \left(\frac{1}{K_j} \right) \sum_{k=1}^{K_j} \hat{p}_{jk} \quad (7)$$

The binary triage output was obtained using a fixed operating threshold θ :

$$\hat{Y}_{Pj} = I(P_j \geq \theta) \quad (8)$$

The threshold was selected exclusively from validation-set patient probabilities using Youden's

Index:

$$Sensitivity(\theta) + Specificity(\theta) - 1 \quad (9)$$

$$\theta_{optimal} = \arg \max_{\theta} J(\theta) \quad (10)$$

The threshold was derived exclusively from validation patient scores using Youden's index and fixed at 0.424084 before evaluation of the selected model on the 31-subject internal holdout. The operating point is a model decision threshold, not a diagnostic cutoff. AUROC was calculated from patient scores independently of this threshold, and calibration was assessed separately [14].

RESULTS AND DISCUSSION

Training Behavior

Training loss decreased, while validation loss stayed higher and rose after the early epochs. Validation patient-level AUROC peaked at 0.8597 at epoch 7 and did not improve consistently thereafter, indicating a train-validation generalization gap and possible overfitting. Accordingly, we retained the epoch-7 checkpoint based on validation patient-level AUROC as the model-selection criterion.

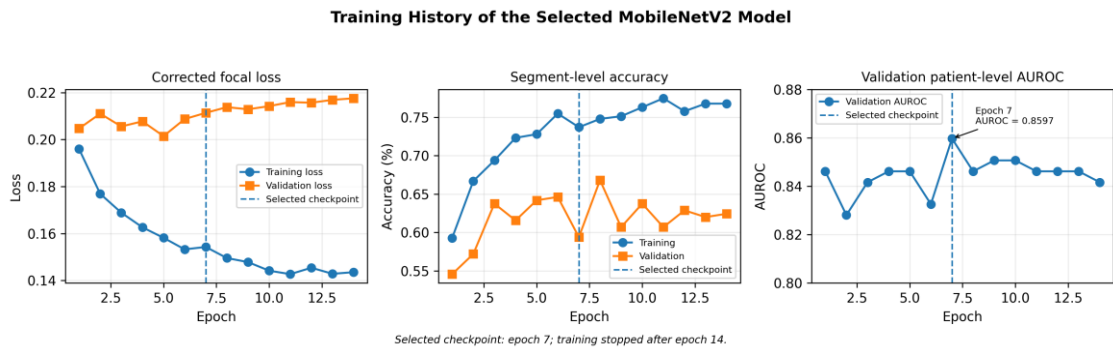


Figure 2. Training history of the selected seed-42 MobileNetV2 model. The checkpoint at epoch 7 was selected by validation patient-level AUROC.

Validation Class-Weight Ablation

Three class-weight settings were compared without changing the subject split, augmentation, architecture, gamma, optimizer, or selection criterion. A configuration was eligible when validation sensitivity was at least 90%; specificity was then prioritized, followed by validation AUROC and TB F1 as tie-breakers. All settings met the sensitivity constraint. The 1.0:1.0 and 1.5:1.0 settings had identical sensitivity and specificity, while 1.5:1.0 achieved the highest AUROC and was fixed before its internal-holdout evaluation.

Table 5. Validation-Only Class-Weight Ablation

Non-TB:TB	Threshold	Sensitivity	Specificity	F1 TB	AUROC
1.0:1.0	0.482491	100.00%	76.47%	86.67%	0.8552

Non-TB:TB	Threshold	Sensitivity	Specificity	F1 TB	AUROC
1.5:1.0	0.424084	100.00%	76.47%	86.67%	0.8597
2.0:1.0	0.432648	100.00%	70.59%	83.87%	0.8552

Patient-Level Triage Performance

At the fixed validation-derived threshold of 0.424084, the selected model correctly classified 24 of 31 internal-holdout participants. AUROC was 0.8866, indicating useful score ranking across thresholds, and sensitivity at the fixed operating point was 71.43%. This performance suggests useful classification at the fixed operating point. Quantify patient-level uncertainty with 2,000 stratified percentile-bootstrap resamples, retaining class sizes.

Table 6. Final Patient-Level Internal-Holdout Performance with 95% Bootstrap Confidence Intervals

Metric	Estimate	95% CI	Count / interpretation
Accuracy	77.42%	61.29%-90.32%	24/31 correctly classified
Sensitivity / TB recall	71.43%	50.00%-92.86%	10/14 TB detected
Specificity	82.35%	64.71%-100.00%	14/17 Non-TB identified
TB precision / PPV	76.92%	58.82%-100.00%	10/13 positive outputs were TB
TB F1	74.07%	56.00%-89.66%	Harmonic mean of precision and recall
NPV	77.78%	64.71%-93.33%	14/18 negative outputs were Non-TB
Balanced accuracy	76.89%	61.55%-90.55%	Mean sensitivity and specificity
Patient-level AUROC	0.8866	0.7479-0.9832	Threshold-independent discrimination

The confusion matrix showed 14 true negatives, 3 false positives, 4 false negatives, and 10 true positives. The model showed promising discrimination but lacked sensitivity for stand-alone clinical screening. The validation-derived threshold defines the reported operating point and should not be interpreted as a universal clinical cutoff.

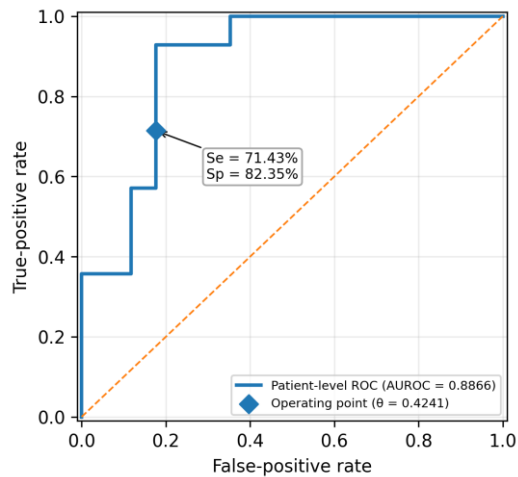


Figure 3. Patient-level ROC on the internal holdout; the marker shows the validation-derived operating point.

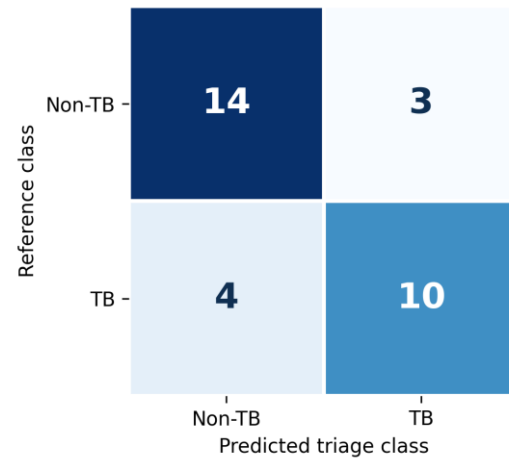


Figure 4. Patient-level confusion matrix at $\theta = 0.424084$ (TB as the positive class).

Robustness, Source-Domain Bias, and Calibration

Across seeds 42, 123, and 2026, validation patient-level AUROC was 0.8718 ± 0.0114 (range 0.8597-0.8824), indicating stable discrimination across training initializations. Validation thresholds ranged from 0.423990 to 0.475388, showing that the operating point remained sensitive to the small validation cohort. Only seed 42 was used for the final reporting run; alternative seeds were not evaluated on the holdout.

Source-stratified analysis showed 92.31% specificity among 13 Coswara Non-TB participants and 50.00% among four pooled TBscreen Non-TB participants. The latter estimate is highly uncertain. More importantly, all TB-positive participants originated from TBscreen Cohort A, whereas Coswara contributed only Non-TB controls; disease label and acquisition source were partially confounded. Exploratory calibration yielded Brier scores of 0.1935 and 0.1903 and expected calibration errors of 0.2630 and 0.2330 on validation and holdout, respectively. With only 30-31 participants, model outputs should be treated as triage scores rather than calibrated clinical probabilities [14].

Computational Feasibility and Limitations

The MobileNetV2 model contained approximately 2.2 million parameters and an 11.7-MB float32 weight file, suggesting potential edge feasibility. Actual on-device latency, memory consumption, power usage, thermal stability, and robustness across microphones remain to be measured; therefore, this study does not claim deployment readiness. Several limitations remain. First, the internal holdout contained only 31 independent participants, producing wide confidence intervals. Second, training curves showed a train-validation generalization gap and possible overfitting despite validation-based checkpoint selection. Third, source and disease labels were partially confounded, preventing claims of source-invariant performance. Fourth, calibration estimates were exploratory. Taken together, the clinical takeaway is that the findings are promising but not yet ready for clinical or public-health use. Larger source-balanced, device-diverse, multi-center prospective cohorts are needed, with prespecified operating points, external validation, calibration assessment, and true edge-device benchmarking before clinical or public-health use [15]-[17].

CONCLUSIONS AND SUGGESTIONS

In summary, RespiScan establishes a methodologically rigorous benchmark for cough-acoustic tuberculosis (TB) triage support. By integrating standardized complex-Morlet CWT scalograms with MobileNetV2 and corrected weighted focal loss, the framework directly addresses common methodological pitfalls through strict subject-level isolation, train-only data augmentation, validation-based thresholding, and comprehensive source-domain auditing. Evaluated on a 31-subject internal holdout set, RespiScan achieved robust internal discrimination (AUROC of 0.8866, 77.42% accuracy, 71.43% sensitivity, and 82.35% specificity at $\theta = 0.424084$). Crucially, while these results confirm strong preliminary efficacy, current limitations in sensitivity, uncertainty, and source-label confounding preclude standalone diagnostic deployment, highlighting the essential need for prospective external validation in larger, source-balanced cohorts.

ACKNOWLEDGMENTS

The authors would like to thank the Computer Engineering Study Program, Faculty of Information Systems, Universitas Dinamika, for academic support and server resources during the development of this research prototype.

DATA AVAILABILITY

The raw cough audio data are not redistributed by the authors and should be obtained from the official TBscreen and Coswara repositories according to their respective access, licensing, and ethical-use requirements. Preprocessing scripts, CWT extraction code, model-training pipeline, evaluation scripts, and trained weights may be shared upon reasonable request where permitted by institutional policy and source-data licenses.

CONFLICT OF INTEREST

The authors declare no conflict of interest regarding the publication of this manuscript.

REFERENCES

- [1] World Health Organization, Global Tuberculosis Report 2024. Geneva, Switzerland: World Health Organization, 2024. [Online]. Available: <https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/global-tuberculosis-report-2024>
- [2] World Health Organization, Target Product Profile for Tuberculosis Diagnosis and Detection of Drug Resistance. Geneva, Switzerland: World Health Organization, 2024. [Online]. Available: <https://www.who.int/publications/i/item/9789240097698>
- [3] World Health Organization, High-Priority Target Product Profiles for New Tuberculosis Diagnostics: Report of a Consensus Meeting. Geneva, Switzerland: World Health Organization, 2014. [Online]. Available: <https://www.who.int/publications/i/item/WHO-HTM-TB-2014.18>
- [4] M. Pahar, M. Klopfer, B. Reeve, R. Warren, G. Theron, and T. Niesler, "Automatic cough classification for tuberculosis screening in a real-world environment," *Physiological Measurement*, vol. 42, no. 10, p. 105014, 2021, doi: 10.1088/1361-6579/ac2fb8.
- [5] M. Sharma et al., "TBscreen: A passive cough classifier for tuberculosis screening with a controlled dataset," *Science Advances*, vol. 10, no. 1, p. eadi0282, 2024, doi: 10.1126/sciadv.adi0282.

- [6] S. Huddart et al., "A dataset of solicited cough sound for tuberculosis triage testing," *Scientific Data*, vol. 11, no. 1, p. 1149, 2024, doi: 10.1038/s41597-024-03972-z.
- [7] D. Jaganath et al., "Accelerating cough-based algorithms for pulmonary tuberculosis screening: Results from the CODA TB DREAM Challenge," *Open Forum Infectious Diseases*, vol. 12, no. 10, p. ofaf572, 2025, doi: 10.1093/ofid/ofaf572.
- [8] D. Bhattacharya et al., "Coswara: A respiratory sounds and symptoms dataset for remote screening of SARS-CoV-2 infection," *Scientific Data*, vol. 10, no. 1, p. 397, 2023, doi: 10.1038/s41597-023-02266-0.
- [9] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. Burlington, MA, USA: Academic Press, 2009.
- [10] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.
- [11] J. Thiemann, N. Ito, and E. Vincent, "The diverse environments multi-channel acoustic noise database (DEMAND): A database of multichannel environmental noise recordings," *Proceedings of Meetings on Acoustics*, vol. 19, no. 1, p. 035081, 2013, doi: 10.1121/1.4799597.
- [12] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.324.
- [13] G. S. Collins et al., "TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods," *BMJ*, vol. 385, p. e078378, 2024, doi: 10.1136/bmj-2023-078378.
- [14] B. Van Calster et al., "Calibration: The Achilles heel of predictive analytics," *BMC Medicine*, vol. 17, p. 230, 2019, doi: 10.1186/s12916-019-1466-7.
- [15] R. D. Riley et al., "Minimum sample size for external validation of a clinical prediction model with a binary outcome," *Statistics in Medicine*, vol. 40, no. 19, pp. 4230-4251, 2021, doi: 10.1002/sim.9025.
- [16] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Medicine*, vol. 17, p. 195, 2019, doi: 10.1186/s12916-019-1426-2.
- [17] A. A. H. de Hond et al., "Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: A scoping review," *npj Digital Medicine*, vol. 5, p. 2, 2022, doi: 10.1038/s41746-021-00549-7.
- [18] C. Infante, D. B. Chamberlain, R. Kodgule, and R. R. Fletcher, "Classification of voluntary coughs applied to the screening of respiratory disease," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2017, pp. 1413-1416, doi: 10.1109/EMBC.2017.8037098.
- [19] K. S. Alqudaihi et al., "Cough sound detection and diagnosis using artificial intelligence techniques: Challenges and opportunities," *IEEE Access*, vol. 9, pp. 102327-102344, 2021, doi: 10.1109/ACCESS.2021.3097559.
- [20] A. E. Edet, S. J. Inyang, I. J. Umoren, and U. E. Etuk, "Machine learning approach for classification of cyber threats actors in web region," *Journal of Technology and Informatics (JoTI)*, vol. 6, no. 1, pp. 70-77, 2024, doi: 10.37802/joti.v6i1.679.
- [21] N. C. I. Natun, M. A. Santhia, and Y. R. Kaesmetan, "Identifikasi pengenalan wajah berdasarkan jenis kelamin menggunakan metode Convolutional Neural Network (CNN),"

Journal of Technology and Informatics (JoTI), vol. 6, no. 1, pp. 50-57, 2024, doi: 10.37802/joti.v6i1.694.

- [22] R. Z. Megantara, P. I. Faot, R. H. Ito, K. F. Annabel, and O. Karnalim, "Pemanfaatan machine learning dalam prediksi rating: Studi kasus pada data abstrak publikasi ilmiah," Journal of Technology and Informatics (JoTI), vol. 7, no. 1, pp. 62-75, 2025, doi: 10.37802/joti.v7i1.999.
- [23] L. A. Sanjani, R. B. M. Putra, and U. L. Yuhana, "Exploring the application of machine learning for automatic inbound email classification in CRM system at XYZ Company," Journal of Technology and Informatics (JoTI), vol. 6, no. 1, pp. 1-7, 2024, doi: 10.37802/joti.v6i1.715.
- [24] A. Aryanti, F. D. Wulandari, M. Rafiif, F. Alghaniyyu, N. Kirana, and M. N. Alfazri, "Segmentasi dan klasifikasi risiko perdarahan dengan algoritma K-Means dan Naive Bayes berdasarkan data klinis dan transfusi," Journal of Technology and Informatics (JoTI), vol. 7, no. 2, pp. 211-219, 2025, doi: 10.37802/joti.v7i2.1124.
- [25] G. H. R. Botha, G. Theron, R. M. Warren, M. Klopper, K. Dheda, P. D. van Helden, and T. R. Niesler, "Detection of tuberculosis by automatic cough sound analysis," Physiological Measurement, vol. 39, no. 4, p. 045005, 2018, doi: 10.1088/1361-6579/aab6d0.