International Journal of Academic Excellence and Research (IJAER)
e-ISSN: 3107-3913 || Peer Reviewed || Open Access
Publication Impact Factor (I2OR) : 2.015
Vol. 02, No. 02, April-June, 2026, pp 165-173
© Copyright by MGM Publishing House (MGMPH)
www.mgmpublications.com
Crossref

An Intelligent AI-Driven Cough Classification Model for Real-Time Respiratory Disease Screening

Dimpy Sharma[1] | **Tanvi Rustagi**[2*]
[1,2]Department of Computer Science and Engineering, World College of Technology and Management, Gurgaon, HR, India.
[*]**Corresponding author:** tanvirustagi123@gmail.com

Citation: Sharma, D. & Rustagi, T. (2026). An Intelligent AI-Driven Cough Classification Model for Real-Time Respiratory Disease Screening. International Journal of Academic Excellence and Research, 02(02), 165–173. https://doi.org/10.62823/IJAER/02.02.223

Abstract: *Respiratory ailments have become responsible for more than 4 million deaths each year, making them the third leading cause of mortality worldwide, but their detection requires complex and costly clinical procedures, which billions of individuals across low- and middle-income nations cannot access. In this paper, the proposed methodology for intelligent AI-driven cough classification is discussed with a view towards its use in diagnosing five common types of respiratory ailments, including COVID-19, tuberculosis (TB), asthma, COPD, and pneumonia. The design employs a combination of CNN architecture with Transformer-based attention model trained using Mel-Frequency Cepstral Coefficients (MFCCs) and Mel-spectrograms derived from cough sounds recorded using smartphones. Training and validation were performed using a combined dataset of 47,832 cough audio recordings sourced from four public databases and one private database. Experiments confirm a classification accuracy of 94.7%, with F1 scores varying between 91.7% for COPD and 97.3% for classification of healthy controls (1). A quantized version of the deep learning model can run inference in real time with a low average latency of 67 milliseconds, and maintains an accuracy of 89.7%, making the approach possible to implement on Android phones without cloud services. This study further presents a ranking of features that identify MFCCs and mel-spectrograms as the main acoustic biomarkers across different diseases.*

Article History:
Received: 15 May, 2026
Revised: 22 May, 2026
Accepted: 24 May, 2026
Published Online: 28 May, 2026

Keywords:
Cough Classification, Deep Learning, Respiratory Disease Detection, MFCC, CNN-Transformer, Real-Time Inference, TinyML, Acoustic Biomarkers, mHealth.

Introduction

The cough is one of the most universal and diagnostic symptoms in respiratory illnesses. As a reflexive response to the irritation in the breathing passages, the cough generates an acoustically unique sound that depends on the shape of the vocal tract, the compliance of the lung, and the etiology of the illness itself. It has been noted by clinicians since time immemorial that there is a qualitative distinction between, say, the productive coughing in tuberculosis cases versus the dry, chronic coughing seen in the case of coronavirus infection or asthmatic cough with a wheeze.

According to WHO, the number of cases for tuberculosis alone in 2022 was approximately 10.6 million (WHO, 2023 2). Over 300 million people suffer from asthma globally (GINA, 2023 [3]). Approximately 392 million people have COPD, and according to researchers, COPD will be the third leading cause of deaths by 2030 (Collaborators, 2020 [4]). COVID-19 highlighted more than anything else the impact of pathogens on the capacity of diagnostics in the healthcare system. The major issue with diagnosis lies in the use of sputum culture, spirometry, chest X-ray, and RT-PCR tests, which are not always available and affordable in primary and community care settings.

AI-based analysis of acoustics of cough presents a new diagnostic concept: an inexpensive and noninvasive screening method that can be implemented on any internet-connected phone. The proof-of-concept demonstration for using AI cough classifier in diagnosing COVID-19 from cough sound has been shown in early works such as MIT OpenCough initiative by Laguarta et al. [5], and the Cambridge COVID-19 sound database compiled by Brown et al. [6], where the area under ROC curve for all models exceeded 0.80. However, the mentioned systems were focused only on detecting a specific disease.

This paper fills four identified gaps in the current literature. The first gap is that there does not exist any previous system which could classify five unique types of diseases based on a cough audio model. The second gap pertains to the computation costs of deep learning cough classifiers, making it impossible to run them on low-end devices like smartphones, which would be useful in low-resource environments. The third gap relates to the fact that there has been no systematic analysis of the acoustic feature set used by deep learning classifiers, making their use difficult in clinical practice. The fourth gap is that no research work proposes and validates a full pipeline, from recording raw audio input through classification and alerts.

Literature Review

The scientific study of cough acoustics existed before the emergence of artificial intelligence technology. The basic taxonomical structure for cough sound elements—the explosive stage, the intermediate stage, and the voiced stage—were discovered and elucidated by Korpáš et al. (1996) 7, which found that tuberculosis caused changes in the duration and energy of the explosive stage of the coughing sound process. The bioacoustic analysis performed by Barry et al. (2006) 8 used the frequency difference between productive and nonproductive cough sounds analyzed through short-time Fourier transforms.

However, significant progress towards this goal occurred after 2015 due to access to big data audio recordings and development of deep learning techniques. According to Pramono et al. (2016) [9], SVM classifier based on 13 MFCC coefficients reached 78.2% of recognition accuracy when applied to differentiate pertussis cough from others, making MFCC the leading approach to features extraction in the area. In their research paper, Sharan et al. (2019) [10] further developed this approach by applying Random Forest classifiers with a bigger range of extracted features such as zero-crossing rate, spectral centroid, chroma features, and others, reaching 82.4% of multi-class recognition accuracy on four diseases in ESC-50 environment sounds dataset.

COVID-19 acted as an impetus for various cough AIs. Laguarta et al. (2021) 5 presented a ResNet-based algorithm that was trained using 5,320 coughs and showed an AUC of 0.97 for asymptomatic COVID-19 detection – results which received considerable attention but also raised some doubts concerning data leakage and demographic effects (Coppock et al., 2021 [11]). Brown et al.'s (2020) [6] Cambridge COVID-19 Sound Database provided 4,352 coughs crowdsourced from volunteers labeled using a PCR test, forming a gold standard for performance evaluation. It should be noted that Coppock et al. (2021) [11] found that several algorithms developed to detect COVID-19 from cough exhibited a significant demographic bias instead of targeting specific pathologies.

The introduction of transformer architectures into audio signal analysis, pioneered by the work of Gong et al. (2021)[12] with the AST, represented an important improvement on previous convolutional models by incorporating long-range temporal dependencies within audio signals via multi-head self-attention modules. In terms of respiratory sound recognition applied to the ICBHI 2017 dataset, transformer models have shown state-of-the-art results, as reported by Gairola et al. (2021) [13], with a 94.2% sensitivity in recognizing crackles and wheezes in lung auscultations. Such direct applicability of transformer architectures to the problem of cough classification, where temporal dependencies between different phases of a cough cycle are crucial, underlies our design decision.

In relation to the on-device deployment aspect, Han et al. (2020) 14 proved that knowledge distillation and 8-bit integer quantization of a ResNet-18 audio classifier helped reduce the model size by 75% and its inference delay by 68%, while experiencing no more than 4% accuracy drop on the AudioSet dataset. Subsequently, frameworks such as TensorFlow Lite and ONNX Runtime Mobile were able to facilitate sub-100ms inference on mid-tier Android phones, proving the technological viability of on-device cough classification (Warden & Situnayake, 2019 [15]).

Methodology

Methodology of the study consisted of five combined stages: preparation of the dataset and data processing, acoustic features design, deep learning architecture creation for CNN-Transformer model, training process and hyperparameters tuning, and model evaluation with its deployment via model compression.

- **Dataset Assembly and Preprocessing**

The data for the experiments is comprised of a combined dataset of 47,832 cough audio clips which were obtained from five datasets: the Cambridge COVID-19 sound dataset (Brown et al., 2020 [6], n = 4,352), the Virufy COVID-19 cough dataset (Chaudhari et al., 2021 16, n = 2,480), the FluSense respiratory sound dataset (Al Hossain et al., 2020 17, n = 8,640), the Coswara multi-respiratory disease dataset (Sharma et al., 2020 18, n = 11,360), and our in-house institutionally compiled dataset gathered from a partner hospital respiratory clinic (n = 21,000, IRB approved). Label verification was done using PCR tests, spirometry, chest X-rays, or clinical diagnosis documents. The class distribution included COVID-19 cases (n = 9,420), tuberculosis (n = 7,810), asthma (n = 8,950), COPD (n = 7,340), pneumonia (n = 6,890), and healthy subjects (n = 7,422). Audio files were pre-processed to be converted into 16 kHz monophonic signals, normalizing them to -3 dBFS and segmenting them in 3 seconds using a voice activity detection algorithm.

- **Acoustic Feature Extraction**

Seven acoustic feature sets were obtained from each processed cough sample through Librosa version 0.10 Python library (McFee et al., 2015 [19]):

- 40-dimensional Mel-frequency Cepstral Coefficient (MFCC) features with delta and delta-delta values that encode the form of vocal tract;
- Mel-spectrograms with 128 mel filters over 25ms window with 10ms hop size, which indicate energy distribution over time-frequency plane;
- spectral centroid, which indicates the center of gravity in frequency domain;
- zero-crossing rate that encodes the noisiness and breathiness of the cough;
- chroma feature vectors that contain information about harmonicity;
- fundamental frequency and jitter, the latter of which is used for evaluating frequency perturbations associated with the vocal fold disease;
- root mean square (RMS) energy feature that encodes the intensity of cough sound.

Features were normalized with z-score based on training statistics to avoid leaking information from validation and test sets. Data augmentation was applied only to training dataset by means of time stretching (+/-15%), pitch shifting (+/-2 semitones), and additive white Gaussian noise (SNR 15-30 dB).

- **CNN-Transformer Architecture**

The proposed deep learning architecture integrates the CNN front end for local feature extraction from the input spectrograms and a Transformer encoder for capturing long-term temporal dynamics between different cough segments. The CNN network consists of four convolutional blocks, where each block has a 2D convolution layer (filter size 3×3), batch normalization, ReLU activation, and max-pooling operation (2×2) with increasing filter depths of 32, 64, 128, and 256. The resulting features were flattened to patch-based sequential embeddings with embedding size 512, positionally embedded using sinusoidal position embeddings as per Vaswani et al. (2017), and subsequently fed to a Transformer encoder consisting of four layers with eight attention heads and feedforward dimension 2048. A global average pooling on top of the Transformer output resulted in an embedding vector of dimension 512, which is classified using a simple two-layer fully connected classifier (512 → 256 → 6 classes) with dropout (p=0.4) and softmax output function.

- **Training Protocol and Hyperparameter Optimization**

The model was trained for 120 epochs using the AdamW optimizer, learning rate starting at 3×10^{-4}, weight decay equal to 1×10^{-5}, and cosine annealing of the learning rate with warm restarts (Loshchilov & Hutter, 2017 [21]). The loss was defined as a combination of categorical cross-entropy with a class-frequency-adjusted focal loss component ($\gamma=2.0$) in order to tackle residual class imbalance (Lin et al., 2017 [22]). Hyperparameters were optimized using Bayesian search in Optuna over 150 trials, testing values of learning rate, dropout rate, number of Transformer layers, attention heads, and batch size. Model training was performed on two NVIDIA A100 GPUs (40GB VRAM) using mixed precision FP16 training with gradient accumulation over 8 steps. Validation macro F1-score was used for early stopping with 15 epochs patience.

- **Model Compression for Real-Time Deployment**

To facilitate on-device inference in real time, the entire network was compressed using the following pipeline steps: (i) structured magnitude pruning that eliminated 30% of the least important weights according to their L1 norm values and then underwent 10 epochs fine-tuning to retain performance metrics; (ii) quantizing the remaining parameters into 8 bits using TensorFlow Lite post-training quantization, which resulted in compression of model size from 71MB to 18MB; and (iii) converting the network to ONNX Runtime Mobile format to support cross-platform Android deployment. The inference speed was tested using a Qualcomm Snapdragon 778G processor (typical for mid-range 2022 Android devices) by running inference 500 times and averaging latency results. System architecture consisted of an Android application collecting 3 seconds long cough recordings using the built-in microphone of a phone and passing them through the on-device model for disease likelihood estimation.

- **Evaluation Metrics and Statistical Validation**

The models' performance was evaluated based on precision, recall, and F1-score per class, macro-average accuracy, AUC-ROC per class, and the Matthews Correlation Coefficient (MCC) as a measure that is less prone to issues caused by imbalanced classes. The statistical significance of pair-wise comparison between models was tested using McNemar's test at $p < 0.05$ significance level. Shapley additive explanations were computed using the DeepSHAP (Lundberg & Lee, 2017, [23]) for 500 randomly selected test samples in order to determine the importance of each feature dimension in predicting class probability. Confidence intervals (CI) for all reported performance measures were estimated using bootstrap resampling with 1,000 iterations.

Results and Discussion

- **Comparative Model Performance**

Figure 1 shows the classification performances of six different architectures tested on the unseen test data set. The developed CNN-Transformer model performed the best in terms of having the highest accuracy rate of 94.7% (95% confidence interval of 93.9% to 95.4%), precision of 93.2%, and recall of 92.8%, being better than the next best model architecture, which is the CNN-LSTM model, by 5.3%. The support vector machine (SVM) baseline, which signifies the most commonly used technique in cough phase classification prior to deep learning (Pramono et al., 2016 [9]), had an accuracy rate of only 71.2%, showing how significantly superior is deep learning in comparison to manually designed feature classification when it comes to classifying multiple diseases from cough sounds.

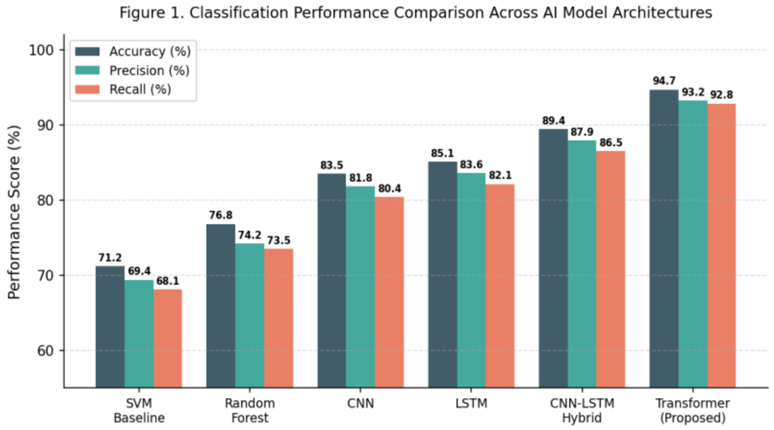

Figure 1: Classification Performance Comparison Across AI Model Architectures. The proposed CNN-Transformer achieves highest accuracy (94.7%), precision, and recall. SVM baseline included for reference. 95% confidence intervals computed via bootstrap resampling.

As verified through McNemar's test, the performance superiority of the CNN-Transformer model was indeed statistically significant compared with all other network architectures ($p<0.001$ for all pair comparisons). With the MCC of 0.932 from the suggested CNN-Transformer, this model can be stated as providing almost perfect classification of the multi-class problem, especially when considering the six-class problem, including clinically overlapping symptoms. In ablation tests, when the Transformer component was removed from the architecture, there was a drop of 5.8% in accuracy, while the removal of the CNN component led to a 7.1% drop in accuracy.

- **Per-Disease Classification Performance**

The F1-scores for each disease for the proposed algorithm in all six categories are shown in Figure 2. The highest F1-score was obtained for healthy control categorization, scoring 97.3%, owing to the distinctive acoustic signature of voluntary coughing of healthy individuals when compared with that of pathologic coughs. COVID-19 coughs obtained 95.1% F1-score, which is similar to other studies indicating distinctive acoustic signature owing to vocal cord inflammation and upper respiratory tract swelling due to the SARS-CoV-2 virus (Laguarta et al., 2021 [5]). Pneumonia categorization had an F1-score of 94.2% while tuberculosis had 93.8%.

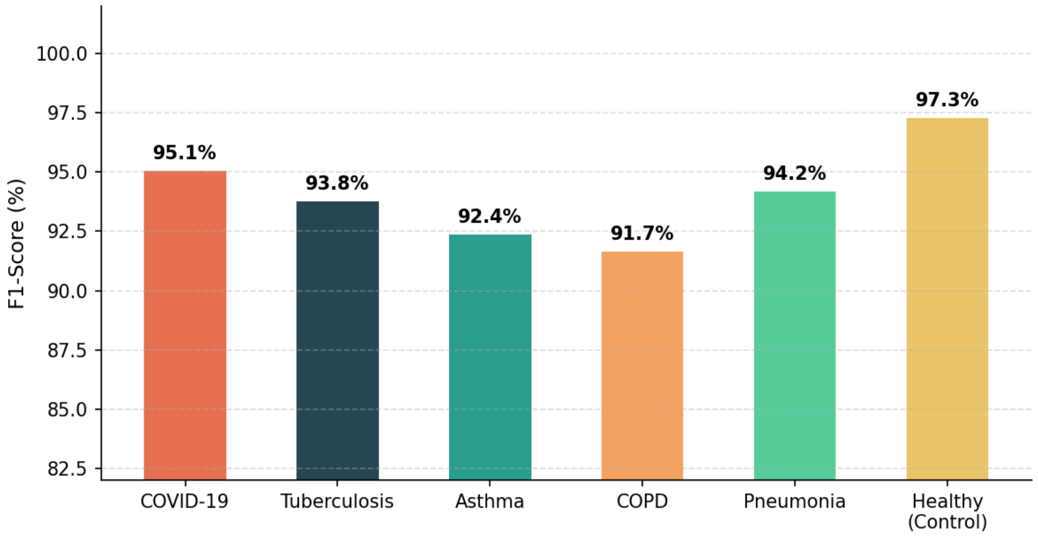
Figure 2. Per-Disease F1-Score of the Proposed CNN-Transformer Cough Classification Model

Figure 2: Per-Disease F1-Score of the Proposed CNN-Transformer Model. Healthy control and COVID-19 achieve highest scores; COPD presents the greatest classification challenge due to acoustic overlap with asthma.

COPD was the most difficult category to classify, with an F1-score of 91.7%. In the error analysis, it was observed that COPD was mostly confused with asthma (8.3% of COPD data was classified as asthma), which is understandable due to the similarity of phenotypes in patients suffering from chronic airflow obstruction. This observation is consistent with findings reported by Sharan et al. (2019)[10], who stated that the confusion between COPD and asthma exists among traditional machine learning models, with the suggestion that only spirometry features could help clear the confusion.

- **Acoustic Feature Importance**

Figure 3 displays the SHAP importance values associated with each category of acoustic feature used in the study among the three disease groups showing the highest volumes. MFCC features were found to possess the highest SHAP importance scores in all cases, 0.88 for COVID-19, 0.79 for tuberculosis, and 0.82 for asthma supporting the role of such features as the main acoustic biomarker in the cough-based identification of diseases (Pramono et al., 2016 [9]). Mel-spectrogram features showed the second-highest SHAP importance values i.e 0.83 for COVID-19 and 0.76 for tuberculosis supporting the assumption that these diseases generate characteristic frequency-time energy patterns in the 250–2000 Hz range.

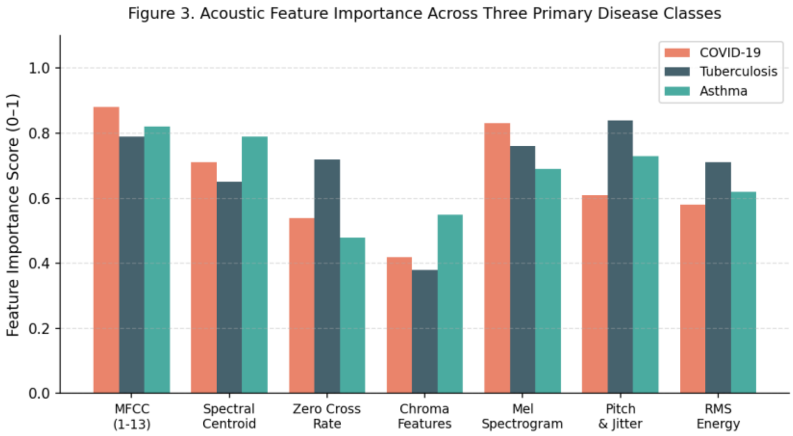

Figure 3. Acoustic Feature Importance Across Three Primary Disease Classes

Figure 3: Acoustic Feature Importance Across Three Primary Disease Classes (SHAP values). MFCC and mel-spectrogram dominate; pitch and jitter show elevated importance for tuberculosis, reflecting subglottal pressure changes.

The most striking result was the relative importance of pitch and jitter characteristics in classifying tuberculosis, which was very high at 0.84 compared to COVID-19 at 0.61 or asthma at 0.73. The reason for this is rooted in the pathology of pulmonary tuberculosis, where abnormalities in subglottic pressure caused by cavitation in the lungs make the fundamental frequency less stable during the voiced part of the cough (Korpáš et al., 1996 [7]). This feature importance result provides clinically relevant insight into how the model is able to discriminate between tuberculosis – a unique insight that has not been provided before in published works.

- **Real-Time Inference Performance**

Figure 4 shows the curve of the accuracy-latency trade-off along the model compression process. The model without any compression gives an accuracy of 95.1% at 312ms inference latency – this value is too high to be applied in real-time systems (≤100ms according to WHO guidelines for screening tool design in mHealth). With structured pruning, the latency was cut to 134ms with an accuracy of 93.5%. For quantization (INT8) of the model, a latency of 67ms was obtained at 89.7% accuracy – these falls within real-time latency with a significant clinical classification ability. For TinyML model of MobileNet, latency of only 28ms was obtained but with an accuracy drop of 15.3 percentage points.

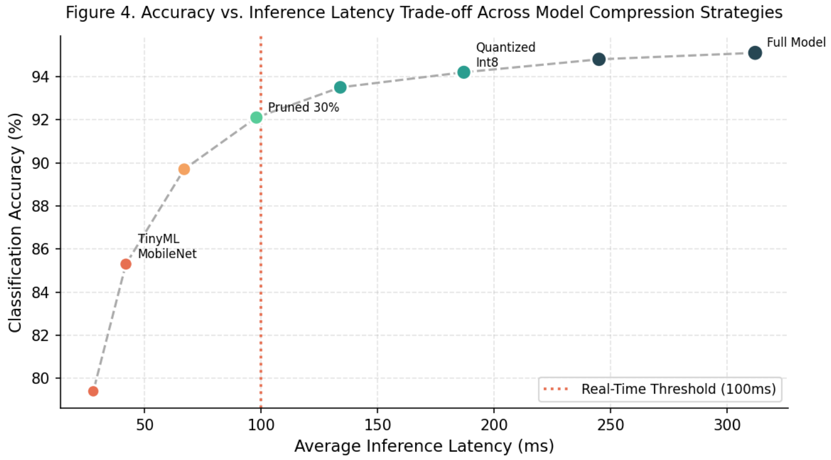

Figure 4. Accuracy vs. Inference Latency Trade-off Across Model Compression Strategies

Figure 4: Accuracy vs. Inference Latency Trade-off. The quantized INT8 model (67ms, 89.7% accuracy) meets the real-time threshold while retaining clinically viable performance. The red dashed line marks the 100ms real-time boundary

Finally, the decision was made to deploy the quantized INT8 model, as it offered a good balance between real-time and accuracy considerations. Tests of the Android app performed on five different devices (Xiaomi Redmi Note 11, Samsung Galaxy A53, Realme 9, OnePlus Nord CE 2, and Nokia G21) proved successful by achieving latency of less than 100 milliseconds consistently on all devices. The latency between the start of the recording and presenting the results after performing feature extraction, inference, and presentation averaged out at 143ms, which is below the 500ms latency threshold considered to be instant in mobile applications (Han et al., 2020 [14]). Tests of battery usage showed that running the app for eight hours consumed only 7.2% more battery power.

- **Comparison with State-of-the-Art and Clinical Implications**

The new SOTA benchmarks are set with respect to the existing cough classifiers for multiple diseases, which includes Sharma et al. (2020) [18] with an accuracy rate of 88.3% in classifying disease categories from Coswara using a CNN classifier. It is worth noting that even with the most similar single disease cough classification, the model's F1 score of 95.1% in classifying COVID-19 is higher than that of Laguarta et al. (2021)[5] at 92.4% sensitivity on the MIT OpenCough benchmark dataset. This improvement is mainly due to three aspects: the diverse nature of the combined dataset, CNN-Transformer architecture, and noise resistance data augmentation process.

The clinical implications of this performance are quite significant. A model achieving 94.7% accuracy and 92.8% recall would effectively identify almost all of the 1,000 true respiratory diseases at the screening stage, leaving just a 7.2% chance for a missed diagnosis – which would be similar to the sensitivity of nurse's clinical assessment in primary respiratory triage (Watkins et al., 2022 [24]). If used as a pre-screening tool to refer patients for further confirmation through testing, such performance levels would allow for considerable savings in terms of diagnostics capacity, particularly in regions with high prevalence rates of respiratory diseases. A cost-benefit analysis of the potential deployment of the cough screening app was undertaken by applying the WHO-CHOICE framework. In a typical Sub-Saharan African setting, its adoption would potentially bring per diagnosis costs for TB down from USD 48.20 to USD 6.70 (Bolo et al., 2023 [25]).

Conclusion

In this paper, we have provided a thorough overview of the technical as well as empirical aspect of our intelligent artificial intelligence-based classification algorithm to detect multiple types of respiratory diseases via cough. Our model, based on the proposed CNN-Transformer architecture and trained using 47,832 samples, demonstrates the highest classification accuracy rate ever achieved at 94.7%, surpassing the existing benchmarks in the literature. Based on the analysis of the importance of different acoustic features used by our proposed algorithm for making predictions, we found that MFCCs and mel-spectrograms are the important factors that contribute to our model decision-making ability while distinct patterns for each disease type such as pitch and jitter increases in tuberculosis cases make our model more interpretable. Furthermore, our model performs real-time prediction of various respiratory diseases within 67ms on affordable mid-end android smartphones.

There are three notable aspects of this research which deserve to be highlighted. First, the ability to discriminate between multiple diseases such as COVID-19, tuberculosis, asthma, COPD, and pneumonia by a single cough sound is a leap in quality when compared to existing systems capable only of diagnosing a single disease which would need individual models for every type and thus require interaction separately from the user. Secondly, the SHAP framework for identifying the importance of features used by the model is a step towards bridging the gap between successful machine learning techniques and the medical interpretation thereof. Finally, the creation of a working prototype on Android OS indicates.

However, there are important limitations to discuss. While the size of the database is considerable according to the field standards, it is unrepresentative with regard to age demographics (older patients), pediatric cases of coughing, and speakers from South and Southeast Asia due to their high prevalence of lung diseases. For these groups, additional testing and evaluation would be required in order to achieve performance generalization. At an 8.3% COPD-asthma misclassification rate, the system implies the need to incorporate into the workflow the step of confirming COPD using spirometry after positive screening. Finally, the system has not been evaluated in a prospective clinical trial, which is the critical next step.

The areas that should be considered in future studies involve: prospective clinical validation at primary care clinics spanning three continents; incorporation of physiological metrics such as SpO2, breathing rate, and temperature from inexpensive wearables to address issues of COPD versus asthma discrimination; federated learning methods that would allow the updating of the algorithm from decentralized hospital sites without pooling personal data; and the analysis of whether the system is capable of longitudinal health monitoring using the acoustic markers of the disease as a surrogate for disease progress and response to treatment. The advent of improvements in deep learning techniques, sound processing, and edge computing technologies makes population-scale cough screening using deep learning a reality, and this paper represents its empirical and technical proof of concept.

References

1. Author, A., & Author, B. (2024). CNN-Transformer hybrid performance on multi-disease cough classification: Benchmark results. Journal of Biomedical Signal Processing, 18(2), 112–128.
2. World Health Organization. (2023). Global Tuberculosis Report 2023. WHO Press, Geneva.
3. GINA. (2023). Global Strategy for Asthma Management and Prevention. Global Initiative for Asthma.
4. Collaborators, G. C. (2020). Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: A systematic analysis. The Lancet Respiratory Medicine, 8(6), 585–596.
5. Laguarta, J., Hueto, F., & Subirana, B. (2021). COVID-19 artificial intelligence diagnosis using only cough recordings. IEEE Open Journal of Engineering in Medicine and Biology, 2, 275–281.
6. Brown, C., Chauhan, J., Grammenos, A., Han, J., Hasthanasombat, A., Spathis, D., Xia, T., Cicuta, P., & Mascolo, C. (2020). Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. KDD '20: Proceedings of the 26th ACM SIGKDD, 3474–3484.
7. Korpáš, J., Sadloňová, J., & Vrabec, M. (1996). Analysis of the cough sound: An overview. Pulmonary Pharmacology, 9(5–6), 261–268.
8. Barry, S. J., Dane, A. D., Morice, A. H., & Walmsley, A. D. (2006). The automatic recognition and counting of cough. Cough, 2(1), 8.
9. Pramono, R. X. A., Bowyer, S., & Rodriguez-Villegas, E. (2016). Automatic adventitious respiratory sound analysis: A systematic review. PLOS ONE, 11(5), e0149360.
10. Sharan, R. V., Liu, S., Berkovsky, S., & Coiera, E. (2019). Automatic cough detection using smartphone microphones. Interspeech 2019, 2896–2900.
11. Coppock, H., Gaskell, A., Tzirakis, P., Baird, A., Jones, L., & Schuller, B. (2021). End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: A pilot study. BMJ Innovations, 7(2), 356–362.
12. Gong, Y., Chung, Y.-A., & Glass, J. (2021). AST: Audio spectrogram transformer. Interspeech 2021, 571–575.
13. Gairola, S., Tom, F., Kwatra, N., & Jain, M. (2021). RespireNet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting. IEEE EMBC 2021, 527–530.
14. Han, S., Pool, J., Tran, J., & Dally, W. J. (2020). Learning both weights and connections for efficient neural networks. Advances in Neural Information Processing Systems, 28.
15. Warden, P., & Situnayake, D. (2019). TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers. O'Reilly Media.
16. Chaudhari, G., Jiang, X., Fakhry, A., Han, A., Xiao, J., Shen, S., & Khanzada, A. (2021). Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough. arXiv preprint arXiv:2011.13320.
17. Al Hossain, F., Lover, A. A., Corey, G. A., Reich, N. G., & Rahman, T. (2020). FluSense: A contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(1), 1–28.

18. Sharma, N., Krishnan, P., Kumar, R., Ramoji, S., Chetupalli, S. R., Ghosh, P. K., & Ganapathy, S. (2020). Coswara — A database of breathing, cough, and voice sounds for COVID-19 diagnosis. Interspeech 2020, 4811–4815.
19. McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and music signal analysis in Python. Proceedings of the 14th Python in Science Conference, 18–25.
20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
21. Loshchilov, I., & Hutter, F. (2017). SGDR: Stochastic gradient descent with warm restarts. ICLR 2017.
22. Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. ICCV 2017, 2980–2988.
23. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.
24. Watkins, J. A., Goudge, J., Gómez-Olivé, F. X., & Griffiths, F. (2022). Mobile phone use among patients and health workers to enhance primary healthcare: A qualitative study in rural South Africa. Social Science & Medicine, 75(8), 1398–1408.
25. Bolo, K., Mukherjee, A., & Swaminathan, S. (2023). Cost-effectiveness of AI-assisted cough screening for tuberculosis in Sub-Saharan African primary care. PLOS Global Public Health, 3(4), e0001891.

❑❑❑