npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-02086-z

AI-driven smartphone screening for acute COPD exacerbations: enhancing health equity in developing regions

Yanbin Gong[1], Chi Xu[1], Caixia Mo[2], Jianhui Wu[3], Huiyong Lin[4], Huaan Su[5], Qingpeng Zhang[6,7] ✉, Qian Zhang[1] ✉ & Shifang Yang[8] ✉

Acute exacerbations of chronic obstructive pulmonary disease (AECOPD) are critical clinical events that necessitate prompt intervention, yet their detection remains challenging in primary care, especially in resource-limited regions where a lack of awareness leads to delayed diagnoses. To address this issue, we developed an AI-based AECOPD detection system that leverages standard smartphone microphones for auscultation, specifically designed for novice users and analyzing sounds to screen for AECOPD without requiring subjective patient-reported scales. Our system demonstrated robust performance, achieving an area under the curve (AUC) of 0.955 (95% CI: 0.929–0.976). A state-transition health-economic model projected per-capita net savings of 456.9 CNY (95% CI: −88.2 to 1,779.3) with a 90.3% probability of positive returns, supporting cost-effective implementation. This research highlights the potential of AI-driven solutions to enhance COPD management in underserved populations, providing a scalable tool to promote health equity where access to pulmonary specialists is constrained.

Chronic obstructive pulmonary disease (COPD) represents a staggering global health burden, ranking as a leading cause of morbidity and mortality worldwide[1]. In China alone, nearly 100 million individuals are affected by COPD, accounting for approximately one-quarter of the global patient population and imposing a substantial economic burden[2].

Acute exacerbations of COPD (AECOPD) accelerate lung-function decline, increase mortality risk and account for over 70% of COPD-related costs via hospitalizations[3,4]. Effective management hinges on timely detection and intervention, yet this remains a formidable challenge in primary care, which is the cornerstone of chronic disease monitoring[5,6]. Studies reveal a significant knowledge gap among primary care providers regarding AECOPD diagnosis and management, with clinical practices often deviating from established GOLD recommendations[7–9]. This issue is particularly acute in developing regions. A 2015 survey in Shanghai, one of China's most developed cities, found that only 19.4% of general practitioners understood the diagnostic distinction between acute and stable COPD states[10].

Patient-related factors significantly complicate the management of COPD. Many COPD patients, especially those with lower educational levels in developing regions, often lack a comprehensive understanding of their condition, which hampers their ability to accurately report symptoms. This gap in knowledge can lead to poor self-management practices, subsequently increasing the risk of acute exacerbations (AEs)[5,6]. While there are symptom assessment tools available, such as EXACT-Pro[11], their effectiveness is often undermined by these patient-related factors and the variability in symptom reporting among individuals[12,13]. Furthermore, a substantial proportion of COPD patients (16–57%) experience varying degrees of cognitive impairment, which can further hinder their ability to self-report accurately[14]. Additionally, patients may under-report symptoms due to factors such as denial, embarrassment, or limited health literacy[15]. These patient-related factors, coupled with disparities in healthcare resource allocation, contribute to significantly higher AECOPD rates in rural areas and secondary hospitals compared to urban tertiary hospitals[16], a disparity also noted in broader healthcare delivery observations.

[1]Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong. [2]Department of Pulmonary and Critical Care Medicine, Leizhou People's Hospital, Leizhou, Guangdong, China. [3]Department of Pulmonary and Critical Care Medicine, Dianbai People's Hospital, Dianbai, Guangdong, China. [4]Department of Pulmonary and Critical Care Medicine, Lufeng People's Hospital, Lufeng, Guangdong, China. [5]Department of Pulmonary and Critical Care Medicine, Jiangmen People's Hospital, Jiangmen, Guangdong, China. [6]HKU Musketeers Foundation Institute of Data Science, The University of Hong Kong, Hong Kong, Hong Kong. [7]Department of Pharmacology and Pharmacy, The University of Hong Kong, Hong Kong, Hong Kong. [8]Department of Pulmonary and Critical Care Medicine, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, China. ✉e-mail: qpzhang@hku.hk; qianzh@cse.ust.hk; yangshifang@gdph.org.cn

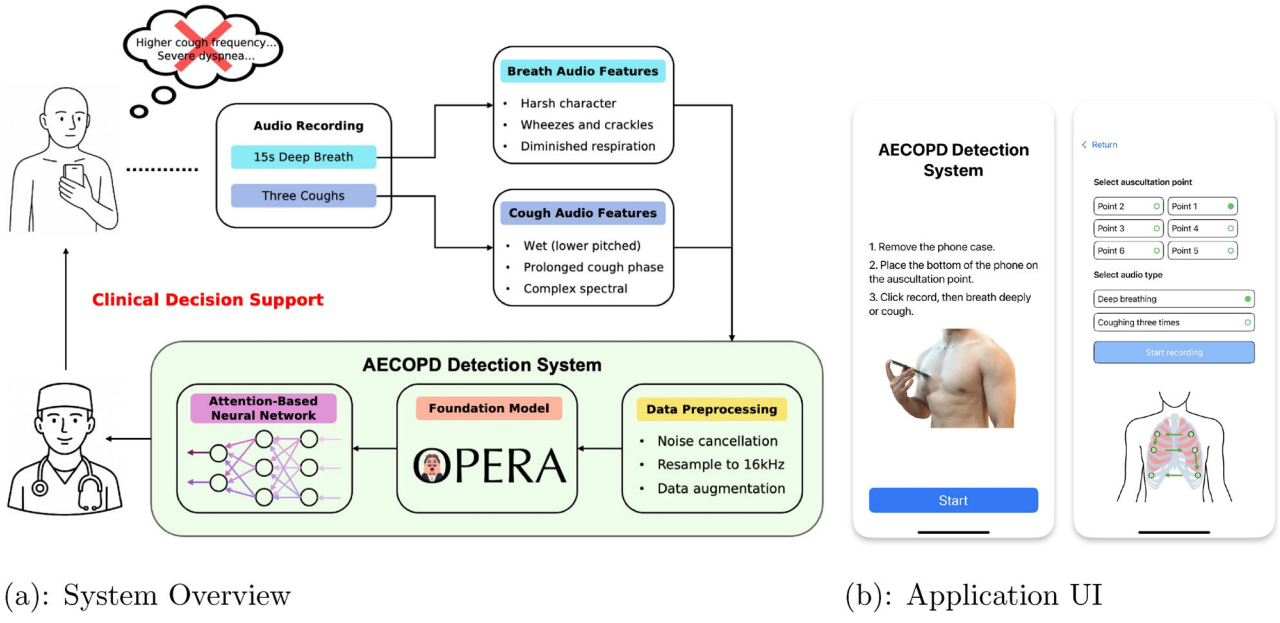


Fig. 1 | AI-system overview and usage process. a Overall system architecture diagram. The process starts with recording the patient's breathing and coughing audio, and then extracting key acoustic features from them. After preprocessing, these data are input into a neural network based on the attention mechanism and the OPERA foundation model for analysis, ultimately providing doctors with clinical decision support for AECOPD treatment. **b** The user interface of the smartphone application for audio collection. The interface design is simple and the operation instructions are clear (such as selecting auscultation points, selecting audio types), demonstrating the ease of use of the tool.

The emergence of telemedicine offers potential for enhancing health equity, particularly in developing areas[17–20]. However, remote management of COPD presents unique challenges. Unlike cardiovascular conditions where patients can report objective measurements, COPD monitoring relies heavily on subjective symptom reporting and availability of experienced physicians for remote assessment[21]. Recently, ResApp (now Pfizer Inc.) developed a smartphone-based algorithm for AECOPD diagnosis using cough audio data combined with patient-reported features such as age, fever, and new cough[22]. However, reliance on patient-reported features can limit applicability in low literacy settings; reported performance was sensitivity 82.6% (95% CI: 72.9–89.9%) and specificity 91.0% (95% CI: 82.4–96.3%).

We hypothesized that deep-breathing and cough sounds captured via consumer smartphones encode acoustic signatures that correlate with physiologic changes in AECOPD. By leveraging a foundational AI model and a multimodal fusion-attention architecture, our system aims to detect AECOPD without relying on self-reported symptoms, thereby advancing health equity in COPD management for underserved populations. As shown in Fig. 1a, patients are guided to perform standardized deep breathing and coughing maneuvers, mirroring auscultation during clinical visits. While previous work demonstrated smartphone-based cardiac auscultation[23,24], lung auscultation presents greater challenges due to lower signal-to-noise ratios[25]. Our system addresses this by integrating clinical physiological insights with a sophisticated AI model. The design is based on the understanding that AECOPD is characterized by alterations in physiological parameters such as dyspnea, respiratory rate, heart rate, oxygen saturation, and inflammatory markers[26]. Although direct measurement of all these parameters via smartphone audio is not feasible, their manifestations are often strongly correlated with discernible acoustic features[27,28]. Deep-breathing auscultation provides greater diagnostic detail compared to cough analysis, yet it presents technical challenges due to inherently lower signal amplitude, potentially resulting in unfavorable signal-to-noise ratios (SNR). Importantly, recent advancements in AI-driven audio signal processing, particularly the OPERA foundation model[29], which has been pre-trained on hundreds of thousands of diverse audio samples, empower our system to extract meaningful patterns from low-SNR smartphone recordings.

Figure 1b illustrates a typical use case of our mobile application. Users are guided to position their smartphones on specific locations of the chest to perform sequential deep-breathing and coughing maneuvers, ensuring a user-friendly experience, particularly for elderly individuals. The system automatically processes recordings and allows re-recording if compromised by misplacement or external disturbances, making it user-friendly for individuals with limited education or cognitive impairments.

To develop and validate this system, we conducted a multi-center study involving one tertiary hospital and two community hospitals in Guangzhou, and eight county-level hospitals. User studies indicate that the system is easy to use for novice users. The system achieved an Area Under the Curve (AUC) of 0.955 (95% CI: 0.929-0.976) for AECOPD detection. We estimate that widespread adoption of this technology could provide substantial health economic benefits by improving health equity and alleviating the healthcare burden in developing regions.

Results

Study population and demographics

This prospective, multi-center study was conducted in 13 hospitals in central and southern China from November 2024 to June 2025. Of 292 enrolled COPD patients that are consecutive sampled, 274 (165 AECOPD and 109 stable) were included for analysis after excluding 18 with missing or damaged data. The baseline demographics and clinical characteristics of the study population are detailed in Table 1. Statistically significant differences were observed between the AECOPD group and the stable group in terms of age, smoking status, and number of AE in the last year (all $P < 0.001$). Specifically, patients in the AECOPD group were older, had a higher percentage of current smokers, and reported more AE in last year compared with the stable group. No significant difference was found in the gender distribution between the two groups.

Dataset composition

Data collection was designed to reflect real-world clinical practice. The process was integrated into routine auscultation of outpatients and performed at the bedside of inpatients in the ward, which are non-isolated clinical environments with inevitable ambient noise. Each patient's

Table 1 | Baseline demographics and clinical characteristics of the study population

Characteristic	All (274)	AECOPD (165)	Stable (109)	*P*-value
Subjects, *n*	274	165	109	
Age (years)				0.001
Mean ± SD	72.9 ± 8.8	74.3 ± 8.5	70.8 ± 8.8	
Median (Q1, Q3)	74.0 (67.0, 79.0)	75.0 (69.0, 81.0)	71.0 (65.0, 78.0)	
Sex, *n* (%)				0.621
Male	221 (80.7%)	131 (79.4%)	90 (82.6%)	
Female	53 (19.3%)	34 (20.6%)	19 (17.4%)	
Smoking status, *n* (%)				0.000
Current smoker	78 (28.5%)	72 (43.6%)	6 (5.5%)	
Former smoker	192 (70.1%)	90 (54.5%)	102 (93.6%)	
Never smoker	4 (1.5%)	3 (1.8%)	1 (0.9%)	
Exacerbation Count				0.000
Median (Q1, Q3)	1.0 (0.0, 2.0)	1.0 (1.0, 2.0)	0.0 (0.0, 1.0)	

condition was diagnosed by two independent medical professionals as stable COPD, mild AE, or moderate to severe AE (requiring hospitalization). During our clinical visits, 142 patients presented with moderate-to-severe AE, which reflects the trend of patients seeking treatment later in the course of AE in our region. Data were collected by researchers at the 2nd, 4th, and 6th intercostal spaces with an average collection time of 4 minutes. The data collection protocol was well tolerated, with 88.0% of patients describing it as simple and convenient, and only 5.1% considering it unsuitable for daily use.

AI system accurately classifies disease states using only audio data

To evaluate the performance of the system and ensure its robustness, this study used five fold stratified cross validation with stratification at the subject visit level, ensuring that all recordings from a given visit were assigned exclusively to either the training or the validation set within each fold, thereby preventing data leakage while maintaining class balance. As shown in Fig. 2a, the system demonstrated excellent discriminative power, with a pooled AUC of 0.955 (95% CI: 0.929–0.976) across all five-fold cross-validations. This performance was consistent across metrics, including 91.5% pooled accuracy (95% CI: 88.2–94.9%), 89.9% sensitivity (95% CI: 84.9–94.5%), and 93.8% specificity (95% CI: 88.8–98.1%). The consistency of the results across the data groups confirmed the stability of the model performance. Although the optimal decision thresholds varied slightly between groups (mean: 0.548 ± 0.128), the average AUC for each group was still as high as 0.951, with minimal differences in accuracy (0.916 ± 0.039) and F1 score (0.924 ± 0.039). Notably, the model achieved 100% specificity in two of the five groups, indicating that it can accurately identify patients with stable disease and minimize false positives. In a stress test with controlled additive Gaussian noise injected into the raw validation data, performance degraded gracefully as noise increased.

To further probe the model's resilience to real-world acoustic variability, we conducted a stress test by injecting controlled levels of additive Gaussian noise into the raw validation data. This experiment was designed to simulate the varying levels of background noise encountered in uncontrolled home or community settings. As illustrated in Fig. 2b, the model's performance degraded gracefully as the noise level increased. Even under significant acoustic interference, the system maintained strong predictive power, demonstrating its robustness and suitability for deployment outside of quiet clinical environments. This resilience is critical for ensuring reliable performance when the tool is used by patients in diverse real-world settings.

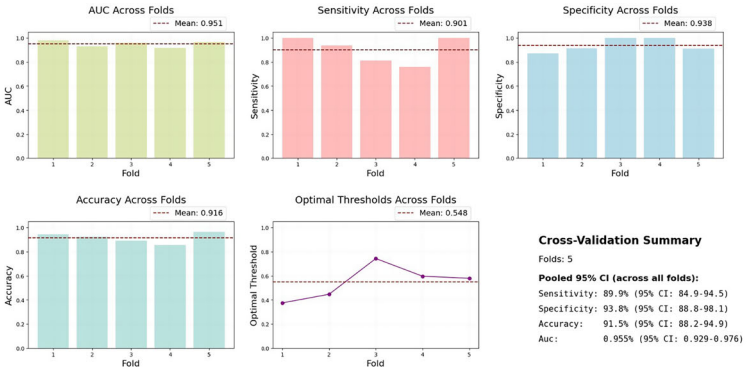
We then comprehensively compare its performance with five baseline models covering different technical paths. As shown in Fig. 2c, our model significantly outperforms all the compared models on all key evaluation metrics. Specifically, we first compare our model with a variant that uses the same OPERA acoustic embedding features but uses a standard multi-layer perceptron (MLP) for classification. Our model achieves significant improvements in AUC (0.955 vs 0.914), sensitivity (0.899 vs 0.855), specificity (0.938 vs 0.867), and accuracy (0.915 vs 0.860). This result strongly demonstrates that the unique architectural design of our model can more effectively learn and extract disease-related discriminative features from OPERA embeddings. In addition, we also compare with the ResApp model, which not only analyzes cough audio but also combines patient self-report data. Despite ResApp's exploitation of additional multimodal information, our audio-only model still demonstrates stronger performance in terms of AUC (0.955 vs 0.890), sensitivity (0.899 vs 0.826), and specificity (0.938 vs 0.910). This highlights the technical advantages and potential of our approach in making accurate diagnoses relying solely on audio signals. Finally, our model outperforms commonly used methods in the field of audio-based disease classification models, such as the fine-tuned Audio Spectrogram Transformer (AST)[30], using convolutional neural networks to process Mel-frequency cepstral coefficients (MFCC+CNN), or using random forests with acoustic statistical features. Our method performs much better than the best baseline model, AST (AUC 0.847, accuracy 0.799), demonstrating its great potential as an innovative disease screening and monitoring tool.

Consistent and equitable performance in heterogeneous patient populations

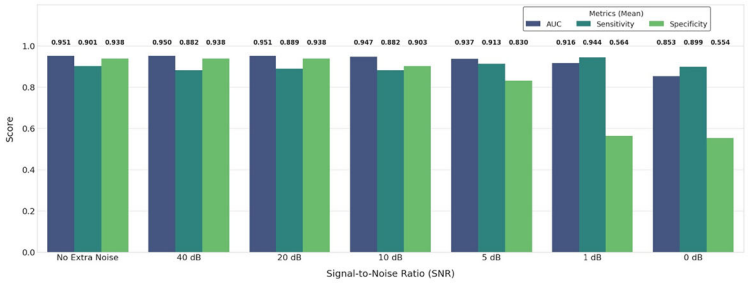
To analyze the generalization ability of the model, a series of pre-specified stratified subgroup analyses were performed on the entire cohort of 274 patients based on key clinical and socioeconomic characteristics. As shown in Fig. 2d, the diagnostic accuracy of the model remained consistent across the range of AECOPD severity. In the cohort diagnosed with AECOPD, the system successfully identified 91.3% of mild AE (n = 23) and 90.1% of moderate to severe AE (n = 142). This consistently high level of performance, especially in mild cases that are difficult to diagnose, highlights its potential for early detection of the disease. The system showed robust performance across patients of different ages. In patients older than 65 years (n = 219), the model had an accuracy of 91.3% (95% CI: 87.2%–94.5%), a sensitivity of 89.1% (95% CI: 83.2%–93.9%), and a specificity of 95.1% (95% CI: 89.8%–98.9%). In the younger cohort (n = 55), performance was comparable, with accuracy, sensitivity, and specificity of 92.7% (95% CI: 85.5%–98.2%), 96.3% (95% CI: 87.5%–100.0%), and 89.3% (95% CI: 75.9%–100.0%), respectively. Analysis of the misclassified samples provided further insights false-negative samples in the older cohort occasionally occurred in patients with no significant abnormalities on clinical auscultation, suggesting that the model may have captured subtle subclinical signs of AE that are often masked by age-related physiological changes. In contrast, the few false-positive samples in the younger cohort were often associated with non-AE of respiratory inflammation, whose acoustic patterns may be similar to acute events. In addition, we also evaluated the application value of the system in different medical ecosystems by stratification by regional GDP per capita. The performance of the model showed excellent stability across these significantly different economic strata. In developed regions (GDP per capita >89,000 CNY; n = 123), the system's accuracy was 93.5% (95% CI: 88.6%–97.6%). In developing regions (n = 151), the system also achieved a comparable accuracy of 90.1% (95% CI: 84.8%–94.7%). The chi-square test analysis showed that the difference in accuracy between the two region groups did not reach a statistically significant level (χ^2 = 1.037, P = 0.309). This result indicates that the performance of the system is independent of regional economic factors and potential confounding variables, thus shows great potential in developing areas.

In summary, these subgroup analyses show that the model is not only accurate across disease severity, age, and socioeconomic status, but also highly robust, reliable, and fair. This provides strong evidence for its widespread and effective deployment in diverse real-world clinical settings.

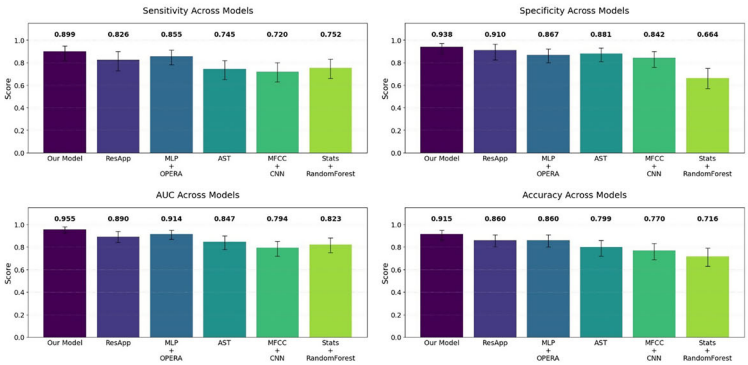
Fig. 2 | Overview of AI system performance. **a** Five-fold cross-validation results showing stable performance across key metrics. **b** Performance degraded gracefully with increasing additive noise. **c** Performance comparison against several baseline models, demonstrating the superiority of our proposed model. **d** Consistent performance across diverse demographic and socioeconomic subgroups supports more equitable healthcare delivery.



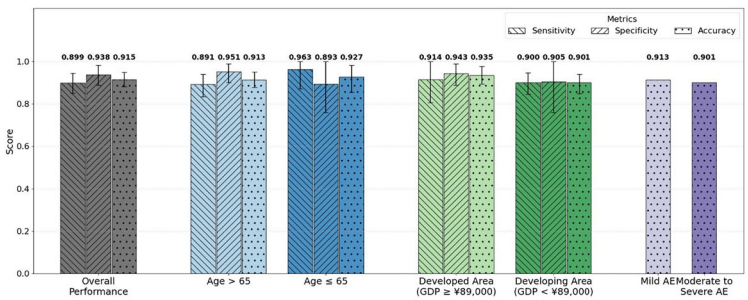
(a): Five-fold cross-validation results.



(b): Performance under different noise levels (averaged across 5 folds).



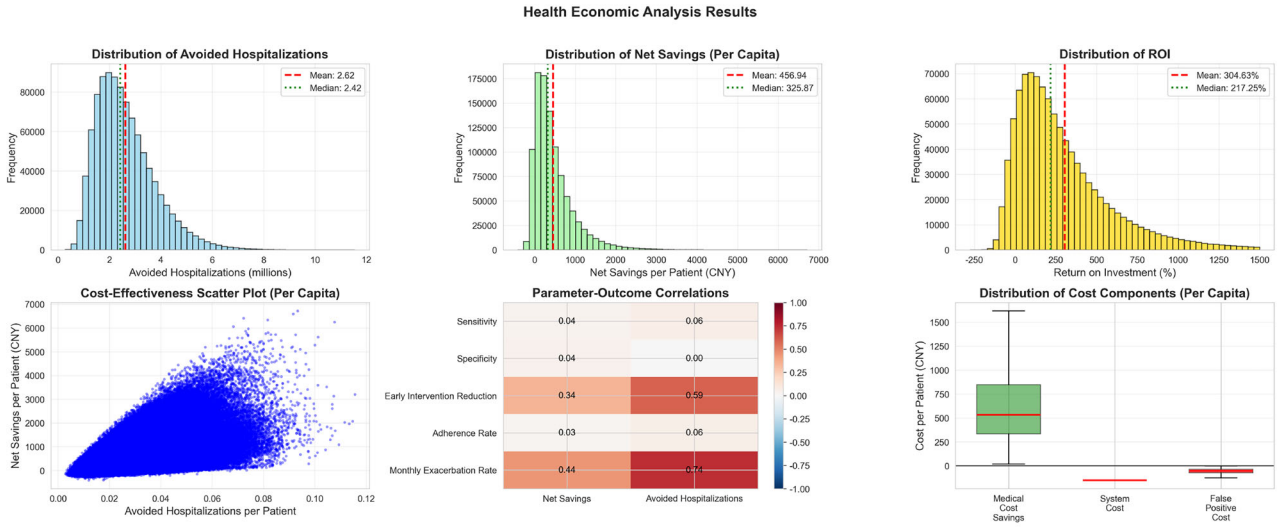
(c): Performance comparison against several baseline models.



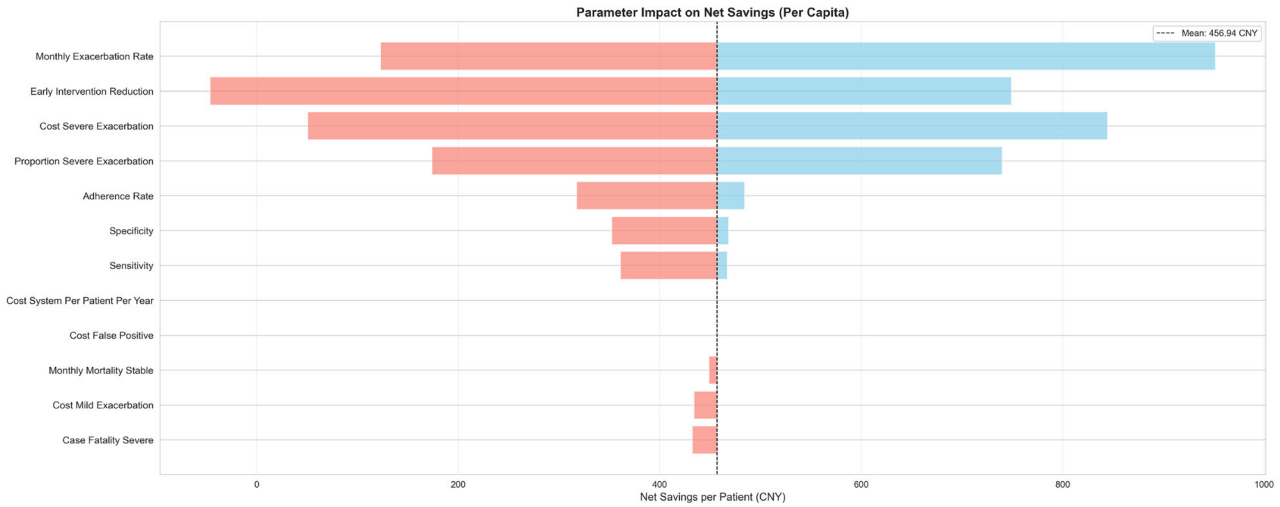
(d): Performance in different subgroups.

Economic modeling indicates substantial cost reductions

Based on the comprehensive probabilistic sensitivity analysis involving one million simulations, our system demonstrates consistent and robust economic value despite parameter uncertainty. As shown in Fig. 3a, the Monte Carlo simulation results demonstrate that the system generates mean per capita net savings of 456.94 CNY (95% CI:–88.15 to 1779.27 CNY), with 90.30% probability of positive returns. Clinical benefits include 0.0262 hospitalizations avoided per patient (95% CI: 0.0102 to 0.0537). The return on investment remains compelling at 304.43% (95% CI:–58.77% to 1186.18%), driven by per capita medical cost savings of 664.52 CNY (95% CI: 126.34 to 1,985.70 CNY) that substantially exceed the combined system cost of 150 CNY and false positive cost of 57.87 CNY (95% CI: 18.28 to



(a): Results of probabilistic sensitivity analysis (PSA)



(b): Tornado plot of one-way sensitivity analysis

Fig. 3 | Health economics of AECOPD detection system. a Results of probabilistic sensitivity analysis (PSA). The probability distribution of the model's economic output was right skewed across per capita outcomes (net savings, hospitalizations avoided). Net savings were strongly positively correlated with the number of hospitalizations avoided. Parameter-output correlation analysis showed that early intervention effects and monthly acute exacerbation rates were key factors in determining the model's output. **b** Tornado plot of the one-way sensitivity analysis. This plot reveals the sensitivity of net savings to key parameters, with severe exacerbation costs, monthly exacerbation rates, and early intervention effectiveness being the most important drivers of economic benefits.

130.19 CNY), ultimately reducing total per capita healthcare costs from 5,362.39 CNY under standard care (95% CI: 1355.21 to 14,408.90 CNY) to 4905.75 CNY with the AI system (95% CI: 1414.17 to 12,779.65 CNY). A one-way sensitivity analysis identified the costs of severe exacerbation, monthly exacerbation rates, and early intervention effectiveness as the most important drivers of economic benefit (Fig. 3b).

We also analyzed the economic outcomes across five distinct scenarios, as detailed in Table 2. The base case was agreed by clinicians and researchers; subsequent scenarios applied progressively conservative assumptions. We see even when the *Early Intervention Reduction* decreases by 5%, *False Positive Event Cost* increases by 100 CNY, and *System Annual Cost* increases by 100 CNY, the net savings are still positive. In the worst case, additional costs are limited relative to the base case benefit.

Table 2 | Net savings under different assumed scenarios

	Different Assumed Scenarios				
Early Intervention Reduction	0.20	0.18	0.15	0.12	0.10
False Positive Event Cost (CNY)	100	150	200	250	300
System Annual Cost (CNY)	150	200	250	300	350
Per Capita Results					
Net Savings (CNY)	456.9	320.3	140.0	−38.0	−185.8
Hospitalizations Avoided	0.0262	0.0236	0.0196	0.0158	0.0132
ROI (%)	304.6	160.1	56.0	−12.7	−53.1

Discussion

In this study, we demonstrate that an AI-driven screening tool, leveraging only the microphones of commercial smartphones, can accurately and objectively detect AECOPD. This work addresses a critical and persistent unmet need in global health: the timely identification of AECOPD in primary care and community settings, particularly where resources are scarce. By achieving a high AUC of 0.955(95% CI: 0.929–0.976), our approach represents a significant advance towards democratizing COPD management, offering a scalable, non-self-report method that can enable earlier, more effective interventions.

Our findings present a potential paradigm shift from current AECOPD detection strategies, which are often hampered by their reliance on subjective patient-reported symptom scales or require resource-intensive clinical assessments like spirometry. The high diagnostic accuracy of our model is clinically significant, but the balance between sensitivity (89.9%) and specificity (93.8%) at the chosen operating point warrants specific discussion, as raised by reviewers. While a higher sensitivity might seem desirable to avoid missing any potential exacerbation, the high specificity of our model is a key strategic advantage in our target resource-limited settings. It minimizes the rate of false positives, thereby reducing unnecessary patient anxiety, costly follow-up tests, and the financial burden on an already strained healthcare system. Preventing alarm fatigue is critical for the long-term adoption and user trust necessary for such a screening program to succeed. Nonetheless, the model's excellent overall discriminative ability allows for the operating threshold to be adjusted based on specific public health goals or clinical contexts.

The clinical potential of our system is strongly reinforced by its economic viability. Our health-economic model, which projects a per-capita net saving of 456.94 CNY, provides a compelling argument for its adoption. We acknowledge that such models are sensitive to their underlying assumptions. To address this, we performed a comprehensive Probabilistic Sensitivity Analysis and incorporated several conservative scenarios, which confirmed the robustness of the positive financial impact. These model-based projections, however, must be validated through real-world implementation studies, which forms a critical part of our future work. This evidence of cost-effectiveness is pivotal, as it moves the conversation about digital health tools from mere technical feasibility to tangible, real-world value for health systems.

A central consideration for a smartphone-based tool is its robustness to real-world variability. Our primary strategy was to embrace this variability by training the model on data collected in non-soundproof clinical environments using various commercial smartphone models. This approach encourages the model to learn signal-invariant features rather than overfitting to a pristine acoustic environment or specific hardware. To further stress-test this, we employed data augmentation strategies with controlled additive noise, which confirmed the model's resilience. This inherent robustness is a key strength, though we recognize that a more systematic evaluation across a wider range of devices and quantified signal-to-noise ratios remains an important direction for future research. Methodologically, the integrity of our performance estimates was ensured by a strict subject-visit-level data partitioning strategy during cross-validation, which prevents data leakage and ensures the model was evaluated on genuinely unseen subjects.

While our study provides a strong proof-of-concept, we acknowledge its limitations, which define our roadmap for future research. The cohort was recruited exclusively from central and southern China. Therefore, the model's external validity in non-Chinese populations with different clinical, genetic, and environmental characteristics remains to be confirmed. Establishing international collaborations to validate and refine the model across diverse ethnic and clinical cohorts is a top priority. Furthermore, our study did not systematically account for respiratory comorbidities like asthma or acute infections that may produce confounding acoustic patterns; future model iterations will need to be trained on broader datasets to improve specificity against these conditions. This cross-sectional study must also be followed by large-scale, prospective, longitudinal validation to assess the system's real-world performance over time and its true impact on clinical outcomes and cost-effectiveness.

To enhance clinical trust and utility, future work will also focus on model interpretability. While our current focus was on establishing diagnostic accuracy, we plan to apply explainable AI (XAI) techniques like SHAP or Grad-CAM to identify the specific acoustic features driving the model's predictions. This could not only increase clinical confidence but also yield new pathophysiological insights into AECOPD. We also recognize the limitation regarding the absence of a formal inter-rater reliability analysis for our ground-truth labels. Although our diagnostic consensus was rigorous and criteria-based, future prospective studies will incorporate independent, blinded assessments prior to consensus to formally quantify the reliability of the diagnostic labeling.

Finally, the responsible and ethical implementation of this technology, particularly in settings with low health literacy, is paramount. We emphasize that this system is designed as a decision-support tool to aid clinical assessment, not to replace it. Mitigating the risk of over-reliance requires a multi-faceted strategy, including clear clinical guidelines, robust user training for non-specialized personnel, and established escalation pathways for managing potential misclassifications, especially false negatives where clinical suspicion remains high. Our study protocol included IRB approval, written informed consent, and data anonymization to protect patient confidentiality, and our fieldwork confirms high smartphone acceptance even in rural areas, supporting the feasibility of this approach. In conclusion, by developing and validating an accurate, objective, and cost-effective AI-driven screening tool on a ubiquitous platform, this work lays the foundation for a more proactive and equitable model of chronic disease management, with the potential to fundamentally reshape care for millions of COPD patients worldwide.

Methods

Study design and setting

We conducted a prospective, multicenter cross-sectional study in central and southern China to develop and validate the AECOPD detection system. The study was conducted in multiple medical centers, including Guangdong Provincial People's Hospital. All participants were fully informed of the purpose, procedures, potential risks and benefits of the study before enrollment and voluntarily signed written informed consent. The subjects of the study were recruited from the respiratory outpatient clinic and inpatient department of each participating center. The recruitment process was arranged during the outpatient consultation and clinical assessment. The data collected included auscultation sounds, basic information of the patients, and a COPD management questionnaire based on the subjective completion of the patients (see Appendix). The inclusion criteria were: 1) aged between 18 and 90 years; 2) confirmed diagnosis of COPD according to the diagnostic criteria of GOLD[31]; 3) able to cooperate with the completion of pulmonary function tests and related questionnaires; 4) able to cooperate with the use of inhaled medications; 5) not participating in other pulmonary rehabilitation programs within six months before enrollment. Exclusion criteria include patients with other severe respiratory diseases (such as severe pneumothorax, pulmonary embolism, etc.), uncontrolled severe systemic diseases (such as unstable cardiovascular disease, untreated tumors, etc.), musculoskeletal or nervous system diseases that affect movement, mental or cognitive disorders, and patients with a history of lung surgery within the past three months. This strict screening criteria are designed to ensure the safety of the research and the homogeneity of the population.

Ethics statement

This study was conducted in compliance with the Declaration of Helsinki and received ethical approval from the Ethics Review Committee of Guangdong Provincial People's Hospital (IRB no. KY2025-422-01). Written informed consent was obtained from all participants. An English translation of the ethics approval documentation is available from the corresponding author upon reasonable request.

Clinical reference standard

To ensure the accuracy and reliability of diagnostic labels, the clinical reference standard was established through a consensus agreement between two experienced respiratory physicians who independently assessed each participant. The clinical evaluations were performed in accordance with the 2021 Rome Proposal[26]. During the evaluation process, experts first distinguished whether the COPD patient was in a stable state or an AE state by whether the patient had dyspnea or worsening cough and sputum within 14 days, which may be accompanied by increased respiratory rate or heart rate. Subsequently, AECOPD patients were further divided into mild or moderate to severe based on a combination of objective physiological indicators, including dyspnea scores, vital signs, oxygen saturation, and C-reactive protein. Conflicts in evaluation opinions were discussed and negotiated to reach a final consensus. This process provides a definitive label for each sample as a benchmark for model development and validation.

Data collection

A custom iOS application was developed for data collection in this study. The application can directly call the microphone at the bottom of the smartphone to collect the raw audio stream, thereby bypassing the interference caused by system-level signal processing optimized for voice calls (such as noise reduction and gain). During data collection, the subject remained seated, and the collector placed the bottom edge of the phone directly on the patient's skin surface and collected respiratory sounds at symmetrical positions in the second, fourth, and sixth intercostal spaces. At each collection point, we recorded 15 seconds of deep breathing sounds and 3 spontaneous cough sounds in sequence.

Model development

To extract discriminative features from complex acoustic signals, we implemented a series of processing steps and model designs. Our model development pipeline consists of audio pre-processing, feature extraction, and a multimodal fusion attention network for classification.

Initially, we applied Wiener filtering and bandpass filtering to preprocess the audio data, effectively removing both device noise and extraneous sounds from the breathing segments. Subsequently, we downsampled the data to 16 kHz and utilized the OPERA model[29] to encode the processed cough and deep breathing sounds into high-dimensional feature embeddings. Specifically, we employed two distinct encoders: OPERA-CT (768 dimensions), based on a Transformer architecture, and OPERA-CE (1280 dimensions), based on convolutional neural networks. This approach allowed us to convert each audio clip into a structured feature vector for subsequent modeling.

To integrate acoustic information from cough and breathing, we designed and implemented a multimodal fusion attention network. The model first maps CT and CE features of different dimensions to a unified feature space through independent projection layers. Then, the spatial dependency of different lung positions is captured by the intra-modal self-attention mechanism[32], and the cross-modal attention mechanism is used to achieve deep information interaction between cough and deep breathing features. Finally, the fused features are fed into a classifier to output the probability of AECOPD.

To ensure an unbiased and robust assessment of our model's performance, we adopted a rigorous 5-fold stratified cross-validation protocol. Critically, stratification was performed at the "subject-visit" level, ensuring that all audio recordings from a single patient visit were exclusively assigned to either the training or the validation set within each fold. This strict separation is essential to prevent any form of data leakage, where the model might learn patient-specific rather than disease-specific features, and thus provides a more realistic estimate of the model's performance on unseen subjects.

In the model inference stage, we did not utilize a fixed classification threshold of 0.5. Instead, we determined the optimal probability threshold for each fold by maximizing Youden's J statistic on its respective validation set, based on the analysis of the receiver operating characteristic (ROC) curve. This data-driven threshold provides a decision-making basis for subsequent performance evaluation (such as sensitivity, specificity, accuracy, etc.) and ensures that the model achieves a better balance between sensitivity and specificity. For robustness analysis, we injected zero mean additive Gaussian noise at multiple levels into the data at evaluation time and recomputed performance metrics.

Health economic analysis

We constructed a health economics analysis framework based on the state-transition Markov model to evaluate the potential economic value of the AECOPD detection system. The model includes four core health states: stable, mild AE, moderate-to-severe AE, and death. We simulated the natural course of the disease and clinical outcomes of the Chinese COPD patient population in a monthly cycle over a 1-year period.

The model compared two scenarios: "standard of care" and "system-assisted management". The key parameters of the model cover multiple dimensions such as epidemiology (such as the total population of COPD, the annual incidence of AE, and background mortality), clinical effectiveness (such as the effectiveness of early intervention in reducing severe AE), system effectiveness (such as the sensitivity and specificity of the system, and patient compliance) and economic costs (such as the cost of treating AE outpatient and inhospital, the cost of using the system, and the cost of false positive events). The baseline values and uncertainty ranges of these parameters are mainly derived from published literature, clinical expert consensus, and real-world data. The cost of treating inhospital AE is based on an empirical distribution constructed based on real hospitalization data from our center to enhance the model's real-world fit.

We conducted a PSA to fully evaluate the impact of parameter uncertainty. In the analysis, all key parameters were assigned specific probability distributions, and 1 million simulations were performed using the Monte Carlo method. The main output indicators of the model include: net cost savings, avoided hospitalizations, avoided deaths, and ROI. In addition, we performed a one-way sensitivity analysis and presented it in the form of a tornado diagram to identify the key drivers that have the greatest impact on the model results.

Statistical analysis

The training, validation, and testing of the models were performed with PyTorch and Scikit-learn. The performance metrics-including AUC, sensitivity, specificity, accuracy, and corresponding 95% confidence intervals were calculated with SciPy and NumPy. Figures were generated with Matplotlib and Seaborn.

Data availability

The datasets generated and analysed during the current study are not publicly available due to the privacy considerations but are available from the corresponding author upon reasonable request after anonymization and subject to institutional approvals and a data use agreement.

Code availability

Custom code used for model training and analysis will be made available to qualified researchers upon reasonable request under a data/code use agreement, consistent with institutional policies and third party model license terms.

Received: 1 July 2025; Accepted: 11 October 2025;
Published online: 21 November 2025

References

1. Momtazmanesh, S. et al. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the global burden of disease study 2019. *EClinMed.* **59**, 101936 (2023).

2. Wang, C. et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. *Lancet* **391**, 1706–1717 (2018).
3. Anzueto, A. Impact of exacerbations on copd. *Eur. Resp. Rev.* **19**, 113–118 (2010).
4. Mathioudakis, A. G. et al. Acute exacerbations of chronic obstructive pulmonary disease: in search of diagnostic biomarkers and treatable traits. *Thorax* **75**, 520–527 (2020).
5. Wedzicha, J. A. & Seemungal, T. A. Copd exacerbations: defining their cause and prevention. *Lancet* **370**, 786–796 (2007).
6. Yawn, B. P., Mintz, M. L. & Doherty, D. E. Gold in practice: chronic obstructive pulmonary disease treatment and management in the primary care setting. *Int. J. Chronic Obstructive Pulmonary Dis.* 16, 289–299 (2021).
7. Vachon, B. et al. Challenges and strategies for improving COPD primary care services in Quebec: results of the experience of the compas+ quality improvement collaborative. *Int. J. Chronic Obstr. Pulmonary Dis.* 17, 259–272 (2022).
8. Criner, R. N. & Han, M. K. Copd care in the 21st century: a public health priority. *Resp. Care* **63**, 591–600 (2018).
9. Liang, L. et al. Trends in hospitalization expenditures for acute exacerbations of COPD in Beijing from 2009 to 2017. *Int. J. Chronic Obstr. Pulmonary Dis.* 15, 1165–1175 (2020).
10. Li, F. et al. The evaluation of general practitioners' awareness/knowledge and adherence to the gold guidelines in a Shanghai suburb. *Asia Pac. J. Public Health* **27**, NP2067–NP2078 (2015).
11. Leidy, N. K. et al. Development of the exacerbations of chronic obstructive pulmonary disease tool (exact): a patient-reported outcome (pro) measure. *Value Health* **13**, 965–975 (2010).
12. Wilkinson, T. M., Donaldson, G. C., Hurst, J. R., Seemungal, T. A. & Wedzicha, J. A. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. *Am. J. Resp. Crit. Care Med.* **169**, 1298–1303 (2004).
13. Wu, P. et al. Pharmaceutical treatment status of patients with COPD in the community based on medical internet of things: a real-world study. *NPJ Prim. Care Resp. Med.* **34**, 10 (2024).
14. Chen, X. et al. Chronic obstructive pulmonary disease as a risk factor for cognitive impairment: a systematic review and meta-analysis. *BMJ Open Resp. Res.* **11**, e001709 (2024).
15. Stellefson, M., Wang, M. Q. & Campbell, O. Factors influencing patient-provider communication about subjective cognitive decline in people with COPD: Insights from a national survey. *Chronic Resp. Dis.* **21**, 14799731241268338 (2024).
16. Yang, T. et al. Exacerbation in patients with stable COPD in China: analysis of a prospective, 52-week, nationwide, observational cohort study (real). *Ther. Adv. Resp. Dis.* **17**, 17534666231167353 (2023).
17. Liu, F., Jiang, Y., Xu, G. & Ding, Z. Effectiveness of telemedicine intervention for chronic obstructive pulmonary disease in China: a systematic review and meta-analysis. *Telemed. e-Health* **26**, 1075–1092 (2020).
18. North, M. et al. A randomised controlled feasibility trial of e-health application supported care vs usual care after exacerbation of COPD: the rescue trial. *NPJ Dig. Med.* **3**, 145 (2020).
19. Liu, X. et al. Improving access to cardiovascular care for 1.4 billion people in China using telehealth. *npj Dig. Med.* **7**, 376 (2024).
20. Enichen, E. J., Heydari, K., Li, B. & Kvedar, J. C. Telemedicine expands cardiovascular care in China–lessons for health equity in the United States. *npj Dig. Med.* **8**, 71 (2025).
21. Velardo, C. et al. Digital health system for personalised COPD long-term management. *BMC Med. Inform. Decis. Mak.* **17**, 1–13 (2017).
22. Claxton, S. et al. Identifying acute exacerbations of chronic obstructive pulmonary disease using patient-reported symptoms and cough feature analysis. *npj Dig. Med.* **4**, 107 (2021).
23. Kang, S.-H. et al. Cardiac auscultation using smartphones: pilot study. *JMIR mHealth uHealth* **6**, e8946 (2018).
24. Luo, H. et al. Smartphone as an electronic stethoscope: factors influencing heart sound quality. *Eur. Heart J.-Dig. Health* **3**, 473–480 (2022).
25. Williams, L. et al. *Auscultation Skills: Breath & Heart Sounds*. (Lippincott Williams & Wilkins, Philadelphia, PA, 2009).
26. Celli, B. R. et al. An updated definition and severity classification of chronic obstructive pulmonary disease exacerbations: the Rome proposal. *Am. J. Resp. Crit. Care Med.* **204**, 1251–1258 (2021).
27. O'donnell, D. & Parker, C. Copd exacerbations· 3: pathophysiology. *Thorax* **61**, 354–361 (2006).
28. Melbye, H., Solis, J. C. A., Jácome, C. & Pasterkamp, H. Inspiratory crackles-early and late-revisited: identifying COPD by crackle characteristics. *BMJ Open Resp. Res.* **8**, e000852 (2021).
29. Zhang, Y. et al. Towards open respiratory acoustic foundation models: Pretraining and benchmarking. *Adv. Neural Inf. Process. Syst.* **37**, 27024–27055 (2024).
30. Gong, Y., Chung, Y.-A. & Glass, J. Ast: Audio spectrogram transformer. *arXiv preprint arXiv:2104.01778* (2021).
31. Venkatesan, P. Gold COPD report: 2024 update. *Lancet Resp. Med.* **12**, 15–16 (2024).
32. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Proces. Syst.* **30** (2017).

Acknowledgments

This research is supported in part by Hong Kong RGC under Contract CERG 16206122, 16204523, HB016, AoE/E-601/22-R, R6021-20, and Contract R8015.

Author contributions

Q.Z., QP.Z. and S.Y. conceived and administered the study. Y.G. performed the mobile application development, model development, baseline implementation, data analysis and curation. C.X. helped part of the model development. Y.G., Q.Z. and S.Y. analyzed the results. S.Y., C.M., J.W., H.L., H.S. and QP.Z. provided important clinical and social knowledge content for the study design. S.Y., C.M., J.W., H.L., H.S. and Y.G. conducted the clinical communication with patients and collected data. S.Y., C.M., J.W., H.L. and H.S. labelled data through a consensus agreement. Y.G. and S.Y. wrote the manuscript. Y.G., Q.Z., QP.Z., S.Y. and C.X. have critically revised the manuscript. All authors have critically read and approved the manuscript.

Competing interests

Qingpeng Zhang is an associate editor of npj Digital Medicine. The concerned author is not part of a peer review process or decision making of the manuscript.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-025-02086-z.

Correspondence and requests for materials should be addressed to Qingpeng Zhang, Qian Zhang or Shifang Yang.